Science Advances
AAAS

Supplementary Materials for

Genomic basis for skin phenotype and cold adaptation in the extinct Steller's sea cow

Diana Le Duc*, Akhil Velluva, Molly Cassatt-Johnstone, Remi-Andre Olsen, Sina Baleka, Chen-Ching Lin, Johannes R. Lemke, John R. Southon, Alexander Burdin, Ming-Shan Wang, Sonja Grunewald, Wilfried Rosendahl, Ulrich Joger, Sereina Rutschmann, Thomas B. Hildebrandt, Guido Fritsch, James A. Estes, Janet Kelso, Love Dalén, Michael Hofreiter*, Beth Shapiro*, Torsten Schöneberg*

*Corresponding author. Email: schoberg@medizin.uni-leipzig.de (T.S.); bashapir@ucsc.edu (B.S.); michael.hofreiter@uni-potsdam.de (M.H.); diana_leduc@eva.mpg.de (D.L.D.)

Published 4 February 2022, *Sci. Adv.* **8**, eabl6496 (2022)
DOI: 10.1126/sciadv.abl6496

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

SM1: Samples, data processing, and ancient DNA authentication

Provenance

1) Fossil collection

Dugong: Tissue sample M.41683.002 was transferred from the Australian Museum to the Swedish Museum of Natural History on 2 May 2017. Both institutions are entitled to the exemption provided by Article VII, paragraph 6, of the CITES Convention.

Hydromalis bone fragments for all samples other than SNMB N51667 were collected from large ribs deposited on the beaches of the Commander Islands, Russia. Co-author Alexander Burdin collected these samples during paleontological expeditions in 2006 and 2007 by drilling small fragments from several dozen large ribs, leaving the ribs on the beaches so as to preserve them *in situ*. No permits were required either for within-country collection or for transfer of subsampled bone powder to the UCSC Paleogenomics Laboratory.

Sample SNMBN51667 was donated to the Braunschweig Natural History Museum in 1910 by Adolph Dattan, a resident of Vladyvostok, Russia. At that time, there was no regulation concerning trade of fossil bones. The specimen has been on public display in the Braunschweig Museum and described previously in several scientific publications.

2) Species-level validation and Radiocarbon Dating

Molecular species designation and radiocarbon dating of fossil fragments of Hydromalis is described below and in the main text. Briefly, authenticity was verified by mapping read fragments to the genome of the dugong using mapping parameters described in what follows. Radiocarbon dates were obtained from each individual fragment at the W. M. Keck Carbon Cycle Accelerator Mass Spectrometry facility at UC Irvine and the Klaus-Tschira-AMS facility in the Curt-Engelhorn-Centre of Archaeometry in Mannheim, Germany.

3) Accessibility

Dugong M.41683.002 is part of the collection at the Australian Museum, 1 William St., Sydney, NSW 2010, Australia. *Hydromalis* SNMBN51667 is part of the permanent collection of the Braunschweig Natural History Museum. Subsamples of *Hydromalis* ribs were deliberately extremely small and, in most instances, were consumed in their entirety for the purposes of genomic analysis. In some cases, small amounts of bone material remains, and are accessioned at the UCSC Paleogenomics Lab with accession details provided in Table S1. When no material remains, access to residual DNA extracts or libraries can be requested from co-author Shapiro.

CT-scanning of the skull of a museum specimen Steller's sea cow (*Hydrodamalis gigas*) – SNMB N51667

To identify the most densely ossified region (MDOR), we performed a radiological assessment of the intact Steller's sea cow skull with a dimension of 490 mm x 330 mm x 220 mm. We used a clinical 320-slice CT volume scanner Aquilion ONE from Toshiba Medical, Japan to scan the entire skull in 0.5 mm steps for a total of 2,721 slices. The scan parameters applied were as follows: 100 kV and 120 kV at 300 mA with a rotation time of 0.75 sec per revolution. We then used the reconstruction cores FC08, FC30, FC35, and FC81. We performed

post-processing for image analysis, including calculation of bone density and a complete 3D reconstruction using a VITREA Advanced Visualization Workstation (version 7.11.5.29) and a MacPro with DICOM Viewer Osirix (MD). After identification of the MDOR in the 3D reconstruction, we calculated and printed a virtual biopsy plan for drilling, so as to minimize destruction of the important museum specimen. In total we collected 1.0 g of bone by drilling.

Radiocarbon dating

Sample SNMB N51667 was radiocarbon dated at the Klaus-Tschira-AMS facility in the Curt-Engelhorn-Centre of Archaeometry in Mannheim, Germany (MAMS; laboratory number GMP 391). All other samples were radiocarbon dated at the Keck AMS laboratory at the University of California Irvine (KCCAMS). The methods were similar for all samples. For ^{14}C and stable isotope determinations we followed methods described in Shammas *et al*. (*32*). Briefly, aliquots of ~200 mg of cortical bone were crushed to mm-sized chips and decalcified overnight with 1N HCl at room temperature, using a measured amount of acid calculated as just sufficient to dissolve all of the bone mineral if no collagen was present. The samples were then washed with ultrapure MQ water, gelatinized at pH 2 at 60°C overnight, and ultrafiltered with precleaned Vivaspin 15 Turbo ultrafiltration devices to select the >30 kDa molecular weight fraction, which was frozen and lyophilized overnight in a vacuum centrifuge.

2 mg aliquots of the lyophilized collagen for ^{14}C dating were transferred to 6 mm quartz combustion tubes. CuO oxidizer and silver wire getter were added and the tubes were sealed under vacuum and combusted at 900°C. CO_2 graphitization and ^{14}C measurement by AMS were carried out at the Keck AMS laboratory at the University of California Irvine. Aliquots containing 0.7 mg of collagen were wrapped in 5 x 9 mm tin foil cups and elemental analyses and $\delta\ ^{13}C$ and $\delta\ ^{15}N$ measurements were carried out using a Fisons NA1500NC elemental analyzer interfaced to a Finnigan Delta Plus isotope ratio mass spectrometer.

Two samples and a ^{14}C-free blank were submitted as bone powder that had been partially pretreated for DNA studies with one or two 24-hour incubations with EDTA and proteinase K. These samples were washed 4 times with MQ water with vortexing, soaked overnight in MQ water, and washed 3 more times with MQ water, to remove residual EDTA and proteinase prior to the normal processing. Following the gelatinization step, we observed that some residual bone powder was present for one sample (SC16.JK045) and the blank, that had received one 24-hour incubation rather than two. Aliquots of this residual material were run as new samples using the protocol outlined above, and both returned ages were consistent with the results from the initial gelatinization.

Data processing

Base calling was done with Bustard (Illumina). We trimmed the adapters and merged overlapping paired-end reads into single sequences using leeHom (*43*). We aligned the reads to the *de novo* assembled dugong (SI2) and the manatee genomes (UCSC GCA_000243295.1 (*9*)) using the Burrows–Wheeler aligner (BWA, version 0.5.10-evan.9–1-g44db244) (*45*) with parameters adjusted for ancient DNA (-n 0.01 –o 2 –l 16500) (*46*). Reads were also aligned to the *H. gigas* mitochondrial genome (*8*). We removed PCR duplicates with bam-rmdup (version 0.6.3) and filtered fragments for read length ($\geq$ 32 bp) and mapping quality (MQ $\geq$ 30) using SAMtools (version 1.3.1) (*44*). We used the ancient DNA genotyping software snpAD (*47*) on scaffolds $\geq$ 100 kbp to infer per scaffold polymorphic positions in the twelve sequenced individuals.

For details on specimens used, data processing, and dating refer to Supplementary Table S1.

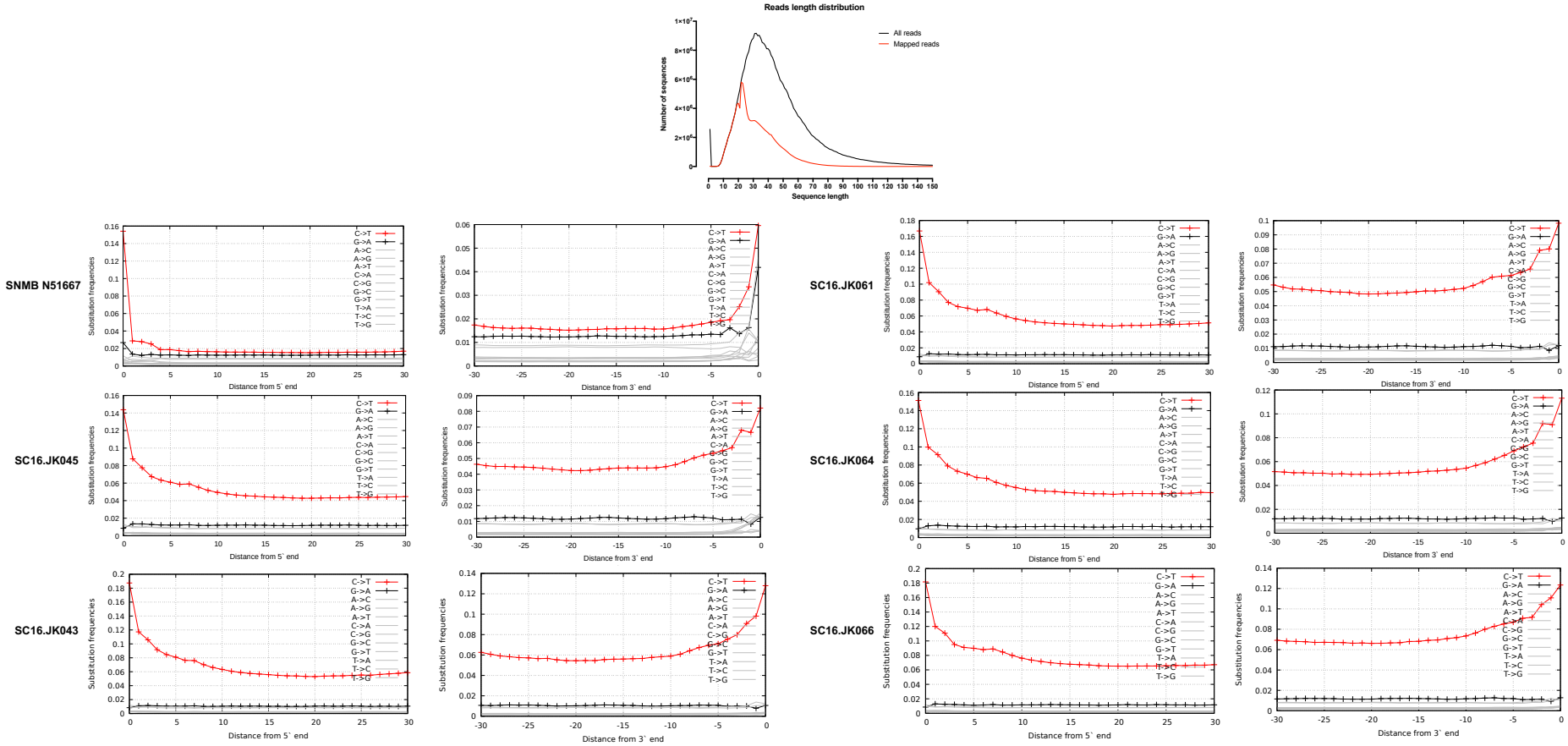
Authentication of ancient DNA

A characteristic of ancient DNA is that cytosines (C) located in proximity of DNA fragment ends undergo deamination to uracils. Thus, in the process of library preparation, DNA polymerases incorporate thymine (T) instead of C. We evaluated the C to T substitutions to authenticate the presence of ancient DNA in our sequences. The read-length distribution and substitution rates at the 5' and 3' ends in our data set (Supplementary Fig. S1) are in accordance with high fragmentation and modification of ancient DNA, respectively.

Supplementary Table S1.

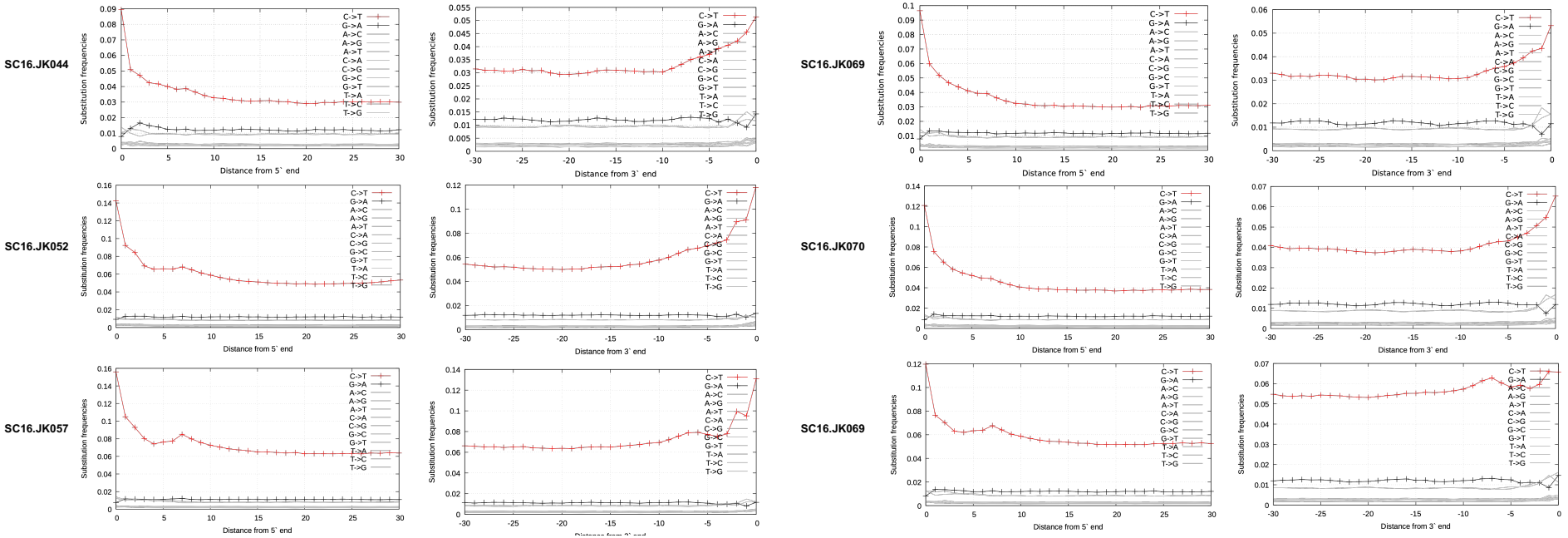
List of used specimens (museum catalog number), sequencing stats, and radiocarbon dating. MQ30 = mapping quality >30, UDG= uracil-DNA glycosylase, UCSC=University of California Santa Cruz.

ID	Location Processed	Library Prep Method	UDG Treated Extract	No. Libraries	Raw Reads	Reads Passing Filters (Merged, Mapped, Length > 32, MQ > 30)	% Reads Mapping	Average Duplicates	Average Read Length MQ30 (bp)	Average Fold Coverage	^{14}C uncalibrated date (years BP)
SC16.JK043	UCSC	Kapp, 2021	N	8	426,518,193	151,963,285	41.31	1.25	58	2.28	1,380±20
SC16.JK044	UCSC	Kapp, 2021	N	6	618,500,715	112,533,135	21.19	1.23	67	1.97	1,250±15
SC16.JK052	UCSC	Kapp, 2021	N	6	501,237,072	115,212,168	27.88	1.26	73	2.14	1,245±15
SC16.JK057	UCSC	Kapp, 2021	N	6	478,650,735	120,501,949	26.37	1.27	75	2.28	1300±15
SC16.JK061	UCSC	Kapp, 2021	N	8	664,459,501	243,889,647	42.89	1.26	61	3.78	N/A
SC16.JK064	UCSC	Kapp, 2021	N	8	727,872,124	188,646,115	29.79	1.27	71	3.38	1,565±15
SC16.JK066	UCSC	Kapp, 2021	N	10	770,625,574	228,951,014	33.67	1.26	68	3.95	2,205±15
SC16.JK069	UCSC	Kapp, 2021	N	6	641,996,782	155,730,543	28.89	1.24	65	2.58	1,155±20
SC16.JK070	UCSC	Kapp, 2021	N	6	599,236,387	170,509,149	33.76	1.26	62	2.69	1,250±15
SC16.JK076	UCSC	Kapp, 2021	N	4	526,920,529	170,200,152	36.59	1.26	64	2.77	1,300±15
SNMB N51667	University of Potsdam	Gansauge & Meyer, 2013	Y	3	10,896,135,224	1,152,095,440	15.4	1.07	46	15.86	1,267±19
SC16.JK045	UCSC	Kapp, 2021	N	12	2,613,737,582	1,007,518,366	48.45	1.21	59	15.63	1,260±15/1,255±15/1,270±20



Supplementary Fig. S1.

Analyses of genomic sequences from the extinct Steller's sea cow individuals. Read-length distribution for sample SNMB N51667 of all reads (black curve) and reads mapped to the dugong genome without a mapping quality threshold (red) show a high fragmentation rate. The substitution patterns from the 5' and 3' ends of the sequenced reads indicate that the mapped reads are of ancient origin. Sample SNMB N51667 was treated with uracil-DNA glycosylase and endonuclease VIII prior to library preparation, effectively reducing DNA damage as shown by the lower C>T substitution frequencies compared to the other samples.



Supplementary Fig. S1 – continued.

Analyses of genomic sequences from the extinct Steller's sea cow individuals. The substitution patterns from the 5' and 3' ends of the sequenced reads indicate that the mapped reads are of ancient origin.

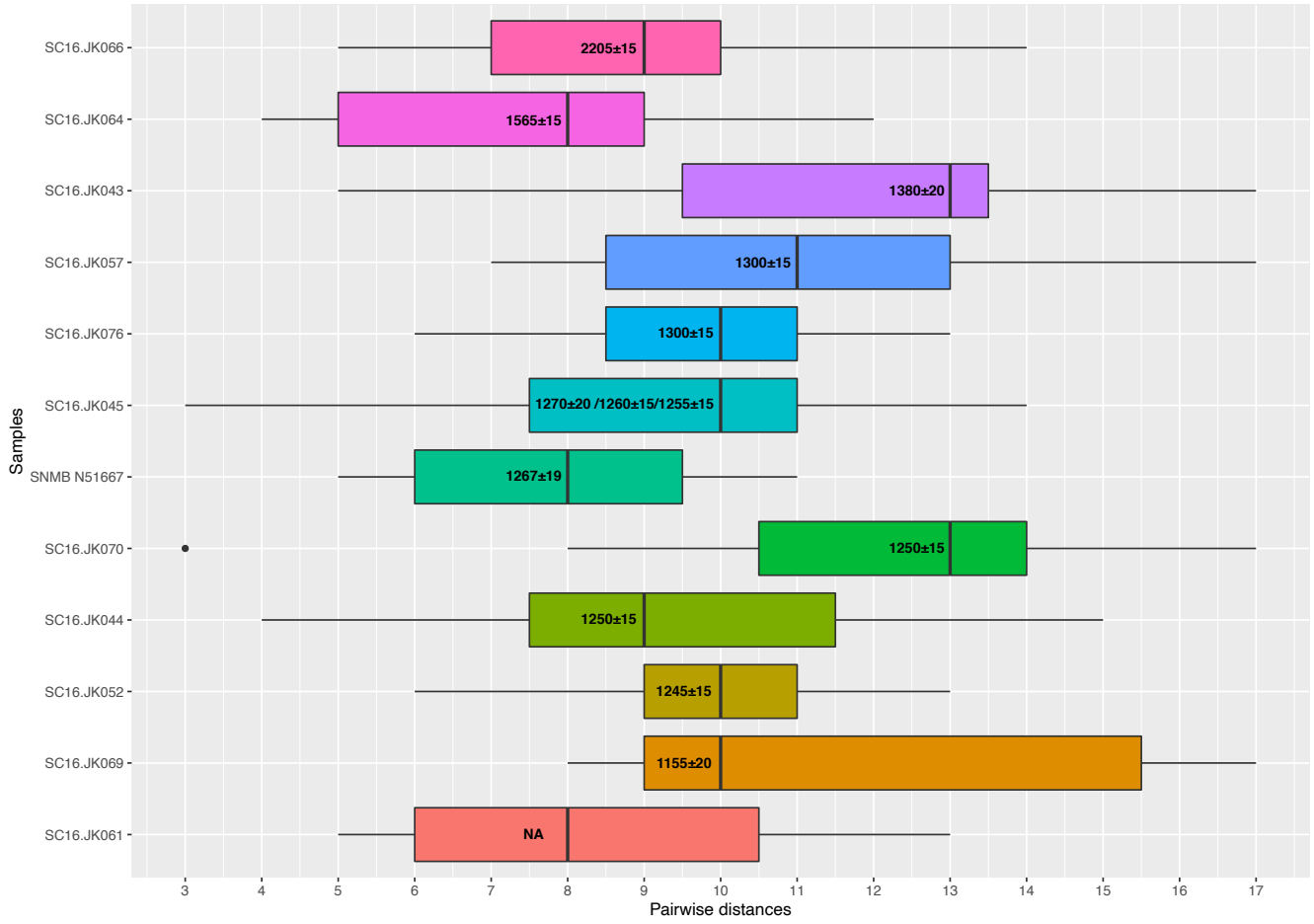
Pairwise distances between sequenced individuals based on mitochondrial genomes

To test whether the 12 sequenced fossils originated from different individuals, we estimated pairwise differences based on their mitochondrial genome sequences. For each sample we called genotypes using samtools mpileup (*44*) on reads mapped to the Steller's sea cow mitochondrial genome (*8*). In the further analyses we included positions with MQ ≥ 30, GQ ≥ 20, and coverage ≥ 10 in all samples. For coverage and read depth statistics refer to Supplementary Table S2. Finally, we created a pairwise distance matrix using the number of polymorphisms between samples. Based on pairwise distances we concluded that samples originated from different individuals (Supplementary Fig. S2).

Supplementary Table S2.

Coverage of the mitochondrial genomes based on reads with mapping quality ≥ 30, genotype quality ≥ 20, and coverage ≥ 10.

Sample ID	Mean depth	No. covered bases	Coverage of mitochondrial genome (%)
SC16.JK045	2,994	16,306	95.5
SNMB N51667	1,349	16,165	94.7
SC16.JK043	886	16,156	94.6
SC16.JK044	236	15,833	92.7
SC16.JK052	863	16,117	94.4
SC16.JK057	1127	16,173	94.7
SC16.JK061	616	16,043	93.9
SC16.JK064	1,037	16,119	94.4
SC16.JK066	2,007	16,220	95.0
SC16.JK069	276	15,859	92.9
SC16.JK070	221	16,014	93.8
SC16.JK076	444	15,951	93.4



Supplementary Fig. S2.

Box plots of the pairwise distances per sample. Numbers on the bars represent the ^{14}C uncalibrated date (years BP). The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points plotted individually.

SM2: *Dugong dugon* genome

De novo assembly

The final *Dugong dugon* assembly is 3,196.6 Mbp in length with a scaffold N50 contiguity of 3110.4 kbp (Supplementary Table S3). The BUSCO (*39*) assessment using 303 eukaryotic orthologs (not shown) found that the additional ARKS+LINKS scaffolding did not disturb the gene completeness of the assembly, but rather decreased the missing gene count and also the duplicate count by a factor of 2.

The BUSCO evaluation of mammalian single-copy orthologs shows the final assembly to be highly gene complete, with few haplotig artefacts: C:93.0% [S:92.1%, D:0.9%], F:4.2%, M:2.8%, n:4104.

Supplementary Table S3.

Assembly metrics of the dugong genome. bp = base pairs.

Total number of scaffolds	16,045
Total number of bases (Mbp)	3,196.9
N50 (Mbp)	3.11
Number of scaffolds > 1 kbp – 10 kbp	16,063 (100 %)
Number of scaffolds >10 kbp – 100 kbp	4,779 (29.7 %)
Number of scaffolds > 100 kbp – 1 Mbp	2,114 (13.2 %)
Number of scaffolds > 1 Mbp	832 (5.2 %)
Total number of contigs	40,386
Number of contigs in scaffolds	27,820
Number of contigs not in scaffolds	12,566
Percentage of assembly in scaffolded contigs	96.8 %
Percentage of assembly in un-scaffolded contigs	3.2 %

Annotation of the dugong and manatee genomes

We used MAKER (*10*) to annotate the dugong (*Dugong dugon*) and manatee (*Trichechus manatus latirostris* (*9*); UCSC GCA_000243295.1) genomes. We included several sources of evidence in the annotation process: *de novo* gene prediction using SNAP (*52*) and protein sequences of three high-quality genomes (human – GRCh38.p13, mouse – GRCm38.p6, and elephant – Loxafr3.0) downloaded from Ensembl v98 (*11*). We used RepeatMasker v4.1.0 (*53*) to identify repeats that matched to entries in standard databases for known repetitive sequences. To increase the likelihood of identifying coding regions, we also used known proteins to identify homologous sequences in our assembly. BLASTX (*54*) was used to align protein data against the raw genomic sequence; if the alignment occurred within the same coordinates with the *de novo* prediction, we considered the region homologous. We applied a threshold of 300 bp for the *ab initio* predicted genes; with additional evidence informed gene prediction, we obtained consensus gene sets of 36,749 and 25,192 genes for dugong and manatee, respectively.

Orthology assignment

Starting from human annotations, orthologs between human, mouse, and elephant were downloaded from Ensembl v98 (*11*). Dugong genes for which the MAKER annotation ortholog was in accordance with the triplet downloaded from Ensembl were included in further analysis. Orthologs transitivity is a major challenge in accurately assigning ortholog groups among multiple species. We also considered only genes that agreed with the orthologous triplets for the manatee. Predicted coding sequences of manatee and dugong were aligned with TBLASTX (*54*) and we filtered only the reciprocal best hits with an e-value $\leq 10^{-10}$. We obtained 4,877 1:1 orthologous genes in the five species (please find here Supplementary Table S4).

SM3: Comparative genomic analyses

We then compared the Steller's sea cow genome to the genomes of extant sirenians – dugong and manatee – and three additional well-characterized genomes: human, mouse, and elephant.

Inference of the Steller's sea cow coding sequence

To identify changes that occurred on the Steller's sea cow lineage, we called variants on reads aligned to the dugong genome using snpAD (*47*). Based on the set of 4,877 1:1 orthologous genes in dugong, manatee, human, mouse, and elephant, we identified the coordinates of the coding sequences in dugong. We lifted those coordinates from the variant called files in the genome of SNMB N51667, since this library showed the lowest level of damage (Supplementary Fig. S1) and, thus, posed the lowest risk of introducing erroneous mutations. We focused only on positions with coverage $\geq 5\times$ and GQ ≥ 20; all other positions were deemed as unresolved and replaced by "N"s (Supplementary Table S4). Variants in manually curated genes were inquired in all individuals.

The six-species orthologs were subjected to codon-based multiple alignments using MACSE v2 (*48*). Regions containing gaps in any of the species were excluded.

Inactivated genes

As pseudogenization is one means of shaping phenotypes, we screened the orthologs for premature stop codons. We manually curated the identified genes by inspecting the multiple sequence alignments for the presence of stop codons in all sequenced individuals. For this purpose, we did not filter the called variants for any quality criteria in the low-coverage individuals and used the number of supporting reads to calculate the probability of error based on the ancient DNA deamination pattern. We identified 10 genes with premature stop codons on the Steller's sea cow lineage (Supplementary Table S5).

For variants that implied C>T or G>A changes, we visualized the reads to exclude that the change occurred at the fragment ends. Furthermore, we calculated the probability that the variant is an error (Supplementary Table S6) by considering the number of reads and the highest error frequency at read ends, i.e. 0.2 (Supplementary Fig. S1). We used Bayes' theorem for each individual as follows:

$P(variant|error = P(error|variant * P(variant)\frac{}{P(error)}$, where $P(error|variant = 0.2$, $P(variant) = 1 - 0.25^{no.of reads with variant}$, and $P(error) = 0.25$). The posterior probability of error was calculated per individual as 1-$P(error) = 0.25$) and the final probability of error was the product of all individual error probabilities (Supplementary Table S6).

For gene ontology enrichment analyses of inactivated genes along the lineage to Steller's sea cow, we downloaded the Gene Ontology (GO) annotation from Ensembl v98 (*11*) and tested for enrichment using the FUNC package (*55*). We ran a hypergeometric test in which the inactivated genes were the set of interest and the rest of the 4,877 annotated 1:1 orthologs were the background. The analysis identified GO categories related to skin function (Supplementary Table S7), which include two arachidonate lipoxygenases (*ALOXE3* and *ALOX12B*).

Supplementary Table S5.

Genes with premature stop codons in Steller's sea cow. The location of the stop codon is given based on the human canonical transcript and the allele frequency in the human population is given based on the genome aggregation database (https://gnomad.broadinstitute.org). * for *CDH24* two bases in the same codon are different in Steller's sea cow compared to the human transcript.

Gene name	Gene description	Premature stop codon (corresponding to human transcript)	Protein length aa (human)	Variation in human population AF (GnomAD)
ALOX12B	arachidonate 12-lipoxygenase, 12R type	NM_001139.3:c.312G>A, p.(Trp104*)	701	0
ALOXE3	arachidonate lipoxygenase 3	NM_001165960.1:c.1578G>A, p.(Trp526*)	711	0
CDH24*	cadherin 24	NM_022478.3:c.382_384delCGGinsTGA, p.(Arg128*)	819	0
EHHADH	enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase	NM_001966.3:c.1674C>A, p.(Tyr558*)	723	0
KRT36	keratin 36	NM_003771.4:c.376C>T, p.(Gln126*)	467	0
NPFFR2	neuropeptide FF receptor 2	NM_004885.2:c.1222G>T, p.(Glu408*)	522	0
PADI2	peptidyl arginine deiminase 2	NM_007365.2:c.1573C>T, p.(Arg525*)	665	7.9e-6
SLC27A6	solute carrier family 27 member 6	NM_014031.3:c.1431G>A, p.(Trp477*)	619	3.9e06
SLPI	secretory leukocyte peptidase inhibitor	NM_003064.3:c.181A>T, p.(Lys61*)	132	0
SMCO2	single-pass membrane protein with coiled-coil domains 2	NM_001145010.1:c.175A>T, p.(Arg59*)	343	0

Supplementary Table S6.

Genes with premature stop codons in Steller's sea cow. For each individual we show the number of reads supporting the stop codon. All positions are homozygous in the sequenced individuals. For C>T or G>A changes we calculated the probability that the change is an error based on the total number of reads from all individuals and the highest error frequency observed at read ends, i.e. 0.2.

Gene	Position	Change	SC16.JK043	SC16.JK044	SC16.JK045	SC16.JK052	SC16.JK057	SC16.JK061	SC16.JK064	SC16.JK066	SC16.JK069	SC16.JK070	SC16.JK076	SNMBN51667	Probability of error
ALOXE3	Scaffold1250:637,303	G → A	4	3	26	2	6	1	2	8	4	1	5	12	2.81955e-08
ALOX12B	Scaffold1250:661,569	G → A	1	5	26	4	9	9	7	8	3	8	3	16	9.43293e-09
NPFFR2	Scaffold41:5,660,445	T → A	3	3	16	0	1	2	1	2	2	4	2	15	
CDH24	scaffold1033:1,142,036	C → T	2	8	32	5	5	9	10	4	5	3	0	9	2.79527e-08
KRT36	scaffold1341:135,061	C → T	4	4	29	2	11	4	11	12	6	5	5	13	5.41105e-09
SMCO2	scaffold288:276,090	C → T	3	1	37	2	5	8	3	12	3	1	3	26	2.62039e-08
EHHADH	scaffold174:953,143	A → T	3	0	27	4	1	6	2	5	4	7	5	25	
PADI2	scaffold758:738,091	G → A	1	0	23	0	1	3	6	1	2	1	4	10	2.21217e-06
SLC27A6	scaffold35:3,237,356	A → T	5	3	18	0	2	3	1	1	3	1	2	19	
SLPI	scaffold956:639,339	T → A	7	3	35	6	3	6	6	10	2	5	2	33	

Supplementary Table S7.

Gene ontology categories, which show enrichment with genes inactivated in Steller's sea cow. Categories in bold are related to skin function. FDR=False discovery rate.

Node name	Node ID	No. genes in node	No. significant genes in node	FDR	Significant genes in node
oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	GO:0016701	10	2	0.016971	*ALOXE3*, *ALOX12B*
oxidoreductase activity, acting on single donors with incorporation of molecular	GO:0016702	9	2	0.020844	*ALOXE3*, *ALOX12B*

Node name	Node ID	No. genes in node	No. significant genes in node	FDR	Significant genes in node
oxygen, incorporation of two atoms of oxygen					
lyase activity	GO:0016829	69	3	0.013267	*ALOXE3*, *ALOX12B*, *EHHADH*
isomerase activity	GO:0016853	59	3	0.012792	*ALOXE3*, *ALOX12B*, *EHHADH*
long-chain fatty acid metabolic process	GO:0001676	31	3	0.011853	*ALOXE3*, *ALOX12B*, *SLC27A6*
fatty acid metabolic process	GO:0006631	123	4	0.017067	*ALOXE3*, *ALOX12B*, *EHHADH*, *SLC27A6*
lipoxygenase pathway	GO:0019372	8	2	0.02308	*ALOXE3*, *ALOX12B*
regulation of water loss via skin	GO:0033561	11	2	0.036389	*ALOXE3*, *ALOX12B*
long-chain fatty acid biosynthetic process	GO:0042759	10	2	0.03259	*ALOXE3*, *ALOX12B*
linoleic acid metabolic process	GO:0043651	6	2	0.01851	*ALOXE3*, *ALOX12B*
hepoxilin metabolic process	GO:0051121	3	2	0.007955	*ALOXE3*, *ALOX12B*
hepoxilin biosynthetic process	GO:0051122	3	2	0.007955	*ALOXE3*, *ALOX12B*
establishment of skin barrier	GO:0061436	8	2	0.02308	*ALOXE3*, *ALOX12B*

ALOX12B and ALOXE3

The premature stop codons in *ALOX12B* and *ALOXE3* were present in all sequenced individuals (Supplementary Table S6). To infer the allele frequency in the population we used the jeffreysci function from the ratesci package in R (*50*). This allows a Bayesian inference of the allele frequency interval using the non-informative Jeffreys prior for a binomial distribution. The estimated allele frequency interval is 0.90–1 with a posterior probability of >0.95.

Furthermore, we inspected the *ALOX12B* and *ALOXE3* read alignments and identified a one-base insertion in *ALOX12B*, which leads to the following frameshift in the corresponding human canonical transcript: NM_001139.3: c.1350dup, p.(Leu451Ala*fs**27). *ALOXE3* also showed signals of faster evolution in Steller's sea cow and cetaceans (Supplementary Table S8).

We therefore inspected the *ALOX12B* and *ALOXE3* sequences in all mammals, including cetaceans and pinnipeds, for which data are available. We used biomartr (*56*) to retrieve all mammalian coding sequences from refseq (accession date 13.04.2021), which we cross checked with the Ensembl v103 entries. Interestingly, in support of the results from Reisch and colleagues (*21*) we could also not find any *ALOX12B* for the available cetaceans, but the gene was intact in pinnipeds (Supplementary Fig. S3). Furthermore, we identified multiple inactivating mutations in *ALOXE3* in cetaceans (Supplementary Fig. S4), while this gene was also intact in the pinnipeds.

Reisch and colleagues suggested that the complete loss of *ALOX12B* gene in cetaceans, but not in other marine, fur-bearing mammals, could be related to their lack of hair (*21*). The same was true for *ALOXE3*, although the pseudogene was still present in the cetacean genomes (*21*). The different skin phenotype of whales and dolphins compared to humans (*16–18*, *57–59*) and mice (*14*, *19*) with *ALOX12B*- and *ALOXE3*-gene loss was considered a potential evidence for different mechanisms involved in epidermal differentiation in marine mammals compared to their terrestrial counterparts (*21*, *60*).

However, Steller described the skin of *H. gigas* to resemble "the bark of an old oak tree" and to be "extremely thick" (*29*), which appears to be in agreement with ichthyotic skin. We were thus prompted to understand what could prevent the formation of calluses especially in dolphins, which also lack the skin lipoxygenase genes (Supplementary Fig. S3 and S4). Cetaceans and particularly dolphins have a very high shedding rate of the stratum corneum of skin, which maintains a smooth surface and limits microbe colonization (*22*, *61*). The outermost cell layer of dolphins can exfoliate up to 12 times a day (*61*). This high turnover rate ensures a smooth body surface and probably helps increase swimming efficiency by reducing drag (*61*). A potential molecular explanation for this observation is the lower number of cell structures specialized for cell-to-cell adhesions, called desomosomes, in the superficial layer of cetacean skin (*22*). Desmosomes are assembled from desmogleins (DSG) and desmocollins (DSC); e.g. the human genome encodes three DSC (DSC1–DSC3) and four DSG (DSG1–DSG4) proteins (*23*). The strongest-binding desmosomes occur in the outermost layers of human skin (DSG1:DSC1 and DSG4:DSC1) (*23*). Yet, cetaceans bear inactivations of *DSC1* and *DSG4*, which prevents the formation of desmosomes and leads to a high exfoliation rate (*22*). Thus, even in the absence of skin lipoxygenase genes, cetaceans cannot develop a hyperkeratotic, scaling skin phenotype. However, sirenians, including Steller's sea cow appear to have functional desmosomes (multiple sequence alignment can be downloaded here) and hence, Steller's description might indeed have depicted an ichthyotic skin.

Our finding of convergent inactivation of skin lipoxygenase genes in cetaceans and Steller's sea cow has several possible implications:

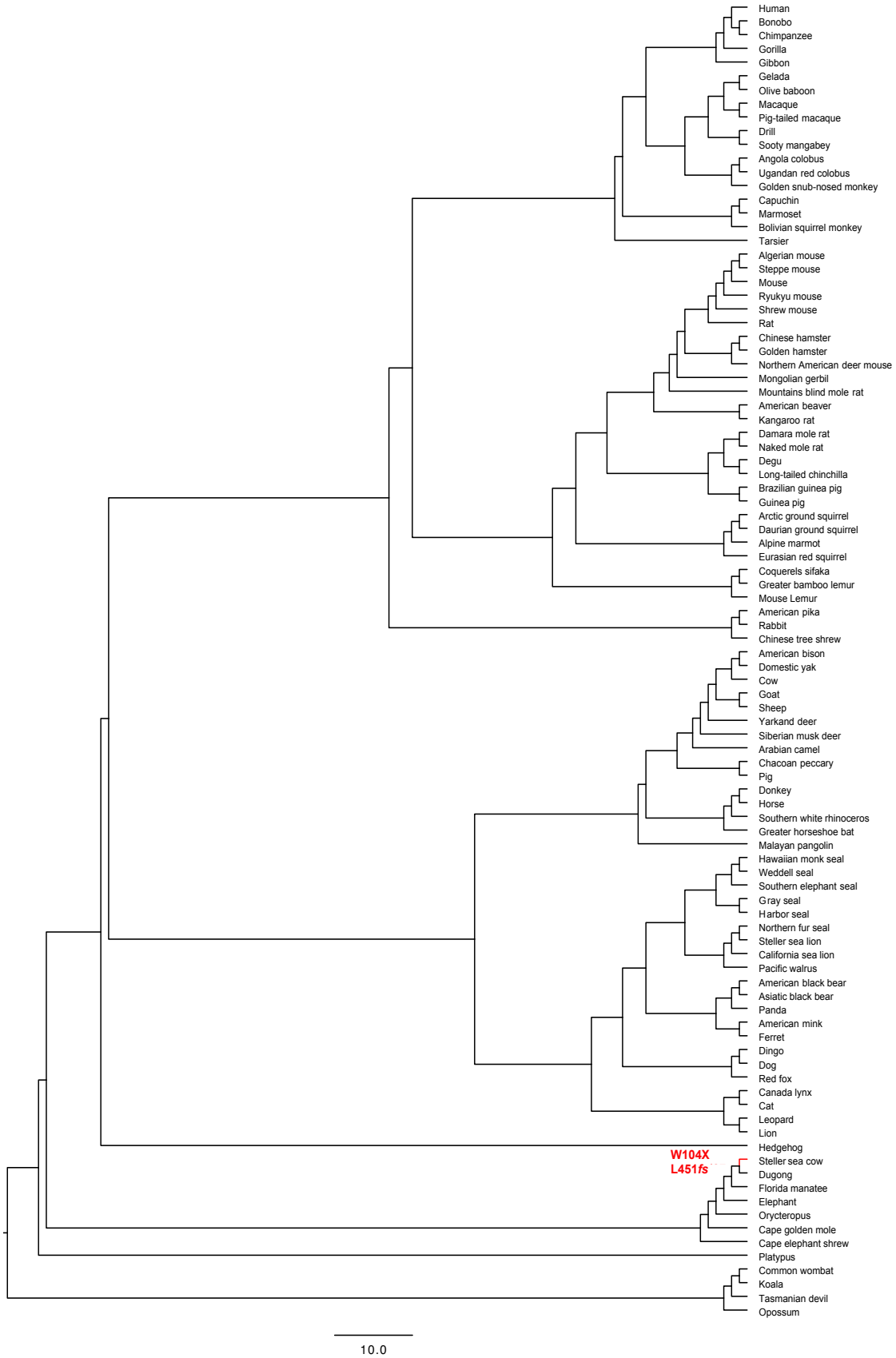
i. The inactivation of desmosome genes predated the *ALOXE3* inactivation and it was thus believed that *ALOXE3* loss is not adaptive, but rather its function is not needed as a result of adaptation through other mechanisms (*22*). However, the inactivation in Steller's sea cow in the presence of functional desmosomes challenges this hypothesis.

ii. *ALOX12B* and *ALOXE3* inactivations in cetaceans have been related to the skin innovations needed at the transition from land to water. The presence of a functional copy in pinnipeds initially suggested that these gene losses may be related to adaptation to a fully aquatic environment. However, we now show that extant sirenians, which are also fully aquatic, have a functional copy of both genes. The loss of the two genes in Steller's sea cow, the only cold adapted sirenian, suggests that these genes could play a role in adaptation to a cold marine environment.

iii. While *ALOX12B* appears to be completely lost from cetacean genomes, *ALOXE3* suffered independent inactivations on this lineage (*22*), as well as on the Steller's sea cow's branch (Supplementary Fig. S4). It is thus possible that the inactivation of *ALOX12B* preceded the one of *ALOXE3*; *ALOX12B* is located upstream of *ALOXE3* in the cascade pathway for epidermic structural lipid formation (*14*). Since this pathway

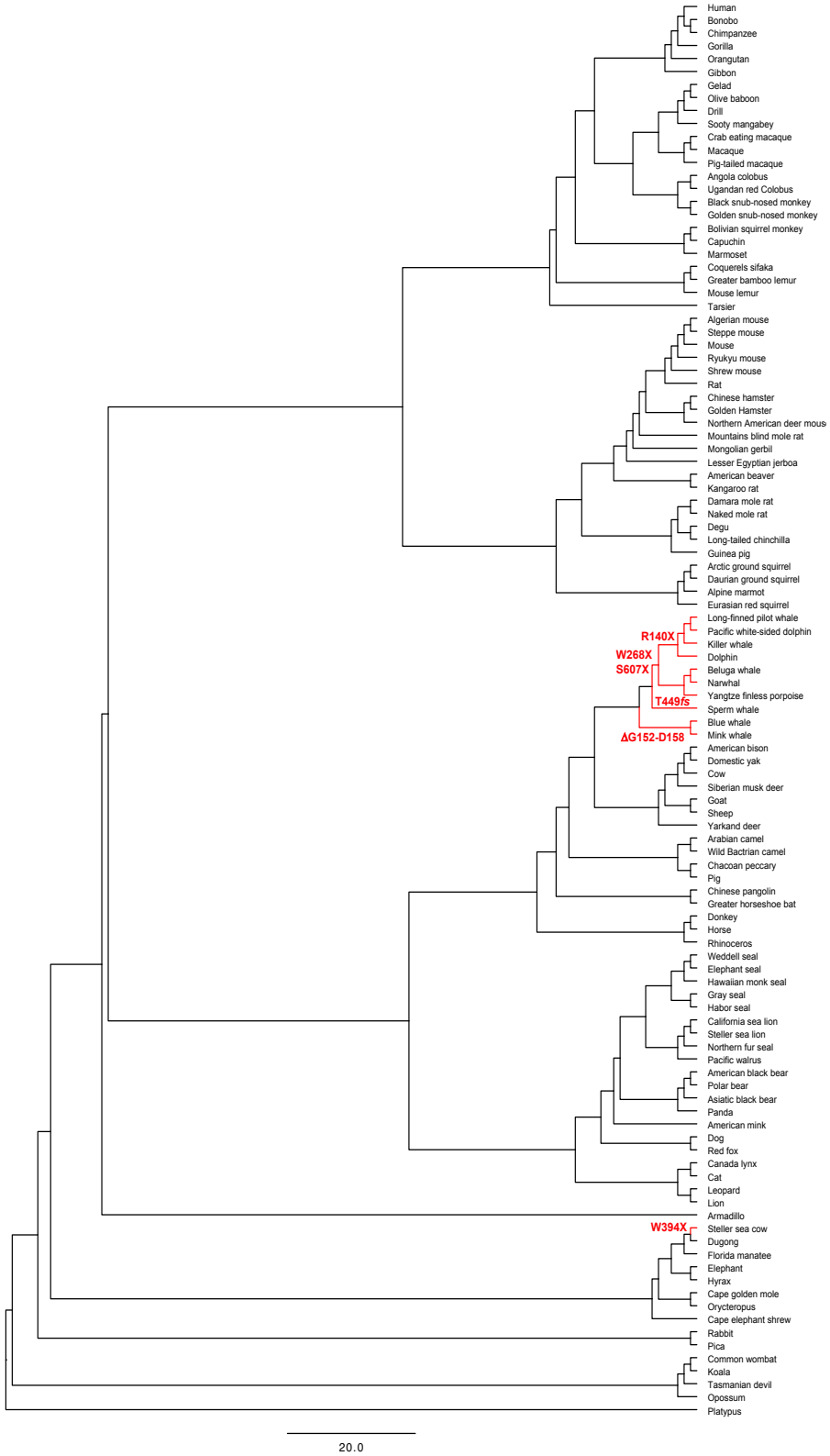
became non-functional, *ALOXE3* could be drifting, which may explain the multiple inactivations in the different cetacean lineages.

While we cannot clearly identify the functional relevance of these inactivations, their convergent molecular evolution suggests a role in the adaptation to cold aquatic life. Steller's sea cow genomic data will thus be an important resource for further delineating signals of parallel evolution and their importance for adaptation to their environment.



Supplementary Fig. S3.

Phylogeny of 97 species for which we inspected the coding sequence of *ALOX12B* (multiple sequence alignment can be downloaded here). We could not identify any sequence for cetaceans in the mammalian coding sequences we downloaded from refseq (accession date 13.04.2021) using biomartr (*56*, *62*, *63*) and cross checked with the Ensembl v103 entries. *ALOX12B* is inactivated in Steller's sea cow marked in red, while all other inquired species have a functional copy. The convergent inactivation in cetaceans and Steller's sea cow suggests a role in the adaptation to cold marine environments.



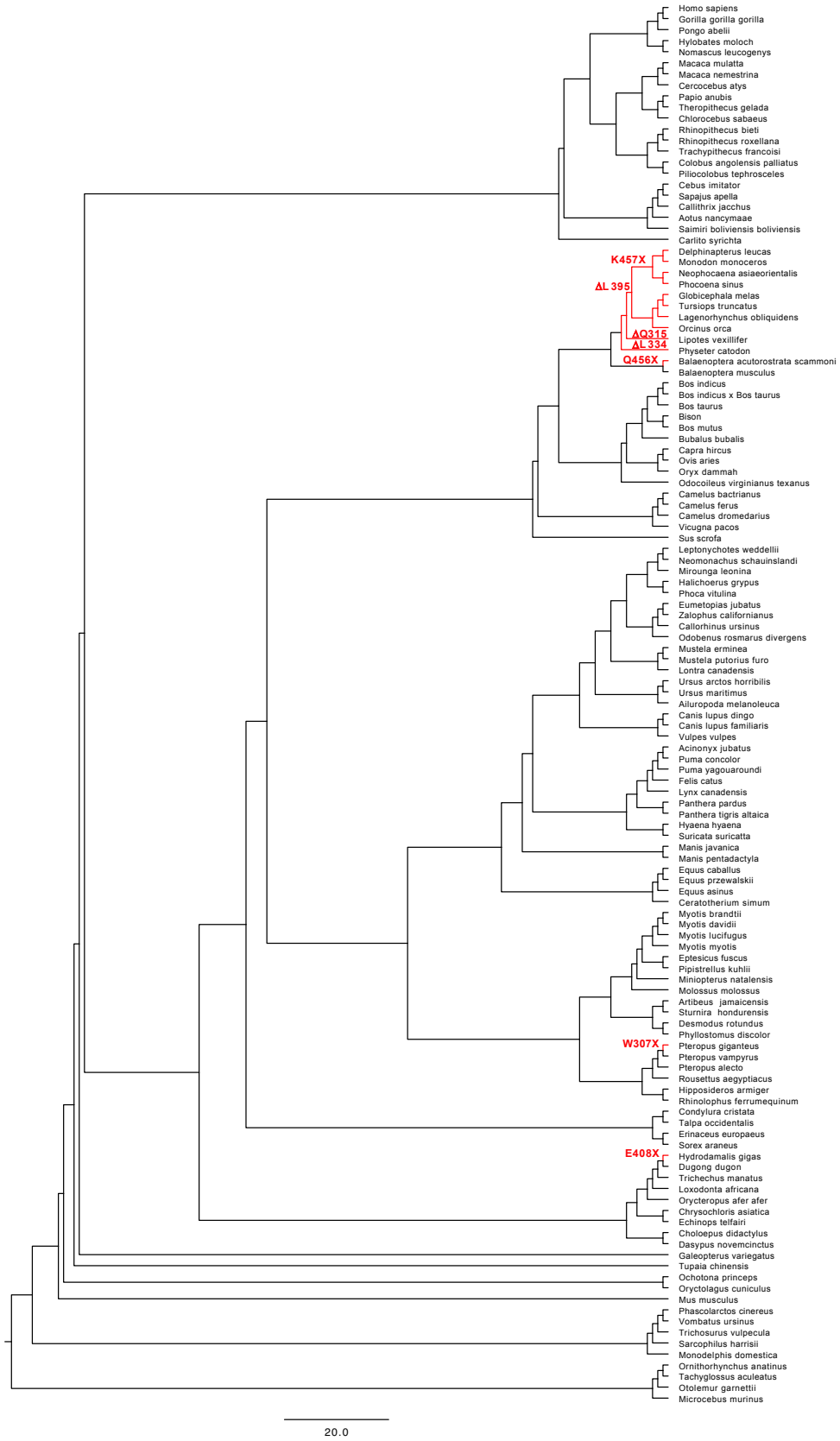
Supplementary Fig. S4.

Phylogeny of 109 species for which we inspected the coding sequence of *ALOXE3* (downloaded from refseq using biomartr (*56*, *62*, *63*) and cross checked with the Ensembl v103 entries – accession date 13.04.2021; multiple sequence alignment can be downloaded here). There are multiple independent inactivation events (in red) in cetaceans and in Steller's sea cow, showing parallel evolution and suggesting a role in the adaptation to cold marine environments.

NPFFR2

We further identified a premature stop codon in *NPFFR2* (scaffold41: 5660445 T>A), which was confirmed in 11 of 12 sequenced individuals (SC16.JK052 had zero reads at this position, Supplementary Table S6), but also in the assembly from Sharko and colleagues (*15*). Based on our 12 sequenced individuals, the estimated allele frequency is 0.91 (range 0.76–0.98 with a 0.95 confidence level).

To detect whether *NPFFR2* occurs specifically on Steller's sea cow branch, we inspected all mammalian coding sequences we could retrieve from refseq using biomartr (*56*) (accession date 13.04.2021; multiple sequence alignment can be downloaded here). Interestingly, we found that *NPFFR2* has also been inactivated multiple times independently in cetaceans, where only the blue whale (*Balaenoptera musculus*) appears to still have an intact ortholog (Supplementary Fig. S5). The single leucine deletions in the different whale and dolphin linages affect positions 100% conserved in all other mammals. Furthermore, L334 and L395 are part of the transmembrane alpha-helices 5 and 6, respectively. Their loss leads to a change in the helical periodicity and, therefore, to a nonfunctional rearrangement of the transmembrane helix bundle. *NPFFR2* is also pseudogenized in the Indian flying fox (*Pteropus giganteus*, Supplementary Fig. S5). Although from an evolutionary perspective we could not identify a common habitat adaptation with the Indian flying fox, the independent inactivation of *NPFFR2* in multiple cetaceans, as well as in Steller's sea cow, suggests that in an aquatic environment there may be a relaxed selection on this gene. However, the mouse *Npffr2* knockout phenotype (*24*) could imply that this gene loss may be beneficial for a capital-based survival strategy, as in species which undergo seasonal fasting, much like the Steller's sea cow.



Supplementary Fig. S5.

Phylogeny of 127 species for which we inspected the coding sequence of *NPFFR2* (downloaded from refseq using biomartr (*56*, *62*, *63*) – accession date 13.04.2021; multiple sequence alignment can be downloaded here). There are multiple independent inactivation events (in red) in cetaceans and in Steller's sea cow, as well as in the Indian flying fox (*Pteropus giganteus*). The independent inactivations may suggest a relaxed selection in an aquatic environment.

Selection analysis

To examine selective constraints along the lineage to the Steller's sea cow, we scanned for differently evolving genes with the CODEML program under a branch model (*49*) using previously curated orthologs in the six species (Steller's sea cow, dugong, manatee, human, mouse, and elephant). We performed likelihood ratio tests (LRTs) to compare evolutionary models on our phylogenetic tree. First, we estimated an average ω across the tree using model = 0. We used the one-ratio model (model = 0, NSsites = 0) to estimate the same ω ratio for all branches in the phylogeny. We then used the two-ratio model (model = 2, NSsites = 0), with a background ω ratio and a different ω on the Steller's sea cow lineage. These two models were compared via a LRT (1 degree of freedom) (*49*). We used a chi-square test to check whether the Steller's sea cow branch is significantly different and if the estimated ω is higher or lower than for the other species, which would imply that these genes are differently evolving (faster or slower). This yielded 197 genes with a faster (Supplementary Table S8) and 41 genes with a slower evolutionary rate (Supplementary Table S9) in the Steller's sea cow lineage compared to the other five species in the tree, which also included extant sirenians.

To identify genes, which could be relevant for aquatic adaptation we tested, using the setup described above and a phylogenetic tree including the same outgroup species (human, mouse, and elephant), 2 cetaceans (blue whale and bottlenose dolphin), and one pinniped (walrus), whether the identified genes show signals of selection on any of the aquatic-adapted branches.

Supplementary Table S8.

Genes with signs of faster evolution on the Steller's sea cow branch. *P*-value calculated with a chi-square test based on the likelihood ratio test between model 0 and model 2 in CODEML (*49*) (1 degree of freedom). Faster evolving genes on the Steller's sea cow lineage associated with energy homeostasis regulation and body weight are marked * (literature search via Pubmed, OMIM and Mouse Genome Informatics (http://www.informatics.jax.org)).

[1]faster evolving in cetaceans; [2]more slowly evolving in cetaceans; [3]faster evolving in pinnipeds; [4]more slowly evolving in pinnipeds.

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
AAMP	angio associated migratory cell protein	0.02	0.02	0.39	1.24e-04
ACP4[1]	acid phosphatase 4	0.17	0.16	1.51	7.84e-03
ACP6	acid phosphatase 6, lysophosphatidic	0.33	0.31	2.38	2.45e-02
ACP7[1]	acid phosphatase 7, tartrate resistant (putative)	0.11	0.1	0.65	7.33e-03
ACSF3*	acyl-CoA synthetase family member 3	0.18	0.17	2.31	2.45e-03
ACTBL2[1]	actin beta like 2	0.09	0.08	0.73	1.17e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
ADGRG7[3]	adhesion G protein-coupled receptor G7	0.5	0.48	2.03	4.50e-02
ALKBH1[1,4]	alkB homolog 1, histone H2A dioxygenase	0.22	0.21	1.95	1.65e-02
ALOXE3*[,1]	arachidonate lipoxygenase 3	0.12	0.11	1.1	5.67e-05
ALS2CL[3]	ALS2 C-terminal like	0.13	0.13	0.35	2.10e-02
AMPD2*	adenosine monophosphate deaminase 2	0.03	0.02	0.1	3.43e-02
APLP1	amyloid beta precursor like protein 1	0.2	0.19	1.08	1.97e-02
APTX	aprataxin	0.24	0.23	1.06	4.52e-02
ARHGAP22	Rho GTPase activating protein 22	0.13	0.12	0.71	2.90e-02
ASNSD1	asparagine synthetase domain containing 1	0.24	0.24	1.53	4.83e-02
ASPHD1	aspartate beta-hydroxylase domain containing 1	0.11	0.11	1.27	1.98e-02
BCAS2	BCAS2 pre-mRNA processing factor	0.07	0.04	0.95	7.79e-03
BCO1*	beta-carotene oxygenase 1	0.12	0.12	1.05	4.11e-02
BMPR1A*	bone morphogenetic protein receptor type 1A	0.01	0.01	0.37	4.07e-02
BNIPL[1]	BCL2 interacting protein like	0.19	0.18	0.93	1.71e-02
BVES	blood vessel epicardial substance	0.16	0.15	0.96	2.22e-02
C1orf116	chromosome 1 open reading frame 116	0.44	0.43	1.58	2.72e-02
CA7	carbonic anhydrase 7	0.1	0.09	0.49	2.23e-02
CACNG1	calcium voltage-gated channel auxiliary subunit gamma 1	0.11	0.1	0.94	3.47e-02
CCDC137[1]	coiled-coil domain containing 137	0.15	0.14	1.22	3.93e-02
CCT4[1]	chaperonin containing TCP1 subunit 4	0.04	0.04	0.24	2.52e-02
CCT7	chaperonin containing TCP1 subunit 7	0.05	0.04	0.25	3.70e-02
CDC23	cell division cycle 23	0.03	0.02	0.24	2.67e-03
CDH24[1,4]	cadherin 24	0.14	0.13	2.36	1.01e-02
CDR2[1,4]	cerebellar degeneration related protein 2	0.16	0.16	0.91	2.75e-02
CFAP52[1]	cilia and flagella associated protein 52	0.09	0.09	0.62	4.34e-02
CLDN4	claudin 4	0.07	0.06	0.52	4.47e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
CMTR2[1]	cap methyltransferase 2	0.19	0.18	1.4	3.83e-02
CNDP1	carnosine dipeptidase 1	0.16	0.16	0.66	4.72e-02
COG1[1]	component of oligomeric golgi complex 1	0.19	0.18	1.04	1.29e-03
COPB1	COPI coat complex subunit beta 1	0.02	0.02	0.24	1.79e-05
CREB3L1	cAMP responsive element binding protein 3 like 1	0.07	0.07	0.37	1.63e-02
CRY1*	cryptochrome circadian regulator 1	0.07	0.06	0.54	9.80e-03
CRYBA1[1]	crystallin beta A1	0.05	0.04	0.59	3.22e-02
CUTC	cutC copper transporter	0.11	0.1	1.23	8.48e-03
CYP2R1*	cytochrome P450 family 2 subfamily R member 1	0.19	0.18	1.99	8.29e-03
DAZAP2	DAZ associated protein 2	0.01	0	0.17	2.98e-02
DCLK3	doublecortin like kinase 3	0.19	0.18	0.63	3.88e-02
DANN2[1,3]	DANN replication helicase/nuclease 2	0.15	0.15	0.42	4.21e-02
DOLK	dolichol kinase	0.11	0.09	0.47	3.69e-03
EHD3	EH domain containing 3	0.05	0.03	2.69	1.93e-11
EHHADH*	enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase	0.28	0.26	0.79	1.63e-02
EIF3M	eukaryotic translation initiation factor 3 subunit M	0.01	0.01	0.73	7.46e-04
EIF4A1[4]	eukaryotic translation initiation factor 4A1	0	0	0.09	5.77e-03
EXOSC2	exosome component 2	0.06	0.05	1.17	4.14e-03
FAN1	FANCD2 and FANCI associated nuclease 1	0.4	0.39	3.06	1.13e-02
FRMD8	FERM domain containing 8	0.1	0.09	0.39	3.09e-02
FXR2	FMR1 autosomal homolog 2	0.03	0.02	0.27	2.65e-02
GABRD[1]	gamma-aminobutyric acid type A receptor subunit delta	0.03	0.03	0.26	1.90e-02
GANAB	glucosidase II alpha subunit	0.11	0.1	0.64	2.10e-03
GAS8[1]	growth arrest specific 8	0.03	0.03	2.16	7.46e-05
GINS3	GINS complex subunit 3	0.08	0.07	0.94	1.78e-02
GJA3	gap junction protein alpha 3	0.12	0.11	0.99	4.44e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
GMEB2	glucocorticoid modulatory element binding protein 2	0.03	0.03	0.19	3.27e-02
GMPPA	GDP-mannose pyrophosphorylase A	0.04	0.03	0.37	3.31e-02
GOLGA5	golgin A5	0.15	0.15	0.67	1.48e-02
GOLGA7[1]	golgin A7	0.03	0.03	4.07	3.66e-02
GOT2	glutamic-oxaloacetic transaminase 2	0.07	0.06	0.76	4.24e-02
GPR183	G protein-coupled receptor 183	0.11	0.1	1.66	2.21e-02
GPR31	G protein-coupled receptor 31	0.19	0.18	4.51	5.49e-03
GPR39*[*,1,3]	G protein-coupled receptor 39	0.12	0.12	1.24	1.51e-02
GPR75*	G protein-coupled receptor 75	0.16	0.15	1.24	3.42e-03
H1-5	H1.5 linker histone, cluster member	0.05	0.04	0.42	4.09e-02
HCFC2[4]	host cell factor C2	0.15	0.15	0.87	2.55e-02
HEXIM2[3]	HEXIM P-TEFb complex subunit 2	0.14	0.13	0.66	3.81e-02
HHIPL2	HHIP like 2	0.19	0.18	0.7	2.16e-02
HSPA9[1]	heat shock protein family A (Hsp70) member 9	0.09	0.08	1.04	2.69e-02
ING1	inhibitor of growth family member 1	0.03	0.03	1.06	3.74e-04
INTS11	integrator complex subunit 11	0.01	0.01	0.26	1.76e-03
ITGA3[1]	integrin subunit alpha 3	0.13	0.12	0.59	4.63e-03
IVD	isovaleryl-CoA dehydrogenase	0.1	0.09	3.44	1.08e-02
KANSL2[1]	KAT8 regulatory NSL complex subunit 2	0.06	0.06	0.72	4.29e-02
KATNBL1	katanin regulatory subunit B1 like 1	0.15	0.15	1.1	4.37e-02
KCNF1[1,3]	potassium voltage-gated channel modifier subfamily F member 1	0.02	0.02	0.16	1.68e-02
KIAA1324L[1]	KIAA1324 like	0.1	0.1	0.79	1.98e-02
KPNA2	karyopherin subunit alpha 2	0.18	0.15	1.48	3.51e-03
KRT14[1]	keratin 14	0.12	0.11	0.83	2.62e-02
KRT18[1]	keratin 18	0.1	0.1	0.86	4.55e-02
KRT32[4]	keratin 32	0.2	0.18	1.08	9.31e-04

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
LENG1[1]	leukocyte receptor cluster member 1	0.17	0.16	1.12	3.00e-02
LIX1[1]	limb and CNS expressed 1	0.05	0.04	0.68	2.48e-02
LMOD1[3]	leiomodin 1	0.16	0.15	1.49	3.39e-02
LPO	lactoperoxidase	0.23	0.22	1.62	3.10e-04
LRRC34[1]	leucine rich repeat containing 34	0.17	0.16	1.3	3.62e-02
LRRC8D	leucine rich repeat containing 8 VRAC subunit D	0.03	0.03	0.24	2.90e-03
MED15[2]	mediator complex subunit 15	0.12	0.11	0.43	2.13e-02
MEIOB	meiosis specific with OB-fold	0.21	0.2	1.2	3.68e-02
MFSD13A	major facilitator superfamily domain containing 13A	0.12	0.11	0.7	3.31e-02
MIOX[1]	myo-inositol oxygenase	0.08	0.07	1.16	4.86e-03
MKKS*	McKusick-Kaufman syndrome	0.34	0.33	1.45	4.37e-02
MMAB	metabolism of cobalamin associated B	0.18	0.17	2.34	2.71e-03
MRGPRF	MAS related GPR family member F	0.08	0.07	0.62	2.90e-02
MRGPRG	MAS related GPR family member G	0.14	0.12	1.03	7.21e-03
MRTO4[1]	MRT4 homolog, ribosome maturation factor	0.08	0.07	0.97	1.80e-02
NAP1L3	nucleosome assembly protein 1 like 3	0.2	0.19	2.38	2.30e-03
NDEL1[1,4]	nudE neurodevelopment protein 1 like 1	0.05	0.04	0.38	4.58e-02
NEUROD4[1]	neuronal differentiation 4	0.14	0.13	1.13	4.04e-02
NIPAL1	NIPA like domain containing 1	0.19	0.19	1.8	2.12e-02
NMD3[1]	NMD3 ribosome export adaptor	0.1	0.09	1.05	2.30e-02
NME9[1]	NME/NM23 family member 9	0.31	0.29	1.48	2.86e-02
NMT2[2]	N-myristoyltransferase 2	0.07	0.08	0.14	4.04e-02
NOC3L[1]	NOC3 like DNA replication regulator	0.12	0.12	0.89	1.03e-02
NSA2[4]	NSA2 ribosome biogenesis factor	0.06	0.05	0.94	1.25e-03
NSUN5	NOP2/Sun RNA methyltransferase 5	0.16	0.15	0.9	3.55e-02
OR13C9	olfactory receptor family 13 subfamily C member 9	0.26	0.22	1.71	4.65e-03

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
OR51F1	olfactory receptor family 51 subfamily F member 1 (gene/pseudogene)	0.23	0.21	1.27	4.35e-03
OR8B3	olfactory receptor family 8 subfamily B member 3	0.21	0.18	1.13	3.06e-03
OSBPL3[1]	oxysterol binding protein like 3	0.15	0.14	0.68	2.35e-02
OSTC	oligosaccharyltransferase complex non-catalytic subunit	0.2	0.17	1.56	4.60e-02
OTUD3[2]	OTU deubiquitinase 3	0.26	0.23	2.76	4.35e-03
OVGP1[3]	oviductal glycoprotein 1	0.38	0.36	1.57	4.19e-02
PANX3	pannexin 3	0.17	0.16	0.94	1.80e-02
PCOLCE2	procollagen C-endopeptidase enhancer 2	0.19	0.18	1.01	3.37e-02
PDILT	protein disulfide isomerase like, testis expressed	0.35	0.34	1.22	4.83e-02
PERP[4]	p53 apoptosis effector related to PMP22	0.08	0.07	0.96	8.52e-03
PHKG2	phosphorylase kinase catalytic subunit gamma 2	0.1	0.09	1.57	3.51e-03
PLA2G4F[1]	phospholipase A2 group IVF	0.24	0.23	0.68	3.59e-02
PLEKHA1	pleckstrin homology domain containing A1	0.09	0.08	0.57	4.34e-02
PLEKHG6[1]	pleckstrin homology and RhoGEF domain containing G6	0.24	0.23	0.81	3.05e-03
POLE3	DNA polymerase epsilon 3, accessory subunit	0.03	0.02	0.33	2.13e-02
PPP2R2D[3]	protein phosphatase 2 regulatory subunit Bdelta	0.04	0.03	0.32	3.90e-02
PRKAG1[4]	protein kinase AMP-activated non-catalytic subunit gamma 1	0.02	0.01	0.43	4.89e-03
PRPF38A	pre-mRNA processing factor 38A	0.1	0.08	0.73	1.53e-04
PSMB4	proteasome 20S subunit beta 4	0.09	0.08	1.13	1.42e-02
PSMD5[1]	proteasome 26S subunit, non-ATPase 5	0.14	0.14	1.29	3.49e-02
PTPN2*	protein tyrosine phosphatase non-receptor type 2	0.17	0.16	1.41	2.86e-02
PTX4	pentraxin 4	0.28	0.27	1.02	3.74e-02
RAD54B	RAD54 homolog B	0.24	0.23	0.94	2.62e-02
RAD54L[1]	RAD54 like	0.07	0.06	1.92	1.85e-04
RASA2[4]	RAS p21 protein activator 2	0.15	0.14	0.88	2.14e-02
RASD1*	ras related dexamethasone induced 1	0.01	0.01	0.31	1.09e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
RASGEF1C[1]	RasGEF domain family member 1C	0.05	0.05	0.4	2.83e-02
RDH5	retinol dehydrogenase 5	0.16	0.15	0.59	4.89e-02
RFX6	regulatory factor X6	0.12	0.12	1.12	3.29e-02
RIOK1	RIO kinase 1	0.26	0.25	1.8	3.05e-03
RPS13	ribosomal protein S13	0.11	0.09	1.16	3.55e-03
RPS3A	ribosomal protein S3A	0.11	0.1	1.09	4.69e-02
RRP12[1]	ribosomal RNA processing 12 homolog	0.09	0.09	0.29	3.64e-02
RSAD1	radical S-adenosyl methionine domain containing 1	0.13	0.12	1.23	1.76e-02
RUNDC1	RUN domain containing 1	0.12	0.11	0.78	2.49e-02
SART3	spliceosome associated factor 3, U4/U6 recycling protein	0.08	0.08	0.45	1.64e-02
SCYL1	SCY1 like pseudokinase 1	0.11	0.11	0.36	2.22e-02
SEC14L3[1,4]	SEC14 like lipid binding 3	0.07	0.06	0.78	8.78e-06
SERINC1	serine incorporator 1	0.06	0.06	1.33	2.79e-03
SERINC3	serine incorporator 3	0.3	0.29	1.28	4.19e-02
SH2D5	SH2 domain containing 5	0.11	0.11	14.02	9.43e-03
SLC16A14	solute carrier family 16 member 14	0.14	0.13	0.57	4.52e-02
SLC22A7	solute carrier family 22 member 7	0.27	0.26	0.78	2.93e-02
SLC24A3*	solute carrier family 24 member 3	0.03	0.01	0.48	1.25e-03
SLC2A2*[,1]	solute carrier family 2 member 2	0.31	0.29	0.68	9.43e-03
SLC30A2	solute carrier family 30 member 2	0.15	0.14	0.54	4.24e-02
SLITRK6[1]	SLIT and NTRK like family member 6	0.11	0.11	0.73	2.37e-02
SMG8	SMG8 nonsense mediated mRNA decay factor	0.06	0.06	0.32	4.55e-02
SMPDL3A[1]	sphingomyelin phosphodiesterase acid like 3A	0.2	0.19	1.48	3.88e-02
SMPDL3B	sphingomyelin phosphodiesterase acid like 3B	0.23	0.23	22.39	4.26e-02
SMYD4[1]	SET and MYND domain containing 4	0.36	0.35	1.59	2.85e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
SNX15	sorting nexin 15	0.12	0.11	1.34	1.47e-02
SRM[1]	spermidine synthase	0.1	0.08	3.86	1.96e-04
SRP68	signal recognition particle 68	0.05	0.04	0.35	2.04e-02
SSTR1	somatostatin receptor 1	0.01	0	0.18	2.42e-03
STMND1[1]	stathmin domain containing 1	0.38	0.36	1.53	4.66e-02
STRN4	striatin 4	0.04	0.04	0.19	2.88e-02
SYMPK[1,4]	symplekin	0.03	0.03	0.14	9.86e-03
SYT12	synaptotagmin 12	0.05	0.05	0.55	8.96e-03
TAAR5	trace amine associated receptor 5	0.15	0.13	2.12	1.05e-03
TAF6	TATA-box binding protein associated factor 6	0.02	0.02	0.22	2.59e-02
TCTE1[1]	t-complex-associated-testis-expressed 1	0.14	0.13	1.27	2.03e-02
TGS1[1]	trimethylguanosine synthase 1	0.37	0.36	1.53	1.12e-02
THOC3	THO complex 3	0.01	0	0.22	1.43e-03
TINAGL1	tubulointerstitial nephritis antigen like 1	0.1	0.09	0.54	3.83e-02
TM6SF1	transmembrane 6 superfamily member 1	0.11	0.11	0.8	1.91e-02
TMPRSS2	transmembrane serine protease 2	0.18	0.17	0.77	1.88e-02
TNFAIP1	TNF alpha induced protein 1	0.03	0.03	0.5	2.32e-02
TRMT5	tRNA methyltransferase 5	0.28	0.27	1.76	4.63e-02
TSHZ2[1]	teashirt zinc finger homeobox 2	0.08	0.07	0.38	3.42e-03
TTC38[1]	tetratricopeptide repeat domain 38	0.15	0.14	0.84	2.93e-02
TWF1	twinfilin actin binding protein 1	0.06	0.05	0.98	6.49e-03
TWISTNB	TWIST neighbor	0.24	0.23	1.21	3.02e-02
TYRP1[4]	tyrosinase related protein 1	0.19	0.19	0.95	4.42e-02
USP14	ubiquitin specific peptidase 14	0.05	0.05	1.03	9.22e-03
USP30[1]	ubiquitin specific peptidase 30	0.1	0.1	0.48	2.34e-02
USP33[1]	ubiquitin specific peptidase 33	0.07	0.07	0.25	4.37e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
USP5[1]	ubiquitin specific peptidase 5	0.03	0.03	0.3	1.69e-02
UTP15[4]	UTP15 small subunit processome component	0.12	0.12	1.09	4.00e-02
VPS16[1,4]	VPS16 core subunit of CORVET and HOPS complexes	0.05	0.05	0.44	7.50e-03
VPS29	VPS29 retromer complex component	0.02	0.01	1.03	1.59e-03
WRNIP1[1]	WRN helicase interacting protein 1	0.03	0.03	0.82	1.01e-02
XPNPEP1	X-prolyl aminopeptidase 1	0.08	0.07	0.45	4.78e-03
XRCC3[1]	X-ray repair cross complementing 3	0.15	0.15	2.4	7.79e-03
YTHDF1[1]	YTH N6-methyladenosine RNA binding protein 1	0.06	0.06	0.22	4.26e-02
ZNF395*	zinc finger protein 395	0.16	0.15	0.76	3.88e-02
ZNF641[2]	zinc finger protein 641	0.16	0.16	1.22	1.83e-02

Supplementary Table S9.

Genes with signs of slower evolution on the Steller's sea cow branch. *P*-value calculated with a chi-square test based on the likelihood ratio test between model 0 and model 2 in CODEML (*49*) (1 degree of freedom). More slowly evolving genes on the Steller's sea cow lineage associated with energy homeostasis regulation and body weight are marked with * (literature search via Pubmed, OMIM and Mouse Genome Informatics (http://www.informatics.jax.org)).

[1]faster evolving in cetaceans; [2]more slowly evolving in cetaceans; [3]faster evolving in pinnipeds; [4]more slowly evolving in pinnipeds.

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
ACSL5*[,1]	acyl-CoA synthetase long chain family member 5	0.29	0.29	0	1.47e-02
ANGPT4[2,4]	angiopoietin 4	0.21	0.21	0	4.09e-02
ATP12A	ATPase H+/K+ transporting non-gastric alpha2 subunit	0.25	0.26	0	2.12e-02
ATP1B3	ATPase Na+/K+ transporting subunit beta 3	0.34	0.35	0	3.37e-02
BCAM[3]	basal cell adhesion molecule (Lutheran blood group)	0.16	0.17	0	5.93e-03
CCDC115[4]	coiled-coil domain containing 115	0.23	0.24	0	4.11e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
CD79B	CD79b molecule	0.28	0.29	0	3.90e-02
CDH3[1,4]	cadherin 3	0.15	0.16	0	1.37e-02
CFAP157	cilia and flagella associated protein 157	0.21	0.21	0	2.45e-02
CHGB	chromogranin B	0.36	0.37	0	4.29e-02
CSNKA2IP	casein kinase 2 subunit alpha' interacting protein	0.49	0.51	0	7.41e-03
DBF4	DBF4 zinc finger	0.33	0.33	0	3.70e-02
DPP9	dipeptidyl peptidase 9	0.05	0.05	0	4.19e-02
FASTKD1	FAST kinase domains 1	0.42	0.43	0.06	2.96e-02
FASTKD2	FAST kinase domains 2	0.48	0.5	0.1	6.74e-03
GFRA3	GDNF family receptor alpha 3	0.22	0.23	0	1.26e-02
GP1BA[2]	glycoprotein Ib platelet subunit alpha	0.51	0.52	0.14	4.97e-02
IARS2	isoleucyl-tRNA synthetase 2, mitochondrial	0.14	0.15	0	3.97e-02
IL4R	interleukin 4 receptor	0.42	0.43	0.11	3.05e-02
IL5[1,4]	interleukin 5	0.59	0.63	0	2.58e-02
IRF5	interferon regulatory factor 5	0.1	0.1	0	1.54e-02
LRIF1	ligand dependent nuclear receptor interacting factor 1	0.64	0.65	0	3.59e-02
LURAP1	leucine rich adaptor protein 1	0.17	0.18	0	2.11e-02
MAEA	macrophage erythroblast attacher	0.01	0.01	0	1.69e-03
MUTYH[1]	mutY DANN glycosylase	0.3	0.3	0	6.10e-03
NEU3	neuraminidase 3	0.34	0.35	0	1.29e-02
NHLRC3	NHL repeat containing 3	0.2	0.21	0	3.81e-02
NPR1	natriuretic peptide receptor 1	0.06	0.06	0	4.31e-02
PEX11A*	peroxisomal biogenesis factor 11 alpha	0.31	0.32	0	3.33e-02
PLA2G2A*	phospholipase A2 group IIA	0.49	0.54	0	2.38e-02
PXMP4[1]	peroxisomal membrane protein 4	0.15	0.16	0	2.07e-04
SLFN14[4]	schlafen family member 14	0.38	0.38	0	4.14e-02

Gene name	Gene description	ω model0	ω background	ω foreground branch	*p*-value
SMPD1*	sphingomyelin phosphodiesterase 1	0.17	0.17	0	3.68e-02
TCTEX1D2	Tctex1 domain containing 2	0.2	0.21	0	4.47e-02
THSD1	thrombospondin type 1 domain containing 1	0.2	0.2	0	2.12e-02
TMEM130[2,4]	transmembrane protein 130	0.19	0.2	0	1.73e-02
TTC16	tetratricopeptide repeat domain 16	0.31	0.31	0.08	2.38e-02
TUBB1[3]	tubulin beta 1 class VI	0.04	0.04	0	4.09e-02
TXNDC17[1,4]	thioredoxin domain containing 17	0.19	0.21	0	4.47e-02
UGT1A1[1]	UDP glucuronosyltransferase family 1 member A1	0.25	0.25	0	1.66e-02
VRTN	vertebrae development associated	0.14	0.14	0.02	3.47e-02

Pathway enrichment analysis

Twenty out of 238 genes that we identified to have significantly different evolutionary rates in Steller's sea cow have also been shown in other studies to modulate body weight and energy homeostasis (marked with * in Supplementary Tables S8 and S9). For example, *BCO1* (faster evolving in Steller's sea cow) is a β-carotenoid-15,15'-oxygenase that generates vitamin A from β-carotene. Interestingly, *BCO1*-deficient mice were more susceptible to high fat diet-induced impairments in fatty acid metabolism, an indirect effect probably related to hypercarotenemia and vitamin A deficiency (*64*). The activity of adenosine monophosphate deaminase 2 (*AMPD2*) (more slowly evolving in Steller's sea cow) positively correlates with hepatic fat synthesis and fat oxidation rates (*65*). Members of the family of bone morphogenetic proteins (BMP) are important regulators of adipogenesis. Genetic variants of the BMP receptor 1A gene (*BMPR1A*) (faster evolving in Steller's sea cow) and its expression levels in adipose tissue have been associated with overweight and obesity (*66*). GPR39 (faster evolving in Steller's sea cow, cetaceans, and pinnipeds) is a constitutively active G protein-coupled receptor and variants have been associated with obesity. Gpr39 deficiency in mice causes increased fat accumulation on a high-fat diet, conceivably due to decreased energy expenditure and adipocyte lipolytic activity and altered diet-induced thermogenesis (*67*, *68*). Recently, GPR75 variants have been found to protect from obesity and mice deficient for GPR75 show resistance to weight gain when fed a high-caloric diet (*69*). The significantly high number of genes associated with energy homeostasis regulation and body weight (hypergeometric test p-value = 0.006) suggests that metabolic pathways involved in the switch of the energetic balance towards fat accumulation contributed to evolutionary adaptations of Steller's sea cow.

To further explore which functions may have experienced different evolutionary pressures along the lineage to Steller's sea cow, we tested GO enrichment using the FUNC package (*55*). We ran a hypergeometric test in which the significantly differently evolving genes were the set of interest. We considered the remaining genes from the 4,877 1:1 orthologs as the background genes.

Several categories with an uncorrected *p*-value < 0.05 fit well with the Steller's sea cow phenotype: e.g. long-chain fatty acid-CoA ligase activity – GO:0004467, low-density lipoprotein particle remodeling – GO:0034374, negative regulation of epidermis development – GO:0045683, regulation of melanin biosynthetic process – GO:0048021 or negative regulation of hair follicle maturation GO:0048817 (Supplementary Table S10).

It is necessary for mammals to maintain constant body temperature in cold environments. In addition to muscle shivering, non-shivering thermogenesis plays a major role in thermal homeostasis. There are two major forms of non-shivering thermogenesis (NST): uncoupling protein 1 (UCP1)-dependent and UCP1-independent. The UCP1-dependent thermogenesis relies on the ability of the brown adipose tissue (BAT) to produce heat based on uncoupling the electron transport from mitochondrial ATP synthesis. UCP1 is absent in several mammals, such as pigs, Steller's sea cow, mammoths, and elephants (*25*); these species must generate heat (if the environment does not provide sufficient temperature) by UCP1-independent mechanisms. There are several thermogenic futile cycles "wasting" ATP by hydrolysis to generate heat, among them the lipolysis/re-esterification cycle, which uses ATP to synthesize acylglycerol (*27*, *28*). Formation of acylglycerols requires CoA-activated fatty acids, which is the ATP-consuming process. *ACSF3* and *ACSL5*, both with significantly different evolutionary rates in Steller's sea cow, encode for acyl-CoA synthetases that activate long chain fatty acids (*72*). It is therefore likely that changes in their function influence this thermogenic futile cycle. Alternatively, fatty acids released from the massive fat depots of the Steller's sea cow could have been activated by both acyl-CoA synthetases and introduced into the β-oxidation to supply ATP, e.g. under fasting conditions. The gene encoding for enoyl-CoA-hydratase:3-hydroxyacyl-CoA dehydrogenase (encoded by *EHHADH*) is also among the faster evolving genes in Steller's sea cow (Supplementary Table S8), supporting this latter point. This enzyme is involved in peroxisomal oxidation of fatty acids. Similarly, the faster evolving isovaleryl-CoA dehydrogenase (encoded by *IVD*, Supplementary Table S8) is a short and medium chain acyl-CoA dehydrogenase also participating in acyl component degradation.

Thus, it seems that differentially evolving genes on the Steller's sea cow lineage are involved in lipid-related metabolic pathways, which could explain their thick blubber, up to 10 cm in some areas of the body (*4*).

Supplementary Table S10.

Gene ontology (GO) (*55*, *70*, *71*) categories enriched with genes with significantly different evolutionary rates on the Steller's sea cow lineage compared to dugong, manatee, human, mouse, and elephant. We considered the GO categories based on uncorrected *p*-values, but none of these categories remained significant after correction for multiple testing.

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
mismatch base pair DNA N-glycosylase activity	GO:0000700	4	1	0.03
purine-specific mismatch base pair DNA N-glycosylase activity	GO:0000701	2	1	0.02

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
oxidized base lesion DNA N-glycosylase activity	GO:0000702	1	1	0.01
peptide receptor activity	GO:0001653	43	2	0.05
aminoacyl-tRNA editing activity	GO:0002161	4	1	0.03
DNA-(apurinic or apyrimidinic site) lyase activity	GO:0003906	3	1	0.02
acyl-CoA ligase activity	GO:0003996	5	1	0.04
exo-alpha-sialidase activity	GO:0004308	1	1	0.01
guanylate cyclase activity	GO:0004383	4	1	0.03
long-chain fatty acid-CoA ligase activity	GO:0004467	5	1	0.04
endonuclease activity	GO:0004519	41	2	0.05
phospholipase activity	GO:0004620	43	2	0.05
sphingomyelin phosphodiesterase activity	GO:0004767	5	1	0.04
transposase activity	GO:0004803	1	1	0.01
isoleucine-tRNA ligase activity	GO:0004822	1	1	0.01
signal transducer activity	GO:0004871	342	7	0.02
transmembrane signaling receptor activity	GO:0004888	247	6	0.02
cytokine receptor activity	GO:0004896	37	2	0.04
interleukin-4 receptor activity	GO:0004913	1	1	0.01
laminin receptor activity	GO:0005055	1	1	0.01
interleukin-5 receptor binding	GO:0005137	1	1	0.01
sodium:potassium-exchanging ATPase activity	GO:0005391	4	2	0
axon guidance receptor activity	GO:0008046	2	1	0.02
G protein-coupled peptide receptor activity	GO:0008528	41	2	0.05
oxidized purine nucleobase lesion DNA N-glycosylase activity	GO:0008534	1	1	0.01
potassium-transporting ATPase activity	GO:0008556	4	2	0
hydrogen:potassium-exchanging ATPase activity	GO:0008900	1	1	0.01
cyclase activity	GO:0009975	6	1	0.05
glucuronosyltransferase activity	GO:0015020	2	1	0.02
thrombin-activated receptor activity	GO:0015057	2	1	0.02
potassium ion transmembrane transporter activity	GO:0015079	38	2	0.04
primary active transmembrane transporter activity	GO:0015399	32	2	0.03
P-P-bond-hydrolysis-driven transmembrane transporter activity	GO:0015405	32	2	0.03
ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	GO:0015662	12	2	0
glial cell-derived neurotrophic factor receptor activity	GO:0016167	1	1	0.01
CoA-ligase activity	GO:0016405	5	1	0.04
oxidoreductase activity, acting on a sulfur group of donors, NAD(P) as acceptor	GO:0016668	5	1	0.04

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
hydrolase activity, acting on glycosyl bonds	GO:0016798	46	3	0.01
hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	GO:0016820	28	2	0.02
phosphorus-oxygen lyase activity	GO:0016849	6	1	0.05
natriuretic peptide receptor activity	GO:0016941	2	1	0.02
alpha-sialidase activity	GO:0016997	1	1	0.01
DNA N-glycosylase activity	GO:0019104	4	1	0.03
kinase activator activity	GO:0019209	27	2	0.02
cation-transporting ATPase activity	GO:0019829	17	2	0.01
active ion transmembrane transporter activity	GO:0022853	33	2	0.03
protein kinase activator activity	GO:0030295	24	2	0.02
transmembrane receptor protein tyrosine kinase activator activity	GO:0030297	2	1	0.02
oxidized DNA binding	GO:0032356	3	1	0.02
oxidized purine DNA binding	GO:0032357	3	1	0.02
mismatch repair complex binding	GO:0032404	4	1	0.03
MutSalpha complex binding	GO:0032407	2	1	0.02
8-oxo-7,8-dihydroguanine DNA N-glycosylase activity	GO:0034039	1	1	0.01
adenine/guanine mispair binding	GO:0035485	1	1	0.01
signaling receptor activity	GO:0038023	308	6	0.04
ATPase coupled ion transmembrane transporter activity	GO:0042625	18	2	0.01
ATPase activity, coupled to transmembrane movement of substances	GO:0042626	28	2	0.02
ATPase activity, coupled to movement of substances	GO:0043492	28	2	0.02
dynein intermediate chain binding	GO:0045505	6	1	0.05
protein-disulfide reductase activity	GO:0047134	2	1	0.02
arachidonate-CoA ligase activity	GO:0047676	4	1	0.03
exo-alpha-(2->3)-sialidase activity	GO:0052794	1	1	0.01
exo-alpha-(2->6)-sialidase activity	GO:0052795	1	1	0.01
exo-alpha-(2->8)-sialidase activity	GO:0052796	1	1	0.01
acid sphingomyelin phosphodiesterase activity	GO:0061750	1	1	0.01
transmembrane receptor activity	GO:0099600	260	6	0.02
decanoate--CoA ligase activity	GO:0102391	4	1	0.03
immunoglobulin production	GO:0002377	38	2	0.04
regulation of immunoglobulin production	GO:0002637	24	2	0.02
positive regulation of immunoglobulin production	GO:0002639	17	2	0.01
positive regulation of production of molecular mediator of immune response	GO:0002702	32	2	0.03
negative regulation of T-helper 1 type immune response	GO:0002826	2	1	0.02

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
positive regulation of type 2 immune response	GO:0002830	6	1	0.05
uronic acid metabolic process	GO:0006063	4	1	0.03
base-excision repair, AP site formation	GO:0006285	3	1	0.02
transposition, DNA-mediated	GO:0006313	1	1	0.01
isoleucyl-tRNA aminoacylation	GO:0006428	1	1	0.01
protein phosphorylation	GO:0006468	575	9	0.04
glycolipid metabolic process	GO:0006664	42	2	0.05
ceramide metabolic process	GO:0006672	37	2	0.04
sphingomyelin catabolic process	GO:0006685	5	1	0.04
glycosphingolipid metabolic process	GO:0006687	30	2	0.03
ganglioside catabolic process	GO:0006689	1	1	0.01
porphyrin-containing compound catabolic process	GO:0006787	4	1	0.03
bilirubin conjugation	GO:0006789	1	1	0.01
cellular sodium ion homeostasis	GO:0006883	6	2	0
regulation of pH	GO:0006885	31	2	0.03
protein catabolic process in the vacuole	GO:0007039	5	1	0.04
lysosomal lumen acidification	GO:0007042	5	1	0.04
receptor guanylyl cyclase signaling pathway	GO:0007168	4	1	0.03
activation of transmembrane receptor protein tyrosine kinase activity	GO:0007171	3	1	0.02
oligosaccharide catabolic process	GO:0009313	4	1	0.03
flavonoid metabolic process	GO:0009812	3	1	0.02
potassium ion import	GO:0010107	14	2	0.01
establishment or maintenance of transmembrane electrochemical gradient	GO:0010248	6	2	0
positive regulation of nuclear cell cycle DNA replication	GO:0010571	4	1	0.03
positive regulation of macrophage derived foam cell differentiation	GO:0010744	4	1	0.03
regulation of cGMP-mediated signaling	GO:0010752	6	1	0.05
positive regulation of cGMP-mediated signaling	GO:0010753	4	1	0.03
positive regulation of sodium ion transport	GO:0010765	5	1	0.04
positive regulation of keratinocyte proliferation	GO:0010838	6	1	0.05
negative regulation of angiogenesis	GO:0016525	34	2	0.03
peroxisome membrane biogenesis	GO:0016557	1	1	0.01
peroxisome fission	GO:0016559	4	1	0.03
biphenyl metabolic process	GO:0018879	2	1	0.02
phenol-containing compound metabolic process	GO:0018958	27	2	0.02
glucuronate metabolic process	GO:0019585	4	1	0.03
developmental maturation	GO:0021700	72	3	0.02
termination of signal transduction	GO:0023021	2	1	0.02

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
cellular monovalent inorganic cation homeostasis	GO:0030004	32	3	0
cellular potassium ion homeostasis	GO:0030007	5	2	0
sphingolipid catabolic process	GO:0030149	14	2	0.01
eosinophil differentiation	GO:0030222	3	1	0.02
regulation of granulocyte differentiation	GO:0030852	6	1	0.05
positive regulation of granulocyte differentiation	GO:0030854	5	1	0.04
transposition	GO:0032196	5	1	0.04
response to peptidoglycan	GO:0032494	4	1	0.03
response to muramyl dipeptide	GO:0032495	6	1	0.05
regulation of monophenol monooxygenase activity	GO:0032771	1	1	0.01
positive regulation of monophenol monooxygenase activity	GO:0032773	1	1	0.01
transforming growth factor beta2 production	GO:0032906	1	1	0.01
regulation of transforming growth factor beta2 production	GO:0032909	1	1	0.01
negative regulation of transforming growth factor beta2 production	GO:0032912	1	1	0.01
positive regulation of mast cell activation involved in immune response	GO:0033008	6	1	0.05
tetrapyrrole catabolic process	GO:0033015	4	1	0.03
negative regulation of myeloid cell apoptotic process	GO:0033033	5	1	0.04
low-density lipoprotein particle remodeling	GO:0034374	5	1	0.04
sodium ion transmembrane transport	GO:0035725	31	2	0.03
interleukin-4-mediated signaling pathway	GO:0035771	3	1	0.02
positive regulation of urine volume	GO:0035810	6	1	0.05
glial cell-derived neurotrophic factor receptor signaling pathway	GO:0035860	2	1	0.02
cellular response to increased oxygen levels	GO:0036295	4	1	0.03
platelet maturation	GO:0036345	2	1	0.02
sodium ion export from cell	GO:0036376	4	2	0
heme catabolic process	GO:0042167	4	1	0.03
xenobiotic catabolic process	GO:0042178	5	1	0.04
pigment metabolic process	GO:0042440	28	2	0.02
negative regulation of hair cycle	GO:0042636	3	1	0.02
catagen	GO:0042637	1	1	0.01
defense response to protozoan	GO:0042832	6	1	0.05
regulation of vascular permeability	GO:0043114	6	1	0.05
positive regulation of mast cell degranulation	GO:0043306	6	1	0.05
enucleate erythrocyte differentiation	GO:0043353	2	1	0.02
regulation of secondary metabolic process	GO:0043455	6	1	0.05
positive regulation of insulin-like growth factor receptor signaling pathway	GO:0043568	5	1	0.04

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
regulation of peroxisome size	GO:0044375	2	1	0.02
regulation of mitochondrial mRNA stability	GO:0044528	2	1	0.02
biological phase	GO:0044848	3	1	0.02
hair cycle phase	GO:0044851	3	1	0.02
depurination	GO:0045007	1	1	0.01
T-helper 2 cell differentiation	GO:0045064	5	1	0.04
regulation of T-helper 1 cell differentiation	GO:0045625	4	1	0.03
negative regulation of T-helper 1 cell differentiation	GO:0045626	1	1	0.01
regulation of T-helper 2 cell differentiation	GO:0045628	4	1	0.03
positive regulation of T-helper 2 cell differentiation	GO:0045630	2	1	0.02
regulation of eosinophil differentiation	GO:0045643	2	1	0.02
positive regulation of eosinophil differentiation	GO:0045645	2	1	0.02
negative regulation of epidermis development	GO:0045683	4	1	0.03
pigment catabolic process	GO:0046149	4	1	0.03
membrane lipid catabolic process	GO:0046466	17	2	0.01
glycosphingolipid catabolic process	GO:0046479	5	1	0.04
regulation of melanin biosynthetic process	GO:0048021	5	1	0.04
positive regulation of melanin biosynthetic process	GO:0048023	3	1	0.02
immunoglobulin secretion	GO:0048305	5	1	0.04
autonomic nervous system development	GO:0048483	6	1	0.05
sympathetic nervous system development	GO:0048485	5	1	0.04
negative regulation of hair follicle maturation	GO:0048817	2	1	0.02
regulation of hair follicle maturation	GO:0048819	3	1	0.02
hair follicle maturation	GO:0048820	6	1	0.05
enucleate erythrocyte development	GO:0048822	2	1	0.02
negative regulation of T cell activation	GO:0050868	36	2	0.04
regulation of immunoglobulin secretion	GO:0051023	5	1	0.04
positive regulation of immunoglobulin secretion	GO:0051024	3	1	0.02
flavone metabolic process	GO:0051552	1	1	0.01
regulation of catagen	GO:0051794	1	1	0.01
negative regulation of catagen	GO:0051796	1	1	0.01
regulation of hair follicle development	GO:0051797	6	1	0.05
negative regulation of hair follicle development	GO:0051799	3	1	0.02
cellular glucuronidation	GO:0052695	3	1	0.02
flavonoid glucuronidation	GO:0052696	2	1	0.02
xenobiotic glucuronidation	GO:0052697	2	1	0.02
monovalent inorganic cation homeostasis	GO:0055067	45	4	0
potassium ion homeostasis	GO:0055075	9	2	0
sodium ion homeostasis	GO:0055078	15	3	0

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
regulation of hair cycle by canonical Wnt signaling pathway	GO:0060901	1	1	0.01
cardiac conduction	GO:0061337	34	2	0.03
mitochondrial ribosome assembly	GO:0061668	5	1	0.04
proton-transporting V-type ATPase complex assembly	GO:0070070	6	1	0.05
vacuolar proton-transporting V-type ATPase complex assembly	GO:0070072	5	1	0.04
thrombin-activated receptor signaling pathway	GO:0070493	3	1	0.02
organelle assembly	GO:0070925	236	5	0.04
biphenyl catabolic process	GO:0070980	1	1	0.01
cellular response to magnesium ion	GO:0071286	3	1	0.02
cellular response to manganese ion	GO:0071287	4	1	0.03
cellular response to ethanol	GO:0071361	4	1	0.03
sodium ion export	GO:0071436	4	2	0
negative regulation of transforming growth factor beta production	GO:0071635	3	1	0.02
podosome assembly	GO:0071800	3	1	0.02
regulation of podosome assembly	GO:0071801	2	1	0.02
positive regulation of podosome assembly	GO:0071803	2	1	0.02
chemokine secretion	GO:0090195	3	1	0.02
regulation of chemokine secretion	GO:0090196	3	1	0.02
positive regulation of chemokine secretion	GO:0090197	3	1	0.02
ammonium ion metabolic process	GO:0097164	65	3	0.02
import into cell	GO:0098657	41	2	0.05
import across plasma membrane	GO:0098739	29	2	0.02
regulation of secondary metabolite biosynthetic process	GO:1900376	5	1	0.04
positive regulation of secondary metabolite biosynthetic process	GO:1900378	3	1	0.02
negative regulation of vasculature development	GO:1901343	37	2	0.04
regulation of myoblast fusion	GO:1901739	6	1	0.05
positive regulation of myoblast fusion	GO:1901741	6	1	0.05
positive regulation of sodium ion transmembrane transport	GO:1902307	3	1	0.02
mitochondrial large ribosomal subunit assembly	GO:1902775	2	1	0.02
regulation of melanosome transport	GO:1902908	2	1	0.02
positive regulation of melanosome transport	GO:1902910	1	1	0.01
negative regulation of leukocyte cell-cell adhesion	GO:1903038	37	2	0.04
regulation of sodium ion export	GO:1903273	1	1	0.01
positive regulation of sodium ion export	GO:1903275	1	1	0.01
regulation of sodium ion export from cell	GO:1903276	1	1	0.01
positive regulation of sodium ion export from cell	GO:1903278	1	1	0.01

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
regulation of potassium ion import	GO:1903286	2	1	0.02
positive regulation of potassium ion import	GO:1903288	1	1	0.01
liposaccharide metabolic process	GO:1903509	42	2	0.05
regulation of cardiac conduction	GO:1903779	17	2	0.01
lysosomal protein catabolic process	GO:1905146	4	1	0.03
potassium ion import across plasma membrane	GO:1990573	13	2	0.01
response to odorant	GO:1990834	2	1	0.02
positive regulation of DNA-dependent DNA replication	GO:2000105	6	1	0.05
negative regulation of blood vessel morphogenesis	GO:2000181	34	2	0.03
sex chromosome	GO:0000803	6	1	0.05
X chromosome	GO:0000805	4	1	0.03
Barr body	GO:0001740	2	1	0.02
extracellular region	GO:0005576	1106	15	0.03
vacuolar lumen	GO:0005775	41	2	0.05
peroxisomal membrane	GO:0005778	22	2	0.01
integral component of peroxisomal membrane	GO:0005779	6	1	0.05
actomyosin contractile ring	GO:0005826	1	1	0.01
cytoplasmic dynein complex	GO:0005868	6	1	0.05
plasma membrane	GO:0005886	1268	17	0.02
integral component of plasma membrane	GO:0005887	428	9	0.01
hydrogen:potassium-exchanging ATPase complex	GO:0005889	1	1	0.01
sodium:potassium-exchanging ATPase complex	GO:0005890	3	1	0.03
external side of plasma membrane	GO:0009897	83	4	0
cell surface	GO:0009986	229	5	0.04
microtubule cytoskeleton	GO:0015630	344	7	0.02
catenin complex	GO:0016342	5	1	0.04
nuclear matrix	GO:0016363	26	2	0.02
vacuolar proton-transporting V-type ATPase complex	GO:0016471	5	1	0.04
immunoglobulin complex	GO:0019814	1	1	0.01
B cell receptor complex	GO:0019815	1	1	0.01
extrinsic component of membrane	GO:0019898	72	3	0.02
interphase microtubule organizing center	GO:0031021	2	1	0.02
anchored component of membrane	GO:0031225	32	2	0.03
intrinsic component of plasma membrane	GO:0031226	445	9	0.01
intrinsic component of peroxisomal membrane	GO:0031231	6	1	0.05
intrinsic component of external side of plasma membrane	GO:0031233	6	1	0.05
anchored component of external side of plasma membrane	GO:0031362	5	1	0.04

Node name	Node ID	No. of genes in node	No. of significant genes in node	Uncorrected *p*-value
Dbf4-dependent protein kinase complex	GO:0031431	1	1	0.01
microbody membrane	GO:0031903	22	2	0.01
proton-transporting V-type ATPase complex	GO:0033176	6	1	0.05
nuclear periphery	GO:0034399	33	2	0.03
centriolar satellite	GO:0034451	33	2	0.03
GID complex	GO:0034657	2	1	0.02
endoplasmic reticulum chaperone complex	GO:0034663	4	1	0.03
extrinsic component of endoplasmic reticulum membrane	GO:0042406	4	1	0.03
lamellar body	GO:0042599	4	1	0.03
receptor complex	GO:0043235	108	4	0.01
extracellular region part	GO:0044421	916	13	0.03
cytoskeletal part	GO:0044430	438	8	0.03
microbody part	GO:0044438	40	2	0.04
peroxisomal part	GO:0044439	40	2	0.04
plasma membrane part	GO:0044459	686	11	0.02
contractile ring	GO:0070938	2	1	0.02
cell periphery	GO:0071944	1303	17	0.03
cation-transporting ATPase complex	GO:0090533	5	2	0
ATPase dependent transmembrane transport complex	GO:0098533	6	2	0
side of membrane	GO:0098552	138	4	0.03
ATPase complex	GO:1904949	6	2	0
cellular protein localization	GO:0034613	553	1	0.04
cellular macromolecule localization	GO:0070727	557	1	0.04

Genes involved in cold adaptation

Next, we explored whether genes previously described to evolve differently in other cold adapted species are evolving faster or slower along the lineage to Steller's sea cows. We used the list of genes collected by Yudin and colleagues that comprises of 4,380 genes selected in at least one of the following species: arctic fox, Yakutian horse, woolly mammoth, polar bear, minke whale, and human populations adapted to frigid environments (*73*). The curated list comprises genes under faster evolution in the cold-adapted species, 416 of which were present in at least two species. We compared the dataset to the 197 genes that we identified as evolving faster in the Steller's sea cow lineage. We found that 51 genes with faster evolution rate in the Steller's sea cow are selected in other cold-adapted species as well (hypergeometric test, p-value < 0.0001) (Supplementary Table S11).

Intriguingly, we identified four genes involved in cold adaptation evolving under natural selection in Steller's sea cow, mammoth (*74*), and polar bear (*75*): *RUNDC1, MRTO4, MKKS,* and *PLEKHG6*. To our knowledge, the functions of *RUNDC1* and *MRT04* have not yet been determined and appeared in the list of genes associated with cold adaptation only because of species-overlapping signatures of selection (*73*). Conversely, mutations in *MKKS*, the McKusick-

Kaufman Syndrome gene, are known to cause two distinct phenotypes in humans: MKKS and Bardet-Biedl Syndrome 6 (BBS6). MKKS is characterized by genitourinary malformations and polydactyly, while BBS6 includes polydactyly, kidney defects, but also obesity and delayed development among its symptoms. Because mice lacking *Mkks* are obese in addition to other phenotypic features typical of BBS, it has been hypothesized that complete loss of *Mkks* leads to BBS6 including obesity, while some mutations may cause MKKS (*76*). The functional advantage of *MKKS* selection in Steller's sea cow is not directly evident, although it is tempting to speculate that this gene may be relevant to the enhanced adipose accumulation in cold environments. The function of *PLEKHG6*, or pleckstrin homology domain-containing family G member 6, is not fully understood, but its expression is positively correlated with lipid droplet size and is up-regulated in adipocytes specialized for lipogenesis (*77*). *PLEKHG6* is up-regulated in brown adipose tissue lacking the transcription factor *IRF4* (*78*), which has been shown to display reduced adaptive thermogenesis compared to the wild type (*79*). Although further analyses will be necessary to test this hypothesis, the upregulation of *PLEKHG6* may be a compensatory mechanism that allows better thermogenesis. This could be even more relevant, since *UCP1*, regarded as essential for NST, is absent in Steller's sea cow and mammoth (*25*).

Supplementary Table S11.

Genes with signs of faster evolution in Steller's sea cow and other arctic-adapted species. Genes under different evolution in at least two other species are marked in red.

Gene ID	Species with faster evolution	Gene ID	Species with faster evolution
PLEKHG6	Mammoth	***SRP68***	Mammoth
RRP12	Mammoth	***PDILT***	Mammoth
MRTO4	Mammoth	***DOLK***	Mammoth
OVGP1	Mammoth	***ACSF3***	Mammoth
GANAB	Mammoth	***ALS2CL***	Mammoth
RASD1	Mammoth	***TSHZ2***	Mammoth
VPS29	Mammoth	***C1orf116***	Mammoth
PERP	Mammoth	***RFX6***	Mammoth
EHHADH	Mammoth	***SH2D5***	Mammoth
GJA3	Mammoth	***RAD54B***	Mammoth
RIOK1	Mammoth	***FAN1***	Mammoth
MKKS	Mammoth	***RUNDC1***	Mammoth
RSAD1	Mammoth	***PTX4***	Mammoth
APTX	Mammoth	***PLEKHG6***	Polar bear
TGS1	Mammoth	***EHD3***	Polar bear
DNA2	Mammoth	***MRTO4***	Polar bear
CDH24	Mammoth	***KRT18***	Polar bear
GAS8	Mammoth	***MKKS***	Polar bear
SCYL1	Mammoth	***SERINC3***	Polar bear

Gene ID	Species with faster evolution	Gene ID	Species with faster evolution
HHIPL2	Mammoth	***ALOXE3***	Polar bear
TCTE1	Mammoth	***SLITRK6***	Polar bear
CNDP1	Mammoth	***SMYD4***	Polar bear
NMT2	Mammoth	***RUNDC1***	Polar bear
MEIOB	Mammoth	***OR8B3***	Polar bear
NIPAL1	Mammoth	***BVES***	Whale
COG1	Mammoth		

SM4: Demographic history reconstruction

Prehistoric population estimates

Prehistoric Steller's sea cow population estimates for regions beyond Bering Island were calculated by extrapolating the number of sea cows per km^2 of habitat at Bering Island to the km^2 of habitat (x) in any region of interest.

Habitat data was obtained from GEBCO (general bathometric chart of the oceans; https://www.gebco.net/data_and_products/gridded_bathymetry_data/).

The number of Steller's sea cows at Bering Island was around 1,000 animals (*2*). We calculated the area between shoreline and 20 m depth contour (approximate depth limit for kelp growth) as a proxy for sea cow habitat. The estimated area for Bering Island is 429 km^2.

From this we obtain the following estimates:

1. For all Alaska:

    - Area of habitat = 42,731 km^2
    - Population estimate = (42,731 km^2 / 429 km^2) * 1,000 = 99,600 sea cows

2. For all North America (Alaska to California):

    - Area of habitat = 57,383 km^2
    - Population estimate = (57,383 km^2 / 429 km^2) * 1,000 = 133,760 sea cows

3. For all North America and Russia (Kamchatka and Kurile Islands)

    - Area of habitat = 87,581 km^2
    - Population estimate = (87,581 km^2 / 429 km^2) * 1,000 = 204,150 sea cows

Inference of effective population size (N_e)

Mutation rate

To estimate the mutation rate in Steller's sea cow, we used reads that were aligned to both dugong and manatee from the two ~15× individuals. Using the snpAD output, we filtered positions with coverage ≥ 5× and < 50×, GQ ≥ 20. To estimate the mutation rate, we counted only polymorphic sites that were fixed in both Steller's sea cow individuals and different from manatee and dugong, respectively. The mutation rate per nucleotide per year was calculated as: $\mu = \frac{no.of derived sites/no.of covered sites}{2t}$, where *t* is the divergence time between the Steller's sea cow and dugong (28.6 million years ago (*7*)), or manatee (41.6 million years ago (*7*)), respectively. Thus, we computed:

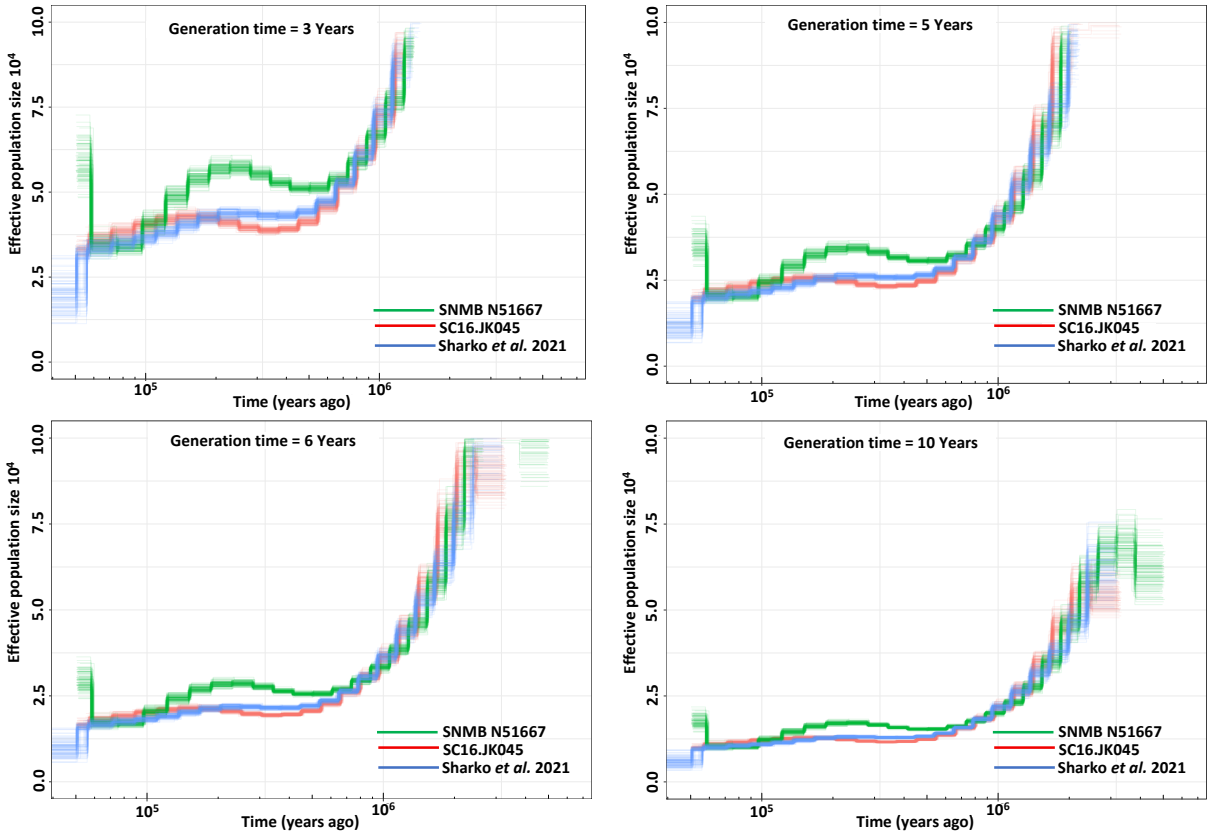
for dugong: $\mu = \frac{30,577,847/2,309,546,751}{2*28.6*10^6} = 2.315*10^{-10}$

for manatee: $\mu = \frac{24,419,358/1,342,627,473}{2*41.6*10^6} = 2.186*10^{-10}$

The mean estimated mutation rate for Steller's sea cow is 0.225e-9, which is comparable to that of woolly mammoths (*80*).

N_e estimation based on pairwise sequentially Markovian coalescent (PSMC) method

PSMC considers the density of heterozygous sites across the diploid genome from an individual to infer the distribution of the two alleles across all chromosomes as a function of time to the most recent common ancestor (*49*). This is then used to estimate effective population size (N_e) variations covering a time span of thousands of generations. Sirenians reach sexual maturity in 3–10 years (https://animaldiversity.org/accounts/Sirenia/); for comparison we considered a generation time of 3, 5, 6, and 10 years for Steller's sea cow (Supplementary Fig. S6). PSMC relies on variant calling methods that are not based on population frequency and without assumption of Hardy–Weinberg equilibrium. We used samtools mpileup (*44*) to call variants for autosomal dugong scaffolds ≥ 100 kbp. The variants were inputted into PSMC with the (-p) parameter set to "4 + 25 × 2 + 4 + 6" and 100 bootstraps were performed to ensure more reliable accuracy of the demographic history trajectory. While the PSMC method has been shown to produce reliable estimates for simulated data in the distant past (> 10,000 years), its accuracy for time periods closer to present is not guaranteed (*49*). Based on the N_e trajectory, the population decline of Steller's sea cow began around 0.5 million years ago, at a time when the North Pacific area could not have been inhabited by humans. Generation time does not influence the overall trajectory or timeline, but shows an effect on the initial estimate for the effective population size (Supplementary Fig. S6).



Supplementary Fig. S6.

PSMC plots of the three Steller's sea cow individuals considering different generation times (sirenians reach sexual maturity in 3–10 years https://animaldiversity.org/accounts/Sirenia/). The generation time shows a negative correlation with the estimated effective population size. Thus, the highest estimated effective population size corresponds to a generation time of 3 years. The trajectory of population size and the time scale are not affected by generation time.

Sex determination of the ~15×-coverage individuals

To determine the sex of the two ~15×-covered individuals, we calculated the ratio of the average coverage of the X chromosome to the average coverage of the autosomes. To this end, we initially determined which scaffolds from the dugong assembly correspond to the X chromosome based on a blast alignment of genes residing on the human chromosome X (GRCh38.p13). This resulted in 153 scaffolds with a total length of 201.3 million bp (Supplementary Table S12). The relative length of the X chromosome in the Indian elephant compared to the length of the entire genome is approximately 5.633 % (*81*); this suggests that for dugong the X-chromosome length would be approximately 181.4 million bp. The sum of the considered scaffolds is slightly larger, but because this represents about 10 % of the actual X-chromosome sequence, the ratio between X-to-autosomes will not be heavily influenced. This ratio is expected to be 1 in a female and 0.5

in a male. We obtained ratios of 0.93 and 1.23 (14.68× and 19.17× compared to the average 15.86× and 15.63×), which suggests that both individuals are females and heterozygosity estimates can be calculated across all scaffolds. To confirm this result, we also checked whether any reads mapped to the *SRY* or *ZFY* dugong genes located on the Y chromosome, and, as expected, we did not identify Steller's sea cow reads corresponding to those regions.

Supplementary Table S12.

Scaffolds from the *de novo* dugong assembly assigned to chromosome X.

Dugong scaffold ID	Scaffold length (bp)	Mean coverage Ind01	Mean coverage Ind02
scaffold7	15,224,240	16.11	20.28
scaffold18	17,251,047	14.65	17.50
scaffold37	14,370,710	16.40	18.14
scaffold79	5,595,303	16.84	18.56
scaffold88	8,460,835	17.58	19.08
scaffold91	6,694,011	17.32	20.56
scaffold108	4,050,669	16.27	18.37
scaffold141	3,630,362	15.57	18.38
scaffold156	8,432,348	16.15	21.90
scaffold198	3,087,677	15.33	19.69
scaffold201	4,768,493	14.90	19.27
scaffold239	3,594,916	16.23	21.99
scaffold298	3,001,275	16.55	21.79
scaffold347	9,823,323	15.73	18.25
scaffold364	4,621,530	16.39	18.92
scaffold376	1,882,354	15.47	18.62
scaffold385	1,847,155	13.03	16.76
scaffold398	1,811,345	15.02	18.66
scaffold416	1,742,563	16.84	20.26
scaffold422	2,074,037	14.52	17.07
scaffold426	1,726,953	12.26	16.77
scaffold478	1,925,199	14.09	18.32
scaffold479	1,575,312	0.07	0.06
scaffold500	1,522,886	16.52	21.68
scaffold537	1,473,645	15.31	19.33
scaffold546	1,410,123	16.34	21.47
scaffold562	1,638,642	17.08	20.92
scaffold584	1,306,290	12.11	16.93

Dugong scaffold ID	Scaffold length (bp)	Mean coverage Ind01	Mean coverage Ind02
scaffold586	1,471,532	14.93	20.57
scaffold614	1,242,202	14.76	18.34
scaffold619	1,228,465	13.38	17.10
scaffold623	1,217,361	15.02	19.56
scaffold630	1,196,204	14.03	18.40
scaffold631	1,195,602	12.51	15.42
scaffold637	1,189,222	18.02	20.49
scaffold647	1,157,931	18.13	21.19
scaffold658	1,148,743	13.37	18.30
scaffold662	2,383,847	11.99	16.00
scaffold675	1,115,131	12.01	16.01
scaffold747	978,951	12.36	17.04
scaffold758	965,813	15.81	19.13
scaffold759	1,178,196	15.64	19.20
scaffold769	949,168	12.31	14.83
scaffold779	935,908	14.25	19.23
scaffold783	931,651	16.44	21.70
scaffold795	912,243	17.55	22.25
scaffold800	906,131	14.01	21.05
scaffold803	901,153	16.04	20.29
scaffold811	923,461	0.05	0.04
scaffold816	1,202,943	12.02	16.14
scaffold825	876,245	16.81	16.94
scaffold844	1,731,182	13.05	17.70
scaffold862	960,403	0.06	0.05
scaffold869	1,106,877	16.02	20.63
scaffold883	1,286,841	15.45	20.16
scaffold972	666,031	15.04	18.92
scaffold980	654,432	11.17	15.38
scaffold984	1,248,677	17.65	22.86
scaffold992	942,668	17.26	21.18
scaffold995	631,532	17.34	22.04
scaffold1001	937,064	13.66	18.52
scaffold1005	620,131	13.40	19.80
scaffold1012	718,877	13.50	19.35
scaffold1018	803,886	14.80	19.16
scaffold1019	612,082	11.78	14.25

Dugong scaffold ID	Scaffold length (bp)	Mean coverage Ind01	Mean coverage Ind02
scaffold1037	591,429	16.17	21.33
scaffold1067	1,423,584	13.14	18.08
scaffold1109	656,714	15.00	20.55
scaffold1112	1,404,205	16.32	19.38
scaffold1114	522,711	17.27	20.01
scaffold1117	537,854	15.76	19.36
scaffold1121	717,088	15.95	20.69
scaffold1123	513,832	11.26	15.78
scaffold1125	913,104	16.31	21.07
scaffold1126	713,031	16.21	21.03
scaffold1131	508,410	15.13	19.92
scaffold1136	536,882	13.78	17.87
scaffold1145	498,065	15.43	19.24
scaffold1160	482,206	14.44	18.88
scaffold1164	472,738	14.29	18.64
scaffold1166	471,438	14.19	18.39
scaffold1168	493,835	13.75	17.69
scaffold1222	430,380	15.24	20.21
scaffold1227	510,886	14.99	19.77
scaffold1235	420,963	15.53	20.42
scaffold1240	416,642	16.01	18.48
scaffold1249	447,741	12.32	17.10
scaffold1255	444,952	14.24	16.50
scaffold1260	484,526	0.03	0.03
scaffold1271	396,544	17.23	21.06
scaffold1272	413,669	12.66	17.40
scaffold1302	372,349	18.03	24.33
scaffold1314	365,108	13.41	16.03
scaffold1318	363,403	15.61	21.56
scaffold1323	358,456	12.85	14.98
scaffold1328	357,130	17.87	21.61
scaffold1343	351,100	14.51	19.44
scaffold1344	350,306	13.18	19.87
scaffold1351	506,834	15.21	21.80
scaffold1354	357,464	16.59	20.19
scaffold1360	340,258	13.23	19.96
scaffold1369	401,922	15.17	19.94

Dugong scaffold ID	Scaffold length (bp)	Mean coverage Ind01	Mean coverage Ind02
scaffold1401	392,625	12.69	18.77
scaffold1407	426,720	15.11	20.00
scaffold1467	294,254	12.90	18.51
scaffold1480	271,984	13.63	18.62
scaffold1508	271,670	17.67	23.62
scaffold1519	431,402	16.12	20.89
scaffold1528	249,864	16.46	21.97
scaffold1536	269,991	15.36	20.43
scaffold1538	256,533	13.35	19.05
scaffold1575	231,741	11.64	15.76
scaffold1576	231,266	16.45	19.98
scaffold1595	229,837	12.57	17.31
scaffold1596	221,616	16.23	21.45
scaffold1613	216,572	14.82	18.61
scaffold1620	213,679	18.89	21.90
scaffold1625	222,770	12.35	17.48
scaffold1637	208,478	0.04	0.03
scaffold1658	199,242	15.71	20.82
scaffold1676	456,324	12.82	17.68
scaffold1707	183,206	15.19	17.07
scaffold1747	171,000	17.38	25.86
scaffold1750	170,304	17.84	24.05
scaffold1799	155,872	14.39	16.61
scaffold1888	133,363	18.62	20.70
scaffold1898	130,930	16.90	25.20
scaffold1960	118,259	17.07	25.72
scaffold1962	117,884	14.63	20.18
scaffold1975	114,250	12.25	17.77
scaffold2006	107,735	15.87	23.04
scaffold2019	105,969	13.76	18.20
scaffold2040	120,537	14.52	19.58
scaffold2085	94,923	15.54	18.95
scaffold2097	93,100	17.16	22.67
scaffold2131	101,433	15.67	20.15
scaffold2168	82,730	15.72	17.82
scaffold2204	78,653	16.22	20.00
scaffold2216	76,809	13.63	20.97

Dugong scaffold ID	Scaffold length (bp)	Mean coverage Ind01	Mean coverage Ind02
scaffold2246	73,297	16.22	22.15
scaffold2260	71,345	18.00	25.03
scaffold2323	104,372	22.67	30.17
scaffold2528	49,333	18.47	24.18
scaffold2649	42,734	19.03	26.44
scaffold2671	41,606	15.03	23.12
scaffold2688	40,587	17.04	26.93
scaffold2907	33,307	13.75	21.16
scaffold3012	31,060	14.86	22.60
scaffold3091	48,139	12.26	18.12
scaffold3181	26,663	11.93	18.70
scaffold3542	20,015	15.95	26.09
scaffold3840	16,285	16.52	25.45
scaffold4549	10,962	16.57	18.26
Total length	201,283,018		

Population diversity estimates

To determine the level of diversity over time in the sampled Steller's sea cow populations, we performed pairwise comparisons of individuals based on their radiocarbon dating results (Supplementary Table S1). We used Consensify (*51*) to generate a consensus pseudohaploid genome sequence. The method controls for errors introduced by differential sequencing coverage. For each position of the dugong genome, three bases were selected from the Steller's sea cow's read stack. If ≥ 2 out of the selected three reads agree, the base is retained. If only two reads are present and in agreement, the base is also retained. If no two reads agree, an "N" is entered for that position. If coverage is < 2, or above twice the average coverage for the library (Supplementary Table S1), then an N is entered for that position. Thus, only libraries with an average coverage > 2 could be included in the pairwise comparisons. Pairwise comparisons were performed for individuals dated to the same time period. We excluded every position which contained an N in any of the two individuals in the pairwise comparison and counted the number of differences within non-overlapping blocks of 50 kbp. For the ~15× individuals we included all scaffolds > 50 kbp, since we could test that they were females. For the low-coverage individuals we included scaffolds > 50 kbp and excluded scaffolds assigned to sex chromosomes, since their sex could not be assigned. We included 10,000 randomly selected blocks of 50 kbp in the pairwise comparisons and used the R ggplot2 density function to plot the differences over these blocks (Fig. 3B) (*82*).

REFERENCES AND NOTES

1. J. A. Estes, A. Burdin, D. F. Doak, Sea otters, kelp forests, and the extinction of Steller's sea cow. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 880–885 (2016).

2. S. T. Turvey, C. L. Risley, Modelling the extinction of Steller's sea cow. *Biol. Lett.* **2**, 94–97 (2006).

3. J. Davis, Steller's sea cow: The first historical extinction of marine mammal at human hands | Natural History Museum; www.nhm.ac.uk/discover/stellers-sea-cow-first-historical-extinction-of-marine-mammal-at-human-hands.html.

4. D. P. Domning, An ecological model for late Tertiary sirenian evolution in the North Pacific Ocean. *Syst. Zool.* **25**, 352–362 (1976).

5. A. Forsten, P. M. Youngman, Hydrodamalis gigas. *Mamm. Species.* **165**, 1 (1982).

6. H. Marsh, T. J. O'Shea, J. E. Reynolds, J. E. Reynolds III, *Ecology and Conservation of the sirenia: Dugongs and Manatees* (Cambridge Univ. Press, 2012), vol. 18.

7. M. S. Springer, A. V. Signore, J. L. Paijmans, J. Vélez-Juarbe, D. P. Domning, C. E. Bauer, K. He, L. Crerar, P. F. Campos, W. J. Murphy, R. W. Meredith, J. Gatesy, E. Willerslev, R. D. MacPhee, M. Hofreiter, K. L. Campbell, Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of sirenia. *Mol. Phylogenet. Evol.* **91**, 178–193 (2015).

8. F. S. Sharko, S. M. Rastorguev, E. S. Boulygina, S. V. Tsygankova, A. S. Ibragimova, A. N. Tikhonov, A. V. Nedoluzhko, Molecular phylogeny of the extinct Steller's sea cow and other sirenia species based on their complete mitochondrial genomes. *Genomics* **111**, 1543–1546 (2019).

9. A. D. Foote, Y. Liu, G. W. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancia, R. Nielsen, X. Qin, J. Qu, B. J. Raney, N. Vijay, J. B. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. Gilbert, R. A. Gibbs, Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).

10. B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).

11. F. Cunningham, P. Achuthan, W. Akanni, J. Allen, M. R. Amode, I. M. Armean, R. Bennett, J. Bhai, K. Billis, S. Boddu, C. Cummins, C. Davidson, K. J. Dodiya, A. Gall, C. G. Girón, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, J. C. Marugán, T. Maurel, A. C. McMahon, B. Moore, J. Morales, J. M. Mudge, M. Nuhn, D. Ogeh, A. Parker, A. Parton, M. Patricio, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, H. Sparrow, E. Stapleton, M. Szuba, K. Taylor, G. Threadgold, A. Thormann, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, A. D. Yates, D. R. Zerbino, P. Flicek, Ensembl 2019. *Nucleic Acids Res.* **47**, D745–D751 (2019).

12. M. R. McGowen, J. Gatesy, D. E. Wildman, Molecular evolution tracks macroevolutionary transitions in Cetacea. *Trends Ecol. Evol.* **29**, 336–346 (2014).

13. P. Krieg, S. Rosenberger, S. de Juanes, S. Latzko, J. Hou, A. Dick, U. Kloz, F. van der Hoeven, I. Hausser, I. Esposito, M. Rauh, H. Schneider, Aloxe3 knockout mice reveal a function of epidermal lipoxygenase-3 as hepoxilin synthase and its pivotal role in barrier formation. *J. Invest. Dermatol.* **133**, 172–180 (2013).

14. F. S. Sharko, E. S. Boulygina, S. V. Tsygankova, N. V. Slobodova, D. A. Alekseev, A. A. Krasivskaya, S. M. Rastorguev, A. N. Tikhonov, A. V. Nedoluzhko, Steller's sea cow genome suggests this species began going extinct before the arrival of Paleolithic humans. *Nat. Commun.* **12**, 2215 (2021).

15. J. K. Simpson, M. Martinez-Queipo, A. Onoufriadis, S. Tso, E. Glass, L. Liu, T. Higashino, W. Scott, C. Tierney, M. A. Simpson, R. Desomchoke, L. Youssefian, A. H. SaeIdian, H. Vahidnezhad, A. Bisquera, J. Ravenscroft, C. Moss, E. A. O'Toole, N. Burrows, S. Leech, E. A. Jones, D. Lim, A. Ilchyshyn, N. Goldstraw, M. J. Cork, S. Darne, J. Uitto, A. E. Martinez, J. E. Mellerio, J. A. McGrath,

Genotype-phenotype correlation in a large English cohort of patients with autosomal recessive ichthyosis. *Br. J. Dermatol.* **182**, 729–737 (2020).

16. H. Bučková, H. Nosková, R. Borská, K. Réblová, B. Pinková, E. Zapletalová, L. Kopečková, O. Horký, J. Němečková, R. Gaillyová, Z. Nagy, K. Veselý, M. Hermanová, K. Stehlíková, L. Fajkusová, Autosomal recessive congenital ichthyoses in the Czech Republic. *Br. J. Dermatol.* **174**, 405–407 (2016).

17. R. Borská, B. Pinková, K. Réblová, H. Bučková, L. Kopečková, J. Němečková, A. Puchmajerová, M. Malíková, M. Hermanová, L. Fajkusová, Inherited ichthyoses: Molecular causes of the disease in Czech patients. *Orphanet J. Rare Dis.* **14**, 92 (2019).

18. S. de Juanes, N. Epp, S. Latzko, M. Neumann, G. Fürstenberger, I. Hausser, H. J. Stark, P. Krieg, Development of an ichthyosiform phenotype in Alox12b-deficient mouse skin transplants. *J. Invest. Dermatol.* **129**, 1429–1436 (2009).

19. M. S. Springer, C. F. Guerrero-Juarez, M. Huelsmann, M. A. Collin, K. Danil, M. R. McGowen, J. W. Oh, R. Ramos, M. Hiller, M. V. Plikus, J. Gatesy, Genomic and anatomical comparisons of skin support independent adaptation to life in water by cetaceans and hippos. *Curr. Biol.* **31**, 2124–2139.e3 (2021).

20. F. Reisch, K. R. Kakularam, S. Stehling, D. Heydeck, H. Kuhn, Eicosanoid biosynthesis in marine mammals. *FEBS J.* **288**, 1387–1406 (2021).

21. V. Sharma, N. Hecker, J. G. Roscito, L. Foerster, B. E. Langer, M. Hiller, A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat. Commun.* **9**, 1215 (2018).

22. O. J. Harrison, J. Brasch, G. Lasso, P. S. Katsamba, G. Ahlsen, B. Honig, L. Shapiro, Structural basis of adhesive binding by desmocollins and desmogleins. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7160–7165 (2016).

23. L. Zhang, C. K. Ip, I. C. J. Lee, Y. Qi, F. Reed, T. Karl, J. K. Low, R. F. Enriquez, N. J. Lee, P. A. Baldock, H. Herzog, Diet-induced adaptive thermogenesis requires neuropeptide FF receptor-2 signalling. *Nat. Commun.* **9**, 4722 (2018).

24. M. J. Gaudry, M. Jastroch, J. R. Treberg, M. Hofreiter, J. L. A. Paijmans, J. Starrett, N. Wales, A. V. Signore, M. S. Springer, K. L. Campbell, Inactivation of thermogenic UCP1 as a historical contingency in multiple placental mammal clades. *Sci. Adv.* **3**, e1602878 (2017).

25. W. Gearty, C. R. McClain, J. L. Payne, Energetic tradeoffs control the size distribution of aquatic mammals. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4194–4199 (2018).

26. A. Roesler, L. Kazak, UCP1-independent thermogenesis. *Biochem. J.* **477**, 709–725 (2020).

27. K. Ikeda, T. Yamada, UCP1 dependent and independent thermogenesis in brown and beige adipocytes. *Front. Endocrinol.* **11**, 498 (2020).

28. G. W. Steller, O. W. Frost, *Journal of a Voyage with Bering, 1741–1742* (Stanford University Press, 1993).

29. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

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

31. N. Shammas, B. Walker, H. Martinez De La Torre, C. Bertrand, J. Southon, Effect of ultrafilter pretreatment, acid strength and decalcification duration on archaeological bone collagen yield. *Nucl. Instrum. Methods Phys. Res. B* **456**, 283–286 (2019).

32. J. Dabney, M. Knapp, I. Glocke, M. T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J. L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).

33. M. T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).

34. P. Korlević, T. Gerber, M.T. Gansauge, M. Hajdinjak, S. Nagel, A. Aximu-Petri, M. Meyer. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).

35. J. D. Kapp, R. E. Green, B. Shapiro, A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *J. Hered.* **112**, 241–249 (2021).

36. N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, D. B. Jaffe, Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

37. L. Coombe, J. Zhang, B. P. Vandervalk, J. Chu, S. D. Jackman, I. Birol, R. L. Warren, ARKS: Chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* **19**, 234 (2018).

38. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

39. A. Mikheenko, G. Valin, A. Prjibelski, V. Saveliev, A. Gurevich, Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics* **32**, 3321–3323 (2016).

40. D. Le Duc, G. Renaud, A. Krishnan, M. S. Almén, L. Huynen, S. J. Prohaska, M. Ongyerth, B. D. Bitarello, H. B. Schiöth, M. Hofreiter, P. F. Stadler, K. Prüfer, D. Lambert, J. Kelso, T. Schöneberg, Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biol.* **16**, 147 (2015).

41. G. Renaud, U. Stenzel, T. Maricic, V. Wiebe, J. Kelso, deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772 (2015).

42. G. Renaud, U. Stenzel, J. Kelso, leeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).

43. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

44. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

45. M. Meyer, M. Kircher, M. T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).

46. K. Prüfer, snpAD: An ancient DNA genotype caller. *Bioinformatics* **34**, 4165–4171 (2018).

47. V. Ranwez, E. J. P. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).

48. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

49. P. J. Laud, ratesci: Confidence Intervals for Comparisons of Binomial or Poisson Rates. R Packag. version 0.3–0 (2018); https//CRAN.R-project.org/package= ratesci.

50. A. Barlow, S. Hartmann, J. Gonzalez, M. Hofreiter, J. L. A. Paijmans, Consensify: A method for generating pseudohaploid genome sequences from palaeogenomic datasets with reduced error rates. *Genes* **11**, 50 (2020).

51. IUCN, The IUCN Red List of Threatened Species. Version 2020–3 (2020).

52. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

53. A. F. Smit, R. Hubley, P. Green, Repeat masker open 4.0; www.repeatmasker.org/.

54. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

55. K. Prüfer, B. Muetzel, H. H. Do, G. Weiss, P. Khaitovich, E. Rahm, S. Pääbo, M. Lachmann, W. Enard, FUNC: A package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* **8**, 41 (2007).

56. H.-G. Drost, J. Paszkowski, Biomartr: Genomic data retrieval with R. *Bioinformatics* **33**, 1216–1217 (2017).

57. R. Cheng, J. Liang, Y. Li, J. Zhang, C. Ni, H. Yu, X. Kong, M. Li, Z. Yao, Next-generation sequencing through multi-gene panel testing for diagnosis of hereditary ichthyosis in Chinese. *Clin. Genet.* **97**, 770–778 (2020).

58. S. Israeli, I. Goldberg, D. Fuchs-Telem, R. Bergman, M. Indelman, O. Bitterman-Deutsch, A. Harel, Y. Mashiach, O. Sarig, E. Sprecher, Non-syndromic autosomal recessive congenital ichthyosis in the Israeli population. *Clin. Exp. Dermatol.* **38**, 911–916 (2013).

59. K. M. Eckl, P. Krieg, W. Küster, H. Traupe, F. André, N. Wittstruck, G. Fürstenberger, H. C. Hennies, Mutation spectrum and functional analysis of epidermis-type lipoxygenases in patients with autosomal recessive congenital ichthyosis. *Hum. Mutat.* **26**, 351–361 (2005).

60. G. Espregueira Themudo, L. Q. Alves, A. M. Machado, M. Lopes-Marques, R. R. da Fonseca, M. Fonseca, R. Ruivo, L. F. C. Castro, *Losing* Genes: The evolutionary remodeling of Cetacea skin. *Front. Mar. Sci.* **7**, 592375 (2020).

61. B. D. Hicks, D. J. St. Aubin, J. R. Geraci, W. R. Brown, Epidermal growth in the bottlenose dolphin, *Tursiops truncatus*. *J. Invest. Dermatol.* **85**, 60–63 (1985).

62. H.-G. Drost, Genomic Data Retrieval, doi:10.1186/1471-2164-10-22.

63. D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. Kasprzyk, BioMart—Biological queries made easy. *BMC Genomics* **10**, 22 (2009).

64. S. Hessel, A. Eichinger, A. Isken, J. Amengual, S. Hunzelmann, U. Hoeller, V. Elste, W. Hunziker, R. Goralczyk, V. Oberhauser, J. Von Lintig, A. Wyss, CMO1 deficiency abolishes vitamin A production from β-carotene and alters lipid metabolism in mice. *J. Biol. Chem.* **282**, 33553–33561 (2007).

65. M. A. Lanaspa, L. E. Epperson, N. Li, C. Cicerchi, G. E. Garcia, C. A. Roncal-Jimenez, J. Trostel, S. Jain, C. T. Mant, C. J. Rivard, T. Ishimoto, M. Shimada, L. G. Sanchez-Lozada, T. Nakagawa, A. Jani, P. Stenvinkel, S. L. Martin, R. J. Johnson, Opposing activity changes in AMP deaminase and AMP-activated protein kinase in the hibernating ground squirrel. *PLOS ONE*. **10**, e0123509 (2015).

66. Y. Böttcher, H. Unbehauen, N. Klöting, K. Ruschke, A. Körner, D. Schleinitz, A. Tönjes, B. Enigk, S. Wolf, K. Dietrich, M. Koriath, G. H. Scholz, Y. H. Tseng, A. Dietrich, M. R. Schön, W. Kiess, M. Stumvoll, M. Blüher, P. Kovacs, Adipose tissue expression and genetic variants of the bone morphogenetic protein receptor 1A gene (BMPR1A) are associated with human obesity. *Diabetes* **58**, 2119–2128 (2009).

67. P. S. Petersen, C. Jin, A. N. Madsen, M. Rasmussen, R. Kuhre, K. L. Egerod, L. B. Nielsen, T. W. Schwartz, B. Holst, Deficiency of the GPR39 receptor is associated with obesity and altered adipocyte metabolism. *FASEB J.* **25**, 3803–3814 (2011).

68. D. Moechars, I. Depoortere, B. Moreaux, B. de Smet, I. Goris, L. Hoskens, G. Daneels, S. Kass, L. Ver Donck, T. Peeters, B. Coulie, Altered gastrointestinal and metabolic function in the GPR39-obestatin receptor-knockout mouse. *Gastroenterology* **131**, 1131–1141 (2006).

69. P. Akbari, A. Gilani, O. Sosina, J. A. Kosmicki, L. Khrimian, Y. Y. Fang, T. Persaud, V. Garcia, D. Sun, A. Li, J. Mbatchou, A. E. Locke, C. Benner, N. Verweij, N. Lin, S. Hossain, K. Agostinucci, J. V. Pascale, E. Dirice, M. Dunn, W. E. Kraus, S. H. Shah, Y. D. I. Chen, J. I. Rotter, D. J. Rader, O. Melander, C. D. Still, T. Mirshahi, D. J. Carey, J. Berumen-Campos, P. Kuri-Morales, J. Alegre-Díaz, J. M. Torres, J. R. Emberson, R. Collins, S. Balasubramanian, A. Hawes, M. Jones, B. Zambrowicz, A. J. Murphy, C. Paulding, G. Coppola, J. D. Overton, J. G. Reid, A. R. Shuldiner, M. Cantor, H. M. Kang, G. R. Abecasis, K. Karalis, A. N. Economides, J. Marchini, G. D. Yancopoulos, M. W. Sleeman, J. Altarejos, G. Della Gatta, R. Tapia-Conyer, M. L. Schwartzman, A. Baras, M. A. R. Ferreira, L. A. Lotta, Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection from obesity. *Science* **373**, eabf8683 (2021).

70. S. Carbon, E. Douglass, N. Dunn, B. Good, N. L. Harris, S. E. Lewis, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A. Muruganujan, X. Huang, S. Poudel, T. Mushayahama, J. C. Hu, S. A. LaBonte, D. A. Siegele, G.

Antonazzo, H. Attrill, N. H. Brown, S. Fexova, P. Garapati, T. E. M. Jones, S. J. Marygold, G. H. Millburn, A. J. Rey, V. Trovisco, G. Dos Santos, D. B. Emmert, K. Falls, P. Zhou, J. L. Goodman, V. B. Strelets, J. Thurmond, M. Courtot, D. S. Osumi, H. Parkinson, P. Roncaglia, M. L. Acencio, M. Kuiper, A. Lreid, C. Logie, R. C. Lovering, R. P. Huntley, P. Denny, N. H. Campbell, B. Kramarz, V. Acquaah, S. H. Ahmad, H. Chen, J. H. Rawson, M. C. Chibucos, M. Giglio, S. Nadendla, R. Tauber, M. J. Duesbury, N. T. Del, B. H. M. Meldal, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, Z. Xie, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bahler, A. Lock, E. R. Bolton, J. De Pons, M. Dwinell, G. T. Hayman, S. J. F. Laulederkind, M. Shimoyama, M. Tutaj, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, G. Binkley, B. L. Dunn, J. M. Cherry, S. R. Engel, F. Gondwe, K. Karra, K. A. MacPherson, S. R. Miyasato, R. S. Nash, P. C. Ng, T. K. Sheppard, A. Shrivatsav Vp, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, E. Bakker, T. Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C. Arighi, A. Auchincloss, K. Axelsen, G. P. Argoud, A. Bateman, B. Bely, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. Casals-Casas, E. Coudert, A. Estreicher, L. Famiglietti, P. Garmiri, G. Georghiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, U. Hinz, C. Hulo, A. Ignatchenko, F. Jungo, G. Keller, K. Laiho, P. Lemercier, D. Lieberherr, Y. Lussi, A. Mac-Dougall, M. Magrane, M. J. Martin, P. Masson, D. A. Natale, N. N. Hyka, I. Pedruzzi, K. Pichler, S. Poux, C. Rivoire, M. Rodriguez-Lopez, T. Sawford, E. Speretta, A. Shypitsyna, A. Stutz, S. Sundaram, M. Tognolli, N. Tyagi, K. Warner, R. Zaru, C. Wu, J. Chan, J. Cho, S. Gao, C. Grove, M. C. Harrison, K. Howe, R. Lee, J. Mendel, H. M. Muller, D. Raciti, K. Van Auken, M. Berriman, L. Stein, P. W. Sternberg, D. Howe, S. Toro, M. Westerfield, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

71. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

72. P. A. Watkins, D. Maiguel, Z. Jia, J. Pevsner, Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J. Lipid Res.* **48**, 2736–2750 (2007).

73. N. S. Yudin, D. M. Larkin, E. V. Ignatieva, A compendium and functional characterization of mammalian genes involved in adaptation to Arctic or Antarctic environments. *BMC Genet.* **18**, 111 (2017).

74. V. J. Lynch, O. C. Bedoya-Reina, A. Ratan, M. Sulak, D. I. Drautz-Moses, G. H. Perry, W. Miller, S. C. Schuster, Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell Rep.* **12**, 217–228 (2015).

75. S. Liu, E. D. Lorenzen, M. Fumagalli, B. Li, K. Harris, Z. Xiong, L. Zhou, T. S. Korneliussen, M. Somel, C. Babbitt, G. Wray, J. Li, W. He, Z. Wang, W. Fu, X. Xiang, C. C. Morgan, A. Doherty, M. J. O'Connell, J. O. McInerney, E. W. Born, L. Dalén, R. Dietz, L. Orlando, C. Sonne, G. Zhang, R. Nielsen, E. Willerslev, J. Wang, Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785–794 (2014).

76. M. A. Fath, R. F. Mullins, C. Searby, D. Y. Nishimura, J. Wei, K. Rahmouni, R. E. Davis, M. K. Tayeh, M. Andrews, B. Yang, C. D. Sigmund, E. M. Stone, V. C. Sheffield, Mkks-null mice have a phenotype resembling Bardet-Biedl syndrome. *Hum. Mol. Genet.* **14**, 1109–1118 (2005).

77. S. Y. Min, A. Desai, Z. Yang, A. Sharma, T. DeSouza, R. M. J. Genga, A. Kucukural, L. M. Lifshitz, S. Nielsen, C. Scheele, R. Maehr, M. Garber, S. Corvera, Diverse repertoire of human adipocyte subtypes develops from transcriptionally distinct mesenchymal progenitor cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 17970–17979 (2019).

78. X. Kong, T. Yao, P. Zhou, L. Kazak, D. Tenen, A. Lyubetskaya, B. A. Dawes, L. Tsai, B. B. Kahn, B. M. Spiegelman, T. Liu, E. D. Rosen, Brown adipose tissue controls skeletal muscle function via the secretion of myostatin. *Cell Metab.* **28**, 631–643.e3 (2018).

79. X. Kong, A. Banks, T. Liu, L. Kazak, R. R. Rao, P. Cohen, X. Wang, S. Yu, J. C. Lo, Y. H. Tseng, A. M. Cypess, R. Xue, S. Kleiner, S. Kang, B. M. Spiegelman, E. D. Rosen, IRF4 is a key thermogenic transcriptional partner of PGC-1α. *Cell* **158**, 69–83 (2014).

80. E. Palkopoulou, S. Mallick, P. Skoglund, J. Enk, N. Rohland, H. Li, A. Omrak, S. Vartanyan, H. Poinar, A. Götherström, D. Reich, L. Dalén, Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).

81. B. Behr, D. Rath, T. B. Hildebrandt, F. Goeritz, S. Blottner, T. J. Portas, B. R. Bryant, B. Sieg, A. Knieriem, S. De Graaf, W. M. C. Maxwell, R. Hermes, Germany/Australia index of sperm sex sortability in elephants and rhinoceros. *Reprod. Domest. Anim.* **44**, 273–277 (2009).

82. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer New York, 2016).