Genomics of adaptive evolution in the woolly mammoth

Díez-del-Molino, David; Dehasque, Marianne; Chacón-Duque, J. Camilo; Penerová, Patrícia; Tikhonov, Alexei; Protopopov, Albert; Plotnikov, Valeri; Kanellidou, Foteini; Nikolskiy, Pavel; Mortensen, Peter; Danilov, Gleb K.; Vartanyan, Sergey; Gilbert, M. Thomas P.; Lister, Adrian M.; Heintzman, Peter D.; van der Valk, Tom; Dalén, Love

Published in:
Current Biology

DOI:
10.1016/j.cub.2023.03.084

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
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 Diez-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

Diez-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.
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. The woolly mammoth had a Holarctic distribution and inhabited terrestrial environments up to 80 degrees north, even during full glacial conditions. Compared to both its extant elephant relatives as well as earlier members of *Mammuthus*,
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. Moreover, genetic analyses have hinted at previously unknown physiological adaptations to the Arctic environment, including genes related to thermal sensation and hemoglobin structure. 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. 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, 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 28.6x (mean of 8.0x) in genome coverage (Figure S1). We also generated additional genomic data for an ~700,000-year-old woolly mammoth sample (Chukochya), increasing its coverage to 2.8x. 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 (2.3–4.1x) and 10 at high coverage (10.4–28.6x), 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. 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 and a background set of all coding genes annotated in the African savannah elephant genome. We found 26 GO terms enriched with p ≤ 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, recombination 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 categories enriched in the Asian elephant (n = 19, p ≤ 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
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 \((\text{Table S5})\). 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” \((\text{STAR Methods})\). 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 \((\text{Figure 3})\). 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 \((\text{Table 1})\).

**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 \((\text{Figure 2}; \text{Table S3})\). In mammals, there are several dozens of genes associated with hair and skin development, 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 \(\text{AHNAK2}\) \((\text{Table 1})\), 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 \(\text{KRT8}\) with nine FdNs mutations\(^{13}\) and \(\text{FLG}\) with seven FdNs mutations.\(^{14}\) We also note that woolly mammoths have four FdNs mutations in the \(\text{LYST}\) gene \((\text{Table S5})\), which has been implicated in the lack of pigmentation 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 wooly mammoth. This includes uncombable...
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 Ppargc1a 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.
vessels and plays a role in eye function. 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. Interestingly, in humans, PDZD2 has also been related to fat tissue and mitochondrial function, 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, in woolly mammoths.

**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.

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. In humans, tumor mutations in both BRCA genes have been associated with the gene TP53, a well-known multi-copy tumor suppressor gene in elephants. 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.

**Thermosensation**

Earlier work has highlighted the importance of thermosensation genes for cold adaptation in the woolly mammoth. 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 and essential for the perception of extreme cold as being painful. In contrast to SCN10A, genes belonging to the TRP channel gene family are activated and inactivated at different temperatures and have previously been implicated in thermosensation in the woolly mammoth. 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. 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
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 \( 15 ^\circ C \).\(^{51,52} \)

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

---

### Table 1. Highly evolved genes in the woolly mammoth

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Fixed derived synonymous</th>
<th>Fixed derived non-synonymous</th>
<th>Aggregated SIFT score</th>
<th>Fixed derived non-synonymous and evolved &lt;700 kya</th>
<th>Predicted phenotype affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHNAK2</td>
<td>8</td>
<td>14</td>
<td>9.73</td>
<td>0 (12)</td>
<td>hair and/or skin development</td>
</tr>
<tr>
<td>SIGLEC14(^a)</td>
<td>6</td>
<td>12</td>
<td>8.98</td>
<td>0 (6)</td>
<td>immune system</td>
</tr>
<tr>
<td>FSIP2(^b)</td>
<td>4</td>
<td>12</td>
<td>6.26</td>
<td>0 (12)</td>
<td>sperm development/function</td>
</tr>
<tr>
<td>KRT8</td>
<td>0</td>
<td>9</td>
<td>6.38</td>
<td>0 (9)</td>
<td>hair and/or skin development</td>
</tr>
<tr>
<td>TEX15</td>
<td>4</td>
<td>8</td>
<td>5.53</td>
<td>0 (6)</td>
<td>sperm development/function</td>
</tr>
<tr>
<td>ADGRV1</td>
<td>5</td>
<td>8</td>
<td>4.71</td>
<td>0 (7)</td>
<td>hearing</td>
</tr>
<tr>
<td>CCNB3</td>
<td>1</td>
<td>7</td>
<td>5.91</td>
<td>0 (5)</td>
<td>pregnancy and embryo develop</td>
</tr>
<tr>
<td>PARP14</td>
<td>3</td>
<td>7</td>
<td>5.57</td>
<td>1 (7)</td>
<td>immune system</td>
</tr>
<tr>
<td>CDK5RAP2(^a)</td>
<td>0</td>
<td>7</td>
<td>5.56</td>
<td>0 (6)</td>
<td>brain size</td>
</tr>
<tr>
<td>FLG</td>
<td>2</td>
<td>7</td>
<td>5.35</td>
<td>5 (5)</td>
<td>ear size and hair and skin development</td>
</tr>
<tr>
<td>BRCA1</td>
<td>1</td>
<td>6</td>
<td>4.47</td>
<td>0 (4)</td>
<td>DNA repair and tumor suppression</td>
</tr>
<tr>
<td>SCN10A</td>
<td>11</td>
<td>6</td>
<td>4.39</td>
<td>0 (6)</td>
<td>thermosensation</td>
</tr>
<tr>
<td>RP11L1</td>
<td>1</td>
<td>6</td>
<td>4.32</td>
<td>0 (4)</td>
<td>vision</td>
</tr>
<tr>
<td>ACDM</td>
<td>0</td>
<td>6</td>
<td>3.24</td>
<td>1 (3)</td>
<td>fat storage and metabolism</td>
</tr>
<tr>
<td>CD1D</td>
<td>2</td>
<td>6</td>
<td>3.2</td>
<td>0 (6)</td>
<td>immune system</td>
</tr>
<tr>
<td>PDZD2(^a)</td>
<td>1</td>
<td>6</td>
<td>3.2</td>
<td>0 (5)</td>
<td>sugar metabolism</td>
</tr>
<tr>
<td>CD4</td>
<td>1</td>
<td>5</td>
<td>3.85</td>
<td>4 (5)</td>
<td>immune system</td>
</tr>
<tr>
<td>TSHZ3</td>
<td>3</td>
<td>5</td>
<td>3.75</td>
<td>2 (5)</td>
<td>urinary tract and kidney development</td>
</tr>
<tr>
<td>ABCC11(^b)</td>
<td>3</td>
<td>5</td>
<td>3.16</td>
<td>2 (5)</td>
<td>ear wax and body odor</td>
</tr>
<tr>
<td>FIGNL1</td>
<td>0</td>
<td>4</td>
<td>1.96</td>
<td>4 (4)</td>
<td>skeletal morphology and body size</td>
</tr>
<tr>
<td>PRSS8</td>
<td>5</td>
<td>3</td>
<td>4.29</td>
<td>3 (3)</td>
<td>hair and/or skin development</td>
</tr>
<tr>
<td>TCHH</td>
<td>1</td>
<td>3</td>
<td>2.63</td>
<td>2 (3)</td>
<td>hair and/or skin development</td>
</tr>
<tr>
<td>FAM214A</td>
<td>0</td>
<td>3</td>
<td>1.74</td>
<td>3 (3)</td>
<td>skeletal morphology and body size</td>
</tr>
<tr>
<td>OFLR94(^a)</td>
<td>0</td>
<td>3</td>
<td>1.33</td>
<td>3 (3)</td>
<td>olfaction</td>
</tr>
<tr>
<td>TMC4</td>
<td>0</td>
<td>3</td>
<td>1.23</td>
<td>3 (3)</td>
<td>salt taste sensitivity</td>
</tr>
<tr>
<td>PCL3</td>
<td>2</td>
<td>3</td>
<td>0.98</td>
<td>3 (3)</td>
<td>skeletal morphology and body size</td>
</tr>
<tr>
<td>CD3E</td>
<td>1</td>
<td>2</td>
<td>1.77</td>
<td>2 (2)</td>
<td>immune system</td>
</tr>
<tr>
<td>IGSF6</td>
<td>2</td>
<td>2</td>
<td>1.57</td>
<td>2 (2)</td>
<td>immune system</td>
</tr>
<tr>
<td>PXMP4</td>
<td>0</td>
<td>2</td>
<td>1.47</td>
<td>2 (2)</td>
<td>fat storage and metabolism</td>
</tr>
<tr>
<td>ADRB2</td>
<td>0</td>
<td>2</td>
<td>1.03</td>
<td>2 (2)</td>
<td>fat storage and metabolism</td>
</tr>
<tr>
<td>NMI</td>
<td>0</td>
<td>2</td>
<td>0.67</td>
<td>2 (2)</td>
<td>skeletal morphology and body size</td>
</tr>
<tr>
<td>ARHGAP21</td>
<td>1</td>
<td>2</td>
<td>0.57</td>
<td>2 (2)</td>
<td>skeletal morphology and body size</td>
</tr>
</tbody>
</table>

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
with repeated abnormalities in sperm flagella.\textsuperscript{54} We also find that woolly mammoths had seven FdNs mutations in the gene \textit{CCNB2}, which has been associated with pregnancy loss in mice.\textsuperscript{55} Similarly, we identified six FdNs mutations in \textit{ZP4} (SIFT score of 2.57), an essential gene for embryo development in rabbits.\textsuperscript{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 \textit{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.\textsuperscript{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 \textit{CDK5RAP2} gene in the human genome. \textit{CDK5RAP2} is a gene that has been identified as controlling human brain size.\textsuperscript{58} In addition, we observed six FdNs mutations in \textit{RP1L1}, which is a gene that is exclusively expressed in retinal photoreceptors and is an important component of the photoreceptor cilium.\textsuperscript{59} Our results also show that the woolly mammoth had eight FdNs mutations in \textit{ADGRV1}. This gene has been associated with hearing loss disorders in both humans and mice.\textsuperscript{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 \textit{ABCC11} (Table S5). As shown in earlier studies,\textsuperscript{5} one of the mutations in the woolly mammoth \textit{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.\textsuperscript{52} 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 Quaternary woolly mammoths.\textsuperscript{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 \textit{FLG} (five FdNs mutations), a gene that encodes for the protein filaggrin that binds to keratin fibers in epithelial cells. Intriguingly, \textit{FLG}-deficient mice develop significantly smaller ears.\textsuperscript{63} We therefore hypothesize that the mutations observed in the \textit{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 \textit{PRSS8}, a gene that is important for hair follicle development and where non-synonymous mutations are responsible for the mutant “frizzy” phenotype in mice.\textsuperscript{54} In addition, we find that two of the three FdNs mutations in \textit{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 \textit{KRTAP4-1}, which encodes for a protein that is essential for the formation of a rigid and resistant hair shaft in mammals.\textsuperscript{55} 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.\textsuperscript{65} For example, the genes \textit{FIGNL1} and \textit{NMI} (from zero to four and two FdNs mutations in the last 700 ka, respectively) are important for osteoblast differentiation and bone density.\textsuperscript{67,68} We also find that the genes \textit{PCLO} and \textit{FAM214A} have changed from the ancestral elephant allele by gaining three FdNs mutations each in the last 700 ka. Mutations in \textit{PCLO} are associated with decreased body size in mice,\textsuperscript{69} whereas expression levels in \textit{FAM214A} affect the waist-hip ratio in humans.\textsuperscript{70} Finally, we find that the gene \textit{ARHGAP21}, which is associated with mandibular prognathism in humans,\textsuperscript{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 \textit{PXMP4}, which in mice affects levels of fatty acid alkyl-diacylglycerol lipids in the liver, and thus likely plays a role in lipid metabolism.\textsuperscript{72} The second gene, \textit{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.\textsuperscript{73}

Finally, our analysis shows that several immunity-related genes have evolved in the last 700 ka. The gene \textit{CD4}, which encodes for a T cell antigen that is an essential component in cell-mediated immunity,\textsuperscript{74} has gained four of its five FdNs mutations in the woolly mammoth during the last 700 ka. Among humans and chimpanzees/bonobos, \textit{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.\textsuperscript{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, \textit{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. **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.

**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, 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 immuinity, 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, 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.

**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. 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,\(^8\) as well as flight behavior in cattle.\(^9\) 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.cell.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 Tryggersons 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 Hélène Muller for assistance with laboratory analyses, to Illumina for partially 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**


**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**


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


## STAR METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>Palkopoulou et al. 84</td>
<td>Wra4.3Ka</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>NSI12.2K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>NSI12.8K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Wra13.9K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>NSI14.3K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>NSI16.0K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu17.0K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu17.6K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu20.0K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu21.3K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu22.0K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>van der Valk et al. 7</td>
<td>Kanchalan</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>van der Valk et al. 78</td>
<td>Wra24.0K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu27.8K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>YaK28.1K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu29.1K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu30.4K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>van der Valk et al. 76</td>
<td>Chu31.9K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>Chu33.7K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>Yamagata et al. 86</td>
<td>YaK39.1K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>Palkopoulou et al. 84</td>
<td>Oim44.2K</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study</td>
<td>YaKinf</td>
</tr>
<tr>
<td>Woolly mammoth</td>
<td>This study, van der Valk et al. 7</td>
<td>Chukochya</td>
</tr>
<tr>
<td>Mastodon</td>
<td>Palkopoulou et al. 86</td>
<td>I</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>0045B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>2981B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>2982B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>2983B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>2984B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>2985B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Palkopoulou et al. 88</td>
<td>B</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Palkopoulou et al. 88</td>
<td>C</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Palkopoulou et al. 88</td>
<td>Christie</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G13</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G14</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G15</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G16</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G17A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G18A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G19A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G20A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G21A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Campbell-Staton et al. 37</td>
<td>G22A</td>
</tr>
<tr>
<td>African savannah elephant</td>
<td>Tollis et al. 88</td>
<td>Hi-Dari</td>
</tr>
</tbody>
</table>

(Continued on next page)
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.\(^8\) 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\(^9\) and the IntCal20 Northern Hemisphere radiocarbon age calibration curve.\(^10\)

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,\(^4,7,84-89,106\) resulting in a final data-set 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 \(\mu\)L 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 \(\mu\)L containing 1X AccuPrime reaction mix (Life Technologies), 0.3 \(\mu\)M of the forward indexing primer, 0.3 \(\mu\)M of the reverse indexing primer, 1.25 U AccuPrime Pfx DNA polymerase (Life Technologies), and 3 \(\mu\)L 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 Vaal 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 Chukochya, 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 (LoxAfR) 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.\(^34\) Next, we merged and sorted BAM files per sample using SAMtools v1.8\(^34\) and removed duplicate sequences using a script that marks duplicates based
on both the start and end position of a fragment. Finally, we extracted only the sequences mapping to the African savannah elephant and woolly mammoth mitogenome using SAMtools. For the previously published sample “Kanchalan”, 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 GeneErode pipeline. Briefly, adapter trimming was done with trimmomatic 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 to call genotypes for all samples individually, only keeping sites with a mapping quality ≥ 30 and a base quality ≥ 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, and repetitive regions which we identified using RepeatModeler v2.0.197 and RepeatMasker v4.0.9.198 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 Slatkin115) we ran Pairwise Sequential Markovian Coalescent (PSMC) analyses on all mammoth genomes above 10x 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 x 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 >10x 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. 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 ≥ 11) and half of the elephant genomes (n ≥ 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 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 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.