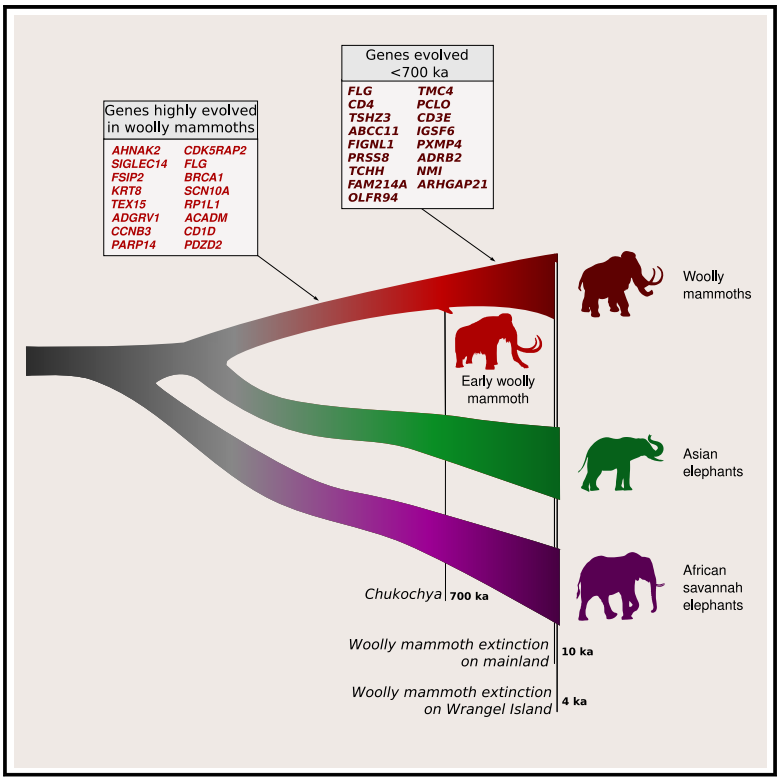
Article

Current Biology

Genomics of adaptive evolution in the woolly mammoth

Graphical abstract



Highlights

- Genomes from 23 woolly mammoths and 28 extant elephants revealed adaptive differences
- Gene ontology suggested enrichment of mammoth genomic adaptations to cold environment
- Highly evolved genes included ones related to hair, skin, fat metabolism, and immunity
- Several key phenotypes appear to have evolved via heterochronous polygenic selection

Authors

David Díez-del-Molino, Marianne Dehasque, J. Camilo Chacón-Duque, ..., Peter D. Heintzman, Tom van der Valk, Love Dalén

Correspondence

diez.molino@gmail.com (D.D.-d.-M.), love.dalen@zoologi.su.se (L.D.)

In brief

Díez-del-Molino et al. analyze unique non-synonymous mutations in 23 woolly mammoth genomes, including a 700,000-year-old specimen. They find that woolly mammoths had highly evolved genes associated with hair and skin development, fat storage and metabolism, immune system function, and body size, some of which evolved during the last 700,000 years.

Díez-del-Molino et al., 2023, Current Biology *33*, 1753–1764
May 8, 2023
https://doi.org/10.1016/j.cub.2023.03.084

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Genomics of adaptive evolution in the woolly mammoth

David Díez-del-Molino,[1,2,3,19,*] Marianne Dehasque,[1,2,3] J. Camilo Chacón-Duque,[1,4] Patrícia Pečnerová,[3,5] Alexei Tikhonov,[6] Albert Protopopov,[7] Valeri Plotnikov,[7] Foteini Kanellidou,[1,8] Pavel Nikolskiy,[9] Peter Mortensen,[10] Gleb K. Danilov,[11] Sergey Vartanyan,[12] M. Thomas P. Gilbert,[13,14] Adrian M. Lister,[15] Peter D. Heintzman,[1,16] Tom van der Valk,[1,3,17,18] and Love Dalén[1,2,3,18,20,21,*]

[1]Centre for Palaeogenetics, 10691 Stockholm, Sweden
[2]Department of Zoology, Stockholm University, 10691 Stockholm, Sweden
[3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 10405 Stockholm, Sweden
[4]Department of Archaeology and Classical Studies, Stockholm University, 10691 Stockholm, Sweden
[5]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark
[6]Zoological Institute of the Russian Academy of Sciences, 190121 Saint Petersburg, Russia
[7]Academy of Sciences of Sakha Republic, 677007 Yakutsk, Russia
[8]Department of Microbiology, Tumor and Cell Biology, Clinical Genomics Facility, Karolinska Institutet, 171 77 Stockholm, Sweden
[9]Geological Institute, Russian Academy of Sciences, 119017 Moscow, Russia
[10]Department of Zoology, Swedish Museum of Natural History, 10405 Stockholm, Sweden
[11]Peter the Great Museum of Anthropology and Ethnography, Kunstkamera, Russian Academy of Sciences, 199034 Saint-Petersburg, Russia
[12]North-East Interdisciplinary Scientific Research Institute N.A. Shilo, Far East Branch, Russian Academy of Sciences (NEISRI FEB RAS), 685000 Magadan, Russia
[13]Center for Evolutionary Hologenomics, GLOBE Institute, Faculty of Health and Medical Sciences, 1353 Copenhagen, Denmark
[14]University Museum NTNU, 7012 Trondheim, Norway
[15]Natural History Museum, SW7 5BD London, UK
[16]Department of Geological Sciences, Stockholm University, 11418 Stockholm, Sweden
[17]Science for Life Laboratory, 17165 Stockholm, Sweden
[18]These authors contributed equally
[19]Twitter: @indianadiez
[20]Twitter: @love_dalen
[21]Lead contact
*Correspondence: diez.molino@gmail.com (D.D.-d.-M.), love.dalen@zoologi.su.se (L.D.)
https://doi.org/10.1016/j.cub.2023.03.084

SUMMARY

Ancient genomes provide a tool to investigate the genetic basis of adaptations in extinct organisms. However, the identification of species-specific fixed genetic variants requires the analysis of genomes from multiple individuals. Moreover, the long-term scale of adaptive evolution coupled with the short-term nature of traditional time series data has made it difficult to assess when different adaptations evolved. Here, we analyze 23 woolly mammoth genomes, including one of the oldest known specimens at 700,000 years old, to identify fixed derived non-synonymous mutations unique to the species and to obtain estimates of when these mutations evolved. We find that at the time of its origin, the woolly mammoth had already acquired a broad spectrum of positively selected genes, including ones associated with hair and skin development, fat storage and metabolism, and immune system function. Our results also suggest that these phenotypes continued to evolve during the last 700,000 years, but through positive selection on different sets of genes. Finally, we also identify additional genes that underwent comparatively recent positive selection, including multiple genes related to skeletal morphology and body size, as well as one gene that may have contributed to the small ear size in Late Quaternary woolly mammoths.

INTRODUCTION

The evolution of mammoths (genus *Mammuthus*) was characterized by a series of morphological transitions defined by increasing specialization to life in cold high-latitude environments with open landscapes and grassy vegetation. This process culminated with the evolution of the woolly mammoth (*Mammuthus primigenius*), which originated in northeastern Siberia during the early stages of the Middle Pleistocene, approximately 700 thousand years ago (kya), and had become extinct by the onset of the Holocene (~12 kya) across the vast majority of its range.[1,2]

The woolly mammoth had a Holarctic distribution and inhabited terrestrial environments up to 80 degrees north, even during full glacial conditions.[2] Compared to both its extant elephant relatives as well as earlier members of *Mammuthus*, it

was uniquely adapted to life in the high Arctic. The exceptional preservation of woolly mammoth remains recovered from permafrost deposits has enabled scientists to identify a wide range of morphological adaptations, such as thick woolly fur, small ears, short tail, and considerable fat deposits.[3] Moreover, genetic analyses have hinted at previously unknown physiological adaptations to the Arctic environment, including genes related to thermal sensation and hemoglobin structure.[4–6] However, recent work has indicated that only a small subset of these adaptations was unique to the woolly mammoth compared to its million-year-old ancestors.[7] Moreover, the small number of mammoth genomes sequenced to date has precluded confident identification of derived mutations that were fixed in the woolly mammoth lineage.

To address this, we here analyze a dataset comprising 22 Late Quaternary (the period encompassing the Late Pleistocene and Holocene) woolly mammoth genomes, an early Middle Pleistocene (700 kya) genome from one of the earliest known woolly mammoths,[7] and 28 genomes from two extant elephant species. We identify genetic variants that had become fixed in the woolly mammoth lineage both prior to 700 kya, around the time that the woolly mammoth originated, and up to the final stages of the last glaciation (i.e., by 50 kya). Based on the woolly mammoth's presence in the high Arctic for hundreds of thousands of years, we hypothesize that a marked proportion of these fixed variants are related to the unique morphology, fat storage and metabolism, thermosensation, and circadian rhythm of the woolly mammoth.

RESULTS

Genome sequencing and variant dataset

We generated 16 new woolly mammoth genomes from specimens collected throughout Eurasia, all of which have been radiocarbon-dated (Table S1). After filtering, these range from 2.3 to 28.6× (mean of 8.0×) in genome coverage (Figure S1). We also generated additional genomic data for an ~700,000-year-old woolly mammoth sample (Chukochya[7]), increasing its coverage to 2.8×. We then merged our genomes with a panel of previously published proboscidean genomes. The final dataset consists of 23 woolly mammoth genomes (22 Late Quaternary and one Middle Pleistocene), with 13 at medium coverage (2.3–4.1×) and 10 at high coverage (10.4–28.6×), together with high-coverage genomes for seven Asian (*Elephas maximus*) and 21 African savannah (*Loxodonta africana*) elephants. Finally, we included the previously published genome of an American mastodon (*Mammut americanum*) for the phylogenetic analyses. Overall, our final dataset comprises 52 proboscidean genomes (Table S2). We excluded Chukochya from the selection analyses involving variants fixed in woolly mammoths due to its deep age compared to the Late Quaternary woolly mammoth genomes. After variant calling and filtering (STAR Methods), our dataset resulted in 58,913,457 high-quality variants segregating from the African savannah elephant reference genome.

Elephantid phylogeny

To confirm that our woolly mammoth genomes form a monophyletic clade sister to Asian elephants to the exclusion of African elephants, and that Chukochya is basal to all other woolly mammoths, we first constructed a genome-wide neighbor-joining (NJ) phylogeny of all 52 proboscideans based on identical-by-state (IBS) distances. The results from this analysis corroborate the established elephantid phylogeny and confirm that woolly mammoths, Asian elephants, and African savannah elephants are all monophyletic clades, with 100% bootstrap support at the relevant nodes (Figure 1A). Chukochya is placed as basal to all other woolly mammoths, which is consistent with its age and previous results based on a more limited genomic dataset.[7] We note that the phylogenetic distribution of Late Quaternary woolly mammoths appears to be independent of genome coverage, geography, and age (see also Figure S1). We next estimated the proportion of genomic windows for which the majority of woolly mammoth alleles match the Asian elephant allelic state to the exclusion of the African savannah elephant, using different window sizes. We find that 98% of the woolly mammoth genome most closely matches the Asian elephant for large window sizes (>100 kb), whereas this percentage falls to 65.8% for single sites (Figures 1B and 1C). This implies that woolly mammoths and African savannah elephants share the ancestral allele at 34.2% of the single sites, which likely originate from either *de novo* mutations exclusive to the Asian elephant branch or incomplete lineage sorting of ancestral polymorphisms. These results highlight the importance of including multiple genomes from two outgroup species to identify mutations unique to the woolly mammoth branch.

GO enrichment

In order to test whether the complete list of genes with derived mutations unique to the woolly mammoth (n = 3,097) is enriched for some relevant functional categories, we performed gene ontology (GO) term enrichment analyses using the software GOrilla[8] and a background set of all coding genes annotated in the African savannah elephant genome. We found 26 GO terms enriched with $p \leq 0.0001$, all with a fold enrichment higher than 1.29 (Figure 2; Table S3). These include expected GO categories related to changes in hair and skin (e.g., cornification, keratinization), fat metabolism (e.g., chylomicron remnant clearance, plasma lipoprotein particle remodeling), DNA repair (e.g., double-strand break repair via homologous recombination, recombinational repair), immune response (leukotriene transport), protein and sugar metabolism/function (e.g., glycosaminoglycan metabolic process, protein activation cascade), and general cell structure and function (e.g., multicellular organismal homeostasis, cytoskeleton organization). To compare these results, we performed the same analyses on the derived alleles unique to the Asian elephant genomes. We find a different set of GO terms enriched in the Asian elephant (n = 19, $p \leq 0.0001$, fold enrichment > 1.31; Table S4), including categories related to eye function, intestinal function, DNA replication and cell division, cell structure and functioning, and brain and nervous system (Figure 2).

Genes highly evolved in the woolly mammoth

In order to investigate molecular adaptations in the woolly mammoth, we annotated all variants segregating from the African savannah elephant. We defined variants as unique in the woolly mammoth if they are derived with respect to the

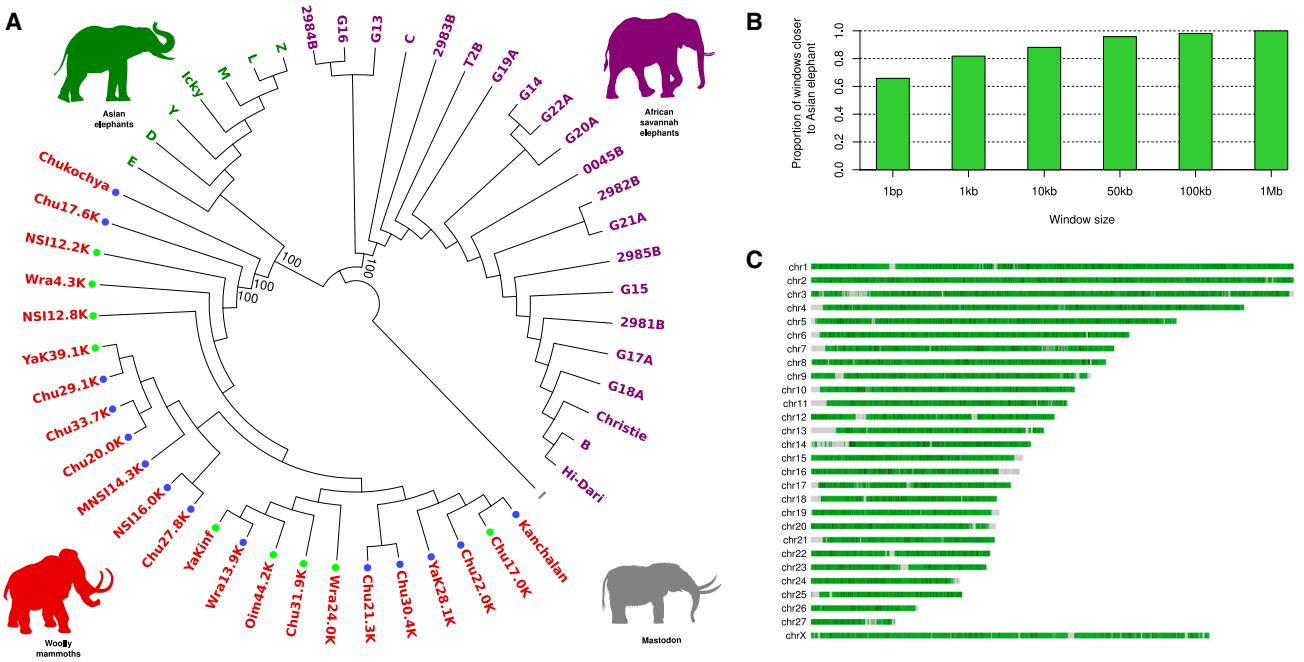


Figure 1. Elephantid phylogeny and genomic window matching

(A) Whole-genome phylogenetic reconstruction of all the elephantid genomes in the dataset based on identical-by-state (IBS) distances. Bootstrap support percentages for the relevant nodes (species defining and placement of Chukochya) are shown. For the woolly mammoths, blue dots represent medium-coverage genomes (2.3–4.1×), while green dots represent high-coverage genomes (>10.4×) (see Tables S1 and S2).

(B) Proportion of genomic windows in which the majority of woolly mammoth alleles match the Asian elephant allele state, for different window sizes.

(C) Distribution of woolly mammoth genomic regions matching the Asian elephant (green) and African savannah elephant (black), for a window size of 10 kb. Gray areas represent windows with less than 10 SNPs for which the African and Asian elephant genomes have different alleles. Each horizontal bar represents a chromosome in the African savannah elephant reference genome.

reference genome and where all African savannah and Asian elephant genomes have only the ancestral allele. We only included variants for which at least half of the woolly mammoth genomes ($n \geq 11$) and half of the elephant genomes ($n \geq 14$) have the site covered. After filtering, we retrieved 1,176,471 high-quality variants fixed for the derived allele and unique to woolly mammoths and checked whether these variants are located within genes. We found 3,097 genes containing fixed-derived non-synonymous mutations (from here on referred to as "FdNs" mutations) that are unique to the woolly mammoth (Table S5). We used SIFT scores to determine the probability of each non-synonymous mutation affecting protein function. To obtain an estimate of the overall impact of all the FdNs mutations in each gene, we summed the SIFT scores of all mutations per gene into a global score called "aggregated SIFT score" (STAR Methods). Finally, we ranked these genes based on the count of non-synonymous mutations and discuss genes with six or more FdNs variants, and an aggregated SIFT score greater than three, in the woolly mammoth genome (Figure 3). Among the mammoth genes enriched for FdNs mutations with predicted high impact on protein function, we found several that are associated with phenotypes such as hair growth, fat storage, lipid metabolism, immune response, modulation of thermal sensation, DNA repair, and reproduction, as well as other metabolic pathways such as protein and sugar function and synthesis (Table 1).

Main phenotypes

Hair and skin development

One of the most distinctive features of the woolly mammoth is its thick fur. Our analyses revealed an enrichment of GO terms associated with hair and skin development (Figure 2; Table S3). In mammals, there are several dozens of genes associated with hair and skin development,[9–11] and a key question is therefore which of these genes have been positively selected in the mammoth lineage. In our dataset, the gene with the second-highest number of FdNs mutations (n = 14) is *AHNAK2* (Table 1), a gene involved in calcium signaling that in humans has an enhanced expression in skin and has been implicated in hair follicle development.[12] Additional genes with multiple FdNs mutations that are highly evolved in mammoths and play a role in hair/skin development include *KRT8* with nine FdNs mutations[13] and *FLG* with seven FdNs mutations.[14] We also note that woolly mammoths have four FdNs mutations in the *LYST* gene (Table S5), which has been implicated in the lack of pigmentation in polar bears[15] as well as beige/lighter coat color in mice and cattle.[16] Although speculative, we hypothesize that the inferred amino acid changes in *LYST* may partly explain the variety of colors observed in different types of woolly mammoth hairs, ranging from colorless to reddish and brown.[17]

Interestingly, we also find that seven out of eight genes associated with abnormal hair development in humans have FdNs mutations in the woolly mammoth. This includes uncombable

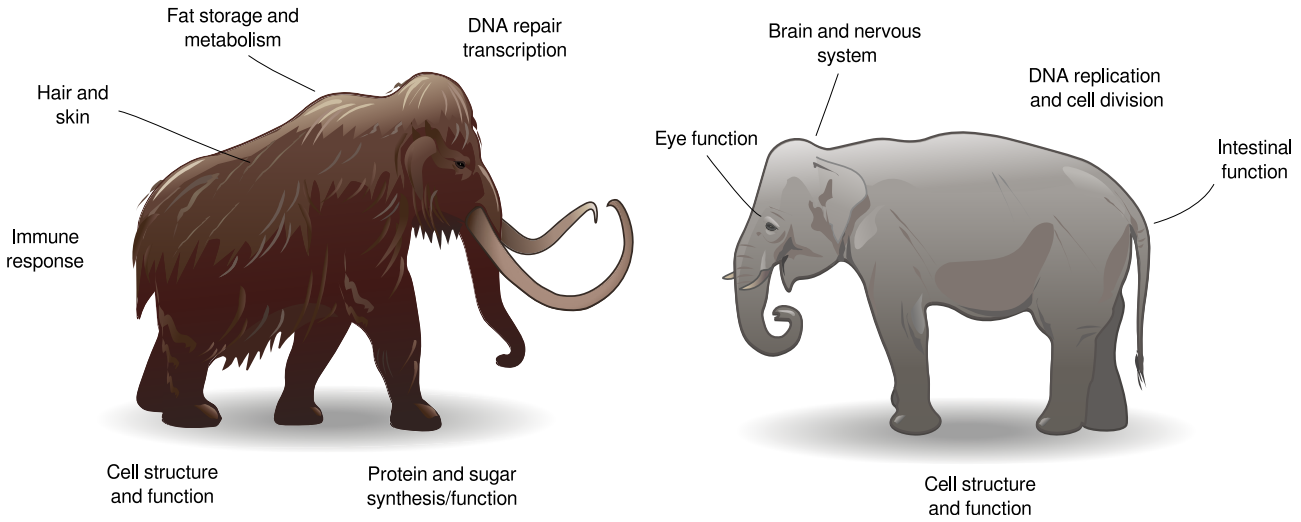


Figure 2. Categories of GO terms significantly enriched in woolly mammoths and Asian elephants

GO term categories based on enrichment of all genes for which either the woolly mammoth (left) or the Asian elephant (right) genomes carry one or more FdNs variants, using annotated African savannah elephant genes as a reference. All categories are based on GO terms enriched in each species with $p \leq 0.0001$. See Tables S3 and S4 for the complete lists of GO terms and their classification by category.

hair syndrome, a genetic disease that is associated with the genes *TCHH*, *PADI3*, and *TGM3*,[18,19] where we find that the first two of these genes have FdNs mutations in the woolly mammoth. Moreover, the main candidate gene for the human genetic disorder called Carvajal syndrome (also known as "woolly hair syndrome") is the gene *DSP*,[20] which has one FdNs mutation in the woolly mammoth. Similarly, we find that the woolly mammoth has FdNs mutations in genes *TP63*, *VPS13B*, *AFF4*, and *SPINK5*. In humans, mutations in these genes cause Rapp-Hodgkin syndrome (wiry, slow-growing, and uncombable hair),[21] Cohen syndrome (thick, bushy hair),[22] CHOPS syndrome (coarse and curly hair),[23] and Netherton syndrome (bamboo hair),[24] respectively.

Fat storage and metabolism

The ability to efficiently metabolize and store lipids is of essential importance for Arctic mammals due to the prolonged lack of food availability during winter.[25] We identified four GO terms related to fat metabolism that are significantly enriched in the woolly mammoth (Table S3). One of the highly evolved genes in the woolly mammoth is *ACADM*, with six FdNs mutations (Figure 3; Table 1). This gene is important for breaking down medium-chain fatty acids and amino acids and has been found to be highly expressed in the liver in cattle.[26] We also find that the woolly mammoth variant of gene *TET1* has five FdNs mutations. Although *TET1* has been implicated in a wide range of functions, including demethylation and cancer resistance,[27] a recent study in mice has shown that *TET1* is an important gene acting as a suppressor of key thermogenic genes, such as the transcriptional coactivator *Ppargc1*α and uncoupling gene *UCP1* that is expressed in beige adipocyte tissue.[28] Beige adipocyte tissue forms in response to cold exposure and results in the increased production of body heat by thermogenesis of stored fat. Reduced expression of *TET1* leads to increased mitochondrial respiration in beige adipocytes and thus improves cold tolerance. We also identified five FdNs mutations in the gene *ACAD10*, which encodes for a protein that catalyzes the oxidation of fatty acyl-CoA derivatives and has been implicated in diabetes in humans.[29] It has also been shown that mice that are deficient in ACAD10 have elevated insulin levels and abnormal glucose tolerance, leading to increased weight gain due to excess fat storage.[30]

Genes related to fat metabolism and storage have also been identified in other Arctic mammals. For example, previous studies have highlighted the gene *APOB* as under positive selection in both reindeer (*Rangifer tarandus*) and polar bears (*Ursus maritimus*), a gene important for lipoprotein transport.[15,31] Our results show that woolly mammoths had three FdNs mutations in *APOB*, further strengthening the hypothesis of this gene being under convergent evolution in Arctic mammals. Moreover, woolly mammoths had one FdNs mutation in *COL5A3*, which has also been highlighted as under positive selection in polar bears.[15] *COL5A3* is a gene that encodes for a glycoprotein that is essential for adipocyte development and is highly expressed in fat tissues.[32] Finally, woolly mammoths also had three FdNs mutations in *FASN*, a gene that encodes for an enzyme responsible for fatty acid synthesis.[33] In earlier studies, *FASN* has been identified as having unique mutations in both reindeer and an Antarctic penguin (*Pygoscelis adeliae*).[31,34]

Protein and sugar synthesis/function

In our functional enrichment analysis (Table S3), two significant GO terms were associated with the synthesis and breakdown of glycosaminoglycans, which are polysaccharide compounds essential for the generation of various mammalian tissues including bone, cartilage, and skin.[35] In addition, we identified a significant GO term related to protein activation cascades, which is a set of reactions that lead to the generation of mature proteins. *VCAN* has eight FdNs mutations (but a SIFT score of 0; Figure 3) and encodes for Versican, a sulfate proteoglycan that is a part of the extracellular matrix in the brain and blood

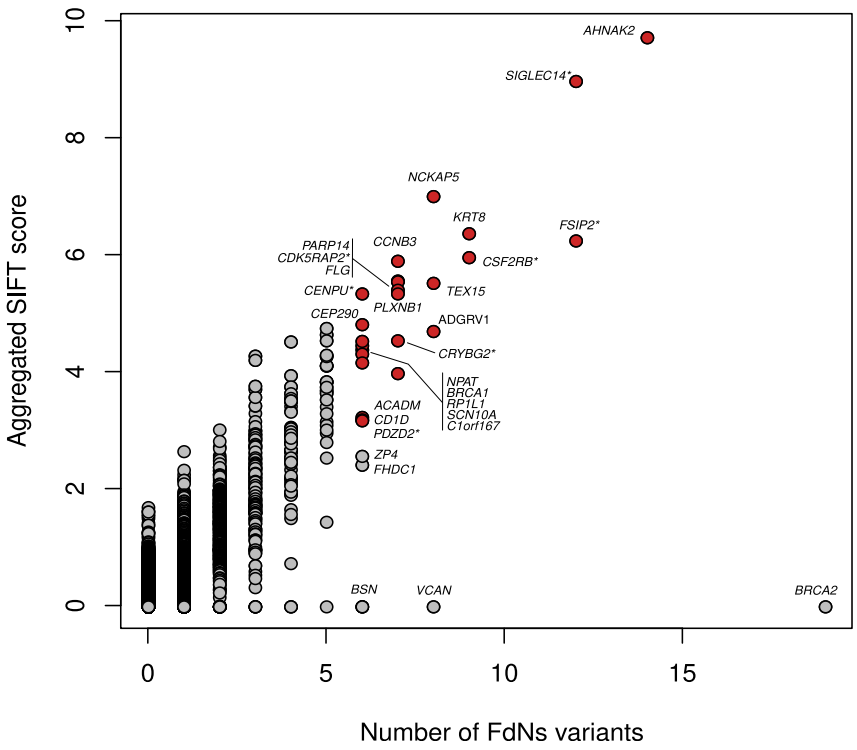


Figure 3. Distribution of woolly mammoth genes arranged by the number of fixed-derived non-synonymous (FdNs) mutations and aggregated SIFT score

Genes with six or more FdNs variants and aggregated SIFT score > 3 are highlighted in red (Table 1). Genes with six or more FdNs variants are labeled. Genes denoted with an asterisk have no name in the African savannah elephant genome annotation but have a match with named genes in the human genome (Table S5).

vessels and plays a role in eye function.[36,37] Related to sugar metabolism we find the gene *PDZD2*, with six FdNs mutations. Albeit not explicitly annotated in the elephant's genome, this gene has a 74.6% match in the human genome annotation. In mice, *PDZD2* is involved in the regulation of pancreatic β cell function and insulin production, hence modulating the levels of glucose in the blood.[38] Interestingly, in humans, *PDZD2* has also been related to fat tissue and mitochondrial function,[39,40] suggesting that mammoths might have evolved a complex metabolic framework that involved interactions between nutrition, fat storage, and heat production.

Immune response

We identified immune response through leukotriene transport as a significant GO term in the enrichment analysis (Table S3). Leukotrienes are part of the immune system mediated through inflammatory reactions. In addition, we also identified several highly evolved genes in our dataset that are associated with the immune system (Table 1). The gene *PARP14* (seven FdNs mutations) promotes differentiation of T helper 2 cells, which are an essential part of the immune system,[41] in particular as a response against extracellular pathogens such as parasitic worms. Our analysis also identified *CD1D*, with six FdNs mutations, as highly evolved in the woolly mammoth. *CD1D* encodes for a molecule that presents antigens to natural killer T cells, which eliminate invasive microorganisms by recognizing microbial lipid agents.[42] We also note that one of the genes with the highest number of FdNs mutations (n = 12) and the highest aggregated SIFT score in our dataset (9.73) has an undetermined annotation but has a 50% match to *SIGLEC14* in the human genome (Table 1). Expression of *SIGLEC14* promotes activation of the inflammasome, thus leading to activation of macrophages.[43]

DNA repair and transcription

Our analyses identified several GO terms associated with DNA repair and transcription (Table S3). The most evolved gene in our entire dataset is *BRCA2*, which has 19 FdNs mutations (Table S5), although it should be noted that its aggregated SIFT score is zero, suggesting that these 19 mutations may all have small effects on protein function. *BRCA1* is also highly evolved in woolly mammoths, with six FdNs mutations, and in contrast to *BRCA2* has a high aggregated SIFT score (4.47). In humans, *BRCA1* and *BRCA2* are well known for their association with breast cancer and play a key role in DNA repair and transcription, acting as tumor-suppression genes.[44] We note that Asian elephants also have a large number of FdNs mutations for these genes (12 and four for *BRCA2* and *BRCA1*, respectively; Table S6), indicating that rapid evolution of *BRCA* genes is common among elephantids, which would be consistent with earlier work on cancer resistance in elephants.[45] In humans, tumor mutations in both *BRCA* genes have been associated with the gene *TP53*, a well-known multicopy tumor suppressor gene in elephants.[46] In addition to the *BRCA* genes, we also identify that the gene *ZNF292*, which has five FdNs mutations in woolly mammoths, encodes for a growth hormone-dependent transcription factor and has been classified as a tumor suppressor gene in humans.[47]

Thermosensation

Earlier work has highlighted the importance of thermosensation genes for cold adaptation in the woolly mammoth.[4] We here identify a gene called *SCN10A* with six FdNs mutations (Table 1), which has not been identified in earlier studies of woolly mammoths. *SCN10A* encodes for a protein forming a Nav1.8 voltage-gated sodium channel that is important in pain sensation[48] and essential for the perception of extreme cold as being painful.[49] In contrast to *SCN10A*, genes belonging to the *TRP* channel gene family are activated and inactivated at different temperatures[50] and have previously been implicated in thermosensation in the woolly mammoth.[4] It is therefore tempting to speculate that a combination of positive selection in both *SCN10A* and the *TRP* channel gene family may have interacted to provide woolly mammoths with a full repertoire of cold sensation adaptations. Lynch et al.[4] identified five *TRP* genes in which woolly mammoths had non-synonymous mutations. Based on our larger dataset, we show that one of these, *TRPV4*, does

Table 1. Highly evolved genes in the woolly mammoth

Gene name	Fixed derived synonymous	Fixed derived non-synonymous	Aggregated SIFT score	Fixed derived non-synonymous and evolved <700 kya	Predicted phenotype affected
AHNAK2	8	14	9.73	0 (12)	hair and/or skin development
SIGLEC14[a]	6	12	8.98	0 (6)	immune system
FSIP2[a]	4	12	6.26	0 (12)	sperm development/function
KRT8	0	9	6.38	0 (9)	hair and/or skin development
TEX15	4	8	5.53	0 (6)	sperm development/function
ADGRV1	5	8	4.71	0 (7)	hearing
CCNB3	1	7	5.91	0 (5)	pregnancy and embryo development
PARP14	3	7	5.57	1 (7)	immune system
CDK5RAP2[a]	0	7	5.56	0 (6)	brain size
FLG	2	7	5.35	5 (5)	ear size and hair and skin development
BRCA1	1	6	4.47	0 (4)	DNA repair and tumor suppression
SCN10A	11	6	4.39	0 (6)	thermosensation
RP1L1	1	6	4.32	0 (4)	vision
ACADM	0	6	3.24	1 (3)	fat storage and metabolism
CD1D	2	6	3.2	0 (6)	immune system
PDZD2[a]	1	6	3.2	0 (5)	sugar metabolism
CD4	1	5	3.85	4 (5)	immune system
TSHZ3	3	5	3.75	2 (5)	urinary tract and kidney development
ABCC11[b]	3	5	3.16	2 (5)	ear wax and body odor
FIGNL1	0	4	1.96	4 (4)	skeletal morphology and body size
PRSS8	5	3	4.29	3 (3)	hair and/or skin development
TCHH	1	3	2.63	2 (3)	hair and/or skin development
FAM214A	0	3	1.74	3 (3)	skeletal morphology and body size
OLFR94[a]	0	3	1.33	3 (3)	olfaction
TMC4	0	3	1.23	3 (3)	salt taste sensitivity
PCLO	2	3	0.98	3 (3)	skeletal morphology and body size
CD3E	1	2	1.77	2 (2)	immune system
IGSF6	2	2	1.57	2 (2)	immune system
PXMP4	0	2	1.47	2 (2)	fat storage and metabolism
ADRB2	0	2	1.03	2 (2)	fat storage and metabolism
NMI	0	2	0.67	2 (2)	skeletal morphology and body size
ARHGAP21	1	2	0.57	2 (2)	skeletal morphology and body size

The 32 genes listed are either those that have a predicted effect on a specific phenotype category, with at least six fixed-derived non-synonymous mutations and an aggregated SIFT score >3 (n = 16), or genes with a specific predicted phenotype where at least two non-synonymous mutations have evolved in the last 700 ka (n = 17). Note that the FLG gene falls into both categories. The number of sites covered in the 700,000-year-old Chukochya genome is shown in parentheses. Genes with unknown, broad, or multiple effects on phenotype (n = 29) are not shown. A full list comprising all genes in the dataset can be found in Tables S5, S6, and S7.

[a]Genes that are not part of the elephant annotation but have a match to genes in the human genome

[b]This gene had a derived loss-of-function mutation fixed in all woolly mammoths

not have any FdNs mutations in the woolly mammoth. Moreover, our results show that only two of the four non-synonymous mutations in *TRPV3* identified by Lynch et al.[4] were fixed in Late Quaternary woolly mammoths. On the other hand, our analyses also identified three additional *TRP* genes with FdNs mutations in the woolly mammoth associated with thermosensation (*TRPV2*, *TRPM5*, and *TRPM2*, with one, one, and three FdNs mutations, respectively) that are activated at temperatures above 15°C.[51,52] Overall, we find that our woolly mammoths have FdNs mutations in eight out of the ten *TRPM*, *TRPV*, and *TRPA* genes identified, implying that woolly mammoths displayed polygenic adaptive evolution to thermosensation.

Reproduction

Among our highly evolved genes in the woolly mammoth, three are associated with reproductive processes (Table 1). One of these, *TEX15* (eight FdNs mutations), is associated with spermatogenesis in humans.[53] Moreover, we identified a gene with 12 FdNs mutations that lacks functional annotation in the African savannah elephant genome but is a 68.3% match to the *FSIP2* gene in the human genome. In humans, *FSIP2* is associated

with repeated abnormalities in sperm flagella.[54] We also find that woolly mammoths had seven FdNs mutations in the gene *CCNB3*, which has been associated with pregnancy loss in mice.[55] Similarly, we identified six FdNs mutations in *ZP4* (SIFT score of 2.57), an essential gene for embryo development in rabbits.[56] Although it is difficult to ascertain the exact reproductive function of these highly evolved genes in the woolly mammoth, we hypothesize that some of these changes may have been related to the transition from a temperate to an Arctic environment, which would entail adapting to a higher degree of seasonality in terms of mating and parturition.

Other phenotypes

Four of the most highly evolved genes in the woolly mammoth are associated with brain function, vision, and hearing (Table 1). For example, gene expression analyses of *FHDC1* in mice (six FdNs mutations in the woolly mammoth) suggest that this gene is involved in physiological processes such as tissue repair, synapse formation, and synapse maintenance, and therefore might be important for brain function.[57] Also, we identified seven FdNs mutations in a gene with unknown annotation in the woolly mammoth, but that has a 75.9% match to the *CDK5RAP2* gene in the human genome. *CDK5RAP2* is a gene that has been identified as controlling human brain size.[58] In addition, we observed six FdNs mutations in *RP1L1*, which is a gene that is exclusively expressed in retinal photoreceptors and is an important component of the photoreceptor cilium.[59] Our results also show that the woolly mammoth had eight FdNs mutations in *ADGRV1*. This gene has been associated with hearing loss disorders in both humans and mice.[60,61] Finally, although not included in our top list of highly evolved genes, we note that there are five FdNs mutations in the gene *ABCC11* (Table S5). As shown in earlier studies,[6] one of the mutations in the woolly mammoth *ABCC11* gene is a stop-gain (loss-of-function) mutation. In humans, the dry wax phenotype, which also is associated with reduced body sweat odor, is caused by a mutation that causes a misfolded protein.[62] The loss-of-function woolly mammoth mutation could thus have led to a similar phenotype.

Adaptive evolution in the last 700 ka

A previous analysis of the 700,000-year-old Chukochya genome, representing one of the earliest known woolly mammoths, indicated that this individual had 88.7% of the protein-coding changes found in Late Quarternary woolly mammoths.[7] Based on our analysis of additional genomic data from Chukochya and FdNs mutations in a much larger panel of Late Quaternary woolly mammoths, we here revise this estimate to 91.7% (3,644 out of 3,972 sites). Conversely, this implies that 8.3% of the non-synonymous mutations in Late Quaternary woolly mammoths became fixed in the last 700 thousand years (ka), providing an opportunity to identify specific genes that underwent adaptive changes after the origin of the woolly mammoth as a species. We identified 30 genes with at least two FdNs mutations in woolly mammoths that became fixed in the last 700 ka, suggesting the presence of important adaptive change during this period (Table 1). The gene that has accumulated the most protein-coding changes in the last 700 ka was *FLG* (five FdNs mutations), a gene that encodes for the protein filaggrin that binds to keratin fibers in epithelial cells. Intriguingly, *FLG*-deficient mice develop significantly smaller ears.[63] We therefore hypothesize that the mutations observed in the *FLG* gene during the last 700 ka may have contributed to the distinct small ear size in Late Quaternary woolly mammoths. This would also imply that the earliest woolly mammoths and their ancestors had larger ears than their Late Quaternary descendants.

We also identified several genes associated with hair growth and structure that have undergone adaptive evolution during the last 700 ka. For example, we identified three recently evolved non-synonymous mutations in *PRSS8*, a gene that is important for hair follicle development and where non-synonymous mutations are responsible for the mutant "frizzy" phenotype in mice.[64] In addition, we find that two of the three FdNs mutations in *TCHH*, one of the genes responsible for uncombable hair syndrome in humans, evolved in the last 700 ka. We also find evidence for the recent evolution of one of the two FdNs mutations in *KRTAP4-1*, which encodes for a protein that is essential for the formation of a rigid and resistant hair shaft in mammals.[65] Taken together, these findings suggest that the development of the woolly mammoth's distinctive pelage was a continuous process during the Pleistocene.

Moreover, our analysis suggests that several genes related to skeletal morphology and body size have been under positive selection in the last 700 ka (Table 1), and we hypothesize that some of these may be related to the observed decrease in mammoth body size in the last million years.[66] For example, the genes *FIGNL1* and *NMI* (from zero to four and two FdNs mutations in the last 700 ka, respectively) are important for osteoblast differentiation and bone density.[67,68] We also find that the genes *PCLO* and *FAM214A* have changed from the ancestral elephant allele by gaining three FdNs mutations each in the last 700 ka. Mutations in *PCLO* are associated with decreased body size in mice,[69] whereas expression levels in *FAM214A* affect the waist-hip ratio in humans.[70] Finally, we find that the gene *ARHGAP21*, which is associated with mandibular prognathism in humans,[71] has accumulated its only two FdNs mutations in the last 700 ka.

We also identified two genes associated with fat storage and metabolism that have undergone protein-coding changes in the last 700 ka, from being identical to those in elephants to having gained two FdNs mutations each (Table 1). One of these is *PXMP4*, which in mice affects levels of fatty acid alkyl-diacylglycerol lipids in the liver, and thus likely plays a role in lipid metabolism.[72] The second gene, *ADRB2*, is of key importance for the regulation of fat metabolism in adipose tissues, where two allele variants, "thrifty" and "energy expense," have been linked to obesity levels in humans.[73]

Finally, our analysis shows that several immunity-related genes have evolved in the last 700 ka. The gene *CD4*, which encodes for a T cell antigen that is an essential component in cell-mediated immunity,[74] has gained four of its five FdNs mutations in the woolly mammoth during the last 700 ka. Among humans and chimpanzees/bonobos, *CD4* differs by 3–4 amino acid changes and has been suggested to have been under positive selection in primates, possibly in response to viral infections.[75] Our results could thus imply that the four recently evolved FdNs mutations in the woolly mammoth are the result of rapid positive selection, possibly as a response to novel pathogens. Two additional immunity-related genes, *IGSF6* and

CD3E, have gained two FdNs mutations each in the last 700 ka. *IGSF6* is a member of the immunoglobulin superfamily that is predicted to play a role in immune response.[76] *CD3E* encodes for a polypeptide that plays an essential role in T cell development that, similarly to *CD4*, has been under positive selection during primate evolution, with 1–2 amino acid differences between humans and chimpanzees/bonobos.[77]

DISCUSSION

Our study builds on an extensive number of modern elephant and ancient mammoth genomes, including one from the early Middle Pleistocene, and provides new insights into the genetic basis for adaptive evolution in the woolly mammoth. The GO analysis revealed an enrichment of genes associated with the development of hair, skin, and fat metabolism in the woolly mammoth, something that we did not observe in our corresponding analysis of the Asian elephant genome. A recent study based on detecting genomic deletions in four woolly mammoth genomes identified private deletions in genes enriched for phenotypes such as fat distribution and hair growth and shape,[78] strengthening the evidence that these are molecular phenotypes unique to mammoths. In addition, our comprehensive dataset of woolly mammoth genomes allowed us to identify genes with multiple fixed-derived non-synonymous mutations and a predicted high impact on protein function. We find that many of these genes are associated with possible adaptations to a cold environment, such as changes to hair pelage, fat metabolism, and thermosensation, as well as several other physiological phenotypes including immune system function, reproduction, and DNA repair mechanisms.

Furthermore, the inclusion of the genome from one of the earliest known woolly mammoths enabled us to discriminate between mutations that arose in earlier forms of *Mammuthus* and those that evolved during the last 700 ka. This analysis offers a tantalizing glimpse of specific genes that in other mammals have been associated with changes in ear size, pelage, skin, body size, fat storage, and metabolism, as well as immunity, which in the woolly mammoth underwent adaptive evolution from the onset of the Middle Pleistocene and thereafter. Interestingly, we note that for several phenotypic traits associated with genes that have experienced adaptive changes since the origin of the *Mammuthus* lineage, different sets of genes appear to have been under positive selection before and after 700 kya. For example, both *AHNAK2* and *KRT8* are associated with hair and skin development and gained 14 and nine FdNs mutations prior to 700 kya, but none after this date. Instead, we find a signature of positive selection in two other genes associated with hair and skin development after 700 kya, *PRSS8* and *TCHH*, which gained three and two FdNs mutations, respectively. Similar patterns are observed in genes related to immune system function, as well as fat storage and metabolism (Table 1). This suggests that evolutionary changes involved in characteristic woolly mammoth adaptations were not governed by single mutations in individual genes, but rather by multiple mutations in several different interrelated genes or combinations of genes. However, instead of changing in concert, it appears that the genes involved in the same phenotype evolved in a temporally stepwise manner.

This study also highlights the advantage of including multiple genomes from the same species when identifying genes that have been subject to adaptive evolution. Positive selection on *de novo* mutations is expected to lead to comparatively rapid fixation of derived alleles,[79] implying that we would expect most adaptive variants that have evolved since mammoths shared an ancestor with elephants to be fixed across all Late Pleistocene woolly mammoths. To identify genes that have evolved through such positive selection, it is therefore important to include a sufficiently large number of genomes to distinguish between polymorphic and fixed variants. In our 22 Late Quaternary woolly mammoth genomes, we identified a total of 21,249,732 sites with derived mutations, but only 1,176,471 (i.e., 5.5%) of these were fixed among all individuals.

Our findings open up several interesting questions that could be addressed in future studies. For example, we hypothesize that the observed non-synonymous mutations could lead to reduced expression in *FLG*, causing small ear size. Similarly, we hypothesize that the genetic changes observed in *TET1* may have improved mammoth tolerance to cold temperatures. Also, finding a significant change in the function of *ABCC11* could imply that mammoths had dry ear wax and reduced body odor. These hypotheses could be tested by expression and functional characterization of the mammoth variant of these genes in cell lines or transgenic mice. It would also be interesting to conduct *in vitro* experiments to assess whether the observed mutations in *ADRB2* led to a higher rate of fat metabolism or a higher rate of fat storage (i.e., whether the observed mammoth variant was more similar to the "thrifty" or "energy-expense" alleles found in humans). Finally, we note that additional woolly mammoth genomes from different stages of the Middle Pleistocene would enable testing of whether the observed adaptive changes in the last 700 ka happened gradually or in a more punctuated manner, for example in response to the repeated climatic shifts that took place during that period, i.e. across multiple glacial cycles.[80]

Limitations of study

We acknowledge that the highly evolved genes with known functional roles identified here provide a conservative and non-exhaustive catalog of the genomic underpinnings of the woolly mammoth phenotype. In our analyses, using the number of FdNs and aggregated SIFT score per gene, we make the assumption that the more evolved the gene, the greater the likelihood of a phenotypic impact. However, as a consequence of our strict criteria, other candidate genes that may have been important for the woolly mammoth phenotype but had fewer FdNs mutations or a lower aggregated SIFT score were not specifically discussed. A list of all genes with at least one FdNs in Late Quaternary woolly mammoths is provided in Table S5. A second limitation in our analyses is that we only identify fixed derived sites in genic regions and therefore do not account for potentially important genomic changes such as in gene copy-number variation, transcription regulation, methylation, or alternative splicing sites.[81] Finally, we identified nine highly evolved genes in the woolly mammoth for which a functional role is either unknown or unclear (Table S5). This includes one unknown gene, three tentatively assigned genes based on their match to the human genome (*CSF2RB*, *CRYBG2*, and *CENPU*), and five

known genes associated with broad or multiple functions (*C1orf167*, *CEP290*, *NCKAP5*, *NPAT*, and *PLXNB1*). For example, *NCKAP5* is one of the most highly evolved genes in our dataset, with eight FdNs mutations, of which three are predicted to be high impact (aggregated SIFT score = 7.02). However, the possible phenotypic role of *NCKAP5* in woolly mammoths is unclear, as this gene has been associated with multiple functions including bipolar disorder, hypersomnia, height, and body mass index in humans,[82] as well as flight behavior in cattle.[83] Our catalog of genes with unknown or unclear functions, therefore, provides prime genomic targets for future studies aimed at disentangling the woolly mammoth phenotype.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Newly sequenced mammoths
  - Dataset
- METHOD DETAILS
  - Laboratory methods
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data processing
  - Variant calling
  - Population genomic analyses
  - Phylogenetic inferences
  - Derived variants and effect-prediction
  - GO term enrichment and phenotype consequences

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.03.084.

ACKNOWLEDGMENTS

D.D.-d.-M., M.D., P.P., and L.D. acknowledge support from the Swedish Research Council (2017-04647 and 2021-00625), Formas (2018-01640), and the Carl Tryggers Foundation (CTS 17:109). T.v.d.V. acknowledges support from the SciLifeLab and Wallenberg Data Driven Life Science Program (KAW 2020.0239). P.D.H. was supported by a Wallenberg Academy Fellowship (KAW 2021.0048). S.V. and G.K.D. were supported by the Russian Science Foundation (project no. 22-27-00082). The authors are grateful to Héloïse Muller for assistance with laboratory analyses, to Illumina for partially co-financing the DNA sequencing, and to Erik Ersmark for the art in Figure 2. The authors also acknowledge support from Science for Life Laboratory (SciLifeLab), the National Genomics Infrastructure (NGI) funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

AUTHOR CONTRIBUTIONS

Conceived the project and designed the study, D.D.-d.-M., M.D., T.v.d.V., and L.D.; collected and provided samples, A.T., P.N., G.K.D., S.V., A.M.L., and L.D.; performed laboratory work in ancient samples, M.D., J.C.C.-D., P.P., and F.K.; processed and mapped the data, D.D.-d.-M. and M.D.; analyzed the data, D.D.-d.-M., M.D., J.C.C.-D., and T.v.d.V.; wrote the manuscript with contributions from all co-authors, D.D.-d.-M., M.D., J.C.C.-D., P.D.H., T.v.d.V., and L.D. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: January 26, 2023
Revised: February 24, 2023
Accepted: March 29, 2023
Published: April 7, 2023

REFERENCES

1. Lister, A.M., and Sher, A.V. (2015). Evolution and dispersal of mammoths across the Northern Hemisphere. Science *350*, 805–809.
2. Dehasque, M., Pečnerová, P., Muller, H., Tikhonov, A., Nikolskiy, P., Tsigankova, V.I., Danilov, G.K., Díez-del-Molino, D., Vartanyan, S., Dalén, L., and Lister, A.M. (2021). Combining Bayesian age models and genetics to investigate population dynamics and extinction of the last mammoths in northern Siberia. Quat. Sci. Rev. *259*, 106913.
3. Boeskorov, G.G., Mashchenko, E.N., Plotnikov, V.V., Shchelchkova, M.V., Protopopov, A.V., and Solomonov, N.G. (2016). Adaptation of the woolly mammoth Mammuthus primigenius (Blumenbach, 1799) to habitat conditions in the glacial period. Contemp. Probl. Ecol. *9*, 544–553. https://doi.org/10.1134/S1995425516050024.
4. Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W., and Schuster, S.C. (2015). Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. Cell Rep. *12*, 217–228.
5. Campbell, K.L., Roberts, J.E.E., Watson, L.N., Stetefeld, J., Sloan, A.M., Signore, A.V., Howatt, J.W., Tame, J.R.H., Rohland, N., Shen, T.-J., et al. (2010). Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. Nat. Genet. *42*, 536–540.
6. Smith, S.D., Kawash, J.K., Karaiskos, S., Biluck, I., and Grigoriev, A. (2017). Evolutionary adaptation revealed by comparative genome analysis of woolly mammoths and elephants. DNA Res. *24*, 359–369.
7. van der Valk, T., Pečnerová, P., Díez-Del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J.A., Dehasque, M., Sağlıcan, E., et al. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature *591*, 265–269. https://doi.org/10.1038/s41586-021-03224-9.
8. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinf. *10*, 48.
9. Duverger, O., and Morasso, M.I. (2014). To grow or not to grow: hair morphogenesis and human genetic hair disorders. Semin. Cell Dev. Biol. *25-26*, 22–33.
10. He, J., Zhao, B., Huang, X., Fu, X., Liu, G., Tian, Y., Wu, C., Mao, J., Liu, J., Gun, S., and Tian, K. (2022). Gene network analysis reveals candidate genes related with the hair follicle development in sheep. BMC Genom. *23*, 428.
11. Edqvist, P.-H.D., Fagerberg, L., Hallström, B.M., Danielsson, A., Edlund, K., Uhlén, M., and Pontén, F. (2015). Expression of human skin-specific genes defined by transcriptomics and antibody-based profiling. J. Histochem. Cytochem. *63*, 129–141.
12. Chovatiya, G., Ghuwalewala, S., Walter, L.D., Cosgrove, B.D., and Tumbar, T. (2021). High-resolution single-cell transcriptomics reveals

heterogeneity of self-renewing hair follicle stem cells. Exp. Dermatol. *30*, 457–471.

13. Schwarz, N., Windoffer, R., Magin, T.M., and Leube, R.E. (2015). Dissection of keratin network formation, turnover and reorganization in living murine embryos. Sci. Rep. *5*, 9007.
14. Weidinger, S., Illig, T., Baurecht, H., Irvine, A.D., Rodriguez, E., Diaz-Lacava, A., Klopp, N., Wagenpfeil, S., Zhao, Y., Liao, H., et al. (2006). Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. J. Allergy Clin. Immunol. *118*, 214–219.
15. Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., et al. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell *157*, 785–794.
16. Reissmann, M., and Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. Semin. Cell Dev. Biol. *24*, 576–586.
17. Tridico, S.R., Rigby, P., Kirkbride, K.P., Haile, J., and Bunce, M. (2014). Megafaunal split ends: microscopical characterisation of hair structure and function in extinct woolly mammoth and woolly rhino. Quat. Sci. Rev. *83*, 68–75.
18. Ü Basmanav, F.B., Cau, L., Tafazzoli, A., Méchin, M.C., Méchin, M.-C., Romano, M.T., Valentin, F., Wiegmann, H., Huchenq, A., Kandil, R., Garcia Bartels, N., et al. (2016). Mutations in three genes encoding proteins involved in hair shaft formation cause uncombable hair syndrome. Am. J. Hum. Genet. *99*, 1292–1304.
19. Méchin, M.C., Takahara, H., and Simon, M. (2020). Deimination and peptidylarginine deiminases in skin physiology and diseases. Int. J. Mol. Sci. *21*, 566. https://doi.org/10.3390/ijms21020566.
20. Al-Owain, M., Wakil, S., Shareef, F., Al-Fatani, A., Hamadah, E., Haider, M., Al-Hindi, H., Awaji, A., Khalifa, O., Baz, B., et al. (2011). Novel homozygous mutation in DSP causing skin fragility-woolly hair syndrome: report of a large family and review of the desmoplakin-related phenotypes. Clin. Genet. *80*, 50–58.
21. Bougeard, G., Hadj-Rabia, S., Faivre, L., Sarafan-Vasseur, N., and Frébourg, T. (2003). The Rapp-Hodgkin syndrome results from mutations of the TP63 gene. Eur. J. Hum. Genet. *11*, 700–704.
22. Rodrigues, J.M., Fernandes, H.D., Caruthers, C., Braddock, S.R., and Knutsen, A.P. (2018). Cohen syndrome: review of the literature. Cureus *10*, e3330.
23. Raible, S.E., Mehta, D., Bettale, C., Fiordaliso, S., Kaur, M., Medne, L., Rio, M., Haan, E., White, S.M., Cusmano-Ozog, K., et al. (2019). Clinical and molecular spectrum of CHOPS syndrome. Am. J. Med. Genet. *179*, 1126–1138.
24. Sarri, C.A., Roussaki-Schulze, A., Vasilopoulos, Y., Zafiriou, E., Patsatsi, A., Stamatis, C., Gidarokosta, P., Sotiriadis, D., Sarafidou, T., and Mamuris, Z. (2017). Netherton syndrome: a genotype-phenotype review. Mol. Diagn. Ther. *21*, 137–152.
25. Blix, A.S. (2016). Adaptations to polar life in mammals and birds. J. Exp. Biol. *219*, 1093–1105.
26. Ji, S., Yang, R., Lu, C., Qiu, Z., Yan, C., and Zhao, Z. (2014). Differential expression of PPARγ, FASN, and ACADM genes in various adipose tissues and longissimus dorsi muscle from yanbian yellow cattle and yan yellow cattle. Asian-Australas. J. Anim. Sci. *27*, 10–18.
27. Liu, W., Wu, G., Xiong, F., and Chen, Y. (2021). Advances in the DNA methylation hydroxylase TET1. Biomark. Res. *9*, 76.
28. Damal Villivalam, S., You, D., Kim, J., Lim, H.W., Xiao, H., Zushin, P.-J.H., Oguri, Y., Amin, P., and Kang, S. (2020). TET1 is a beige adipocyte-selective epigenetic suppressor of thermogenesis. Nat. Commun. *11*, 4313.
29. Bian, L., Hanson, R.L., Muller, Y.L., Ma, L., Baier, L.J., MAGIC Investigators, Kobes, S., Knowler, W.C., and Bogardus, C. (2010). Variants in ACAD10 are associated with type 2 diabetes, insulin resistance and lipid oxidation in Pima Indians. Diabetologia *53*, 1349–1353.
30. Bloom, K., Mohsen, A.-W., Karunanidhi, A., El Demellawy, D., Reyes-Múgica, M., Wang, Y., Ghaloul-Gonzalez, L., Otsubo, C., Tobita, K., Muzumdar, R., et al. (2018). Investigating the link of ACAD10 deficiency to type 2 diabetes mellitus. J. Inherit. Metab. Dis. *41*, 49–57.
31. Lin, Z., Chen, L., Chen, X., Zhong, Y., Yang, Y., Xia, W., Liu, C., Zhu, W., Wang, H., Yan, B., et al. (2019). Biological adaptations in the Arctic cervid, the reindeer (Rangifer tarandus). Science *364*, eaav6312. https://doi.org/10.1126/science.aav6312.
32. Yousuf, S., Li, A., Feng, H., Lui, T., Huang, W., Zhang, X., Xie, L., and Miao, X. (2022). Genome-wide expression profiling and networking reveals an imperative role of IMF-associated novel CircRNAs as ceRNA in Pigs. Cells *11*, 2638. https://doi.org/10.3390/cells11172638.
33. Maier, T., Jenni, S., and Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 å resolution. Science *311*, 1258–1262. https://doi.org/10.1126/science.1123248.
34. Li, C., Zhang, Y., Li, J., Kong, L., Hu, H., Pan, H., Xu, L., Deng, Y., Li, Q., Jin, L., et al. (2014). Two Antarctic penguin genomes reveal insights into their evolutionary history and molecular changes related to the Antarctic environment. GigaScience *3*, 27.
35. Köwitsch, A., Zhou, G., and Groth, T. (2018). Medical application of glycosaminoglycans: a review. J. Tissue Eng. Regen. Med. *12*, e23–e41.
36. Islam, S., and Watanabe, H. (2020). Versican: a dynamic regulator of the extracellular matrix. J. Histochem. Cytochem. *68*, 763–775.
37. Parra, M.M., Spoth, E., Ronquillo, C.C., Henderson, R., and Hartnett, M.E. (2022). Multimodal retinal imaging findings in two cousins with VCAN-related vitreoretinopathy or wagner disease. Ophthalmic Surg. Lasers Imaging Retina *53*, 639–643.
38. Tsang, S.W., Shao, D., Cheah, K.S.E., Okuse, K., Leung, P.S., and Yao, K.-M. (2010). Increased basal insulin secretion in Pdzd2-deficient mice. Mol. Cell. Endocrinol. *315*, 263–270.
39. Chen, X., Ayala, I., Shannon, C., Fourcaudot, M., Acharya, N.K., Jenkinson, C.P., Heikkinen, S., and Norton, L. (2018). The diabetes gene and Wnt pathway effector TCF7L2 regulates adipocyte development and function. Diabetes *67*, 554–568.
40. Arroyo, J.D., Jourdain, A.A., Calvo, S.E., Ballarano, C.A., Doench, J.G., Root, D.E., and Mootha, V.K. (2016). A genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. Cell Metab. *24*, 875–885.
41. Riley, J.P., Kulkarni, A., Mehrotra, P., Koh, B., Perumal, N.B., Kaplan, M.H., and Goenka, S. (2013). PARP-14 binds specific DNA sequences to promote Th2 cell gene expression. PLoS One *8*, e83127.
42. Speak, A.O., Cerundolo, V., and Platt, F.M. (2008). CD1d presentation of glycolipids. Immunol. Cell Biol. *86*, 588–597.
43. Tsai, C.-M., Riestra, A.M., Ali, S.R., Fong, J.J., Liu, J.Z., Hughes, G., Varki, A., and Nizet, V. (2020). Siglec-14 enhances NLRP3-inflammasome activation in macrophages. J. Innate Immun. *12*, 333–343.
44. Narod, S.A., and Foulkes, W.D. (2004). BRCA1 and BRCA2: 1994 and beyond. Nat. Rev. Cancer *4*, 665–676.
45. Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife *5*, e11994. https://doi.org/10.7554/eLife.11994.
46. Voskarides, K., and Giannopoulou, N. (2023). The role of TP53 in adaptation and evolution. Cells *12*, 512. https://doi.org/10.3390/cells12030512.
47. Lee, J.H., Song, S.Y., Kim, M.S., Yoo, N.J., and Lee, S.H. (2016). Frameshift mutations of a tumor suppressor gene ZNF292 in gastric and colorectal cancers with high microsatellite instability. APMIS *124*, 556–560.
48. Xue, Y., Chidiac, C., Herault, Y., and Gaveriaux-Ruff, C. (2021). Pain behavior in SCN9A (Nav1.7) and SCN10A (Nav1.8) mutant rodent models. Neurosci. Lett. *753*, 135844.
49. Zimmermann, K., Leffler, A., Babes, A., Cendan, C.M., Carr, R.W., Kobayashi, J.-I., Nau, C., Wood, J.N., and Reeh, P.W. (2007). Sensory

neuron sodium channel Nav1.8 is essential for pain at low temperatures. Nature *447*, 856–859. https://doi.org/10.1038/nature05880.

50. Foulkes, T., and Wood, J.N. (2007). Mechanisms of cold pain. Channels *1*, 154–160.
51. Talavera, K., Yasumatsu, K., Voets, T., Droogmans, G., Shigemura, N., Ninomiya, Y., Margolskee, R.F., and Nilius, B. (2005). Heat activation of TRPM5 underlies thermal sensitivity of sweet taste. Nature *438*, 1022–1025.
52. Bandell, M., Macpherson, L.J., and Patapoutian, A. (2007). From chills to chilis: mechanisms for thermosensation and chemesthesis via thermoTRPs. Curr. Opin. Neurobiol. *17*, 490–497.
53. Okutman, O., Muller, J., Baert, Y., Serdarogullari, M., Gultomruk, M., Piton, A., Rombaut, C., Benkhalifa, M., Teletin, M., Skory, V., et al. (2015). Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. Hum. Mol. Genet. *24*, 5581–5588.
54. Martinez, G., Kherraf, Z.-E., Zouari, R., Fourati Ben Mustapha, S., Saut, A., Pernet-Gallay, K., Bertrand, A., Bidart, M., Hograindleur, J.P., Amiri-Yekta, A., et al. (2018). Whole-exome sequencing identifies mutations in FSIP2 as a recurrent cause of multiple morphological abnormalities of the sperm flagella. Hum. Reprod. *33*, 1973–1984.
55. Chotiner, J.Y., Leu, N.A., Xu, Y., and Wang, P.J. (2022). Recurrent pregnancy loss in mice lacking the X-linked Ccnb3 gene. Biol. Reprod. *106*, 382–384.
56. Lamas-Toranzo, I., Fonseca Balvís, N., Querejeta-Fernández, A., Izquierdo-Rico, M.J., González-Brusi, L., Lorenzo, P.L., García-Rebollar, P., Avilés, M., and Bermejo-Álvarez, P. (2019). ZP4 confers structural properties to the zona pellucida essential for embryo development. eLife *8*, e48904. https://doi.org/10.7554/eLife.48904.
57. Dutta, P., and Maiti, S. (2015). Expression of multiple formins in adult tissues and during developmental stages of mouse brain. Gene Expr. Patterns *19*, 52–59.
58. Bond, J., Roberts, E., Springell, K., Lizarraga, S.B., Scott, S., Higgins, J., Hampshire, D.J., Morrison, E.E., Leal, G.F., Silva, E.O., et al. (2005). A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. Nat. Genet. *37*, 353–355.
59. Noel, N.C.L., and MacDonald, I.M. (2020). RP1L1 and inherited photoreceptor disease: A review. Surv. Ophthalmol. *65*, 725–739.
60. Faletra, F., Morgan, A., Ghiselli, S., Murru, F.M., and Girotto, G. (2020). Hearing loss and brain abnormalities due to pathogenic mutations in ADGRV1 gene: a case report. Hear. Bal. Commun. *18*, 196–198.
61. Yan, W., Long, P., Chen, T., Liu, W., Yao, L., Ren, Z., Li, X., Wang, J., Xue, J., Tao, Y., et al. (2018). A natural occurring mouse model with Adgrv1 mutation of Usher syndrome 2C and characterization of its recombinant inbred strains. Cell. Physiol. Biochem. *47*, 1883–1897.
62. Toyoda, Y., Sakurai, A., Mitani, Y., Nakashima, M., Yoshiura, K.-I., Nakagawa, H., Sakai, Y., Ota, I., Lezhava, A., Hayashizaki, Y., et al. (2009). Earwax, osmidrosis, and breast cancer: why does one SNP (538G>A) in the human ABC transporter ABCC11 gene determine earwax type? FASEB J *23*, 2001–2013.
63. Pau, H., Fuchs, H., de Angelis, M.H., and Steel, K.P. (2005). Hush puppy: a new mouse mutant with pinna, ossicle, and inner ear defects. Laryngoscope *115*, 116–124.
64. Spacek, D.V., Perez, A.F., Ferranti, K.M., Wu, L.K.-L., Moy, D.M., Magnan, D.R., and King, T.R. (2010). The mouse frizzy (fr) and rat "hairless" (frCR) mutations are natural variants of protease serine S1 family member 8 (Prss8). Exp. Dermatol. *19*, 527–532.
65. Wu, D.-D., Irwin, D.M., and Zhang, Y.-P. (2008). Molecular evolution of the keratin associated protein gene family in mammals, role in the evolution of mammalian hair. BMC Evol. Biol. *8*, 241.
66. Lister, A.M. (2022). Mammoth evolution in the late Middle Pleistocene: The Mammuthus trogontherii-primigenius transition in Europe. Quat. Sci. Rev. *294*, 107693.
67. Park, S.J., Kim, S.J., Rhee, Y., Byun, J.H., Kim, S.H., Kim, M.H., Lee, E.J., and Lim, S.-K. (2007). Fidgetin-like 1 gene inhibited by basic fibroblast growth factor regulates the proliferation and differentiation of osteoblasts. J. Bone Miner. Res. *22*, 889–896.
68. Venit, T., Dzijak, R., Kalendová, A., Kahle, M., Rohožková, J., Schmidt, V., Rülicke, T., Rathkolb, B., Hans, W., Bohla, A., et al. (2013). Mouse nuclear myosin I knock-out shows interchangeability and redundancy of myosin isoforms in the cell nucleus. PLoS One *8*, e61406.
69. Mukherjee, K., Yang, X., Gerber, S.H., Kwon, H.-B., Ho, A., Castillo, P.E., Liu, X., and Südhof, T.C. (2010). Piccolo and bassoon maintain synaptic vesicle clustering without directly participating in vesicle exocytosis. Proc. Natl. Acad. Sci. USA *107*, 6504–6509.
70. Bhattacharya, A., Freedman, A.N., Avula, V., Harris, R., Liu, W., Pan, C., Lusis, A.J., Joseph, R.M., Smeester, L., Hartwell, H.J., et al. (2022). Placental genomics mediates genetic associations with complex health traits and disease. Nat. Commun. *13*, 706.
71. Perillo, L., Monsurrò, A., Bonci, E., Torella, A., Mutarelli, M., and Nigro, V. (2015). Genetic association of ARHGAP21 gene variant with mandibular prognathism. J. Dent. Res. *94*, 569–576.
72. Blankestijn, M., Bloks, V.W., Struik, D., Huijkman, N., Kloosterhuis, N., Wolters, J.C., Wanders, R.J.A., Vaz, F.M., Islinger, M., Kuipers, F., et al. (2022). Mice with a deficiency in peroxisomal membrane protein 4 (PXMP4) display mild changes in hepatic lipid metabolism. Sci. Rep. *12*, 2512.
73. Takenaka, A., Nakamura, S., Mitsunaga, F., Inoue-Murayama, M., Udono, T., and Suryobroto, B. (2012). Human-specific SNP in obesity genes, adrenergic receptor beta2 (ADRB2), Beta3 (ADRB3), and PPAR γ2 (PPARG), during primate evolution. PLoS One *7*, e43461.
74. Koskinen, R., Salomonsen, J., Tregaskes, C.A., Young, J.R., Goodchild, M., Bumstead, N., and Vainio, O. (2002). The chicken CD4 gene has remained conserved in evolution. Immunogenetics *54*, 520–525.
75. Zhang, Z.D., Weinstock, G., and Gerstein, M. (2008). Rapid evolution by positive Darwinian selection in T-cell antigen CD4 in primates. J. Mol. Evol. *66*, 446–456.
76. Ikeda, K., Yamaguchi, K., Tanaka, T., Mizuno, Y., Hijikata, A., Ohara, O., Takada, H., Kusuhara, K., and Hara, T. (2010). Unique activation status of peripheral blood mononuclear cells at acute phase of Kawasaki disease. Clin. Exp. Immunol. *160*, 246–255.
77. Ohtani, H., Nakajima, T., Akari, H., Ishida, T., and Kimura, A. (2011). Molecular evolution of immunoglobulin superfamily genes in primates. Immunogenetics *63*, 417–428.
78. van der Valk, T., Dehasque, M., Chacón-Duque, J.C., Oskolkov, N., Vartanyan, S., Heintzman, P.D., Pečnerová, P., Díez-Del-Molino, D., and Dalén, L. (2022). Evolutionary consequences of genomic deletions and insertions in the woolly mammoth genome. iScience *25*, 104826.
79. Otto, S.P., and Whitlock, M.C. (2006). Fixation probabilities and times. Encyclopedia of Life Sciences (Wiley). https://doi.org/10.1038/npg.els.0005464.
80. Lisiecki, L.E., and Raymo, M.E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanogr. Paleoclimatology *20*, https://doi.org/10.1029/2004pa001071.
81. Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., and Slate, J. (2010). Adaptation genomics: the next generation. Trends Ecol. Evol. *25*, 705–712.
82. Rausch, J.C., Lavine, J.E., Chalasani, N., Guo, X., Kwon, S., Schwimmer, J.B., Molleston, J.P., Loomba, R., Brunt, E.M., Chen, Y.-D.I., et al. (2018). Genetic variants associated with obesity and insulin resistance in Hispanic boys with nonalcoholic fatty liver disease. J. Pediatr. Gastroenterol. Nutr. *66*, 789–796.
83. Valente, T.S., Baldi, F., Sant'Anna, A.C., Albuquerque, L.G., and Paranhos da Costa, M.J.R. (2016). Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. PLoS One *11*, e0156956.

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

85. Yamagata, K., Nagai, K., Miyamoto, H., Anzai, M., Kato, H., Miyamoto, K., Kurosaka, S., Azuma, R., Kolodeznikov, I.I., Protopopov, A.V., et al. (2019). Signs of biological activities of 28,000-year-old mammoth nuclei in mouse oocytes visualized by live-cell imaging. Sci. Rep. *9*, 4050.

86. Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A.M., To, T.-H., Kortschak, R.D., et al. (2018). A comprehensive genomic history of extinct and living elephants. Proc. Natl. Acad. Sci. USA *115*, E2566–E2574.

87. Campbell-Staton, S.C., Arnold, B.J., Gonçalves, D., Granli, P., Poole, J., Long, R.A., and Pringle, R.M. (2021). Ivory poaching and the rapid evolution of tusklessness in African elephants. Science *374*, 483–487.

88. Tollis, M., Ferris, E., Campbell, M.S., Harris, V.K., Rupp, S.M., Harrison, T.M., Kiso, W.K., Schmitt, D.L., Garner, M.M., Aktipis, C.A., et al. (2021). Elephant genomes reveal accelerated evolution in mechanisms underlying disease defenses. Mol. Biol. Evol. *38*, 3606–3620.

89. Reddy, P.C., Sinha, I., Kelkar, A., Habib, F., Pradhan, S.J., Sukumar, R., and Galande, S. (2015). Comparative sequence analyses of genome and transcriptome reveal novel transcripts and variants in the Asian elephant Elephas maximus. J. Biosci. *40*, 891–907.

90. Illumina (2019). bcl2fastq conversion software. http://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html?langsel=/ch/.

91. Kutschera, V.E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D.W.G., Khoonsari, P.E., Nystedt, B., et al. (2022). GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. BMC Bioinf. *23*, 228.

92. St. John, J. (2011). SeqPrep: tool for stripping adaptors and/or merging paired reads with overlap into single reads. https://githubcom/jstjohn/SeqPrep.

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

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

95. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

96. Danecek, P., McCarthy, S., and Li, H. (2015). bcftools—utilities for variant calling and manipulating vcfs and bcfs. https://samtools.github.io/bcftools/bcftools.html.

97. Smit, A., and Hubley, R. (2015). RepeatModeler Open-1.0. https://www.repeatmasker.org/RepeatModeler/.

98. Smit, A.F.A., Hubley, R., and Green, P. (2017). RepeatMasker Open-3.0. https://www.repeatmasker.org/.

99. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

100. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinf. *15*, 356.

101. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. *38*, 3022–3027.

102. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814.

103. Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C.L., and Bult, C.J. (2021). Mouse Genome Database (MGD): knowledgebase for mouse–human comparative biology. Nucleic Acids Res. *49*, D981–D987.

104. Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360.

105. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). Radiocarbon *62*, 725–757.

106. Yamagata, K., Nagai, K., Miyamoto, H., Anzai, M., Kato, H., Miyamoto, K., Kurosaka, S., Azuma, R., Kolodeznikov, I., Protopopov, A.V., et al. (2019). Researchers revive 28,000-year-old woolly mammoth DNA. Sci. Rep. *9*, 4050.

107. Dehasque, M., Pečnerová, P., Kempe Lagerholm, V., Ersmark, E., Danilov, G.K., Mortensen, P., Vartanyan, S., and Dalén, L. (2022). Development and optimization of a silica column-based extraction protocol for ancient DNA. Genes *13*, 687. https://doi.org/10.3390/genes13040687.

108. Sinding, M.-H.S., Arneborg, J., Nyegaard, G., and Gilbert, M.T.P. (2015). Ancient DNA unravels the truth behind the controversial GUS Greenlandic Norse fur samples: the bison was a horse, and the muskox and bears were goats. J. Archaeol. Sci. *53*, 297–303.

109. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*. db.prot5448.

110. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87.

111. Pečnerová, P., Palkopoulou, E., Wheat, C.W., Skoglund, P., Vartanyan, S., Tikhonov, A., Nikolskiy, P., van der Plicht, J., Díez-Del-Molino, D., and Dalén, L. (2017). Mitogenome evolution in the last surviving woolly mammoth population reveals neutral and functional consequences of small population size. Evol. Lett. *1*, 292–303.

112. Feuerborn, T.R., Palkopoulou, E., van der Valk, T., von Seth, J., Munters, A.R., Pečnerová, P., Dehasque, M., Ureña, I., Ersmark, E., Lagerholm, V.K., et al. (2020). Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets. BMC Genom. *21*, 844.

113. Rogers, R.L., and Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet. *13*. e1006601–16.

114. Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. PLoS One *8*, e79667.

115. van der Valk, T., Díez-Del-Molino, D., Marques-Bonet, T., Guschanski, K., and Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. Curr. Biol. *29*, 165–170.e6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Woolly mammoth	Palkopoulou et al.[84]	Wra4.3Ka
Woolly mammoth	This study	NSI12.2K
Woolly mammoth	This study	NSI12.8K
Woolly mammoth	This study	Wra13.9K
Woolly mammoth	This study	NSI14.3K
Woolly mammoth	This study	NSI16.0K
Woolly mammoth	This study	Chu17.0K
Woolly mammoth	This study	Chu17.6K
Woolly mammoth	This study	Chu20.0K
Woolly mammoth	This study	Chu21.3K
Woolly mammoth	This study	Chu22.0K
Woolly mammoth	van der Valk et al.[7]	Kanchalan
Woolly mammoth	van der Valk et al.[78]	Wra24.0K
Woolly mammoth	This study	Chu27.8K
Woolly mammoth	This study	YaK28.1K
Woolly mammoth	This study	Chu29.1K
Woolly mammoth	This study	Chu30.4K
Woolly mammoth	van der Valk et al.[78]	Chu31.9K
Woolly mammoth	This study	Chu33.7K
Woolly mammoth	Yamagata et al.[85]	YaK39.1K
Woolly mammoth	Palkopoulou et al.[84]	Oim44.2K
Woolly mammoth	This study	YaKinf
Woolly mammoth	This study, van der Valk et al.[7]	Chukochya
Mastodon	Palkopoulou et al.[86]	I
African savannah elephant	Campbell-Staton et al.[87]	0045B
African savannah elephant	Campbell-Staton et al.[87]	2981B
African savannah elephant	Campbell-Staton et al.[87]	2982B
African savannah elephant	Campbell-Staton et al.[87]	2983B
African savannah elephant	Campbell-Staton et al.[87]	2984B
African savannah elephant	Campbell-Staton et al.[87]	2985B
African savannah elephant	Palkopoulou et al.[86]	B
African savannah elephant	Palkopoulou et al.[86]	C
African savannah elephant	Palkopoulou et al.[88]	Christie
African savannah elephant	Campbell-Staton et al.[87]	G13
African savannah elephant	Campbell-Staton et al.[87]	G14
African savannah elephant	Campbell-Staton et al.[87]	G15
African savannah elephant	Campbell-Staton et al.[87]	G16
African savannah elephant	Campbell-Staton et al.[87]	G17A
African savannah elephant	Campbell-Staton et al.[87]	G18A
African savannah elephant	Campbell-Staton et al.[87]	G19A
African savannah elephant	Campbell-Staton et al.[87]	G20A
African savannah elephant	Campbell-Staton et al.[87]	G21A
African savannah elephant	Campbell-Staton et al.[87]	G22A
African savannah elephant	Tollis et al.[88]	Hi-Dari

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
African savannah elephant	Campbell-Staton et al.[87]	T2B
Asian elephant	Palkopoulou et al.[86]	D
Asian elephant	Palkopoulou et al.[86]	E
Asian elephant	Tollis et al.[88]	Icky
Asian elephant	Lynch et al.[4]	L
Asian elephant	Lynch et al.[4]	M
Asian elephant	Lynch et al.[4]	Y
Asian elephant	Reddy et al.[89]	Z
Chemicals, peptides, and recombinant proteins		
USER enzyme	New England Biolabs	NEB #M5508
AccuPrime reaction mix	Life Technologies	Cat #12344040
AccuPrime Pfx DNA polymerase	Life Technologies	Cat #12344024
Agencourt AMPure XP beads	Beckman Coulter	Cat #10136224
Critical commercial assays		
MinElute purification columns	QIAGEN	Cat #28115
Deposited data		
Raw sequencing data	This study	PRJEB59491
Software and algorithms		
bcl2Fastq v1.8.3	Illumina[90]	https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
GenErode	Kutschera et al.[91]	https://github.com/NBISweden/GenErode
Seqprep v1.2	St. John et al.[92]	https://github.com/jstjohn/SeqPrep
BWA v0.7.17	Li and Durbin[93]	https://bio-bwa.sourceforge.net/
SAMtools v1.8	Li et al.[94]	http://www.htslib.org/
Trimmomatic v0.39	Bolger et al.[95]	https://github.com/usadellab/Trimmomatic
bcftools v1.8	Danecek et al.[96]	http://www.htslib.org/
RepeatModeler v2.0.1	Smit and Hubley[97]	https://www.repeatmasker.org/RepeatModeler/
RepeatMasker v4.0.9	Smit et al.[98]	https://www.repeatmasker.org/RepeatMasker/
Pairwise Sequential Markovian Coalescent (PSMC)	Li and Durbin[99]	https://github.com/lh3/psmc
ANGSD v0.939	Korneliussen et al.[100]	https://github.com/ANGSD/angsd
MEGA11	Tamura et al.[101]	https://www.megasoftware.net/
SIFT	Ng and Henikoff[102]	https://sift.bii.a-star.edu.sg/
GOrilla	Eden et al.[8]	http://cbl-gorilla.cs.technion.ac.il/
Mouse Genome Informatics database	Blake et al.[103]	https://www.informatics.jax.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact: Love Dalén (love.dalen@zoologi.su.se).

Materials availability

This study did not generate new reagents.

Data and code availability

Raw sequencing data of the newly generated mammoth genomes can be found on the European Nucleotide Archive (ENA: PRJEB59491). Sample specific accession numbers are provided in Table S1. This paper also analyzes existing, publicly available data. The accession numbers for these data are listed in Table S2. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Newly sequenced mammoths

We generated whole genome sequencing data for 16 new woolly mammoth samples (Table S1). These samples were opportunistically collected during various expeditions in Siberia over the past 25 years. We also generated additional sequencing data for a previously published woolly mammoth, Chukochya, dating to ~700 kya.[7] We generated radiocarbon dates for three woolly mammoth samples. Dating was done at either the Oxford Radiocarbon Accelerator Unit (ORAU) (^{14}C lab identification: OxA), or NEISRI FEB RAS, Magadan (^{14}C lab identification: MAG). All radiocarbon dates were calibrated using Oxcal 4.4[104] and the IntCal20 Northern Hemisphere radiocarbon age calibration curve.[105]

Dataset

We additionally downloaded the raw sequencing data from previously published elephantid genomes, including six woolly mammoths, seven Asian elephants, 21 African savannah elephants, and one American mastodon,[4,7,78,84,86–89,106] resulting in a final dataset of 52 elephantid genomes (Table S2).

METHOD DETAILS

Laboratory methods

For the 16 new woolly mammoth samples, DNA extractions, and library preparations were performed following standard ancient DNA practices in either the dedicated ancient DNA lab facilities at the Swedish Museum of Natural History or the Centre for Palaeogenetics, both located in Stockholm, Sweden. Briefly, we first collected 50-200 mg of bone or tooth powder using a Dremel drill. For all but two samples, we then carried out DNA extractions using the final silica column protocol presented in Dehasque et al.[107] For sample Chu20.0K, which was part of a protocol development study,[107] we performed an additional predigestion treatment to improve sequencing efficiency (Table S1). For sample YaKinf, a skin sample, we used a digestion buffer suited for keratin-rich tissue.[108] After overnight digestion, the extraction protocol was continued from day two as described in Dehasque et al.[107] We then prepared double-stranded sequencing libraries following the protocol by Meyer and Kircher,[109] including treatment with either 3 or 6 μL of USER (New England Biolabs) as described in Dehasque et al.[107] The USER enzyme, a mixture of uracil–DNA–glycosylase (UDG) and endonuclease VIII (endoVIII), excises uracil bases incorporated as a consequence of post-mortem damage, except at CpG sites.[110] Clean-up steps were performed using MinElute purification columns (QIAGEN). We amplified and double-indexed with unique barcodes the sequencing libraries in a PCR reaction volume of 25 μL containing 1X AccuPrime reaction mix (Life Technologies), 0.3 μM of the forward indexing primer, 0.3 μM of the reverse indexing primer, 1.25 U AccuPrime Pfx DNA polymerase (Life Technologies), and 3 μL of DNA library. PCR reaction conditions were the same as described in Pečnerová et al.[111] The number of PCR cycles varied from 6 to 16, depending on DNA quantity as indicated by the Ct values of a qPCR. We generated between 3 to 32 independent amplified libraries for each sample to minimize sequence clonality during sequencing. We removed both too-short (bead-to-mix ratio 1.8) and too-long fragments (bead-to-mix ratio 0.5) from the amplified libraries using magnetic Agencourt AMPure XP beads (Beckman Coulter). Additionally, we further sequenced a previously-published early Middle Pleistocene mammoth sample, Chukochya, in order to increase its genomic coverage. We used an existing DNA extract from van der Valk et al.[7] to prepare 13 new independent amplified genomic libraries following the same laboratory protocols as described above. The final amplified libraries were sent to the National Genomics Infrastructure (NGI Stockholm) for sequencing using the Illumina NovaSeq platform (S4, 2x100 or 2x150).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data processing steps, statistical methods, software used, as well as randomization and/or strategies to assess significance are described in detail below. We did not use any methods where it was necessary to determine whether the data met the assumptions of the statistical approach.

Data processing

For the newly sequenced samples, bcl2Fastq v1.8.3[90] was used to demultiplex raw sequencing reads and convert them from Bcl to Fastq (CASAVA software suite). For all ancient samples, we mapped the Fastq reads of each library and processed subsequent BAM files using the "historical track" of an in-development version of the GenErode pipeline[91] with minor modifications. The raw sequencing data from the previously published ancient genomes in the dataset was downloaded and processed using the same pipeline. Briefly, SeqPrep v1.2[92] was used to merge paired-end reads and trim adapter sequences, with a small modification to the source code to keep bases with the highest quality score in overlapping regions. Only fragments with a minimum length of 30 bp were kept for all samples except for Chuckochya, where only sequencing fragments $\geq$35 bp were retained.[7] We then mapped the reads against a concatenated reference genome that included the African savannah elephant (LoxAfr4) and human (hg19) nuclear genomes, and the woolly mammoth mitogenome (Krause mammoth, DQ188829), in order to exclude possible contamination.[112] Mapping was done using BWA aln v0.7.17[93] with ancient DNA-specific settings as in Palkopoulou et al.[84] Next, we merged and sorted BAM files per sample using SAMtools v1.8[94] and removed duplicate sequences using a script that marks duplicates based

on both the start and end position of a fragment.[84] Finally, we extracted only the sequences mapping to the African savannah elephant and woolly mammoth mitogenome using SAMtools. For the previously published sample "Kanchalan",[7] which was treated with Afu uracil-DNA glycosylase leaving post-mortem DNA damage at the ends of the molecules, 2 bp were trimmed from both ends of the molecules before mapping. For the modern samples, all previously published, we downloaded the raw reads and processed the data using the "modern track" of the GenErode pipeline. Briefly, adapter trimming was done with trimmomatic[95] also keeping only fragments at least 30 bp in length. We then mapped the sequences to the same concatenated reference genome as above and merged and sorted BAM files, and removed possible duplicates using SAMtools.

Variant calling

We used bcftools v1.8[96] to call genotypes for all samples individually, only keeping sites with a mapping quality $\geq$30 and a base quality $\geq$30. To avoid erroneous variants due to mismapping around indels, we discarded all indel sites and those within five base pairs. We also excluded CpG sites, since these can be protected from the USER enzyme and therefore enriched for errors caused by post-mortem damage,[110] and repetitive regions which we identified using RepeatModeler v2.0.1[97] and RepeatMasker v4.0.9.[98] In order to avoid bias toward reference alleles in sites with low coverage, the resulting BCF files were scanned for genotype consistency based on each site's depth of coverage (the number of reads covering a specific site). Genotypes of sites covered by less than four reads and with the presence of both reference and alternative alleles, which are prone to reference bias, were corrected based on the number of high-quality alleles present for each allele.

Population genomic analyses

As a genome quality check and to ensure that the mammoth genomes do not contain unexpected diversity as a result of mismapping and contamination (i.e. Rogers and Slatkin[113]) we ran Pairwise Sequential Markovian Coalescent (PSMC)[99] analyses on all mammoth genomes above 10× coverage. We first filtered out all CpG sites as well as genomic regions below or above three times the average genome wide coverage. We used the standard parameters -N25 -t15 -r5 -p "4 + 25*2 + 4+6". The PSMC output was rescaled to years using a generation time of 25 years and a mutation rate of 2.5×10^{-8} per site per generation as in Palkopoulou et al.[84] We additionally checked genome-wide autosomal heterozygosity for all individuals with average genome coverage >10× using realSFS as implemented in ANGSD. We considered only uniquely mapping reads (-uniqueOnly 1) and bases with quality score >19 (-minQ 20).[100,114] ANGSD uses genotype-likelihoods allowing for the incorporation of statistical uncertainty in low-coverage data and shows high accuracy in estimating heterozygosity for genomes at different coverages.[115] Overall, these analyses provided results (see Figures S1 and S2) that are consistent with previous work on woolly mammoth genomics.[84]

Phylogenetic inferences

We reconstructed phylogenetic trees on the basis of pairwise genetic distances between all individuals. First, we randomly sampled an allele at each site for each genome using ANGSD.[100] We then calculated genome-wide pairwise differences for all samples, as well as 100 resampling replicates based on 100,000 sites each. A Neighbor Joining (NJ) tree based on the pairwise differences was then obtained using MEGA11.[101] In addition, we used variable window sizes across the genome (1 kb, 10 kb, 50 kb, 100 kb and 1 Mb) to identify regions of the mammoth genome that are closer to the African savannah elephant than to the Asian elephant by randomly sampling a single allele from each sample group (woolly mammoths, Asian elephants, and African savannah elephants) and estimating the sequence divergence between these groups for each window.

Derived variants and effect-prediction

We identified sites where all Late Quaternary woolly mammoth genomes (n = 22) are homozygous for the alternative allele with respect to the African savannah elephant reference genome, and all African savannah and Asian elephant genomes are homozygous for the reference allele. We included only sites for which at least half of the Late Quaternary woolly mammoth genomes ($n \geq 11$) and half of the elephant genomes ($n \geq 14$) had a genotype call (after filtering). For comparison, we also identified sites for which all Asian elephant genomes have a fixed homozygous alternative genotype, and all the other species' genomes are fixed for the reference allele. Next, we used SIFT[102] to annotate all identified variants and obtained SIFT scores for each of the protein-coding changing mutations. SIFT scores range between 0 and 1 with lower scores indicating a higher likelihood of having an impact on the protein-function. However, to aid interpretation, we calculated an aggregated SIFT score for each gene as the sum of (1 - SIFT) for each protein-coding variant in that gene so that the higher the aggregated SIFT score the more likely it is that the protein function is altered. Finally, for all woolly mammoth-specific variants, we assessed the presence of the ancestral or derived allele in the 700 ka Chukochya genome.

GO term enrichment and phenotype consequences

We conducted a gene ontology (GO) enrichment using GOrilla[8] on all genes for which the woolly mammoth genomes carried one or more FdNs variants. We used all annotated African savannah elephant genes as a background set and their respective GO terms where obtained using the human database. Additionally, we obtained phenotypic data of mouse knockouts for all genes with woolly mammoth variants using the Mouse Genome Informatics database.[103]