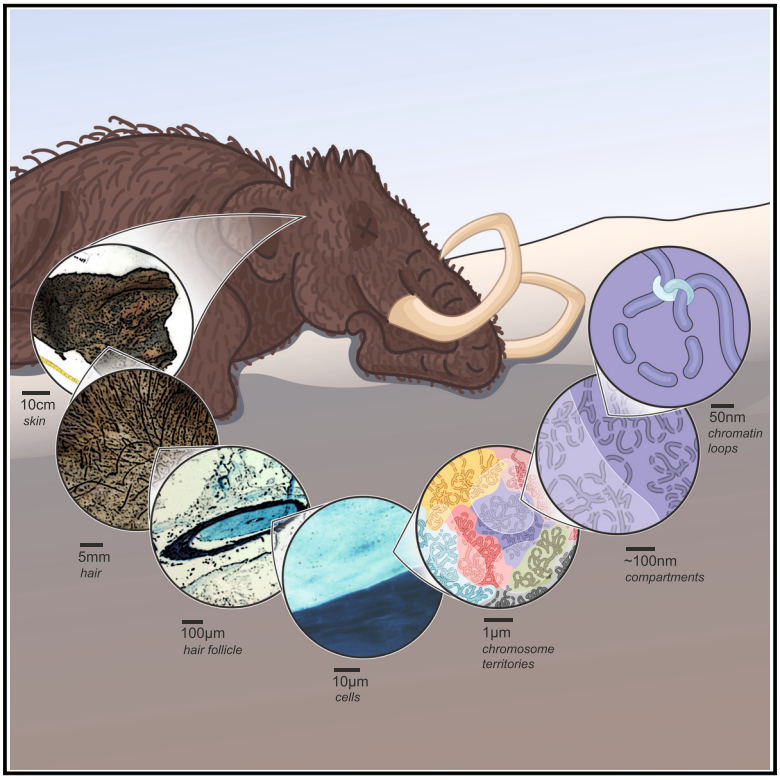
Cell

Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample

Graphical abstract



Highlights

- 3D genome architecture is preserved in a 52,000-year-old woolly mammoth sample
- PaleoHi-C makes it possible to assemble the woolly mammoth's genome
- Chromatin compartments also persist, enabling study of mammoth gene expression
- We propose that dehydration led to a glass transition arresting molecular movement

Authors

Marcela Sandoval-Velasco,
Olga Dudchenko,
Juan Antonio Rodríguez,
Cynthia Pérez Estrada, ...,
Marc A. Marti-Renom,
M. Thomas P. Gilbert,
Erez Lieberman Aiden

Correspondence

olga.dudchenko@bcm.edu (O.D.),
martirenom@cnag.eu (M.A.M.-R.),
tgilbert@sund.ku.dk (M.T.P.G.),
erez@erez.com (E.L.A.)

In brief

Three-dimensional genome architecture is shown to be preserved in a 52,000-year-old woolly mammoth sample, enabling genome assembly and analysis of gene expression. Spontaneous freeze-drying of the mammoth carcass in the cold Siberian environment potentially led to a glass transition that preserved sample morphology across eight orders of magnitude in length, from the 3-m carcass to 50-nm chromatin loops.

Sandoval-Velasco et al., 2024, Cell *187*, 3541–3562
July 11, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2024.06.002

CellPress

Cell

CellPress
OPEN ACCESS

Article

Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample

Marcela Sandoval-Velasco,[1,34] Olga Dudchenko,[2,3,34,*] Juan Antonio Rodríguez,[1,4,34] Cynthia Pérez Estrada,[2,3,34] Marianne Dehasque,[5,6,7] Claudia Fontsere,[1] Sarah S.T. Mak,[1] Ruqayya Khan,[2] Vinícius G. Contessoto,[3] Antonio B. Oliveira Junior,[3] Achyuth Kalluchi,[8] Bernardo J. Zubillaga Herrera,[9,10] Jiyun Jeong,[2] Renata P. Roy,[2,3,11] Ishawnia Christopher,[2] David Weisz,[2] Arina D. Omer,[2] Sanjit S. Batra,[2] Muhammad S. Shamim,[2] Neva C. Durand,[2,12] Brendan O'Connell,[13,14] Alfred L. Roca,[15] Maksim V. Plikus,[16] Mariya A. Kusliy,[17] Svetlana A. Romanenko,[17] Natalya A. Lemskaya,[17] Natalya A. Serdyukova,[17] Svetlana A. Modina,[17] Polina L. Perelman,[17] Elena A. Kizilova,[18] Sergei I. Baiborodin,[18] Nikolai B. Rubtsov,[18] Gur Machol,[2] Krisha Rath,[2] Ragini Mahajan,[2,3,19] Parwinder Kaur,[20] Andreas Gnirke,[12] Isabel Garcia-Treviño,[21] Rob Coke,[21] Joseph P. Flanagan,[22] Kelcie Pletch,[22] Aurora Ruiz-Herrera,[23] Valerii Plotnikov,[24] Innokentiy S. Pavlov,[24] Naryya I. Pavlova,[25] Albert V. Protopopov,[24,26] Michele Di Pierro,[9,10] Alexander S. Graphodatsky,[17] Eric S. Lander,[12,27,28] M. Jordan Rowley,[8] Peter G. Wolynes,[3,19,29] José N. Onuchic,[3,19,29] Love Dalén,[5,6,7] Marc A. Marti-Renom,[4,30,31,32,*] M. Thomas P. Gilbert,[1,33,*] and Erez Lieberman Aiden[2,3,12,35,*]

[1]Center for Evolutionary Hologenomics, University of Copenhagen, DK-1353 Copenhagen, Denmark
[2]The Center for Genome Architecture and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[3]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA
[4]Centre Nacional d'Anàlisi Genòmica, CNAG, 08028 Barcelona, Spain
[5]Centre for Palaeogenetics, SE-106 91 Stockholm, Sweden
[6]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, 10405 Stockholm, Sweden
[7]Department of Zoology, Stockholm University, SE-106 91 Stockholm, Sweden
[8]Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE 68198, USA
[9]Department of Physics, Northeastern University, Boston, MA 02115, USA
[10]Center for Theoretical Biological Physics, Northeastern University, Boston, MA 02215, USA
[11]Departments of Biology and Physics, Texas Southern University, Houston, TX 77004, USA
[12]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[13]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA
[14]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR 97239, USA
[15]Department of Animal Sciences and Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[16]Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697, USA
[17]Institute of Molecular and Cellular Biology SB RAS, Novosibirsk 630090, Russia
[18]Institute of Cytology and Genetics SB RAS, Novosibirsk 630090, Russia
[19]Department of Biosciences, Rice University, Houston, TX 77005, USA
[20]UWA School of Agriculture and Environment, University of Western Australia, Perth, WA 6009, Australia
[21]San Antonio Zoo, San Antonio, TX 78212, USA
[22]Houston Zoo, Houston, TX 77030, USA
[23]Departament de Biologia Cel·lular, Fisiologia i Immunologia and Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain
[24]Academy of Sciences of Sakha Republic, Yakutsk 677000, Russia
[25]Institute of Biological Problems of Cryolitezone SB RAS, Yakutsk 677000, Russia
[26]North-Eastern Federal University, Yakutsk 677027, Russia
[27]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[28]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[29]Departments of Physics, Astronomy, & Chemistry, Rice University, Houston, TX 77005, USA
[30]Centre for Genomic Regulation, The Barcelona Institute for Science and Technology, 08003 Barcelona, Spain
[31]ICREA, 08010 Barcelona, Spain
[32]Universitat Pompeu Fabra, 08002 Barcelona, Spain
[33]University Museum NTNU, 7012 Trondheim, Norway
[34]These authors contributed equally
[35]Lead contact
*Correspondence: olga.dudchenko@bcm.edu (O.D.), martirenom@cnag.eu (M.A.M.-R.), tgilbert@sund.ku.dk (M.T.P.G.), erez@erez.com (E.L.A.)
https://doi.org/10.1016/j.cell.2024.06.002

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

SUMMARY

Analyses of ancient DNA typically involve sequencing the surviving short oligonucleotides and aligning to genome assemblies from related, modern species. Here, we report that skin from a female woolly mammoth (†*Mammuthus primigenius*) that died 52,000 years ago retained its ancient genome architecture. We use PaleoHi-C to map chromatin contacts and assemble its genome, yielding 28 chromosome-length scaffolds. Chromosome territories, compartments, loops, Barr bodies, and inactive X chromosome (Xi) superdomains persist. The active and inactive genome compartments in mammoth skin more closely resemble Asian elephant skin than other elephant tissues. Our analyses uncover new biology. Differences in compartmentalization reveal genes whose transcription was potentially altered in mammoths vs. elephants. Mammoth Xi has a tetradic architecture, not bipartite like human and mouse. We hypothesize that, shortly after this mammoth's death, the sample spontaneously freeze-dried in the Siberian cold, leading to a glass transition that preserved subfossils of ancient chromosomes at nanometer scale.

INTRODUCTION

Study of ancient DNA (aDNA) began with the sequencing of mitochondrial fragments from historic samples[1,2] and has subsequently undergone a remarkable expansion. Current paleogenomic analyses involve DNA sequencing of whole genomes (aDNA-Seq) from a plethora of extinct and archaic humans (e.g., Green et al., Rasmussen et al.[3,4]), animals (e.g., Miller et al., Orlando et al.[5,6]), plants (e.g., Martin et al., Ramos-Madrigal et al.[7,8]), and pathogens (e.g., Bos et al., Smith et al.[9,10]) spanning over one million years.[11]

Yet given the fragmentary nature of typical aDNA molecules,[12,13] such analyses invariably rely on mapping short reads to a modern reference genome. This enables the identification of single-nucleotide polymorphisms (SNPs) and small indels for phylogenomic and population analyses, and identification of variants with functional consequences.[3–5] But such methods overlook larger-scale differences, such as chromosomal rearrangements. In fact, the only large-scale differences that can be identified are cases where a modern sequence is absent in the ancient genome.[7]

aDNA studies have also explored epigenetic differences by recovering methylated cytosines from ancient templates[14–16] or through DNA decay patterns, which can indicate nucleosome positioning.[15] Yet there is no genome-wide epigenetic data for many ancient species of interest, such as the woolly mammoth (†*Mammuthus primigenius*). Here, too, the short length of aDNA makes it difficult to place epigenetic features in their genomic context.

One useful class of techniques for generating larger-scale insights into genomes is based on Hi-C,[17] which interrogates the shape of whole chromosomes. Hi-C exploits DNA-DNA proximity ligation[18,19] to ligate DNA sequences that are nearby in 3D within the cell nucleus, even if the sequences are distant along the 1D contour of the chromosome. Contact data can also determine the relative positioning of sequences along chromosomes, since sequences that tend to be in contact are more likely to be nearby in 1D. Consequently, Hi-C facilitates genome assembly, finds errors in draft contigs (contiguous sequences) that turn out to be incorrectly assembled, and reliably orders and orients contigs into chromosome-length scaffolds.[20–22] In addition, contact patterns reflect the epigenetic state and activity level of loci genome-wide.[17,23] Thus, generating Hi-C data for ancient samples would be valuable.

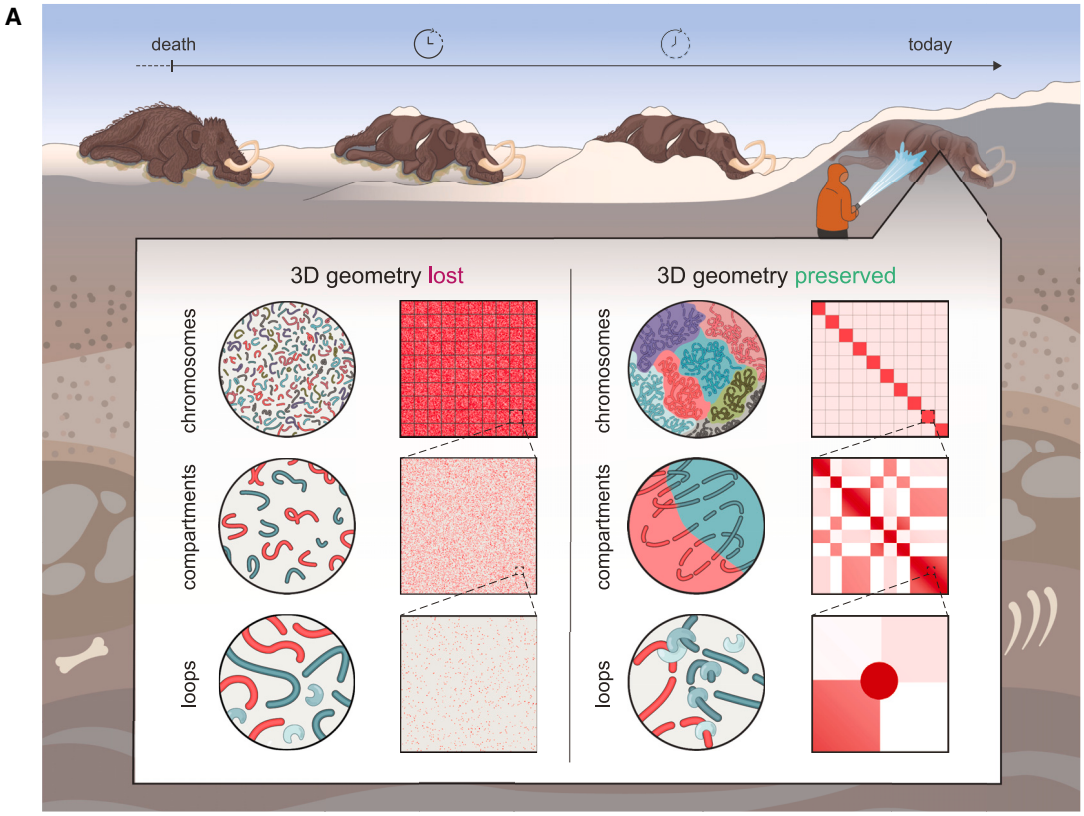
However, Hi-C might be impossible in ancient samples, because the surviving aDNA fragments are so short that they may diffuse through space, gradually erasing the ancient chromosomal morphology (Figure 1A).

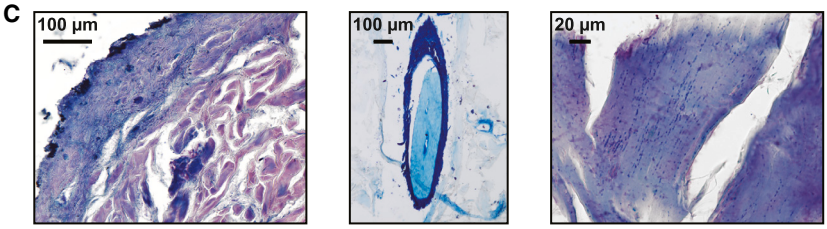
The purpose of this paper is, therefore, to answer three questions.

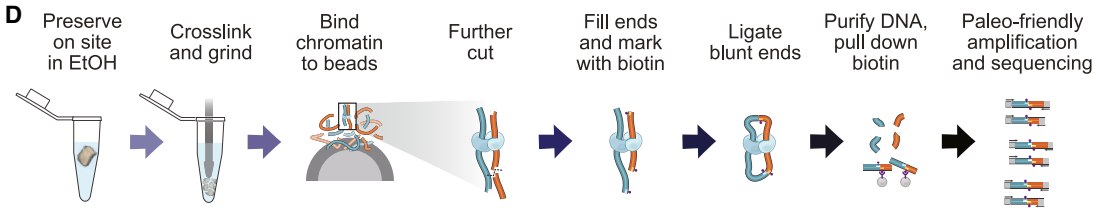
First, to what extent can the 3D morphology of chromosomes be preserved in ancient samples? To answer this question, we developed PaleoHi-C, a variant of the *in situ* Hi-C protocol[23] adapted for ancient samples (Figure 1). We applied PaleoHi-C to a permafrost-preserved woolly mammoth skin sample that is 52,000 years old. We demonstrate that this sample exhibits all the architectural features that are typically visible in modern Hi-C maps—chromosome territories, Barr bodies, active and inactive chromatin compartments, domains, and loops. We confirm the presence of several of these properties in a second sample that is 39,000 years old. Our results demonstrate that the morphological features of chromosomes can remain intact on scales as short as 50 nanometers (nm), forming non-mineralized fossils, or subfossils, that can persist for many millennia.

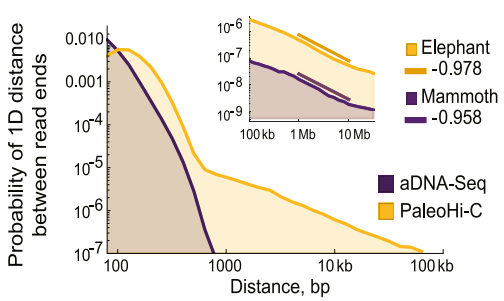
Second, can the 3D morphology of ancient chromosomes illuminate the biology of ancient species? To answer this question, we use PaleoHi-C data to generate a genome assembly for the woolly mammoth, including the reconstruction of its karyotype. We then compare contact patterns in woolly mammoth skin to those in elephant skin and other mammals. We show that it is possible to identify genes whose transcriptional state is altered in woolly mammoth as compared to elephant. We also show that the woolly mammoth's Xi exhibits a tetradic architecture that is distinct from the bipartite Xi architecture in human and mouse and confirm that this architecture is also present in modern elephants.

Third, how is it possible that the morphology of chromosomes is not erased by the action of diffusion over such long periods of time? We show that the behavior of the ancient chromatin is consistent with a glass transition, a physical process whereby









F

Total read count	4,444,894,354
Unique paired alignments (%)	24,415,411 (0.55%)
Short range (%)	10,641,052 (0.24%)
Long range, intrachromosomal (%)	1,763,225 (0.04%)
Long range, interchromosomal (%)	2,797,415 (0.06%)

(legend on next page)

factors like cooling and dehydration can effectively arrest the diffusion of the molecules in a material. We hypothesize that this glass transition was induced by spontaneous freeze-drying of the woolly mammoth tissue, shortly after death, in the cold Siberian climate. We provide some experimental evidence that supports this hypothesis.

Taken together, our findings highlight new approaches to the study of genome architecture, epigenetics, and gene expression in ancient samples and open the door to generating *de novo* genome assemblies for ancient and extinct species. They also demonstrate that the morphology of ancient biosamples can be well preserved at scales as fine as 50 nm.

RESULTS

We applied PaleoHi-C to skin from a 52,000-year-old woolly mammoth

We collected skin samples from the head of a Late Pleistocene woolly mammoth, ID IN18-032, near Belaya Gora, Sakha Republic, Siberia (N68.57887, E147.16055; Figures 1B and S1A). Conventional aDNA sequencing data has been published from this specimen under the ID "YakInf."[24] Carbon dating implied the mammoth had lived more than 45,000 years ago,[24] and mitochondrial DNA analysis suggested 52,000 years (see STAR Methods).

Histology showed some preserved skin morphology, including an apparent basal layer of keratinocytes and underlying dermis, with subdermal skeletal muscle. We also observed intact hair follicles, including the hair shaft and outer root sheath. On longitudinal sections, some skeletal muscle morphology was also preserved, with what appeared to be mammoth nuclei distributed along myofibers. The putative nuclei were visible when stained with DAPI and had a wrinkled appearance, consistent with dehydration (Figures 1C and S1B). These findings are in line with histological studies of other well-preserved mammoths.[25,26] We concluded that IN18-032 was well-preserved at both macroscopic and microscopic scales.

We applied a modified *in situ* Hi-C protocol,[23] dubbed PaleoHi-C, to generate chromatin contact data for IN18-032. The principal modifications were (1) immediate preservation in the field using ethanol to reduce microbial activity, (2) sequestration of chromatin bodies on beads, (3) omission of shearing and size selection given the lack of high molecular weight DNA, and (4) library amplification using a paleo-compatible workflow (see Figure 1D and STAR Methods). Twenty-six experiments were conducted, generating ~4.4 billion read pairs (Figure 1F). To minimize contamination, experiments were performed in an aDNA laboratory that had not previously processed modern elephant samples.

We generated *de novo* genome assemblies of the African and Asian elephants with chromosome-length scaffolds

To evaluate the PaleoHi-C data, it would be useful to have chromosome-length genome assemblies for the woolly mammoth's closest living relative, the Asian elephant (*Elephas maximus*), and its next-closest relatives, the African elephants (genus *Loxodonta*). However, the published assemblies for these taxa are fragmentary.

We therefore used published[27] and newly generated *in situ* Hi-C data to produce genome assemblies with chromosome-length scaffolds both for the African savanna elephant (*Loxodonta africana*; hereafter referred to as "African elephant"), Loxafr3.0_HiC (upgrading the assembly in Palkopoulou et al.[28]), and for the Asian elephant, ASM1433276v1_HiC (upgrading Tollis et al.[29]) (Figure S1D; Table S1, #1–2). The assemblies showed high conservation of synteny, suggesting that chromosomal rearrangements are rare in elephantids (STAR Methods).

PaleoHi-C read pairs aligning to elephant derive from authentic ancient mammoth DNA

Next, we attempted to align the PaleoHi-C data to the African elephant genome assembly, Loxafr3.0_HiC. As is common for aDNA, most of the 4.4 billion read pairs aligned to contaminants (Figure S1E). However, 2.55% (113,233,202 read pairs) aligned to the elephant. The vast majority of these were not unique, or aligned poorly or near one another in the elephant (<20 kb [kilobases] apart), but ~0.1% of the total (4.6 million) were candidate long-range chromatin contacts. For Hi-C performed on a modern Asian elephant sample, this proportion was 42.6% (Table S1, #3–4; STAR Methods).

To establish that the reads reflected aDNA, we used PMDTools[30] to look for patterns of DNA damage, such as cytosine deamination in single-stranded 5′ overhangs, that are common in aDNA.[31,32] Damage was elevated in all mammoth

Figure 1. PaleoHi-C reveals that the morphology of chromosomes is preserved in a 52,000-year-old sample of woolly mammoth skin

(A) Ancient DNA is fragmented. The fragments may be free to diffuse through space, erasing the 3D morphology of ancient chromosomes (left), including chromosome territories (top), chromatin compartments (middle), and point-to-point loops (bottom). If so, assays like Hi-C will fail. If diffusion is limited, these features may survive (right) and potentially be examined using Hi-C.

(B) Primary mammoth sample used in this study.

(C) Histology of skin (left), hair follicles (center), and subdermal muscle (right).

(D) PaleoHi-C overview. Samples are collected into ethanol in the field. In the laboratory, samples are crosslinked with formaldehyde. Tissue is disrupted and ground. Instead of isolating nuclei, as in *in situ* Hi-C, chromatin bodies are bound to beads and cut with a restriction enzyme. Ends are repaired and blunted, introducing biotin. After DNA-DNA proximity ligation, junctions are captured onto streptavidin beads, built into libraries using a paleo-compatible workflow, and sequenced.

(E) PaleoHi-C data (yellow) and aDNA-Seq data (purple) for the same sample were aligned to African elephant. The log-log histogram shows relative contact probability vs. 1D genomic distance between read ends. A fraction of PaleoHi-C reads reflects contacts between loci that lie far away in 1D. Not so for aDNA-Seq. For the (E) inset, similar plot comparing PaleoHi-C on woolly mammoth skin vs. *in situ* Hi-C on Asian elephant skin. Power laws are seen in the 1–10 Mb distance range (solid lines), with nearly identical scalings (mammoth: −0.96; elephant: −0.98).

(F) PaleoHi-C summary statistics.

See also Figure S1.

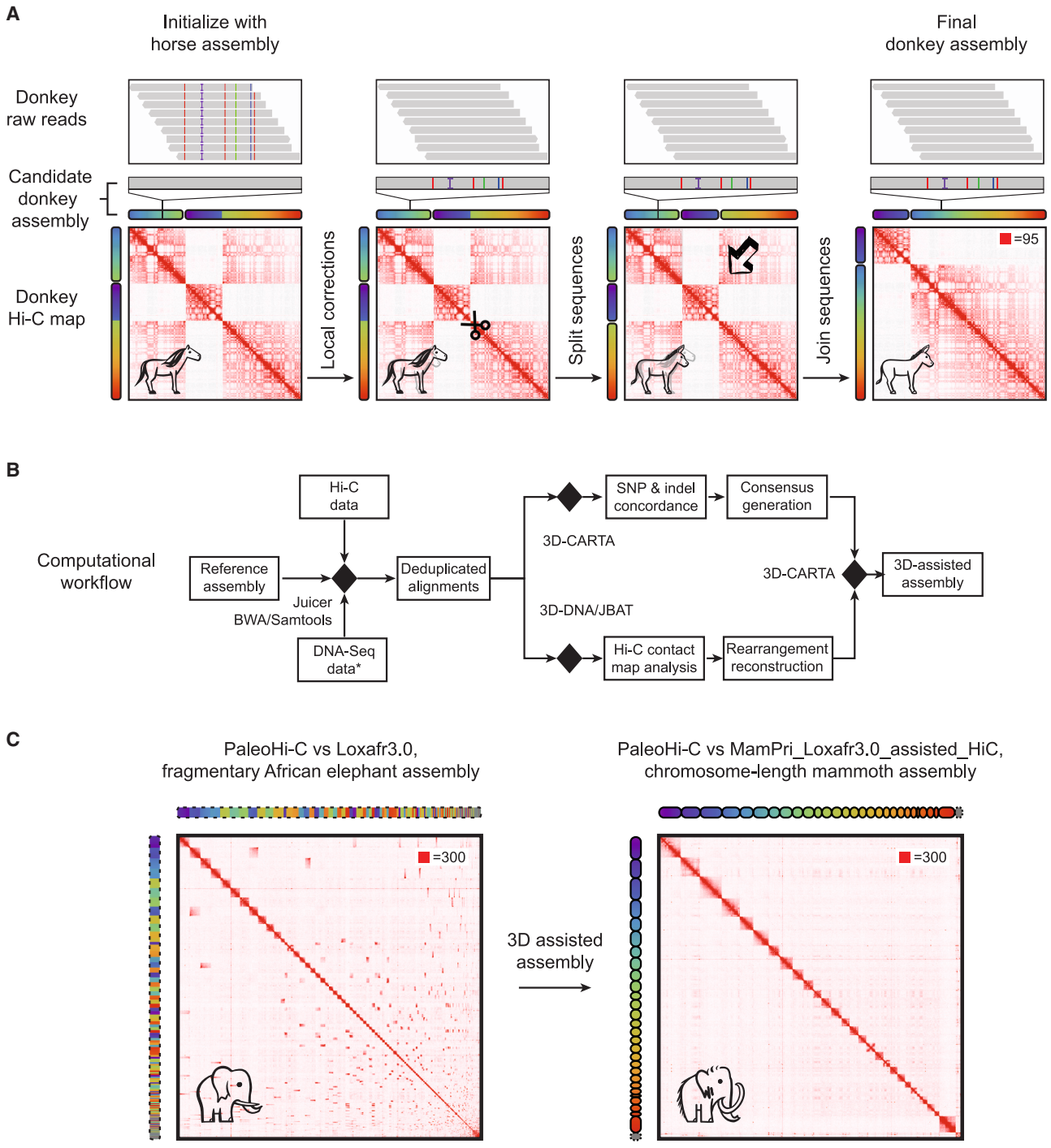


Figure 2. We developed reference-assisted 3D genome assembly and used it to assemble a woolly mammoth genome with chromosome-length scaffolds

(A) Schematic showing 3D genome assembly of donkey using horse (EquCab2.0) to assist. We initially assume that the donkey genome is identical to EquCab2.0 (leftmost column), and correct this assumption step-by-step, generating intermediate assemblies (columns 2 and 3), and finally a donkey genome assembly with accurate local sequence and chromosome-length scaffolds (column 4). We illustrate using two representative horse sequences, chr8 and chr24. Bottom row: Contact maps showing reads from donkey aligned to genome assemblies from each step. 2nd-to-last row: Chromograms illustrating locus order, with respect to the final (correct) order, which begins with purple and ends with red. 2nd row: Zooming in on DNA sequence for each step. Gray indicates that the current assembly sequence matches the initial, horse, sequence. Mismatches are shown in color: A (green), C (blue), G (yellow) and T (red). The "I" icon indicates insertions. Top row: Donkey sequence reads aligned to the current draft. Differences shown as in the 3rd row. The procedure for 3D-assisted assembly is as

(legend continued on next page)

libraries as compared to modern elephant libraries (Figure S1F; Table S1, #6; STAR Methods). This remained true when we restricted our analysis to candidate long-range chromatin contacts. Notably, the frequency of damage was low for an ancient sample, consistent with good preservation.

To confirm that these read pairs came from mammoths rather than elephants, we assigned reads to species based on SNPs that are fixed in woolly mammoth but absent in elephants or vice versa.[24] Reads from mammoth libraries overwhelmingly matched mammoth rather than elephant (99.8% vs. 0.2%), whereas reads from elephant libraries overwhelmingly matched elephant (99.9% vs. 0.1%) (see Figure S1F; Table S1, #7; STAR Methods). This remained true when we restricted the analysis to candidate long-range chromatin contacts. The few discordant assignments occurred at rates consistent with sequencing errors.

We conclude that the read pairs aligning to the elephant genome assembly in the PaleoHi-C dataset derived from ancient mammoth DNA, including the ~4.6 million candidate long-range mammoth chromatin contacts.

Candidate long-range mammoth chromatin contacts reflect ancient chromatin conformation, which has not been erased by subsequent molecular movement

We next evaluated whether the contacts between mammoth DNA sequences were ancient. By "ancient contacts," "ancient conformation," etc., we mean that they reflect the 3D structure at the time of the mammoth's death. This is important because, in principle, mammoth DNA could have diffused so extensively over the millennia that the ancient chromosome conformation was completely erased.

To do so, we examined the relationship between 1D distance along the contour of the elephant chromosome, s, and the contact probability $p(s)$. Hi-C experiments exhibit power-law scalings, $p(s) \propto s^{\gamma}$, reflecting structural features of chromatin across scales.[17,23] These scalings manifest as a linear decline when $p(s)$ is examined on a log-log plot, whose slope indicates the exponent of the power law.

The functions $p(s)$ for woolly mammoth skin and modern Asian elephant skin were almost identical. For example, the power law for distances between 1 megabase (Mb) and 10 Mb was $p(s) \propto s^{-0.96}$ for mammoth skin vs. $p(s) \propto s^{-0.98}$ for Asian elephant skin (Figures 1E and S1G). This is consistent with the preservation of the ancient chromosome conformation in the woolly mammoth sample.

In addition to the power law, representing bulk properties of the Hi-C assay, we also wanted to examine whether individual loci exhibited the ancient conformation. To test this, we attempted to upgrade the published fragmentary genome assembly of the African elephant (Loxafr3.0) into a chromosome-length assembly, by ordering and orienting the fragments. Specifically, we compared using the 4.6 million candidate long-range read pairs from the PaleoHi-C data vs. a comparable number of long-range contacts obtained by downsampling *in situ* Hi-C data from the African elephant. The resulting order and orientation were nearly identical and closely matched the chromosome-length genome assembly of the African elephant generated from a much larger modern dataset (see we generated *de novo* genome assemblies of the African and Asian elephants with chromosome-length scaffolds and STAR Methods).

These results provide strong evidence that the ancient chromosome conformation persists in the woolly mammoth sample and that PaleoHi-C is able to assay the resulting contacts.

Reference-assisted 3D genome assembly of a woolly mammoth yields chromosome-length scaffolds

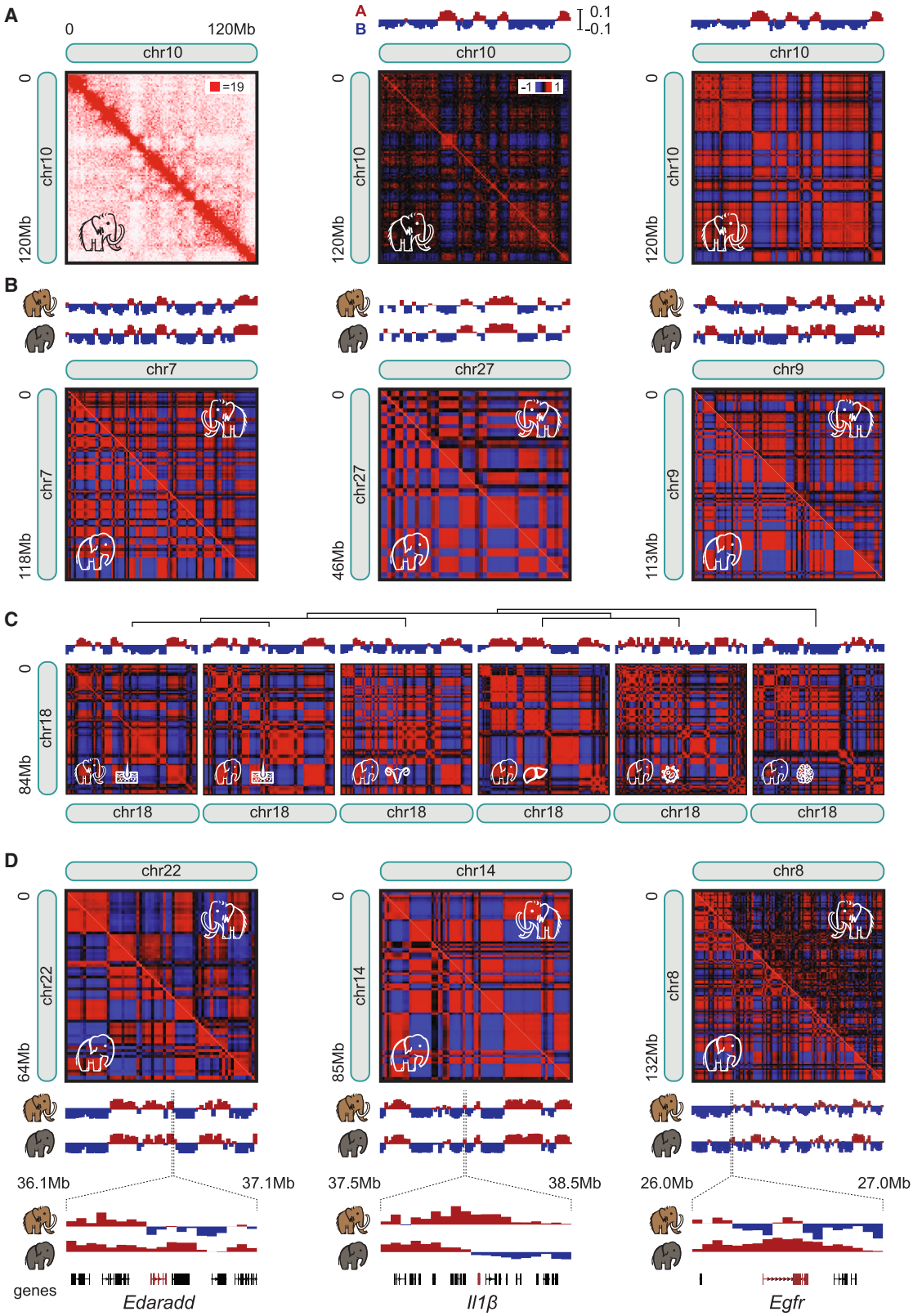
We next sought to assemble a genome for the woolly mammoth. Because published methods for *de novo* assembly of mammalian genomes with Hi-C typically require hundreds of millions of contacts, we developed a genome assembly method, dubbed "reference-assisted 3D genome assembly," that yields chromosome-length scaffolds using far fewer contacts (Figure 2A). The method begins by aligning the Hi-C data for a target species to an "assisting" reference assembly for a related species. The Hi-C data are compared to the assisting assembly to correct the local sequence, yielding an assembly that matches the target species at the sequence level, but with scaffolds drawn from the assisting assembly. (This step can also utilize aDNA-Seq.) Next, the contact map is examined to identify large-scale evolutionary changes (and errors in the assisting assembly). We make corresponding modifications to the scaffolding, such as splitting sequences and changing their ordering and orientation. This yields a modified genome assembly with chromosome-length scaffolds that accurately reflects the target genome, even if the assisting genome assembly is highly fragmented.

follows. 1st column: DNA-Seq and Hi-C reads from the donkey are aligned to the assisting horse genome assembly, EquCab2.0 (top two rows), making a contact map for the donkey with respect to EquCab2.0 (bottom row). Small-scale differences between aligned reads and EquCab2.0 are apparent (top two rows). 2nd column: Individual read alignments are examined and corresponding changes are introduced, as in traditional DNA resequencing. The result is a locally corrected genome assembly. Subsequent steps focus on constructing accurate chromosome-length scaffolds for donkey by looking for inconsistencies between donkey Hi-C and the candidate donkey genome assembly. In the bottom row, an inconsistency is highlighted with a pair of scissors: contacts are rare between two long sequences, which are adjacent to each other on a single chromosome in EquCab2.0 (see chromogram), indicating positions on different chromosomes in donkey. As such, they should be placed on separate scaffolds, and this is done in the next step. The resulting assembly (3rd column) contains three long scaffolds. The Hi-C map indicates that the first and last of these scaffolds are in frequent 3D contact, indicating that they are adjacent in the donkey genome. This change is made in the final donkey genome assembly (4th column). This strategy requires very little Hi-C data, compatible with PaleoHi-C, and yields a genome assembly that matches the true donkey genome both at single-base and chromosome scales.

(B) Workflow for reference-assisted 3D assembly. Hi-C data (and, optionally, DNA-Seq) are aligned to assisting assembly and deduplicated. Alignments are analyzed with 3D-DNA,[22] Juicebox Assembly Tools,[33] and 3D-CARTA (see STAR Methods) to correct scaffolds and local sequences.

(C) Combining PaleoHi-C data and African elephant draft genome assembly Loxafr3.0 (left) enables assembly of woolly mammoth genome with 28 chromosome-length scaffolds (right). Interactive map at https://t.3dg.io/3d-mammoth-Fig-2C.

See also Figure S2.



(legend on next page)

We validated the approach by comparing a genome assembly of the donkey, *Equus asinus*, that we created *de novo*, to a genome assembly that we created using a genome assembly for the horse (*Equus caballus*) to assist. The resulting genome assemblies were extremely similar (see STAR Methods and Figures S2A–S2C).

We then applied this strategy to the woolly mammoth, using PaleoHi-C data in combination with shotgun aDNA sequencing from the same sample,[24] to produce a genome assembly, MamPri_Loxafr3.0_assisted_HiC, assisted by the Loxafr3.0 African elephant genome assembly.[28] Despite the fact that Loxafr3.0 is fragmented, the resulting woolly mammoth assembly exhibited 28 chromosome-length scaffolds comprising 3.04 Gb of sequence (97.7% of total sequence), with a contig N50 length of 54 kb (Figure 2C; Table S2, #8). Overall, the genome assemblies we obtained for the woolly mammoth and the African and Asian elephant showed high conservation of synteny (Figure S2D). The results were similar when we used an assembly of the Asian elephant to assist (Figure S2E; Table S2, #11).

To annotate genes in MamPri_Loxafr3.0_assisted_HiC, we ran TOGA.[34] We assessed gene annotation using the fraction of complete Benchmarking Universal Single-Copy Orthologs.[35] We detected 94.6% of genes (Table S2, #9), a value that until recently was obtainable mostly in model organisms.[36]

When aligned to MamPri_Loxafr3.0_assisted_HiC, the PaleoHi-C data reflected a total of 4.8 million long-range contacts (defined as non-duplicate PaleoHi-C read pairs with two unique alignments to MamPri_Loxafr3.0_assisted_HiC that are not within 20 kb of one another; Table S3, #1). This contact data makes it possible to study the genome architecture and epigenetics of woolly mammoth skin.

Chromosome territories persist in ancient woolly mammoth skin

We began by looking for common architectural features visible in Hi-C data from modern samples. For example, at the coarsest resolution, eukaryotic genomes tend to fold into one of two architectural types.[37] Type I exhibits features associated with a telomere-to-centromere axis, such as centromere and telomere clustering, as well as chromosomal hairpin structures. Type II lacks these features. Instead, prominent chromosome territories are seen, i.e., chromosomes occupy discrete subvolumes of the nucleus.

The PaleoHi-C mammoth data clearly exhibited chromosome territories and a Type II architecture (Figures 2C and S2H), matching those seen in the African and Asian elephant and in line with other mammals with typical chromosome sizes.[37]

Cell-type-specific segregation between chromatin from the active (A) and inactive (B) compartments persists in ancient woolly mammoth skin

Zooming in, we looked for the segregation of the active (A) and inactive (B) chromatin seen in numerous species. This phenomenon, called compartmentalization, leads to a plaid pattern in the Hi-C contact matrix, with elevated contact frequencies between loci in the same compartment, and diminished contact frequencies between loci in different compartments. The pattern can be enhanced by rescaling to account for 1D sequence proximity and calculating the 1st- or 2nd-order autocorrelation matrix.[17]

Remarkably, we observed the plaid pattern in PaleoHi-C maps from the woolly mammoth. These findings are consistent with the presence of A and B compartments in woolly mammoth skin nuclei (Figures 3A and S3A).

A and B compartment assignments for individual loci can be determined by calculating the first principal component of the contact matrix. We did so for the woolly mammoth at 1-Mb resolution and compared the results to those for a panel of tissues derived from Asian elephants: skin, ovary, peripheral blood monocytes (PBMCs), liver, and brain. Strikingly, the assignments in elephant skin were consistently more similar to mammoth skin than to other elephant tissues (Pearson's *r* for elephant skin vs. mammoth skin: 0.93, vs. elephant ovary: 0.86, vs. elephant liver: 0.83, vs. elephant brain: 0.80, vs. elephant PBMCs: 0.76; Figures 3B, 3C, and S3B). This implies that the ancient woolly mammoth sample retains cell-type-specific architectural features, reflecting ancient chromatin activation. We discuss these data more extensively below.

Point-to-point chromatin loops persist in ancient woolly mammoth skin

Zooming in further, we looked for point-to-point chromatin loops in the woolly mammoth skin. Such loops bring pairs of genomic sites that lie far apart along the linear genome into close physical proximity, facilitating regulation of genes by distal enhancers.[23] They form via extrusion, a process by which large loops are generated by first creating a small loop between two adjacent DNA sites and then sliding these tether points in opposite directions along the 1D genome.[38–41] This process is mediated by

Figure 3. Segregation between active (A) and inactive (B) genome compartments can survive in ancient samples, enabling comparison of gene activity in woolly mammoth and Asian elephant skin

To facilitate comparisons, panels are with respect to the African elephant genome assembly Loxafr3.0_HiC.

(A) Raw contact data corresponding to chr10 (left) as well as 1st- (center) and 2nd-order (right) autocorrelation matrices, which can help enhance signal. The principal eigenvector of the autocorrelation matrix (above) can distinguish between A (active) and B (inactive) chromatin compartments. We adopt the sign convention where positive values (red) correspond to A and negative (blue) to B.

(B) 2nd-order autocorrelation matrices and A/B eigenvectors for chr7, chr27, and chr9 in skin from mammoth (above diagonal) and modern Asian elephant (below diagonal) are highly similar.

(C) 2nd-order autocorrelation matrices and A/B eigenvectors for chr18. From left to right: woolly mammoth skin PaleoHi-C, and Asian elephant *in situ* Hi-C for skin, ovary, liver, blood, and brain. Above: dendrogram reflecting Euclidean distances between eigenvectors.

(D) Comparison of skin from mammoth (above diagonal) and Asian elephant (below diagonal) for chr22, chr14, and chr8. Zoom-in: CRUSH annotations of A/B compartments at 50-kb resolution using PaleoHi-C reflect differences at genes *Edaradd* (on chr22, left), *Il1β* (chr14, center), and *Egfr* (chr8, right). Other matrices and eigenvectors are 1-Mb resolution. Interactive map at https://t.3dg.io/3d-mammoth-Fig-3.

See also Figure S3.

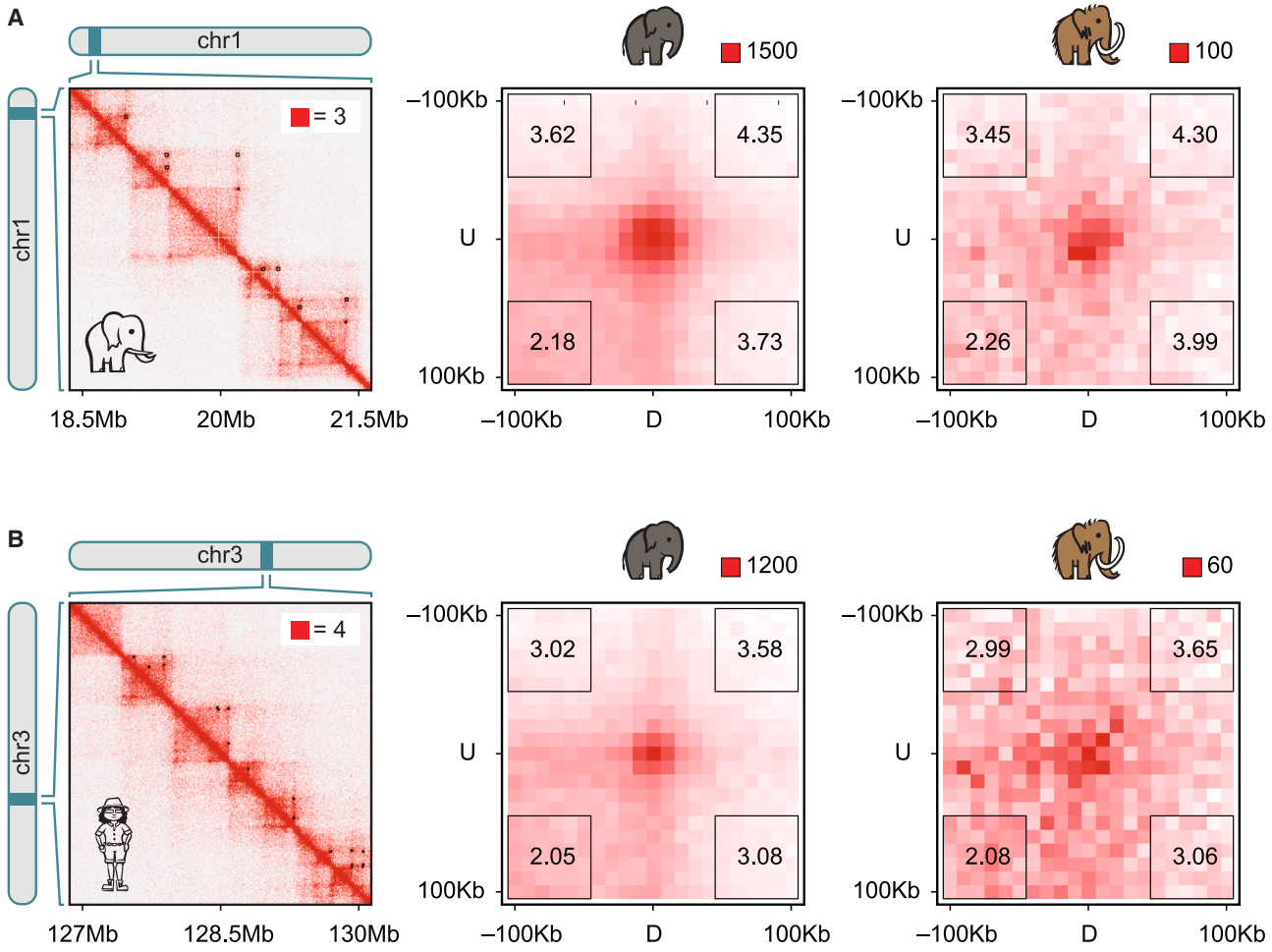


Figure 4. Chromatin loops can survive after 52,000 years in permafrost

(A) Left: Representative loop calls (black boxes) in a Hi-C map from African elephant primary fibroblasts,[27] chr1:18.5–21.5 Mb, available interactively at https://t.3dg.io/3d-mammoth-Fig-4A. Center: Aggregate peak analysis (APA) of *in situ* Hi-C data from Asian elephant skin using African elephant loop list at 10-kb resolution. The central bin sums contacts from 10-kb-by-10-kb pixels surrounding each loop; other bins correspond to translations of the loop pixel set. Enrichments of the central bin vs. the average for each corner (black box) are shown, indicating presence of loops in aggregate. Right: APA using the same loop list repeated on mammoth PaleoHi-C data.

(B) Left: Representative loop calls (black boxes) from chr3:127-130 Mb in Hi-C maps from human skin, downloaded from ENCODE (https://www.encodeproject.org), available interactively at https://t.3dg.io/3d-mammoth-Fig-4B. Center: APA for Asian elephant skin Hi-C using loops from human skin. Right: APA for mammoth skin PaleoHi-C using the same loop set.

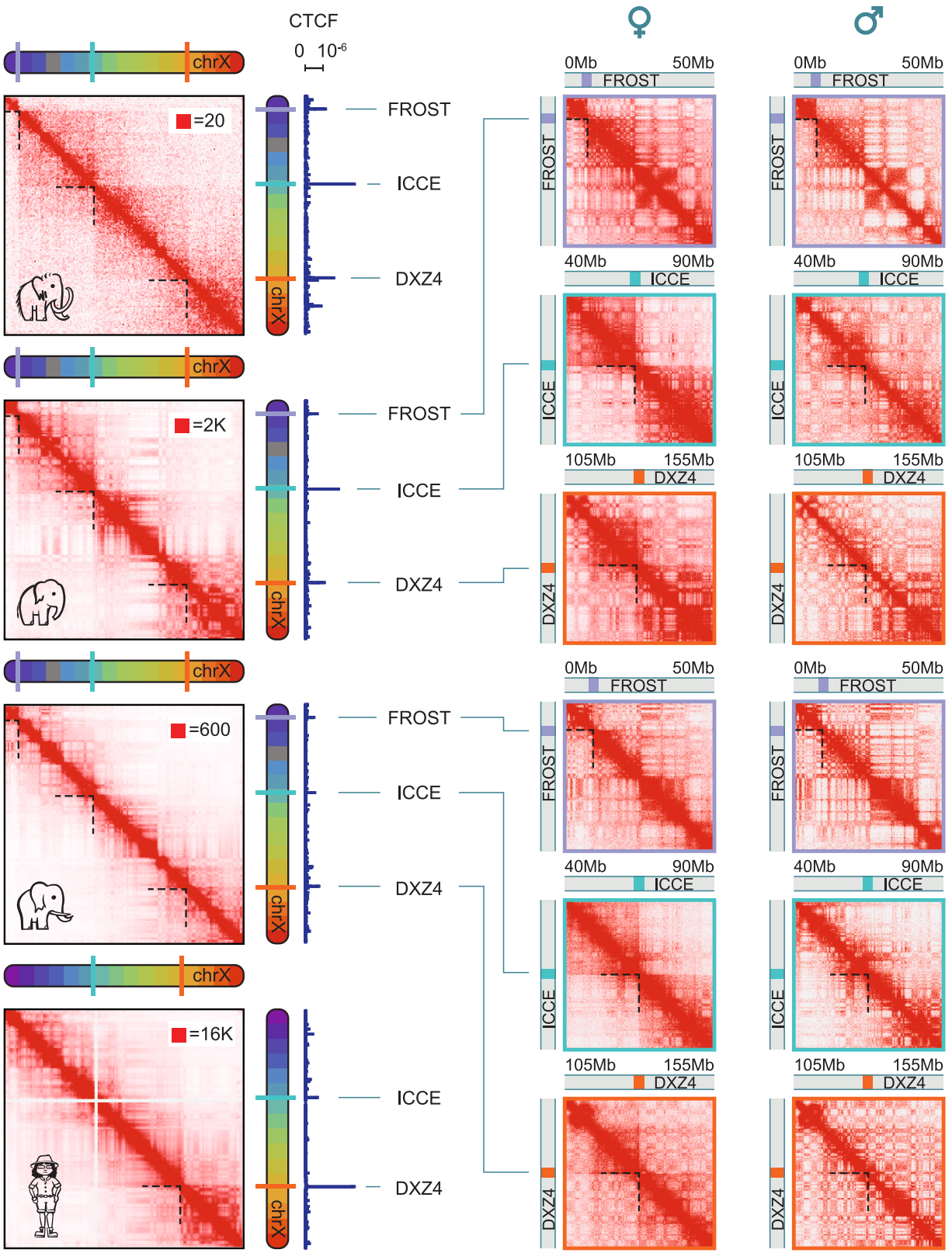
See also Figure S4.

cohesin and other structural maintenance of chromosomes (SMC) complexes and can be arrested by the protein CTCF (CCCTC-binding factor), leading to the formation of persistent chromatin loops and domains. Loops lead to very close spatial proximity between their anchor points (~50 nm, the diameter of cohesin[42]).

Because the mammoth contact map is too sparse to observe individual loops, which usually requires over a billion contacts, we looked for evidence of chromatin loops in aggregate. To do so, we identified 3,578 point-to-point loops in skin-derived fibroblasts from the African elephant[27] (see Figure 4A and STAR Methods). We then performed aggregate peak analysis (APA),[23] comparing the aggregate enrichment of the signal from the corresponding positions in woolly mammoth to the enrichment when these positions are translated in any direction (Figures 4A and S4B). We observed a pronounced enrichment in the center of the APA plot (2.26-fold enrichment relative to the lower-left corner). This enrichment was very similar to the enrichment seen for the same set of loops using *in situ* Hi-C data from Asian elephant skin (2.18).

This analysis indicates that many of the elephant loops are still present, at corresponding genomic positions, in woolly mammoth skin. Furthermore, once we account for the relative rarity of mammoth contacts in PaleoHi-C, loops are as visible in ancient samples as in modern samples.

We repeated this analysis using loops from human skin annotated by ENCODE (https://www.encodeproject.org/; see STAR Methods). The results were similar (Figures 4B and S4C),



(legend on next page)

demonstrating that loops can persist over vast evolutionary distances (the most recent common ancestor of humans and woolly mammoths lived 105 million years ago) as well as over millennia in permafrost.

These results also imply that, given enough PaleoHi-C data, it should be possible to observe individual chromatin loops in woolly mammoth samples.

Barr bodies and DXZ4 superdomains persist in ancient woolly mammoth skin

In female eutherian mammals, one of the two X chromosomes (designated Xi) is inactivated and adopts a conformation known as the Barr body. The Barr body is physically compacted and segregated from other chromosomes.[43,44] This segregation manifests in Hi-C maps as a reduced interchromosomal contact frequency.[23]

Notably, reduced X chromosome contact frequency is also apparent in mammoth PaleoHi-C data (Figure 2C). This observation implies that our sample derives from a female, matching the findings of an earlier study based on aDNA-Seq coverage.[24]

In humans and mice, the Xi exhibits a bipartite structure, with two large superdomains. Thought to form via loop extrusion, the boundary between these superdomains lies at, and depends on, the CTCF-binding tandem repeat element DXZ4.[23,38–41,45–47]

In the woolly mammoth, we clearly observed a superdomain boundary at DXZ4, orthologous to the human boundary (Figure 5).

Taken together, the results of the preceding sections demonstrate that chromosome territories, compartments, loops, Barr bodies, and Xi superdomains are apparent in PaleoHi-C data. It is not surprising that woolly mammoths exhibited these structures when they were alive, but it is notable that these features, as small as 50 nm in size, persist after 52,000 years in permafrost. These observations highlight the fact that the overall morphology of chromosomes can be well-preserved even while the DNA itself is highly fragmented. They also illustrate that this architecture can be reliably interrogated by PaleoHi-C.

We next asked whether it was possible to use PaleoHi-C data to learn new biology about the woolly mammoth.

PaleoHi-C reveals apparent differences in gene activity in mammoth skin vs. Asian elephant skin

We showed above that cell-type-specific patterns of chromatin activation persisted in the woolly mammoth skin sample. We next sought to identify species-specific epigenetic changes by closely comparing the compartmentalization patterns in woolly mammoth skin with those seen in its closest living relative, the Asian elephant.

To interrogate compartments at single-gene resolution, we used CRUSH (Compartment Refinement for the Ultraprecise Stratification of Hi-C), a method similar to the A-B index,[49] but designed for sparse Hi-C datasets, to generate A/B annotations in both woolly mammoth and Asian elephant at 50-kb resolution. We then looked for loci where the compartment score was substantially different in woolly mammoth skin vs. Asian elephant skin (see STAR Methods). This revealed 1,428 mammoth altered regions, dubbed MARs, spanning 820 genes (Table S3, #5). Of these genes, 425 were active in mammoth but not in elephant ("M > E genes"); and 395 were active in elephant but not in mammoth ("M < E genes").

For example, *Edaradd* switched from the active compartment in elephant skin to the inactive compartment in mammoth skin (Figure 3D). Human *EDARADD* protein interacts with *EDAR* through a domain that is highly conserved across mammals,[50] suggesting that its function is well-preserved in the clade. *EDAR* is strongly implicated in the development of hair follicles and sweat glands, and mutations in *EDAR* and *EDARADD* can give rise to hypohidrotic ectodermal dysplasia in humans, which is characterized by sparse and thin hair and the absence of sweat glands.[51] Interestingly, *EDAR* is under strong selective pressure in the Nunavik Inuit, an Arctic Indigenous people,[52] suggesting that this pathway may be relevant for adaptation to a cold environment.

A second example is *Il1β*, interleukin-1 beta, a regulator of wound healing and other immune and inflammatory responses.[53] It switched to the active compartment in the mammoth (Figure 3D).

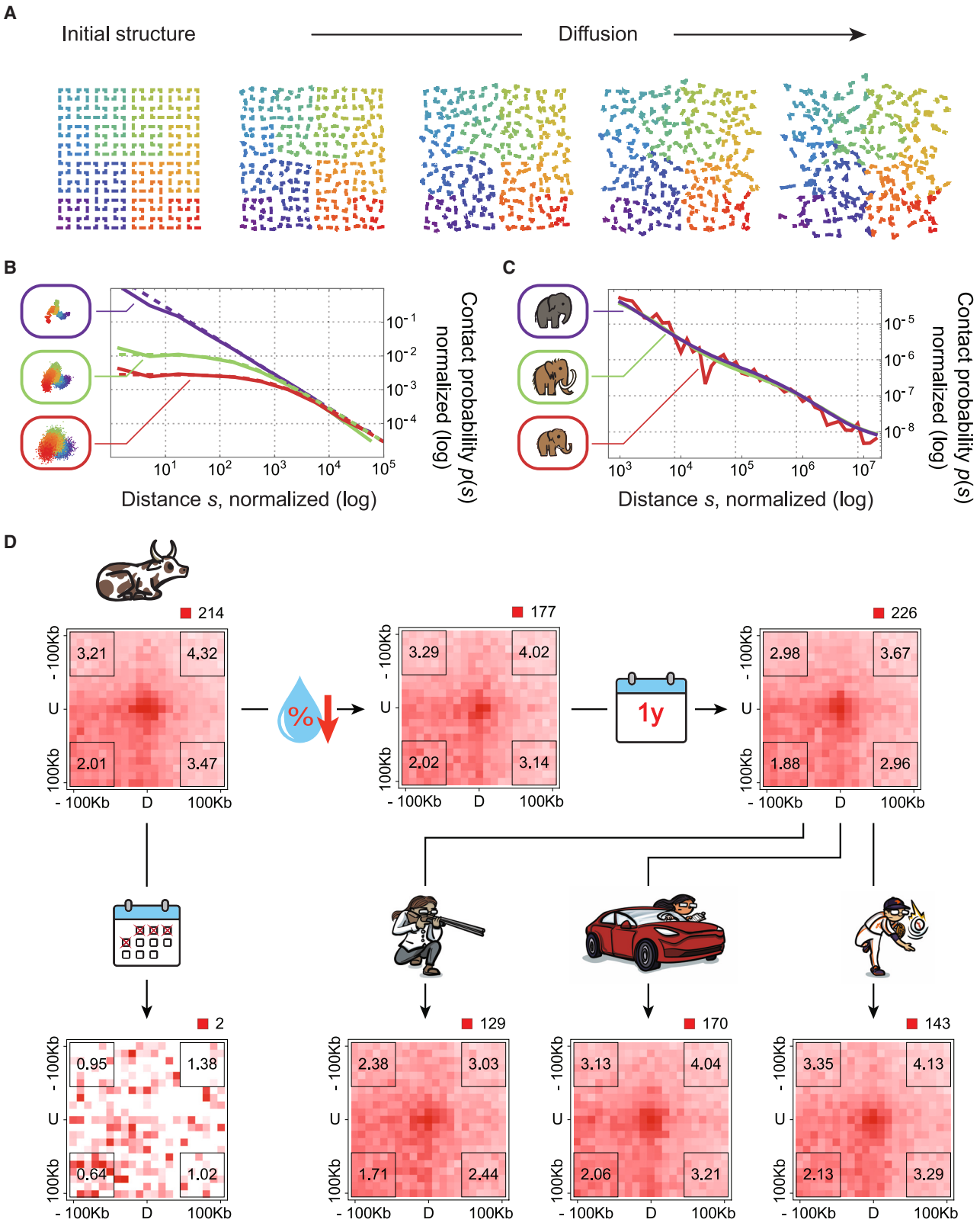
Finally, the epidermal growth factor receptor *Egfr*, known to be a key regulator of skin and hair homeostasis,[54] switched to the inactive compartment in mammoth (Figure 3D). Inhibition of *EGFR* in humans can cause eyelash trichomegaly, a condition in which eyelashes increase in both length and thickness, and more generally hypertrichosis (excessive hair growth).[55] By contrast, infusion with epidermal growth factor inhibited wool fiber production in sheep.[56]

We performed ontology analysis on the M > E and M < E gene sets using GOrilla.[57] We observed 39 significantly enriched terms (p value < 0.0001; STAR Methods). Interestingly, these include "hair follicle development" (GO:0001942, 6.56-fold enrichment), which includes many genes with a role in both hair development and homeostasis (Figures S3E and S3F; Table S3, #6–7). The enriched gene ontology (GO) terms also include nine others related more generally to cell fate in the epidermis (e.g., GO:0008544 for epidermis development).

These findings show that the woolly mammoth genome compartmentalization reflects species-specific activity states that suggest differentially regulated genes and pathways.

Figure 5. The inactive X chromosome (Xi) in woolly mammoths exhibits a tetradic structure, with 4 superdomains, distinct from the bipartite structure in humans

1st column, top to bottom: Contact maps for female mammoth, Asian elephant blood, African elephant fibroblasts,[27] and human GM12878 B-lymphoblastoid cells.[23] Chromosome orientation and chromogram color scheme (indicating locus order) are based on human chrX, purple (p-terminus) to red (q-terminus). (X chromosome locus order is highly conserved from human to elephantids.) Purple, cyan, and orange ticks indicate FROST, ICCE, and DXZ4 repeats, respectively. Tracks show how many reads align to the DXZ4/ICCE-associated CTCF binding motif.[48] Peaks correspond to CTCF-binding tandem repeats. Dashed lines indicate superdomain boundaries. Resolution: 1 Mb. 2nd and 3rd columns: Zoom-in on FROST, ICCE, and DXZ4 in Asian and African elephant females (2nd column) and males (3rd column). The superdomain boundaries are female specific. Interactive maps: https://t.3dg.io/3d-mammoth-Fig-5.
See also Figure S5.



(legend on next page)

The Xi in woolly mammoth and other elephantids exhibits a tetradic structure, distinct from the bipartite structure seen in humans and mice

As noted above, the human and mouse Xi exhibit a bipartite structure, with a prominent boundary at the CTCF-binding tandem repeat element DXZ4 that is conserved in woolly mammoth. Surprisingly, the woolly mammoth Xi does not exhibit a bipartite structure. Rather, we detected a tetradic Xi structure, with three boundary elements giving rise to four superdomains (Figure 5). In addition to DXZ4, two new boundary elements were evident.

One is a CTCF-binding tandem repeat called the inactive-X CTCF-binding contact element (ICCE).[41,48] Notably, ICCE is present in humans but does not form a superdomain boundary. When we estimated the number of CTCF binding sites associated with ICCE, we found that, unlike in humans, DXZ4 is not the predominant CTCF-binding locus on the mammoth X chromosome (Figure 5; STAR Methods). We hypothesize that when ICCE is longer, it is bound by more CTCF proteins, enhancing its ability to arrest extruding SMC complexes and giving rise to an additional superdomain boundary.

The third superdomain boundary, which we dubbed FROST, or funky repeat outlining superdomain topology, lies ~45 kb downstream of *Shroom2* in MamPri_Loxafr3.0_assisted_HiC. FROST lies at the boundary of the pseudoautosomal region (Figure S5A), which is much larger in mammoth and other elephantids as compared to humans.[58] To the best of our knowledge, FROST is not orthologous to a tandem repeat that has been annotated in humans.

The tetradic Xi is also seen in *in situ* Hi-C maps for female African and Asian elephants (Figure 5). As expected, the superdomain boundaries are absent in males of both species (Figures 5 and S5A). This is in line with the conservation of tetradic Xi architecture across elephantids.

Taken together, our data is consistent with a model whereby long CTCF-binding repeats, which can arrest extrusion by SMC-family complexes, emerge and disappear on the X chromosome during eutherian evolution, thereby repartitioning the Xi.[41,47,48]

The mammoth DNA in our sample has diffused minimally

As ancient chromosomes degrade into short DNA fragments, the fragments, like other small molecules, may diffuse through space. Such movement has the potential to erase the features of the ancient genome architecture, including power-law relationships at scales where displacement due to diffusion exceeds the initial 3D distance between the loci (Figures 6A, 6B, and S6A–S6F; STAR Methods). We therefore sought to measure the extent of mammoth DNA diffusion.

To do so, we plotted the contact probability $p(s)$ against 1D distance, s, along the mammoth chromosome. We also plotted the results for modern Asian elephant skin, as a proxy for the mammoth's initial state. The functions are nearly identical. No erasure of the power laws is seen, down to the smallest assayable distance of 500 base pairs (bp). (These experiments cannot reliably measure $p(s)$ for $s < 500$ bp because the restriction enzymes used cut too infrequently.)

Because 500 bp of nuclear chromatin has a spatial extent of ~50 nm, the preservation of power-law scaling down to 500 bp rules out diffusion with a root mean squared displacement (RMSD) of >50 nm. In fact, simulation studies suggest the RMSD must be several-fold lower (Figure S7A). These estimates are in line with the preservation of loops reported above, since 50 nm is roughly the 3D distance between loop anchor loci; it is also consistent with the persistence of other, larger features, such as chromosome territories (1–2 μm) and Barr bodies (1 μm).[59,60]

These results imply that the mammoth DNA in our sample has diffused minimally.

The 3D genome architecture of dehydrated tissue remained resilient after a year at room temperature

We wondered how genome architecture could be so well-preserved in a 52,000-year-old sample. We hypothesized that this might be due to dehydration, which is often used to preserve foods.[61,62]

To explore this hypothesis, we compared the results of *in situ* Hi-C on (i) fresh beef, (ii) beef after 96 h at room temperature, and (iii) dehydrated beef after a year at room temperature (STAR Methods). We observed chromosome territories, compartments, and loops in both the fresh and dehydrated beef. These features were absent in the beef exposed at room temperature without dehydration. Strikingly, chromatin architecture in a year-old dehydrated beef was robust to many perturbations, like getting run

Figure 6. The mammoth DNA in our sample has diffused minimally, in line with the resilience of chromatin architecture in modern dehydrated samples

(A) To theoretically model how chromosome architecture in ancient remains is affected by the passage of millennia, we begin with an initial, large chromatin chain; degradation is modeled as dissection of this chain into many short chains, whose subsequent diffusion is modeled as simple diffusion of phantom chains. This schematic illustrates the diffusion process. Here, we use a space-filling Hilbert curve as the initial chromatin conformation. 1D position along the trajectory is indicated using color, from purple to red. As particles diffuse, fine structure is disturbed, but coarse-scale co-localization persists, visible as monochromatic clusters. Eventually, the initial conformation is erased at all scales.

(B) The effects of diffusion are visible using contact data. Here, the initial chromatin conformation is modeled as a 3D random walk. Without diffusion (purple), random walks yield a power law relating distance along the polymer chain, measured in monomers, s, to the probability that two monomers are in contact, $p(s)$. Diffusion erases the power law up to the scale of the RMSD (lower: green; higher: red). Theoretical calculations (dashed) match simulations (solid). As particles diffuse, they again erase the initial conformation at increasingly large scales.

(C) Scalings for Asian elephant *in situ* Hi-C (purple), IN18-032 PaleoHi-C (green), and Yuka PaleoHi-C (red) are well-preserved down to 500 bp, the smallest distance assayable using our current methods.

(D) Genome architecture in modern dehydrated samples is extremely resilient. For example, APA shows that chromatin loops present in fresh cow liver (top left), are gone after 96 h at room temperature (bottom left). But after dehydration loops remain (top center), even after a year at room temperature (top right), and even if (bottom right) the sample is subsequently blasted with a shotgun, run over by a car, or hit with a fastball.

See also Figures S6 and S7.

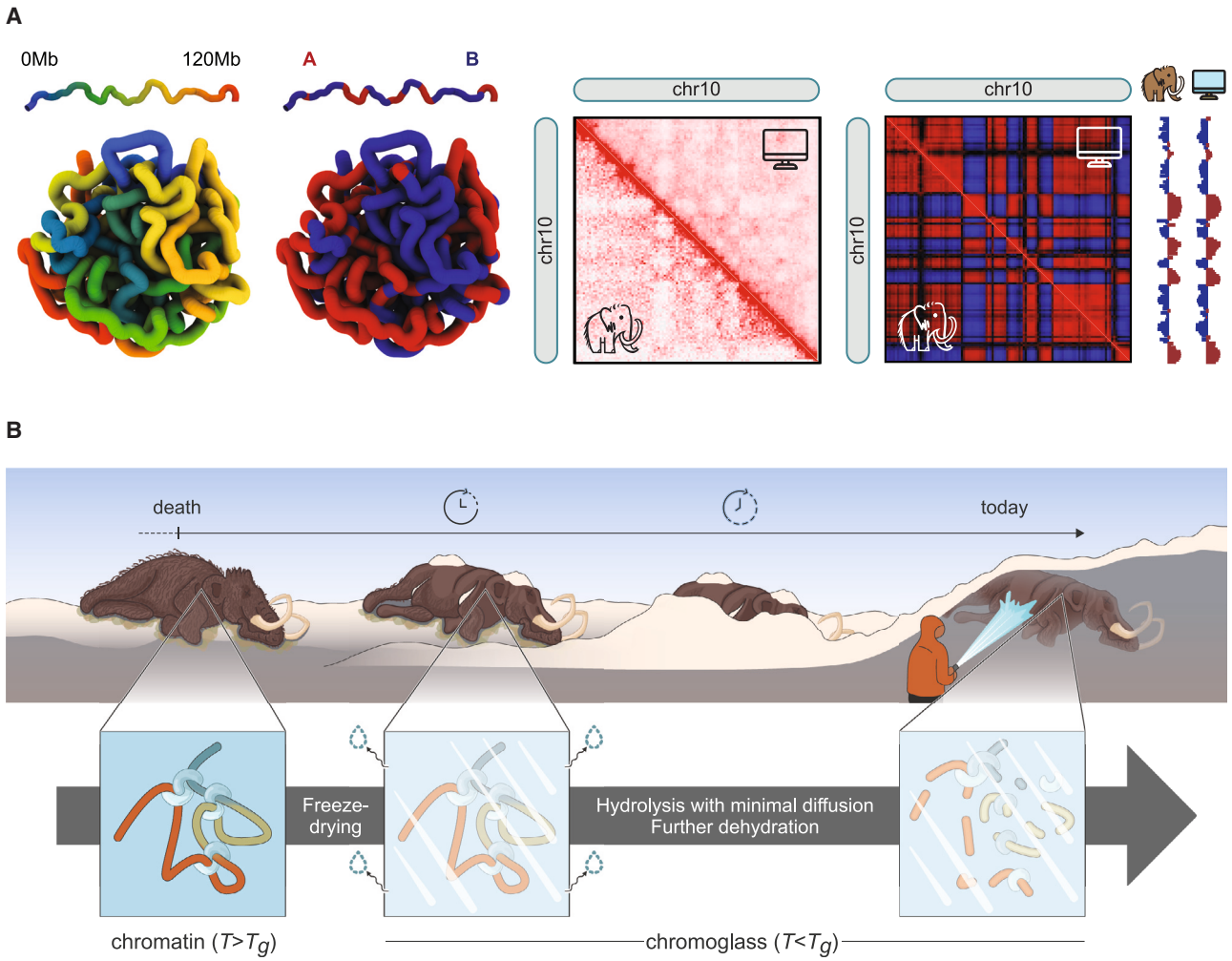


Figure 7. We hypothesize that the woolly mammoth samples studied here contain chromoglass—chromatin trapped in a glassy state, where molecular diffusion is minimal

(A) We combined polymer physics and PaleoHi-C data to infer 3D structures of woolly mammoth chromosomes. Shown is a representative structure from the simulated ensemble generated for mammoth chr10, colored by genomic position (1st panel) and A/B type (2nd panel). A contact map (3rd panel) and Pearson's autocorrelation matrix (4th panel) from the simulated structures (above diagonal) is compared to mammoth PaleoHi-C (below diagonal). Resolution: 1 Mb. The structures can be explored using Spacewalk at https://t.3dg.io/3d-mammoth-Fig-7A.

(B) We also propose a physical model for how the morphology of chromosomes can persist in ancient samples: the chromatin is contained in a glassy, non-crystalline, solid state. In this state, molecular diffusion, such as the diffusion of short aDNA fragments, is minimal. The glass transition might have been brought about by spontaneous freeze-drying: the gradual sublimation of the sample's water into the cold Siberian atmosphere. This could explain the preservation of morphological features across at least eight orders of magnitude in size: from the mammoth carcass (meters in length), to histological features (10s of microns), to loops (50 nm).

See also Figure S8.

over by a car, immersion in acid, and being pulverized with a shotgun (Figures 6D and S7B; Table S4, #1; Data S1; Video S1).

Combining PaleoHi-C data with physical modeling makes it possible to infer the 3D structure of woolly mammoth chromosomes

Hi-C data can be used to infer the three-dimensional structure of modern chromosomes using polymer physics.[63–65] To demonstrate this for ancient chromosomes, we constructed an ensemble of 3D structures for mammoth chromosome 10 that is maximally compatible with every individual PaleoHi-C contact, a method known as direct inversion[63] (see Figures 7A and S8; STAR Methods). The resulting ensemble, generated using the OpenMiChroM package,[65] closely recapitulates the PaleoHi-C map and exhibits the globular structures typical for mammalian chromosomes. This ensemble can be explored using the Spacewalk genome browser at https://t.3dg.io/3d-mammoth-Fig-7A.

Genome architecture persists in skin from a second woolly mammoth

To demonstrate that other samples are also compatible with PaleoHi-C, we studied a second woolly mammoth sample,

derived from Yuka, a juvenile female from 39,000 years ago. Yuka is considered one of the best-preserved woolly mammoths ever found.[66] A single PaleoHi-C library was sequenced to a depth of 7.2 million reads, yielding 14,669 contacts (Table S4, #3). We found that nuclear architecture persists in Yuka, including both chromosome territories and, setting aside noise due to low depth of coverage, power law scaling (Figures 6C and S6G). Other features could not be interrogated due to the sparsity of the map.

DISCUSSION

Many were surprised by the observation that intact DNA oligonucleotides can be preserved in ancient remains.[1,2] A further surprise has now emerged: morphological features of chromosomes, features that are not encoded in covalent bonds, persisted in a sample of woolly mammoth skin for a period of 52,000 years.

This allowed us to develop PaleoHi-C, a variant of *in situ* Hi-C designed for ancient samples, thereby enabling detailed genetic, epigenetic, and nucleomic studies of the woolly mammoth sample. We find that the sample's genome architecture is remarkably well-preserved, including chromosome territories, the cell-type-specific compartmentalization of active (A) and inactive (B) chromatin, domains, loops, Barr bodies, and Xi superdomains. The persistence of these structures may bear on the survival of the protein machinery that gives rise to them. For instance, the survival of loops may indicate that the underlying protein complexes, including CTCF and cohesin, are still present, in some form, in the ancient sample. Examining these ancient protein structures is a natural goal for future studies.

We show that PaleoHi-C reveals information about the biology of the woolly mammoth. For instance, combined with a novel algorithm for reference-assisted 3D genome assembly, PaleoHi-C allowed us to generate a reference genome for the woolly mammoth with chromosome-length scaffolds. This assembly approach could be useful for many extinct species. With deeper PaleoHi-C data, it should be possible to create assemblies of ancient genomes with chromosome-length scaffolds entirely *de novo*. We also observed distinct gene activation patterns in mammoth skin as compared to skin from its closest living relative, Asian elephant. These patterns enabled us to identify key regulators of hair growth, like *Edaradd* and *Egfr*, whose activity appears to be altered in mammoth skin vs. elephant skin. PaleoHi-C can also illuminate processes in mammalian evolution more broadly, such as the dynamic partitioning of the Xi, which is bipartite in humans and mice but tetradic in mammoths and other elephantids. These partitions are associated with the expansion and contraction of long tandem CTCF-binding repeats.

The preservation of 3D chromosome morphology is remarkable because, in an ancient biosample, the chromosomal polymer becomes highly discontinuous due to the breakdown of DNA's phosphate backbone. The short DNA oligonucleotides that remain might, in principle, diffuse independently through space. Why doesn't such diffusion erase the underlying genome architecture?

We propose that this is related to the concept of a glass transition (also called vitrification). Certain materials, such as glass, become extremely viscous when their temperature falls below a material-specific threshold known as the glass transition temperature (T_g). Although the resulting "amorphous" or "glassy" material lacks a repetitive crystal structure, it nevertheless appears solid and is characterized, at the molecular level, by extremely slow diffusion. Widely studied by food scientists, the glass transition—responsible for shelf-stable provisions ranging from tortilla chips to instant coffee—is critical for food preservation since it inhibits both cellular breakdown and the ability of microorganisms to enzymatically attack food.[61,62] Moreover, vitrification of cellular contents is a widespread and essential protective mechanism for anhydrobiotes—organisms that can survive for years in a desiccated state. These include microbial spores, plant seeds, tardigrades, and brine shrimp.[67–69]

We hypothesize that the woolly mammoth samples assayed in the present study contain chromatin trapped in a glassy state—chromatin glass, or "chromoglass" for short (Figure 7B).

There are several reasons to think that such a glass transition has occurred. First, our PaleoHi-C data shows that diffusion in the samples has been minimal. Second, it is well known from food science that animal flesh, when sufficiently cooled, undergoes a glass transition.[70] Third, the maximum temperature of permafrost is 0°C. The glass transition temperature of fresh beef, −13°C, is not far below this value,[71] and multiple factors, such as dehydration, could easily lead to vitrification at a much higher temperature. This is why dehydration has been used for thousands of years to make products like prosciutto legs and chuño stable at room temperature.[72,73] DNA-DNA crosslinks and protein scaffolding could also allow glass transition at a higher temperature.

Interestingly, once a material has entered a glassy state, it can gradually become more robust to heating, not only because processes such as dehydration might continue, but also because the glass transition temperature rises as a glass ages.[74] All this makes it plausible that the samples studied here have been held continuously below their glass transition temperature for tens of thousands of years. In particular, these factors might help explain how chromosome morphology was preserved during the process of sample excavation, when it was exposed to temperatures above freezing as it was flushed out of the soil using ambient temperature water.

We propose that the chromoglass might have formed by spontaneous freeze-drying: the gradual sublimation of the sample's water into the cold atmosphere of Late Pleistocene Siberia. This would be similar to modern industrial freeze-drying, although in the latter case, sublimation takes place in a vacuum chamber. The resulting material functioned as a time capsule for the conformation of the ancient biomolecules, including DNA, locking its architecture in place even as its phosphate backbone degraded. *In toto*, the glass transition appears to have facilitated preservation of morphological features across eight orders of magnitude in length: from the 3-m mammoth carcass (in the case of Yuka) to histological features at the scale of tens of microns and down to 50-nm loops. It is possible that much smaller features are preserved but cannot be detected using our present methods.

The abundance of chromosome subfossils within the broader fossil record is unknown. One key issue is how long DNA

architecture might be preserved in the chromoglass state. Ancient DNA studies have demonstrated that, in an appropriate context, short DNA oligonucleotides can survive for >2 million years.[75] Our estimates suggest that, if preserved in a glassy state, the architecture of aDNA, too, might remain well-preserved in samples of this age (see STAR Methods).

Ultimately, it will be useful to test PaleoHi-C on a wide range of biological materials representative of ancient and historic DNA to characterize the preservation conditions in which PaleoHi-C may be applicable. Freeze-dried samples from cold environments, such as Siberia, are likely to be good candidates. But samples from warmer environments that have undergone hot-air drying might also retain 3D genome architecture. Indeed, we observe comparable architecture preservation in freeze-dried and hot-air-dried samples (see Figures 6D and S7B; Data S1). Similarly, artificial dehydration processes, such as mummification, might lead to the formation of chromoglass. If so, a wide range of ancient Egyptian samples, from pharaohs to victual mummies, could contain chromoglass. Ancient Egyptian priests believed that mummification prepared a person or animal to be reanimated in a future life. They may have been closer to the mark than has been realized.

Limitations of the study

This work shows that genome architecture is preserved in certain ancient samples. How often this is so for other samples is unclear.

Genome architecture reflects differences in gene activity between mammoth skin and elephant skin. However, the samples were not obtained from the same anatomic location. Doing so is difficult, since elephants are endangered, and mammoths, extinct. We also cannot tell if the differences we observe are due to changes for the same cell type in mammoth vs. elephant, or if they reflect changes in cellular composition for mammoth vs. elephant skin. The results might also be influenced by differences in the likelihood of nuclear degradation for different cell types.

Gene ontology analysis is primarily based on work in mice and humans and may be less reliable in elephantids.

Finally, we make assumptions while modeling how genome architecture can be preserved in a decaying sample. First, we frequently describe the aDNA fragments as discrete molecules. Owing to ectopic DNA-DNA and DNA-protein crosslinks, the molecular structures seen in ancient samples may be more complex. Second, we model the movement of aDNA fragments as ordinary diffusion. This makes the model more tractable but is undoubtedly a simplified description. More sophisticated models are needed.

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 STUDY PARTICIPANT DETAILS
  - Woolly mammoth sample IN18-032
  - Yuka woolly mammoth sample
  - Asian elephant samples
  - Donkey sample
  - Bovine samples
  - Other samples
- METHOD DETAILS
  - PaleoHi-C protocol
  - Description of IN18-032 PaleoHi-C datasets
  - Yuka PaleoHi-C
  - Ancient DNA-Seq without USER treatment
  - Modern sample Hi-C library preparation and sequencing
  - Histological examination of sample IN18-032
  - Mammoth nuclei isolation and staining
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Dating of IN18-032
  - Sequencing and *de novo* assembly of modern elephants
  - IN18-032 PaleoHi-C preliminary analysis
  - Reference-assisted 3D genome assembly of the woolly mammoth
  - 3D genome architecture persists in IN18-032
  - Apparent differences in gene activity in mammoth skin vs. Asian elephant skin
  - Tetradic structure of woolly mammoth and modern elephant X chromosomes
  - Modeling the diffusion of DNA fragments in samples of ancient woolly mammoth skin
  - After drying, the 3D genome architecture of a tissue can remain stable for months, even at room temperature
  - 3D modeling of ancient chromosomes
  - 3D genome architecture persists in Yuka
  - Theoretical predictions suggest that DNA architecture in chromoglass could survive for far longer than the oldest known ancient DNA samples

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.06.002.

ACKNOWLEDGMENTS

We thank Beth Shapiro and Richard Green (protocol development advice); Richard Mitchell, Peter Hotez, Aliaksandr Astrowski, Aviva Aiden, Sirui Zhou, Susannah Selber-Hnatiw, Guy Rouleau, Emil Karpinski, George Church, Saul Godinez, Zane Colaric, Shaiza Pasha, Galina Aglyamova, Jefferson Sinson, Anat Vivante, Sergei Kliver, Dimoklis Gkountaroulis, Camilo Chacón-Duque, Douglass Turner, Joel Cracraft, and Paul Sweet (discussions); Adam Fotos and Scistories (figures); Mahdi Sadr (videography); the Houston Zoo veterinary team, Dan Fisher, Hojun Song, Brandon Lyons, Ray Riley, and the Rosenthal Meat Science and Technology Center team at Texas A&M, Mary Thompson, and Stephen O'Brien (samples); Judah Aiden and Thomas Griggs (experiment assistance); and Ron Mathis (baseball pitching). E.L.A. acknowledges the McNair Medical Institute, NIH ENCODE (UM1HG009375), US-Israel Binational Science Foundation (2019276), and NSF-DBI-2021795. Center for Theoretical Biological Physics is supported by NSF-PHY-2019745, PHY-2210291 (to J.N.O.) and hardware donated by AMD. The Welch Foundation supported E.L.A. (Q-1866), J.N.O. (C-1792), and A.B.O.J. E.L.A. and M.A.M.-R. acknowledge NHGRI-RM1HG011016. M.A.M.-R. acknowledges Spanish Ministerio de Ciencia e Innovación (PID2020-115696RB-I00) and ERC-609989 under the 7th Framework Program FP7/2007-2013. M.T.P.G. and J.A.R. acknowledge ERC-681396, DNRF-143, and NNF-21OC0070726. M.J.R. acknowledges NIH-R35-GM147467. L.D. acknowledges Swedish Research Council (2017-04647 and 2021-00625) and ERC 101054984 PrimiGenomes. A.R.-H. acknowledges PID2020-112557GB-I00 and CGL2017-83802-P. M.D.P. and B.J.Z.H. acknowledge NIGMS/R35-GM146852. A.L.R. acknowledges USFWS-AFE2129-

F22AP01215, UIUC College of ACES Seed Grant, and Fulbright Denmark. M.V.P. acknowledges NSF-DMS1763272, NIH-R01-AR079150, LEO Foundation (LF-AW-RAM-19-400008, LF-OC-20-000611), and Keck Foundation WMKF-5634988. Genome assembly was performed in association with the DNA Zoo Consortium (www.dnazoo.org), which acknowledges support from Illumina, IBM, and Pawsey Supercomputing Center.

AUTHOR CONTRIBUTIONS

M.T.P.G. and E.L.A. conceived of the project. M.A.M.-R., M.T.P.G., and E.L.A. oversaw analysis. V.P., I.S.P., N.I.P., A.V.P., and L.D. provided mammoth samples. M.S.-V., C.P.E., and O.D. developed PaleoHi-C, with contributions from S.S.T.M., B.O., A.G., E.S.L., M.T.P.G., and E.L.A. M.S.-V., C.P.E., and M.A.K. performed PaleoHi-C, with assistance from S.S.T.M., S.A.M. and P.L.P. S.A.R., N.A.L., N.A.S., P.L.P., E.A.K., and S.I.B. performed imaging experiments with coordination by O.D., M.V.P., N.B.R., and A.S.G. I.G.-T., R.C., J.P.F., and K.P. provided elephant samples. O.D., R.K., J.J., R.P.R., and A.D.O. performed Hi-C on modern samples. O.D. led development of genome assembly algorithms and performed assembly with assistance from D.W., S.S.B., M.S.S., J.A.R., M.A.M.-R., N.C.D., P.K., and A.R.-H. O.D., J.A.R., M.D., and C.F. performed bioinformatics analyses with assistance from M.V.P., D.W., M.S.S., G.M., and K.R. A.K. and M.J.R. performed CRUSH. O.D., E.L.A., and P.G.W. developed the chromoglass model with contributions from J.N.O., M.D.P., V.G.C., A.B.O.J., I.C., B.J.Z.H., R.M., and M.A.M.-R. Simulations were performed by I.C. (simple diffusion), B.J.Z.H. (degradation), and V.G.C. and A.B.O.J. (structure inference) with supervision by O.D. and E.L.A., M.D.P., and J.N.O., respectively. O.D., J.A.R., M.S.-V., C.P.E., A.L.R., V.G.C., J.N.O., P.G.W., E.S.L., M.A.M.-R., M.T.P.G., and E.L.A. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

E.L.A., M.T.P.G., and L.D. are on the scientific advisory board of Colossal Biosciences and hold stock options. From 2021 to 2023, M.A.M.-R. received consulting honoraria from Acuity Spatial Genomics. E.L.A. and O.D. are inventors on US provisional patent applications 16/308,386 (E.L.A. and O.D., filed 12/7/2018), 16/247,502 (E.L.A. and O.D., 1/14/2019), and PCT/US2020/064704 (E.L.A., 12/11/2020) by the Baylor College of Medicine and the Broad Institute, relating to the assembly methods in this manuscript.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors used generative AI to help perform, check, and typeset supplemental calculations (ChatGPT) and to create a soundtrack for Video S1 (ChatGPT and Suno.com). Afterward, the authors reviewed and edited as needed, and they take full responsibility for the publication's contents.

Received: June 30, 2023
Revised: March 7, 2024
Accepted: June 3, 2024
Published: July 11, 2024

REFERENCES

1. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., and Wilson, A.C. (1984). Dna-Sequences From the Quagga, an Extinct Member of the Horse Family. Nature *312*, 282–284.
2. Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. Nature *314*, 644–645.
3. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.
4. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.
5. Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F., Sher, A., et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. Nature *456*, 387–390.
6. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.
7. Martin, M.D., Cappellini, E., Samaniego, J.A., Zepeda, M.L., Campos, P.F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S.Y.W., Dietrich, F.S., et al. (2013). Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. Nat. Commun. *4*, 2172.
8. Ramos-Madrigal, J., Smith, B.D., Moreno-Mayar, J.V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M.T.P., and Wales, N. (2016). Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. Curr. Biol. *26*, 3195–3201.
9. Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., et al. (2011). A draft genome of Yersinia pestis from victims of the Black Death. Nature *478*, 506–510.
10. Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J., and Allaby, R.G. (2014). A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. Sci. Rep. *4*, 4003.
11. 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.
12. Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T.P., Willerslev, E., et al. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc. Biol. Sci. *279*, 4724–4733.
13. Kistler, L., Ware, R., Smith, O., Collins, M., and Allaby, R.G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Res. *45*, 6310–6320.
14. Llamas, B., Holland, M.L., Chen, K., Cropley, J.E., Cooper, A., and Suter, C.M. (2012). High-resolution analysis of cytosine methylation in ancient DNA. PLoS One *7*, e30226.
15. Pedersen, J.S., Valen, E., Velazquez, A.M.V., Parker, B.J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D.J., Kelly, T.K., Vang, S., et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. *24*, 454–466.
16. Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S., Meshorer, E., and Carmel, L. (2014). Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science *344*, 523–527.
17. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.
18. Cullen, K.E., Kladde, M.P., and Seyfred, M.A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. Science *261*, 203–206.
19. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.
20. Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125.
21. Kaplan, N., and Dekker, J. (2013). High-throughput genome scaffolding from in vivo DNA interaction frequency. Nat. Biotechnol. *31*, 1143–1147.

22. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95.

23. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

24. Díez-del-Molino, D., Dehasque, M., Chacón-Duque, J.C., Pečnerová, P., Tikhonov, A., Protopopov, A., Plotnikov, V., Kanellidou, F., Nikolskiy, P., Mortensen, P., et al. (2023). Genomics of adaptive evolution in the woolly mammoth. Curr. Biol. *33*, 1753–1764.e4.

25. Repin, V.E., Taranov, O.S., Ryabchikova, Tikhonov, A.N., and Pugachev, V.G. (2004). Sebaceous glands of the woolly mammoth, Mammothus primigenius Blum: histological evidence. Dokl. Biol. Sci. *398*, 382–384.

26. Kato, H., Anzai, M., Mitani, T., Morita, M., Nishiyama, Y., Nakao, A., Kondo, K., Lazarev, P.A., Ohtani, T., Shibata, Y., and Iritani, A. (2009). Recovery of cell nuclei from 15,000 years old mammoth tissues and its injection into mouse enucleated matured oocytes. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *85*, 240–247.

27. Álvarez-González, L., Arias-Sardá, C., Montes-Espuña, L., Marín-Gual, L., Vara, C., Lister, N.C., Cuartero, Y., Garcia, F., Deakin, J., Renfree, M.B., et al. (2022). Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals. Cell Rep. *41*, 111839.

28. 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.

29. 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.

30. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

31. Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.

32. Brotherton, P., Endicott, P., Sanchez, J.J., Beaumont, M., Barnett, R., Austin, J., and Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Res. *35*, 5717–5728.

33. Dudchenko, O., Shamim, M.S., Batra, S.S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., St Hilaire, B.G., Yao, W., Stamenova, E., et al. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at bioRxiv *x*, x. https://doi.org/10.1101/254797.

34. Kirilenko, B.M., Munegowda, C., Osipova, E., Jebb, D., Sharma, V., Blumer, M., Morales, A.E., Ahmed, A.-W., Kontopoulos, D.-G., Hilgers, L., et al. (2023). Integrating gene annotation with orthology inference at scale. Science *380*, eabn3107.

35. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. *38*, 4647–4654.

36. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In Gene Prediction: Methods and Protocols, M. Kollmar, ed. (Springer), pp. 227–245.

37. Hoencamp, C., Dudchenko, O., Elbatsh, A.M.O., Brahmachari, S., Raaijmakers, J.A., van Schaik, T., Sedeño Cacciatore, Á., Contessoto, V.G., van Heesbeen, R.G.H.P., van den Broek, B., et al. (2021). 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science *372*, 984–989.

38. Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456–E6465.

39. Nichols, M.H., and Corces, V.G. (2015). A CTCF Code for 3D Genome Architecture. Cell *162*, 703–705.

40. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

41. Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., Shamim, M., et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. USA *113*, E4504–E4512.

42. Anderson, D.E., Losada, A., Erickson, H.P., and Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. J. Cell Biol. *156*, 419–424.

43. Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

44. Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515–526.

45. Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., and Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579.

46. Jégu, T., Aeby, E., and Lee, J.T. (2017). The X chromosome in space. Nat. Rev. Genet. *18*, 377–389.

47. Froberg, J.E., Pinter, S.F., Kriz, A.J., Jégu, T., and Lee, J.T. (2018). Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. Nat. Commun. *9*, 5004.

48. Westervelt, N., and Chadwick, B.P. (2018). Characterization of the ICCE Repeat in Mammals Reveals an Evolutionary Relationship with the DXZ4 Macrosatellite through Conserved CTCF Binding Motifs. Genome Biol. Evol. *10*, 2190–2204.

49. Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol. Cell *67*, 837–852.e7.

50. Bryk, J., Hardouin, E., Pugach, I., Hughes, D., Strotmann, R., Stoneking, M., and Myles, S. (2008). Positive selection in East Asians for an EDAR allele that enhances NF-kappaB activation. PLoS One *3*, e2209.

51. Deshmukh, S., and Prashanth, S. (2012). Ectodermal dysplasia: a genetic review. Int. J. Clin. Pediatr. Dent. *5*, 197–202.

52. Zhou, S., Xie, P., Quoibion, A., Ambalavanan, A., Dionne-Laporte, A., Spiegelman, D., Bourassa, C.V., Xiong, L., Dion, P.A., and Rouleau, G.A. (2019). Genetic architecture and adaptations of Nunavik Inuit. Proc. Natl. Acad. Sci. USA *116*, 16012–16017.

53. Cai, Y., Xue, F., Quan, C., Qu, M., Liu, N., Zhang, Y., Fleming, C., Hu, X., Zhang, H.-G., Weichselbaum, R., et al. (2019). A Critical Role of the IL-1β-IL-1R Signaling Pathway in Skin Inflammation and Psoriasis Pathogenesis. J. Invest. Dermatol. *139*, 146–156.

54. Pastore, S., Mascia, F., Mariani, V., and Girolomoni, G. (2008). The epidermal growth factor receptor system in skin repair and inflammation. J. Invest. Dermatol. *128*, 1365–1374.

55. Alexandrescu, D.T., Kauffman, C.L., and Dasanu, C.A. (2009). Persistent hair growth during treatment with the EGFR inhibitor erlotinib. Dermatol. Online J. *15*, 4.

56. Moore, G.P., Panaretto, B.A., and Carter, N.B. (1985). Epidermal hyperplasia and wool follicle regression in sheep infused with epidermal growth factor. J. Invest. Dermatol. *84*, 172–175.

57. 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.

58. Raudsepp, T., and Chowdhary, B.P. (2015). The Eutherian Pseudoautosomal Region. Genome Res. *147*, 81–94.

59. Parada, L.A., McQueen, P.G., and Misteli, T. (2004). Tissue-specific spatial organization of genomes. Genome Biol. *5*, R44.

60. Mittwoch, U. (2014). Sex Chromosomes (Academic Press).

61. Ratti, C. (2001). Hot air and freeze-drying of high-value foods: a review. J. Food Eng. *49*, 311–319.

62. Nowak, D., and Jakubczyk, E. (2020). The Freeze-Drying of Foods-The Characteristic of the Process Course and the Effect of Its Parameters on the Physical Properties of Food Materials. Foods *9*, 1488. https://doi.org/10.3390/foods9101488.

63. Zhang, B., and Wolynes, P.G. (2015). Topology, structures, and energy landscapes of human chromosomes. Proc. Natl. Acad. Sci. USA *112*, 6062–6067.

64. Di Pierro, M., Zhang, B., Aiden, E.L., Wolynes, P.G., and Onuchic, J.N. (2016). Transferable model for chromosome architecture. Proc. Natl. Acad. Sci. USA *113*, 12168–12173.

65. Oliveira Junior, A.B., Contessoto, V.G., Mello, M.F., and Onuchic, J.N. (2021). A Scalable Computational Approach for Simulating Complexes of Multiple Chromosomes. J. Mol. Biol. *433*, 166700.

66. Rudaya, N., Protopopov, A., Trofimova, S., Plotnikov, V., and Zhilich, S. (2015). Landscapes of the "Yuka" mammoth habitat: A palaeobotanical approach. Rev. Palaeobot. Palynol. *214*, 1–8.

67. Crowe, J.H., Hoekstra, F.A., and Crowe, L.M. (1992). Anhydrobiosis. Annu. Rev. Physiol. *54*, 579–599.

68. Sallon, S., Cherif, E., Chabrillange, N., Solowey, E., Gros-Balthazard, M., Ivorra, S., Terral, J.-F., Egli, M., and Aberlenc, F. (2020). Origins and insights into the historic Judean date palm based on genetic analysis of germinated ancient seeds and morphometric studies. Sci. Adv. *6*, eaax0384.

69. Romero-Perez, P.S., Dorone, Y., Flores, E., Sukenik, S., and Boeynaems, S. (2023). When Phased without Water: Biophysics of Cellular Desiccation, from Biomolecules to Condensates. Chem. Rev. *123*, 9010–9035.

70. Roos, Y.H. (2010). Glass transition temperature and its relevance in food processing. Annu. Rev. Food Sci. Technol. *1*, 469–496.

71. Akköse, A., and Aktaş, N. (2008). Determination of glass transition temperature of beef and effects of various cryoprotective agents on some chemical changes. Meat Sci. *80*, 875–878.

72. Cato, M.P. (1998). De Agricultura (Prospect Books).

73. Haines, H.R., and Sammells, C.A. (2011). Adventures in Eating (University Press of Colorado).

74. Joule, J.P. (1867). Observations On The Alteration Of Thee Freezing Point In Thermometers. Am. J. Pharm. *1835–1907*, 420.

75. Kjær, K.H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T.S., Michelsen, C.S., Sand, K.K., Jelavić, S., Ruter, A.H., Schmidt, A.M.A., et al. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature *612*, 283–291.

76. Houck, M.L., Kumamoto, A.T., Gallagher, D.S., Jr., and Benirschke, K. (2001). Comparative cytogenetics of the African elephant (Loxodonta africana) and Asiatic elephant (Elephas maximus). Cytogenet. Cell Genet. *93*, 249–252.

77. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.

78. Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. Science *326*, 865–867.

79. Renaud, G., Petersen, B., Seguin-Orlando, A., Bertelsen, M.F., Waller, A., Newton, R., Paillot, R., Bryant, N., Vaudin, M., Librado, P., and Orlando, L. (2018). Improved de novo genomic assembly for the domestic donkey. Sci. Adv. *4*, eaaq0392.

80. Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA (The Pennsylvania State University).

81. Frönicke, L., Wienberg, J., Stone, G., Adams, L., and Stanyon, R. (2003). Towards the delineation of the ancestral eutherian genome organization: comparative genome maps of human and the African elephant (Loxodonta africana) generated by chromosome painting. Proc. Biol. Sci. *270*, 1331–1340.

82. Harris, H.L., Gu, H., Olshansky, M., Wang, A., Farabella, I., Eliaz, Y., Kalluchi, A., Krishna, A., Jacobs, M., Cauer, G., et al. (2023). Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. Nat. Commun. *14*, 3303.

83. Maschenko, E.N., Potapova, O.R., Heintzman, P.D., Kapp, J.D., Shapiro, B., Protopopov, A.V., Boeskorov, G.G., Pavlov, I.S., Plotnikov, V.V., Kolesov, S.D., et al. (2021). Morphology, Individual Age, DNA and Sex of the Yuka Mammoth (Mammuthus primigenius) from Northern Yakutia, Russia. Paleontol. J. *55*, 1230–1259.

84. Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., et al. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. *27*, 849–864.

85. Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.-C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., et al. (2011). Modernizing reference genome assemblies. PLoS Biol. *9*, e1001091.

86. Shumate, A., Zimin, A.V., Sherman, R.M., Puiu, D., Wagner, J.M., Olson, N.D., Pertea, M., Salit, M.L., Zook, J.M., and Salzberg, S.L. (2020). Assembly and annotation of an Ashkenazi human reference genome. Genome Biol. *21*, 129.

87. Steinberg, K.M., Schneider, V.A., Graves-Lindsay, T.A., Fulton, R.S., Agarwala, R., Huddleston, J., Shiryev, S.A., Morgulis, A., Surti, U., Warren, W.C., et al. (2014). Single haplotype assembly of the human genome from a hydatidiform mole. Genome Res. *24*, 2066–2076.

88. Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Tseng, E., Rowan, T.N., Low, W.Y., Zimin, A., Couldrey, C., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience *9*, giaa021. https://doi.org/10.1093/gigascience/giaa021.

89. Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebersberger, I., Pääbo, S., and Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature *439*, 724–727.

90. Rogaev, E.I., Moliaka, Y.K., Malyarchuk, B.A., Kondrashov, F.A., Derenko, M.V., Chumakov, I., and Grigorenko, A.P. (2006). Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius. PLoS Biol. *4*, e73.

91. Jung, W.J., Kim, S.W., Giri, S.S., Kim, H.J., Kim, S.G., Kang, J.W., Kwon, J., Lee, S.B., Oh, W.T., Jun, J.W., and Park, S.C. (2021). Janthinobacterium tructae sp. nov., Isolated from Kidney of Rainbow Trout

(Oncorhynchus mykiss). Pathogens *10*, 229. https://doi.org/10.3390/pathogens10020229.

92. Shideler, S.M., Goff, K., Gauthier, J., Levesque, R.C., and Lewenza, S. (2023). Complete genome sequence of Pseudomonas veronii strain OST1911 isolated from oil sand tailing pond water in Alberta, Canada. Microbiol. Resour. Announc. *12*, e0064623.
93. Chaudhary, D.K., Kim, D.-U., Kim, D., and Kim, J. (2019). Flavobacterium petrolei sp. nov., a novel psychrophilic, diesel-degrading bacterium isolated from oil-contaminated Arctic soil. Sci. Rep. *9*, 4134.
94. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.
95. Horakova, A.H., Calabrese, J.M., McLaughlin, C.R., Tremblay, D.C., Magnuson, T., and Chadwick, B.P. (2012). The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. Genome Biol. *13*, R70.
96. Astashyn, A., Tvedte, E.S., Sweeney, D., Sapojnikov, V., Bouk, N., Joukov, V., Mozes, E., Strope, P.K., Sylla, P.M., Wagner, L., et al. (2023). Rapid and sensitive detection of genome contamination at scale with FCS-GX. Preprint at bioRxiv *x*, x. https://doi.org/10.1101/2023.06.02.543519.
97. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101.
98. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics *26*, 589–595.
99. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv *x*, x. https://doi.org/10.48550/arXiv.1303.3997.
100. Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P., and Aiden, E.L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Syst. *6*, 256–258.e1.
101. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
102. Rowley, M.J., Poulet, A., Nichols, M.H., Bixler, B.J., Sanborn, A.L., Brouhard, E.A., Hermetz, K., Linsenbaum, H., Csankovszki, G., Lieberman Aiden, E., and Corces, V.G. (2020). Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. elegans to mammals. Genome Res. *30*, 447–458.
103. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
104. Gouy, M., Tannier, E., Comte, N., and Parsons, D.P. (2021). Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. Methods Mol. Biol. *2231*, 241–260.
105. Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.
106. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. *67*, 901–904.
107. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
108. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinf. *15*, 356.
109. Kwon, D., Lee, J., and Kim, J. (2019). GMASS: a novel measure for genome assembly structural similarity. BMC Bioinf. *20*, 147.
110. Tempel, S. (2012). Using and Understanding RepeatMasker. In Mobile Genetic Elements: Protocols and Genomic Applications, Y. Bigot, ed. (Humana Press), pp. 29–51.
111. Morgulis, A., Gertz, E.M., Schäffer, A.A., and Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. Bioinformatics *22*, 134–141.
112. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
113. Osipova, E., Hecker, N., and Hiller, M. (2019). RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. GigaScience *8*, giz132. https://doi.org/10.1093/gigascience/giz132.
114. Suarez, H.G., Langer, B.E., Ladde, P., and Hiller, M. (2017). chainCleaner improves genome alignment specificity and sensitivity. Bioinformatics *33*, 1596–1603.
115. Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. USA *100*, 11484–11489.
116. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.
117. Tange, O. (2018). GNU Parallel https://doi.org/10.5281/zenodo.1146014.
118. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. Methods Ecol. Evol. *9*, 410–419.
119. 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.
120. 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.
121. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.
122. 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.
123. Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., and Suchard, M.A. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. *30*, 713–724.
124. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.
125. Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA *86*, 1939–1943.
126. Hansen, A., Willerslev, E., Wiuf, C., Mourier, T., and Arctander, P. (2001). Statistical evidence for miscoding lesions in ancient DNA templates. Mol. Biol. Evol. *18*, 262–265.
127. 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.
128. Pop, M., Phillippy, A., Delcher, A.L., and Salzberg, S.L. (2004). Comparative genome assembly. Brief. Bioinform. *5*, 237–248.

129. Kim, J., Larkin, D.M., Cai, Q., Lewin, H.A., Ma, J., Zhang, Y., Zhang, Y., Ge, R.L., Auvil, L., Capitanu, B., and Zhang, G. (2013). Reference-assisted chromosome assembly. Proc. Natl. Acad. Sci. USA *110*, 1785–1790.

130. Kolmogorov, M., Raney, B., Paten, B., and Pham, S. (2014). Ragout-a reference-assisted assembly tool for bacterial genomes. Bioinformatics *30*, i302–i309.

131. Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. *23*, 258.

132. Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner (Lawrence Berkeley National Lab.(LBNL)).

133. Goenka, S.D., Turakhia, Y., Paten, B., and Horowitz, M. (2020). SegAlign: A Scalable GPU-Based Whole Genome Aligner. In SC20: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1–13.

134. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

135. Yang, F., Alkalaeva, E.Z., Perelman, P.L., Pardini, A.T., Harrison, W.R., O'Brien, P.C.M., Fu, B., Graphodatsky, A.S., Ferguson-Smith, M.A., and Robinson, T.J. (2003). Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. Proc. Natl. Acad. Sci. USA *100*, 1062–1066.

136. Trifonov, V.A., Vorobieva, N.V., Serdyukova, N.A., and Rens, W. (2017). FISH with and Without COT1 DNA. In Fluorescence In Situ Hybridization (FISH): Application Guide, T. Liehr, ed. (Springer Berlin Heidelberg), pp. 123–133.

137. Nohara, K., Chen, Z., and Yoo, S.-H. (2017). A Filtration-based Method of Preparing High-quality Nuclei from Cross-linked Skeletal Muscle for Chromatin Immunoprecipitation. J. Vis. Exp. *10*, 3791–56013.

138. 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.

139. Rabl, C. (1885). Uber zellthilung. Morphol. Jahrb *10*, 214–330.

140. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J.A., Ferenc, K., Kumar, V., Lemma, R.B., Lucas, J., Chèneby, J., Baranasic, D., et al. (2024). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *52*, D174–D182. https://doi.org/10.1093/nar/gkad1059.

141. Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M.C., and Vitale, L. (2016). GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. Database *2016*, baw153. https://doi.org/10.1093/database/baw153.

142. Kalluchi, A., Harris, H.L., Reznicek, T.E., and Rowley, M.J. (2023). Considerations and caveats for analyzing chromatin compartments. Front. Mol. Biosci. *10*, 1168562.

143. Zhou, J. (2022). Sequence-based modeling of three-dimensional genome architecture from kilobase to chromosome scale. Nat. Genet. *54*, 725–734.

144. Ramani, V., Shendure, J., and Duan, Z. (2016). Understanding Spatial Genome Organization: Methods and Insights. Dev. Reprod. Biol. *14*, 7–20.

145. Sander, G., Bawden, C.S., Hynd, P.I., Nesci, A., Rogers, G., and Powell, B.C. (2000). Expression of the homeobox gene, Barx2, in wool follicle development. J. Invest. Dermatol. *115*, 753–756.

146. Sennett, R., Wang, Z., Rezza, A., Grisanti, L., Roitershtein, N., Sicchio, C., Mok, K.W., Heitman, N.J., Clavel, C., Ma'ayan, A., and Rendl, M. (2015). An Integrated Transcriptome Atlas of Embryonic Hair Follicle Progenitors, Their Niche, and the Developing Skin. Dev. Cell *34*, 577–591.

147. Rezza, A., Wang, Z., Sennett, R., Qiao, W., Wang, D., Heitman, N., Mok, K.W., Clavel, C., Yi, R., Zandstra, P., et al. (2016). Signaling Networks among Stem Cell Precursors, Transit-Amplifying Progenitors, and their Niche in Developing Hair Follicles. Cell Rep. *14*, 3001–3018.

148. Chen, W., Xiao Liu, Z., Oh, J.-E., Shin, K.-H., Kim, R.H., Jiang, M., Park, N.-H., and Kang, M.K. (2012). Grainyhead-like 2 (GRHL2) inhibits keratinocyte differentiation through epigenetic mechanism. Cell Death Dis. *3*, e450.

149. Tsuji, G., Ito, T., Chiba, T., Mitoma, C., Nakahara, T., Uchi, H., and Furue, M. (2018). The role of the OVOL1-OVOL2 axis in normal and diseased human skin. J. Dermatol. Sci. *90*, 227–231.

150. Choi, H.-R., Kim, W.K., Kim, E.Y., Han, B.S., Min, J.-K., Chi, S.-W., Park, S.G., Bae, K.-H., and Lee, S.C. (2013). Dual-specificity phosphatase 10 controls brown adipocyte differentiation by modulating the phosphorylation of p38 mitogen-activated protein kinase. PLoS One *8*, e72340.

151. Ishiguro, I., Naito, J., Saito, K., and Nagamura, Y. (1993). Skin L-tryptophan-2,3-dioxygenase and rat hair growth. FEBS Lett. *329*, 178–182.

152. Hong, G., Zhang, W., Li, H., Shen, X., and Guo, Z. (2014). Separate enrichment analysis of pathways for up- and downregulated genes. J. R. Soc. Interface *11*, 20130950.

153. Amberg, N., Sotiropoulou, P.A., Heller, G., Lichtenberger, B.M., Holcmann, M., Camurdanoglu, B., Baykuscheva-Gentscheva, T., Blanpain, C., and Sibilia, M. (2019). EGFR Controls Hair Shaft Differentiation in a p53-Independent Manner. iScience *15*, 243–256.

154. Headon, D.J., Emmal, S.A., Ferguson, B.M., Tucker, A.S., Justice, M.J., Sharpe, P.T., Zonana, J., and Overbeek, P.A. (2001). Gene defect in ectodermal dysplasia implicates a death domain adapter in development. Nature *414*, 913–916.

155. Polkoff, K.M., Gupta, N.K., Green, A.J., Murphy, Y., Chung, J., Gleason, K.L., Simpson, S.G., Walker, D.M., Collins, B., and Piedrahita, J.A. (2022). LGR5 is a conserved marker of hair follicle stem cells in multiple species and is present early and throughout follicle morphogenesis. Sci. Rep. *12*, 9104.

156. Hibino, T., and Nishiyama, T. (2004). Role of TGF-beta2 in the human hair cycle. J. Dermatol. Sci. *35*, 9–18.

157. Xu, P., Dang, Y., Wang, L., Liu, X., Ren, X., Gu, J., Liu, M., Dai, X., and Ye, X. (2016). Lgr4 is crucial for skin carcinogenesis by regulating MEK/ERK and Wnt/β-catenin signaling pathways. Cancer Lett. *383*, 161–170.

158. Romano, R.-A., Smalley, K., Liu, S., and Sinha, S. (2010). Abnormal hair follicle development and altered cell fate of follicular keratinocytes in transgenic mice expressing DeltaNp63alpha. Development *137*, 1431–1439.

159. Kretzschmar, K., Boonekamp, K.E., Bleijs, M., Asra, P., Koomen, M., Chuva de Sousa Lopes, S.M., Giovannone, B., and Clevers, H. (2021). Troy/Tnfrsf19 marks epidermal cells that govern interfollicular epidermal renewal and cornification. Stem Cell Rep. *16*, 2379–2394.

160. de Gennes, P.-G. (1979). Scaling Concepts in Polymer Physics (Cornell University Press).

161. Vivante, A., Bronshtein, I., and Garini, Y. (2020). Chromatin Viscoelasticity Measured by Local Dynamic Analysis. Biophys. J. *118*, 2258–2267.

162. Sutherland, W. (1905). LXXV. A dynamical theory of diffusion for non-electrolytes and the molecular mass of albumin. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science *9*, 781–785. https://doi.org/10.1080/14786440509463331.

163. Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann. Phys. *322*, 549–560. https://doi.org/10.1002/andp.19053220806.

164. von Smoluchowski, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Ann. Phys. *326*, 756–780. https://doi.org/10.1002/andp.19063261405.

165. Debenedetti, P.G., and Stillinger, F.H. (2001). Supercooled liquids and the glass transition. Nature *410*, 259–267.

166. Devireddy, R., and Thirumala, S. (2011). Preservation protocols for human adipose tissue-derived adult stem cells. Methods Mol. Biol. *702*, 369–394.

167. Isachenko, E., Isachenko, V., Katkov, I.I., Dessole, S., and Nawroth, F. (2003). Vitrification of mammalian spermatozoa in the absence of cryoprotectants: from past practical difficulties to present success. Reprod. Biomed. Online *6*, 191–200.

168. Hart Hansen, J.P. (1998). Bodies from cold regions. In Mummies, Disease and Ancient Cultures (Cambridge University Press), pp. 336–350.

169. de Gennes, P.G., and Leger, L. (1982). Dynamics of Entangled Polymer Chains. Annu. Rev. Phys. Chem. *33*, 49–61.

170. Zhang, B., and Wolynes, P.G. (2017). Genomic Energy Landscapes. Biophys. J. *112*, 427–433.

171. Oliveira Junior, A.B., Estrada, C.P., Aiden, E.L., Contessoto, V.G., and Onuchic, J.N. (2021). Chromosome Modeling on Downsampled Hi-C Maps Enhances the Compartmentalization Signal. J. Phys. Chem. B *125*, 8757–8767.

172. Contessoto, V.G., Dudchenko, O., Aiden, E.L., Wolynes, P.G., Onuchic, J.N., and Di Pierro, M. (2023). Interphase chromosomes of the Aedes aegypti mosquito are liquid crystalline and can sense mechanical cues. Nat. Commun. *14*, 326.

173. Hardin, C., Eastwood, M.P., Luthey-Schulten, Z., and Wolynes, P.G. (2000). Associative memory hamiltonians for structure prediction without homology: alpha-helical proteins. Proc. Natl. Acad. Sci. USA *97*, 14235–14240.

174. Oliveira, A.B., Jr., Yang, H., Whitford, P.C., and Leite, V.B.P. (2019). Distinguishing Biomolecular Pathways and Metastable States. J. Chem. Theory Comput. *15*, 6482–6490.

175. Cheng, R.R., Contessoto, V.G., Lieberman Aiden, E., Wolynes, P.G., Di Pierro, M., and Onuchic, J.N. (2020). Exploring chromosomal structural heterogeneity across multiple cell lines. Elife *9*, e60312. https://doi.org/10.7554/eLife.60312.

176. Rycroft, C. (2009). Voro++: A Three-Dimensional Voronoi Cell Library in C++ (Office of Scientific and Technical Information (OSTI)) https://doi.org/10.2172/946741.

177. Ramasubramani, V., Dice, B.D., Harper, E.S., Spellings, M.P., Anderson, J.A., and Glotzer, S.C. (2020). freud: A software suite for high throughput analysis of particle simulation data. Comput. Phys. Commun. *254*, 107275.

178. Ghesquière, P., Mineva, T., Talbi, D., Theulé, P., Noble, J.A., and Chiavassa, T. (2015). Diffusion of molecules in the bulk of a low density amorphous ice from molecular dynamics simulations. Phys. Chem. Chem. Phys. *17*, 11455–11468.

179. Hertzog, M., and Erdel, F. (2023). The Material Properties of the Cell Nucleus: A Matter of Scale. Cells *12*, 1958. https://doi.org/10.3390/cells12151958.

180. Arceri, F., Landes, F.P., Berthier, L., and Biroli, G. (2020). A Statistical Mechanics Perspective on Glasses and Aging. In Encyclopedia of Complexity and Systems Science, R.A. Meyers, ed. (Springer Berlin Heidelberg), pp. 1–68.

181. Gould, S.J. (1989). Wonderful Life: The Burgess Shale and the Nature of History (W. W. Norton & Company).

182. Gray, B. Project Pluto. https://projectpluto.com/.

183. Bardyn, A., Baklouti, D., Cottin, H., Fray, N., Briois, C., Paquette, J., Stenzel, O., Engrand, C., Fischer, H., Hornung, K., et al. (2017). Carbon-rich dust in comet 67P/Churyumov-Gerasimenko measured by COSIMA/Rosetta. Mon. Not. R. Astron. Soc. *469*, S712–S722.

184. Olson, L.E., Zhang, J., Taylor, H., Rose, D.W., and Rosenfeld, M.G. (2005). Barx2 functions through distinct corepressor classes to regulate hair follicle remodeling. Proc. Natl. Acad. Sci. U.S.A. *102*, 3708–3713. https://doi.org/10.1073/pnas.0500519102.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Mammoth skin sample, preserved in EtOH	This study, Díez-del-Molino et al.[24]	SRA: SAMN41184652, related to IN18-032; Yakinf; EBI: SAMEA14392564
Mammoth skin sample, handled "as is"	This study, Díez-del-Molino et al.[24]	SRA: SAMN41184653, related to IN18-032; Yakinf; EBI: SAMEA14392564
Mammoth skin sample used for imaging	This study, Díez-del-Molino et al.[24]	IN18-032; Yakinf; EBI: SAMEA14392564
Yuka sample	This study, Rudaya et al.[66]	SRA: SAMN41184654, related to SRA: SAMN26794180
Asian elephant skin sample, epidermis enriched	This study	SRA: SAMN35989145
Asian elephant skin sample, derma enriched	This study	SRA: SAMN35989145
Asian elephant blood sample (F)	This study	SRA: SAMN14122059
Asian elephant blood sample (M)	This study	SRA: SAMN14122065
Asian elephant brain sample	This study	SRA: SAMN35989142
Asian elephant liver sample	This study	SRA: SAMN35989143
Asian elephant ovary sample	This study	SRA: SAMN35989144
Donkey blood sample	This study	SRA: SAMN10973769
Bovine liver	This study	SRA: SAMN41403634
Bovine liver, freeze-dried	This study	SRA: SAMN41403635
Bovine heart	This study	SRA: SAMN41403636
Bovine heart jerky	This study	SRA: SAMN41403638
Bovine heart, freeze-dried	This study	SRA: SAMN41403641, SRA: SAMN41403642
Bovine cheek muscle	This study	SRA: SAMN41403633
Bovine sirloin	This study	SRA: SAMN41403637, SRA: SAMN41403639, SRA: SAMN42012258
Deposited data		
Raw and analyzed data	This study	GEO: GSE268050, BioProject: PRJNA512907
Histology archive	This study	https://doi.org/10.5281/zenodo.11193546
IN18-032 whole genome aDNA-Seq	Díez-del-Molino et al.[24]	EBI: ERS11989967
African elephant fibroblast *in situ* Hi-C data	Álvarez-González et al.[27]	SRA: SRR19650826
Yuka trunk aDNA-Seq	Maschenko et al.[83]	SRA: SRR18349585
Yuka hair aDNA-Seq	Maschenko et al.[83]	SRA: SRR18349586
Domestic donkey DNA-Seq data	Orlando et al.[6]	SRA: SRR873443, SRA: SRR873444, SRA: SRR873445
African elephant draft genome assembly Loxafr3.0	Palkopoulou et al.[28]	GenBank: GCA_000001905.1; Loxafr3.0
Asian elephant draft genome assembly ASM1433276v1	Tollis et al.[29]	GenBank: GCA_014332765.1; ASM1433276v1
Domestic horse genome assembly	Wade et al.[78]	RefSeq: GCF_000002305.2; EquCab2.0
Domestic donkey draft genome assembly	Renaud et al.[79]	GenBank: GCA_003033725.1: ASM303372v1
Human genome assembly hg38	Schneider et al.[84]	GenBank: GCA_000001405; GRCh38.p14; hg38
Human genome assembly hg19	Church et al.[85]	GenBank: GCA_000001405.14; GRCh37.p13; hg19
Human genome assembly Ash1_v2.2	Shumate et al.[86]	GenBank: GCA_011064465.2; Ash1_v2.2
Human genome assembly CHM1_1.1	Steinberg et al.[87]	GenBank: GCA_000306695.2; CHM1_1.1
Bovine genome assembly bosTau9	Rosen et al.[88]	RefSeq: GCF_002263795.3; ARS-UCD2.0
Woolly mammoth reference mitogenome	Krause et al.[89]	GenBank: NC_007596

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Asian elephant mitogenome	Rogaev et al.[90]	GenBank: NC_005129
Janthinobacterium tructae genome assembly	Jung et al.[91]	RefSeq: GCF_006517255.1; ASM651725v1
Pseudomonas veronii genome assembly	Shideler et al.[92]	RefSeq: GCF_030436055.1; ASM3043605v1
Flavobacterium petrolei genome assembly	Chaudhary et al.[93]	RefSeq: GCF_003314435.2; ASM331443v2
Loxafr3.0 RefSeq gene annotations	O'Leary et al.[94]	RefSeq: GCF_000001905.1
Human skin chromatin loop list mapped by ENCODE	https://www.encodeproject.org	https://www.encodeproject.org/files/ENCFF531FEY
GM12878 *in situ* Hi-C chromatin loop list	Rao et al.[23]	GEO: GSE63525
CTCF PWM matrix	https://jaspar.genereg.net/	JASPAR: MA0139.1
DXZ4/ICCE CTCF consensus motif	Horakova et al.[95]; Westervelt and Chadwick[48]	Horakova et al.[95]; Westervelt and Chadwick[48]
GX database build-date 2023-01-24	Astashyn et al.[96]	https://github.com/ncbi/fcs-gx/
BUSCO database build	Manni et al.[35]	Eutheria odb10.2021-02-19
GOrilla database from Mar 6, 2021	Eden et al.[57]	https://cbl-gorilla.cs.technion.ac.il/
Compilation of previously published elephantid mitogenomes	van der Valk et al.[11]	van der Valk et al.[11]
Nucleosome-resolution chromatin structures	Harris et al.[82]	Harris et al.[82]
Experimental models: Cell lines		
Primary fibroblast cell line LAF-12	Mary Thompson, Stephen O'Brien	LAF-12
Software and algorithms		
3D-CARTA	This study	https://github.com/aidenlab/3d-carta; https://doi.org/10.5281/zenodo.11193689
Simple diffusion modeling and visualization	This study	https://doi.org/10.5281/zenodo.11242201
Juicer and Juicer Tools including HiCCUPS and APA	Rao et al.[23]; Durand et al.[77]	https://github.com/aidenlab/juicer
BWA	Li and Durbin[98]; Li[99]	https://github.com/lh3/bwa
3D-DNA	Dudchenko et al.[22]	https://github.com/aidenlab/3d-dna/tree/phasing
Juicebox	Durand et al.[97]	https://github.com/aidenlab/Juicebox/wiki/Download
Juicebox Assembly Tools	Dudchenko et al.[33]	https://github.com/aidenlab/Juicebox/wiki/Download
Juicebox.js	Robinson et al.[100]	https://igv.org/doc/juiceboxjs.html
PMDTools	Skoglund et al.[30]	https://github.com/pontussk/PMDtools
SAMtools	Danecek et al.[101]	https://github.com/samtools/
SIP	Rowley et al.[102]	https://github.com/PouletAxel/SIP
SIPMeta	Rowley et al.[102]	https://github.com/JRowleyLab/SIPMeta
GOrilla	Eden et al.[57]	https://cbl-gorilla.cs.technion.ac.il/
CRUSH	https://github.com/JRowleyLab/CRUSH	https://github.com/JRowleyLab/CRUSH
OpenMiChroM	Oliveira et al.[65]	https://github.com/junioreif/OpenMiChroM
DNA degradation simulator	https://github.com/DiPierroLab/Bernardo/tree/master/DNA_degradation_src	https://github.com/DiPierroLab/Bernardo/tree/master/DNA_degradation_src
Spacewalk	https://github.com/igvteam/spacewalk/	https://github.com/igvteam/spacewalk/
MIA	Green et al.[3]	https://github.com/mpieva/mapping-iterative-assembler

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Muscle v3.8.31	Edgar[103]	https://drive5.com/muscle/downloads_v3.htm
SeaView v5.0.5	Gouy et al.[104]	https://doua.prabi.fr/software/seaview
BEAST v1.10.4	Suchard et al.[105]	https://github.com/beast-dev/beast-mcmc
Tracer v1.7.2	Rambaut et al.[106]	https://github.com/beast-dev/tracer/releases
GX (gx-version: build:Jan 19 2023 15:50:20; git:v0.3.0–151-g9aad15db)	Astashyn et al.[96]	https://github.com/ncbi/fcs-gx/
VCFtools	Danecek et al.[107]	https://vcftools.sourceforge.net/
picard v2.15.0	Picard Toolkit 2019	http://broadinstitute.github.io/picard/
ANGSD v0.940	Korneliussen et al.[108]	https://github.com/ANGSD/angsd/releases
GMASS	Kwon et al.[109]	https://github.com/jkimlab/GMASS
LASTZ	Harris[80]	https://github.com/lastz/lastz
BBMap (version 38.79)	Bushnell[98]	https://bioweb.pasteur.fr/packages/pack@bbmap@38.79
TOGA	Kirilenko et al.[34]	https://github.com/hillerlab/TOGA
MAVR (v0.1)	https://github.com/mahajrod/MAVR	https://github.com/mahajrod/MAVR
RepeatMasker, version open-4.0.7	Tempel[110]	https://www.repeatmasker.org/
Tandem Repeat Finder trf, version 4.02	Benson[134]	https://github.com/Benson-Genomics-Lab/TRF
Windowmasker 1.0.0	Morgulis et al.[111]	https://github.com/goeckslab/WindowMasker
SegAlign	Goenka et al.[112]	https://github.com/gsneha26/SegAlign
RepeatFiller	Osipova et al.[113]	https://github.com/hillerlab/GenomeAlignmentTools
chainCleaner	Suarez et al.[114]	https://github.com/hillerlab/GenomeAlignmentTools
kentUtil	Kent et al.[115]	https://github.com/ucscGenomeBrowser/kent/
BUSCO v5.4.6	Manni et al.[35]	https://busco.ezlab.org/
FIMO	Grant et al.[116]	https://meme-suite.org/meme/doc/fimo.html
bedtools	Quinlan and Hall[112]	https://bedtools.readthedocs.io/
POSSUMM	Harris et al.[82]	https://github.com/aidenlab/EigenVector
GNU Parallel	Tange[117]	https://www.gnu.org/software/parallel/
ChatGPT	OpenAI	https://chat.openai.com
Wolfram Mathematica	Wolfram Research, Inc.	https://www.wolfram.com/mathematica/
Other		
Interactive Figure 2C	This study	https://t.3dg.io/3d-mammoth-Fig-2C
Interactive Figure 3	This study	https://t.3dg.io/3d-mammoth-Fig-3
Interactive Figure 4A	This study	https://t.3dg.io/3d-mammoth-Fig-4A
Interactive Figure 4B	This study	https://t.3dg.io/3d-mammoth-Fig-4B
Interactive Figure 5	This study	https://t.3dg.io/3d-mammoth-Fig-5
Interactive Figure 7A	This study	https://t.3dg.io/3d-mammoth-Fig-7A
Interactive Figure S1D	This study	https://t.3dg.io/3d-mammoth-Fig-S1D
Interactive Figure S2AB	This study	https://t.3dg.io/3d-mammoth-Fig-S2AB
Interactive Figure S7A	This study	https://t.3dg.io/3d-mammoth-Fig-S7A
Interactive Figure S7B	This study	https://t.3dg.io/3d-mammoth-Fig-S7B
Interactive Figure S8B	This study	https://t.3dg.io/3d-mammoth-Fig-S8B
Interactive Figure S8E	This study	https://t.3dg.io/3d-mammoth-Fig-S8E
Interactive African elephant contact map with SIP loop annotations	This study	https://t.3dg.io/3d-mammoth-SIP
Interactive Data S1 Figure M1	This study	https://t.3dg.io/3d-mammoth-Fig-M1
Interactive Data S1 Figure M3	This study	https://t.3dg.io/3d-mammoth-Fig-M3

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Interactive Data S1 Figure M5	This study	https://t.3dg.io/3d-mammoth-Fig-M5
Interactive Data S1 Figure M6	This study	https://t.3dg.io/3d-mammoth-Fig-M6
Interactive Data S1 Figure M7	This study	https://t.3dg.io/3d-mammoth-Fig-M7
Interactive Data S1 Figure M8	This study	https://t.3dg.io/3d-mammoth-Fig-M8
Interactive Data S1 Figure M9	This study	https://t.3dg.io/3d-mammoth-Fig-M9
Interactive Data S1 Figure M10	This study	https://t.3dg.io/3d-mammoth-Fig-M10
African elephant interactive genome assembly page	This study	https://www.dnazoo.org/assemblies/Loxodonta_africana
Asian elephant interactive genome assembly page	This study	https://www.dnazoo.org/assemblies/Elephas_maximus
Domestic donkey interactive genome assembly page	This study	https://www.dnazoo.org/assemblies/Equus_asinus
Woolly mammoth interactive genome assembly page	This study	https://www.dnazoo.org/assemblies/Mammuthus_primigenius

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Erez Lieberman Aiden (erez@erez.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- PaleoHi-C and *in situ* Hi-C data reported in this paper has been deposited at GEO and SRA (bioproject PRJNA512907). Accession numbers are listed in the key resources table. Original histology images have been deposited at Zenodo and are publicly available as of the date of publication. The DOI is listed in the key resources table.
- All original code is available on GitHub (https://github.com/aidenlab/3d-carta) and has been deposited to Zenodo. The code is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Woolly mammoth sample IN18-032

The specimen (ID IN18-032) was collected in September 2018 near Belaya Gora, Sakha Republic, Siberia (N68.57887, E147.16055) during an international expedition to Tirekhtyakh, Indigirka tributary. It comprised a fragment of a skin from a woolly mammoth head, with hair still attached and an intact external ear (see Figure S1A).

Two fragments from the rim of the skin sample were collected for research purposes several hours after the sample was extracted from permafrost. The first sample was placed into 95% ethanol after removing the inner and outer surfaces to reduce exterior contaminants. This fragment was primarily used for PaleoHi-C experiments as well as for generating the shotgun Illumina data (aDNA-Seq) published in Díez-del-Molino et al.[24] A second larger fragment collected from the same location (see Figure S1A) was not placed in ethanol, and instead transported at ambient temperature until it could be stored frozen for longer term storage. This sample was primarily intended for radiocarbon dating. Leftovers from radiocarbon dating analysis have been used for additional PaleoHi-C experiments. A closeup photo on Figure S1A shows the second (no-ethanol) fragment.

Yuka woolly mammoth sample

We also applied PaleoHi-C to a second woolly mammoth sample, a juvenile female woolly mammoth nicknamed 'Yuka'. The mammoth, estimated to have died about 39,000 years ago, was found in August 2010 in the Oyogos-Yar area that belongs to the coastline of the Laptev Sea in the Arctic Yakutia (N72°40′49.44″, E142°50′38.35″) by Vasiliy Gorokhov, the head of the nearby Yukagir village. More details about its discovery can be found in Rudaya et al.[66] A fragment of tissue for PaleoHi-C was cut from a fold of the trunk 14 cm below the tusk alveolus on the right.

Asian elephant samples

We generated Hi-C data for five different Asian elephant tissues: peripheral blood mononuclear cells (PBMCs), liver, ovary, brain and skin.

PBMCs were obtained from two different Asian elephant individuals, one female and one male. Ear vein blood samples for both preparations were opportunistically collected during routine veterinary procedures at the Houston Zoo. The skin, liver, ovary and brain tissue samples were collected during a necropsy of a single female Asian elephant individual (*Elephas maximus indicus*) performed in 2004 at the San Antonio Zoo. The samples were obtained under Baylor College of Medicine protocol AN-6832.

Donkey sample

The donkey Hi-C data for the prototype reference-assisted 3D genome assembly was generated from peripheral blood mononuclear cells that were extracted from a blood sample purchased from Animal Technologies, Inc. (https://www.animaltechnologies.com/).

Bovine samples

The proof-of-concept Hi-C experiments to assess how drying affects the capacity to conserve 3D genome architecture, even at elevated temperatures, were done using beef samples.

A fresh liver sample was obtained from the Rosenthal Meat Science and Technology Center at Texas A&M University. Vacuum-sealed heart and cheek as well as the plastic-wrapped sirloin steak samples were purchased at a local grocery store. Freeze-dried liver (Freeze Dried Beef Liver, single ingredient dog treats from Natural Rapport, moisture content: 8.5%), freeze-dried heart (small-batch Pets Freeze-Dried Beef Heart Treats for Dogs and Cats, moisture content: 7%; Stella & Chewy's Freeze-Dried Raw Beef Heart, moisture content: 8%), and heart jerky (Beef Heart Jerky, single ingredient dog treats from Natural Rapport, moisture content: 12%) were purchased from an online retailer.

Other samples

The primary fibroblast cell line LAF-12 used in nuclear staining experiments was obtained from Mary Thompson and Stephen O'Brien (NIH).

METHOD DETAILS

PaleoHi-C protocol

Sample preparation, crosslinking and aDNA-chromatin extraction

Mammoth tissue was divided into 80 mg subsamples, which were processed separately. The procedure for each sample was as follows. The sample was first coarsely crushed using a scalpel. Then, 500 μL of 1% formaldehyde was added as crosslinking solution, and the sample was left to incubate at room temperature (RT) for 10 min in slow rotation. 40 μL of 2.5 M glycine solution was added to the reaction, vortexed and incubated for 5 min at RT. The sample was centrifuged for 2 min at 6000 x g and the supernatant was discarded into appropriate chemical waste. 500 μL of Wash Buffer 1 (10 mM Tris-HCl pH 8, 50 mM NaCl and 0.01% Tween) was added and mixed to wash. Sample was centrifuged for 2 min at 6000 x g and the supernatant was discarded. This wash was repeated for a total of two washes.

After washing the crosslinked tissue, the sample was ground using a plastic pestle. While keeping the sample in a cold rack, 550 μL of freshly prepared Lysis buffer (10 mM HEPES pH 8, 10 mM NaCl, 0.2% IGEPAL 630, 1X Protease inhibitors) were added and the reaction was vortexed thoroughly. The reaction was left to incubate in a cold rack inside the fridge at 4°C for 30 min. The reaction was vortexed again and homogenized with a plastic pestle with 30 up-and-down strokes inside the tube. Samples were spun at 2500 x g for 5 min at RT. (The supernatant from this step was collected to prepare additional datasets, see section description of IN18-032 PaleoHi-C datasets.) Then 500 μL of Wash Buffer 2 (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA) were added without disturbing the tissue pellet. The sample was spun at 2500 x g for 5 min at RT to re-collapse the tissue pellet between washes, and the supernatant was discarded. The sample was washed a second time with 500 μL of Wash Buffer 2, spun at 2500 x g for 5 min at RT, and the supernatant was discarded. The sample was resuspended in 250 μL of Wash Buffer 2 and 100 μL of 2% SDS were added. The reaction was vortexed to mix everything thoroughly and left to incubate with shaking for 10 min at 50°C. After incubation, the supernatant containing crosslinked aDNA-chromatin was collected and an estimate of DNA recovered was made using a Qubit instrument with dsDNA HS assay.

Bead binding, chromatin digestion, end repair and biotin incorporation, end ligation

Carboxylated beads (2X) were added to the supernatant collected previously. The reaction was mixed by pipetting 10 times up and down. (Different beads such as SPRI, Ampure XP and MagBio beads were used to bind aDNA-chromatin in solution from different tissue replicates. No significant correlation between library quality and bead type was found; see Table S1, #3.)

The reaction was incubated for 5 min at RT. The mix was placed on a magnet for 5 min, and the supernatant was discarded. The bead-bound samples were then washed with 500 μL of Wash Buffer 1. Again, the supernatant was discarded, and the beads were washed a second time with 250 μL of Wash Buffer 1.

After removing the supernatant from the last wash, 70 μL of restriction enzyme digestion mix containing 5U of DpnII was added to the beads and the reaction was left incubating at 1000 rpm for 1 h at 37°C. The sample was then placed on a magnet for 5 min, the supernatant was discarded, and the beads were washed twice with 500 μL of Wash Buffer 1.

Next, 50 μL of a reaction mix containing 10X NEBuffer 2, 10 mM each of dATP, dTTP and dGTP, 0.4 mM Biotin-14-dCTP and 5 U/μL Klenow Polymerase Large Fragment was added to each sample and the reaction was left to incubate at 1000 rpm for 30 min at 25°C. Afterward the reaction was placed on a magnet, the supernatant was discarded, and the bead-bound sample was washed twice with Wash Buffer 1. Finally, 500 μL of ligation reaction mix containing 10X T4 DNA ligase buffer, 20 mg/mL BSA, 10% Triton X-100 and T4 DNA Ligase 400 U/μL was added to each sample, and reactions were left to incubate at 500 rpm for 2 h at RT.

Nucleotide exchange, crosslink reversal, and DNA purification

While keeping the tubes on a thermoblock, 2.5 μL of 10 mM dNTP mix and 2.5 μL of T4 DNA polymerase were added to each sample, and the reactions were left incubating further at 1000 rpm for 15 min at 16°C. Samples were then placed on a magnet, supernatant was discarded, and samples were washed once with 250 μL Wash Buffer 1. Next, a crosslink reversal reaction to separate DNA from proteins and degrade the leftover proteins was carried out by adding 100 μL of CRMix containing 50 mM Tris-HCl pH 8, 1% SDS, 0.25 mM $CaCl_2$ and 1 mg/mL Proteinase K. The reactions were left to incubate at 1000 rpm for 15 min at 55°C followed by 45 min at 68°C. After incubation, samples were placed on a magnet, and the supernatant was transferred to new 1.5 mL Eppendorf tubes. 200 μL of carboxylated beads (SPRI, Ampure XP or MagBio) were added and mixed by pipetting. After incubating for 5 min at RT, samples were placed on a magnet, the supernatant was discarded, and the bead-bound samples were washed with 250 μL of freshly prepared 80% EtOH while still on a magnet. EtOH was removed and samples were washed again for a total of 2 washes. After the second wash, samples were left to air-dry for 5 min. 58 μL of EB Buffer were added, and samples were incubated for 5 min at 37°C.

Library preparation, cleanup and amplification

Purified DNA was prepared for sequencing on two platforms, Illumina and DNBSEQ (BGI). 27 μL for each library were aliquoted and libraries were prepared following the BEST protocol.[118] After the fill-in step, samples were cleaned using C1 Streptavidin beads. 60 μL of C1 beads per sample were washed and resuspended in 130 μL of 2X NTB Buffer (10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA). The mix was then added to each sample and the reaction was left to incubate at RT for 30 min with slow movement (300 rpm). After incubation, samples were placed on a magnet, supernatant was discarded and the beads were washed twice with 200 μL of Buffer NWB (10 mM Tris-HCl, 1 M NaCl, 1 mM EDTA, 0.05% Tween 20), followed by two washes with 200 μL of TWB buffer (10 mM Tris-HCl, 0.5 mM EDTA, 0.05% Tween 20). While on a magnet, supernatant was removed and beads were re-suspended in 94 μL of Amplification Master Mix (1X Kapa U+ HotStart ReadyMix, 10 mg/mL BSA). The mix was split into two 47 μL reactions that were each supplemented with 0.3 μM of each Fwd. and Rev. indexes to a final volume of 50 μL per reaction. All experiments up to this step were performed at the dedicated aDNA clean laboratory at the GLOBE Institute (University of Copenhagen, Denmark). The replicates were then taken to be PCR-amplified for 16–20 cycles. The reactions were then cleaned using 1.2X SPRI beads and eluted into a final volume of 32 μL with EB Buffer. Following quantification and visualization of amplified PaleoHi-C libraries on an Agilent TapeStation instrument, samples were submitted to BGI Europe for sequencing on either DNBSEQ-G400 or Illumina NovaSeq 6000 (see description of IN18-032 PaleoHi-C datasets section).

Description of IN18-032 PaleoHi-C datasets

Overall, 26 PaleoHi-C datasets were generated for IN18-032: six primary libraries from the ethanol-preserved sample and 20 supplementary datasets (see Table S1, #3).

The primary datasets were generated as follows: the ethanol-preserved tissue was split into two biological replicates (~80mg each). The PaleoHi-C protocol was carried out on these as described in the PaleoHi-C protocol section until the step when crosslinks were reversed and the DNA purified. The resulting material was split in two, with one-half prepared for downstream sequencing on the Illumina platform, resulting in two libraries (one library from each of the tissue replicates), and the other half for the BGI sequencing platform. The material being prepared for BGI sequencing was split again before the amplification step and processed as two separate PCR reactions, resulting in four libraries, two for each biological tissue replicate. The resulting datasets are designated as ''EtOH precipitate'' in Table S1, #3.

Supernatant material from the tissue lysis step (see PaleoHi-C protocol section) was collected and taken through the steps downstream of lysis, as described in the protocol section, to convert into a set of supplementary libraries. Similar to the primary set, the material was split after DNA purification to be prepared for the Illumina and BGI sequencing platforms. When preparing the BGI libraries, DNA was split in two for PCR amplification. This resulted in six ''EtOH supernatant'' supplementary datasets (Table S1, #3).

We additionally explored the viability of the PaleoHi-C method on leftovers of the second skin sample that had been collected for 14C carbon dating in the field, and which thus was not preserved in ethanol. Four replicate subsamples were taken from this sample, and the PaleoHi-C protocol was carried out as described above, starting with the crosslinking step. Purified DNA was prepared for sequencing on the BGI sequencing platform. All reactions but one were split before the amplification step and processed as two separate libraries. This procedure resulted in seven supplementary datasets (designated ''No EtOH precipitate'' in Table S1, #3).

Finally, post tissue-lysis supernatant material from the no-ethanol sample experiments was converted into ''No EtOH supernatant'' supplementary libraries; see Table S1, #3. Just as with the corresponding precipitate datasets, purified DNA was prepared for sequencing on the BGI sequencing platform. All reactions but one were split before the amplification step and processed as two separate libraries.

Yuka PaleoHi-C

The Yuka PaleoHi-C experiment was performed at the Animal Cytogenetics Laboratory at IMCB as described in the PaleoHi-C protocol section with the following modifications due to reagent availability. First, Kzo9I, a MboI isoschizomer, was used in place of MboI. Second, the *Nucleotide exchange* step was skipped. Finally, the library preparation was done using the Illumina TruSeq Nano kit with Taq amplification (20 cycles) rather than following the BEST protocol. The resulting library was sequenced on a NextSeq 550 Illumina instrument (see Table S4, #3).

Ancient DNA-Seq without USER treatment

An extensive aDNA-Seq dataset for sample IN18-032 was published in Díez-del-Molino et al.[24] We supplemented the published sequencing reads with a small additional dataset. Just as in the case of the data published in Díez-del-Molino et al.,[24] DNA was extracted from a piece of the ethanol preserved tissue sample at the ancient DNA lab at the Swedish Museum of Natural History, Stockholm, Sweden. The tissue was digested overnight with the lysis buffer described in Sinding et al.[119] Next, the lysate was concentrated and purified as described in the supplementary protocol of Dehasque et al.[120] A double-stranded sequencing library was prepared following the protocol of Meyer et al.[121] with one exception. Specifically, we omitted the USER treatment during the blunt-end repair step which is typically used in aDNA-Seq to excise uracil bases from the DNA fragments, as described in Pečnerová et al.[122] The resulting library was sequenced on an Illumina NovaSeq 4000 instrument (see Table S1, #5).

Modern sample Hi-C library preparation and sequencing

The *in situ* Hi-C libraries for the PBMCs isolated from the donkey and Asian elephant blood samples were prepared as described in Rao et al.,[23] and the MboI restriction enzyme was used for chromatin digestion. The libraries were sequenced on the Illumina sequencing platform (Table S1, #1; Table S2, #2). Deep sequencing of the male and female Asian elephant PBMC *in situ* Hi-C libraries for detailed analysis of Xi superdomains was done on the Ultima Genomics platform (Table S3, #11).

Nuclei isolated from solid Asian elephant tissues were processed as described in Rao et al.[23] using a combination of Csp6I and MseI restriction enzymes for chromatin digestion. For the skin, two Hi-C libraries were generated. In addition to the first library (dubbed "Necropsy skin sample") for which a skin nuclei extract was processed with a combination of Csp6I and MseI following Rao et al.,[23] an additional second library, dubbed "Necropsy skin sample, dermis" and targeting the more interior skin layers, was prepared with MboI. This latter library did not start with a nuclear extract. Instead, the sample was dounced and the resulting chromatin bodies were bound to the beads following the PaleoHi-C protocol up to *Library preparation*. The final steps were completed as described in Rao et al.[23] The following components from NEBNext Ultra II DNA Library Prep Kit for Illumina were used for end repair and adaptor ligation for the dermis library: NEB E7646AA, E7647AA, E7648AA, E7374AA. All libraries were sequenced on the Illumina sequencing platform (Table S1, #1).

Bovine *in situ* Hi-C libraries were prepared as described above for the "Necropsy skin sample, dermis" and sequenced on Illumina NextSeq 550 instrument (Table S4, #1).

Histological examination of sample IN18-032

Small pieces of mammoth skin, measuring 0.5 × 1 cm, were soaked in PBS (Phosphate Buffered Saline, pH 7.2–7.4) for 24 h. Samples were then fixed in formalin (10% pH-neutral solution in PBS) for 12 h. The fixed samples were washed in PBS and distilled water (3×, 2 h/wash). The fixed tissues were dehydrated by incubating them in ethanol solutions of increasing concentrations and then cleared with xylene, after which tissues were embedded in paraffin using standard histological techniques. The paraffin blocks were cut to 3.5–7 μm-thick sections using a MICROM HM 355S microtome. The sections were mounted on SuperFrost glasses.

Sections were then stained with Giemsa and Van Geison's stains and examined using ZEISS Scope.A1 microscope. Separately, staining with DAPI was performed and stained samples were examined using a ZEISS Axios Scope.A1 as well as an Olympus FV 3000 laser scanning confocal microscope, see Figures S1B and S1C.

Mammoth nuclei isolation and staining

Ethidium bromide (EB) staining

Mammoth skin tissue (approximately 100 mg) was soaked in buffer A (0.35 M sucrose, 0.05 M Tris-HCl pH 7.5, 0.066 M EDTA, 3 mM $CaCl_2$, 0.025 M KCl) and minced with forceps and scissors in a culture dish. The minced tissue was transferred into a Dounce homogenizer together with 2 mL of buffer A supplemented with 1% Triton X-100 and homogenized with ~25 passes of the loose pestle and ~25 passes of the tight pestle. The resulting solution was further mixed with a narrow metal spatula. The homogenate was transferred onto two layers of nylon filter (pore size ~100 μm). The filtrate was centrifuged at 600 g at 4°C for 5 min, and the supernatant was removed. The pellet was washed in another 2 mL of Buffer A, and centrifuged at 400 g at 4°C for 5 min to remove the supernatant. Finally the pellet was resuspended in 100 μL of Buffer A. A total of 8 μL of the suspension were smeared across the slide and dried at room temperature. The slide was stained with EB (manufacturer: Sigma-Aldrich) at a final concentration of 0.01 μg/μL (1 μL of the stock EB solution (0.5 mg/mL) in 50 μL DABCO), and photographed immediately after staining (see Figure S2G).

In parallel, we stained a nuclear preparation from a primary fibroblast cell line LAF-12. The cell line, established from a skin biopsy of an African elephant at the National Zoological Park in Washington, D.C., was cultured at 37°C in an incubator supplied with 5% CO_2 in α-MEM with 15% FBS, 100 IU/mL penicillin and 100 μg/mL streptomycin. We pelleted cells from a single 90%-confluent T150

flask, resuspended them in 100 μL buffer A and moved into a Dounce homogenizer. After ~20 passes of each pestle the resulting homogenate was filtered, cleaned and stained as described above for the mammoth tissue homogenate.

Propidium iodide (PI) staining

A variant of the experiment described above was done using crosslinked mammoth tissue as input. Specifically, ~100 mg of mammoth skin tissue were soaked in buffer A and minced with forceps and scissors in a culture dish. Tissue bits were crosslinked in 1% formaldehyde solution in phosphate-buffered saline (PBS) for 10 min at room temperature. The crosslinking solution was removed via centrifugation (600 g, 4°C, 5 min), and the tissue was washed twice in 2 mL of ice-cold 0.2 M glycine to quench the formaldehyde. The sample was centrifuged at 600 g at 4°C for 5 min between the washes. Crosslinked tissue was homogenized and filtered as described in the ethidium bromide staining section. The final pellet was resuspended in 100 μL of Hi-C lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 0.2% IGEPAL). 8 μL of the solution were stained with Propidium Iodide at a final concentration of 0.07 μg/μL in DABCO (20 mg/mL, 100 mM Tris HCl pH 7.5 in glycerol, Sigma). See Figure S2G.

QUANTIFICATION AND STATISTICAL ANALYSIS

Dating of IN18-032

Radiocarbon dating of IN18-032

The sample has been dated using radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU), which yielded an infinite age estimate (>44,900 years, OxA-38763).[24]

Phylogenetic dating of IN18-032

To obtain a more precise estimate of the sample's age, we used a Bayesian molecular dating method based on the mitochondrial genome as previously published.[11] First, using publicly available aDNA-Seq data from Díez-del-Molino et al.,[24] the mitogenome was assembled for IN18-032 from merged and trimmed FASTQ files containing DNA fragments over 35 base pairs using mapping-iterative-assembler (MIA)[3] with the Asian elephant mitogenome (GenBank: NC_005129[90]) as a reference. Given the vast amount of whole-genome sequencing data, only one library (accession: ERR10173648) was used for this analysis. This resulted in a mitogenome coverage of 800×. Positions with less than 3× coverage or with a sequence agreement of less than 67% were assigned as missing data.

Next, the assembled mitogenome was aligned to previously published elephantid mitogenomes compiled by van der Valk et al., 2021[11] with Muscle v3.8.31.[103] The obtained multiple sequence alignment (containing a total of 172 mammoths, nine African savanna elephants, two Asian elephants and four American mastodons) was visually inspected using SeaView v5.0.5.[104] Considering the assembly and alignment limitations on the hypervariable control region, the VNTR region as identified in the woolly mammoth reference mitogenome (GenBank: NC_007596,[89] positions 16157 to 16476) was removed from the alignment. Additionally, sequences supported by only one sample in the control region (and assigned as gaps in the remaining samples) were also removed.

Next, we performed a Bayesian phylogenetic analysis using BEAST v1.10.4.[105] In order to avoid over-parameterising the model for tip dating and to reduce the uncertainty of the age estimate (Chacón-Duque et al., in prep), we removed all samples without finite calibrated radiocarbon ages from the alignment, in order to obtain a tip date estimate only for the sample of interest. This filtering removed all mastodons and left a total of 121 radiocarbon-dated mammoths. Following van der Valk et al., 2021,[11] the alignment was split into six partitions: tRNA, rRNA, first, second, and third codon positions, and the control region. All were assigned the HKY + Gamma + Invariant substitution model, except for the tRNA partition, for which the HKY + Invariant model was used instead.[11] Also, the same log-normal prior for the divergence between *Loxodonta* and *Elephas/Mammuthus* of 5.3 Mya (million years ago) was used with a strict molecular clock and a flexible skygrid coalescent model.[123] Two Markov Chain Monte Carlo (MCMC) chains using a uniform tip prior (date: 75 kya [thousand years ago], range: 10 kya - 2 Mya) were run for 100,000,000 iterations, sampling each 10,000 and discarding the first 10% as burn-in. We checked convergence with Tracer v1.7.2[106] and combined the chains with LogCombiner v1.10.4.[105] This resulted in a final age estimate of 52.3 ka (thousand years) (95% HPD: 14.4–86.9 ka), or, as we describe it in the main text, roughly 52,000 years.

Sequencing and *de novo* assembly of modern elephants

Hi-C data analysis and de novo chromosome-length genome assembly of the Asian elephant

The *in situ* Hi-C libraries for various Asian elephant samples were analyzed using the Juicer pipeline[77] using ASM1433276v1 (GenBank: GCA_014332765.1), a draft Asian elephant genome reference from Tollis et al.[29] as reference. Juicer version 1.6 with BWA 0.7.17-r1188 was run with the "-s none" flag. Quality metrics for Asian elephant datasets, as generated by Juicer, can be found in Table S1, #1.

Deduplicated alignments from HIC1382, HIC12818, HIC16296, HIC16297 and HIC16298 were used as input for the 3D-DNA ("3D *de novo* assembly") pipeline[22] and Juicebox Assembly Tools (JBAT)[33,97] to perform misjoin correction, anchor, order and orient sequences in ASM1433276v1 and produce a chromosome-length reference ASM1433276v1_HiC.

The chromosome-length sequences were ordered, oriented and labeled (chr1 through chr27 and chrX) as described in the Elephantidae chromosome assignment section. Contact maps built with respect to the original draft and the chromosome-length reference are shown in Figure S1D (top).

Hi-C data analysis and de novo chromosome-length assembly of the African elephant

To generate a chromosome-length genome assembly for an African savanna elephant (hereafter referred to as "African elephant") we used our previously published Hi-C dataset, SRA: SRR19650826.[27] The data was generated from two fibroblast cell lines, one derived from a male, and another one from a female individual. For the purposes of this study the two datasets were processed separately. The data was aligned to Loxafr3.0[28] and deduplicated using Juicer.[77] The alignment and deduplication statistics are included in Table S1, #2.

Only the LAF1789 dataset was used for genome assembly. The corresponding merged_nodups.txt file was used as input to the 3D-DNA pipeline[22] to produce a candidate genome assembly. The assembly was reviewed and polished using JBAT.[33]

The chromosome-length sequences were assigned chromosome labels (chr1 through chr27 and chrX) based on karyotyping data as described in the Elephantidae chromosome assignment section. Contact maps generated with respect to the draft Loxafr3.0 genome assembly and the resulting chromosome-length Loxafr3.0_HiC are shown in Figure S1D (bottom).

IN18-032 PaleoHi-C preliminary analysis

PaleoHi-C sequencing data alignment and metagenomic analysis

Detailed statistics associated with all 26 datasets, as generated by the Juicer Hi-C processing pipeline[77] can be found in Table S1, #3 (Juicer version 1.6; BWA 0.7.17-r1188; 8 threads; openjdk version "1.8.0_222-ea"; Juicer Tools Version 1.9.9). Data was aligned to the African elephant genome reference Loxafr3.0_HiC, and the pipeline was run with "-s none" flag to account for non-enzymatic fragmentation of the sample. The distribution of the percent of total alignable read pairs in the primary and supplementary datasets is shown in Figure S1E (top, BGI data).

In order to better characterize the sequences present in the sample we used GX (gx-version: build:Jan 19 2023 15:50:20; git:v0.3.0–151-g9aad15db), a new cross-species aligner, to align 10 million paired-end reads from each of the 22 BGI libraries representing all six biological replicates to a curated reference database comprising 47,754 taxa (database build-date 2023-01-24). GX is part of the NCBI Foreign Contamination Screen (FCS) tool suite.[96] The results are shown in Figure S1E (bottom).

Comparison with ancient DNA-Seq without USER treatment

The non-USER aDNA-Seq data was aligned to Loxafr3.0 and Loxafr3.0_HiC using Juicer version 1.6/BWA 0.7.17-r1188.[77] Alignment statistics can be found in Table S1, #5.

PaleoHi-C read pairs aligning to elephant derive from ancient DNA

DNA damage is well documented in ancient and degraded DNA samples, and is most frequently evident in DNA sequence data in the form of cytosine (C) deamination,[124–126] which are then converted to uracil (U) and its analogues, thus revealed as cytosine to thymine (T) transitions in the resulting sequence data.

To find the proportion of PaleoHi-C reads bearing this damage pattern we used PMDtools,[30] https://github.com/pontussk/PMDtools. The software assigns each read a score (PMDscore) for which positive values indicate postmortem damage, with the higher scores associated with higher degrees of damage. PMDtools was run with default parameters to compute the percentage of damaged reads in generated datasets (reads with PMDscore >3). The analysis was done using PaleoHi-C data alignments to Loxafr3.0.[28]

In addition to the PaleoHi-C dataset we analyzed one of the *in situ* Hi-C libraries for the Asian elephant (a skin necropsy sample) and the African elephant *in situ* Hi-C dataset from Álvarez-González et al.,[27] SRR19650826. We also computed the percentage of reads with damage in a USER-treated vs. non-USER-treated shotgun whole-genome sequencing (aDNA-Seq) experiment for the same mammoth individual. The USER-treated data was published in Díez-del-Molino et al.[24]

The mammoth-derived PaleoHi-C reads have an elevated degradation signature, on par with the non-USER treated aDNA-Seq data (Figure S1F (top); Table S1, #6). The damage is higher than in any of the modern elephant datasets. The result remained true when analysis was run specifically for intrachromosomal long-range 3D contacts (i.e., using only reads that represent interactions between loci occurring on the same Loxafr3.0 scaffold and separated by 1D distance greater than 20 kb).

PaleoHi-C read pairs aligning to elephant derive from mammoth DNA

To ascertain the authenticity of our mammoth sample, we ran a diagnostic test using a set of SNPs to decisively rule out the possibility of contamination from modern, closely related species of *Elephantidae* in our mammoth PaleoHi-C sample.

For this, we first obtained mammoth diagnostic positions from a SNP panel from a recent study[24] of 29 elephants (7 Asian elephants and 21 African elephants) and 23 woolly mammoths using VCFtools.[107] We consider a site to be diagnostic when all elephants have the reference allele, and the mammoths always show the alternative allele, and requiring that at least 4 mammoths and 14 elephants have data for the position. In total, we used 1,034,282 diagnostic positions.

We then proceeded to analyze the alignments at the diagnostic positions in the PaleoHi-C dataset. In addition to the PaleoHi-C dataset we analyzed one of the *in situ* Hi-C libraries generated for the Asian elephant (skin necropsy sample) and the African elephant *in situ* Hi-C dataset from Álvarez-González et al.,[27] SRR19650826.

Together with our mammoth Hi-C data, all datasets were mapped to the African elephant genome assembly Loxafr3.0[28] in a single-end mode to recover as many reads as possible. We applied bwa-0.7.17 mem for mapping the African and Asian elephant Hi-C libraries and bwa aln with parameters "-l 16500 -n 0.01 -o 2" for mapping the mammoth Illumina and BGI PaleoHi-C.[98,99] We removed duplicates for each library with picard v2.15.0 (http://broadinstitute.github.io/picard/) MarkDuplicates. BAM files from the same sample were merged with picard v2.15.0 MergeSamFiles. We filtered for primary alignments using samtools view -F 260.[101]

To account for variable coverage depth between different datasets, we performed haploid calling by randomly sampling one base at each diagnostic position using ANGSD v0.940[108] with the following parameters: ''-doHaploCall 1 -doCounts -rf positions_mammoth_fixed_angsd.bed -uniqueOnly 1 -remove_bads 1 -minMapQ 30''. We then counted the total covered sites per sample and the positions showing the mammoth or the elephant allele.

We found that reads from mammoth libraries - when they could be assigned - were overwhelmingly assigned to mammoth (99.8% of the time) and almost never assigned to elephant (0.2%), whereas reads from elephant libraries were overwhelmingly assigned to elephant (99.9%) and almost never assigned to mammoth (0.1%). The same result can be obtained using only a subset of reads that represent long-range intrachromosomal 3D contacts for the mammoth and modern elephants, see Figure S1F (bottom) and Table S1, #7. The few cases of discordant assignment could be explained by sequencing errors. We conclude that the signal in PaleoHi-C experiments derives from ancient mammoth DNA, rather than from modern contaminants.

PaleoHi-C read separation probability as a function of genomic distance

In order to analyze the decay of read separation probability with genomic distance in (i) aDNA-Seq, (ii) PaleoHi-C experiments, and (iii) *in situ* Hi-C data generated using a modern Asian elephant skin sample, we aligned all datasets to Loxafr3.0_HiC using Juicer (version "1.6"[77]). The pipeline was run with ''-s none'' flag for all datasets and, in the case of aDNA-Seq data, with an additional ''-j'' flag (stands for ''just exact duplicates excluded at the deduplication step'') supplied for handling very large low complexity datasets.

Only high-quality alignments (mapping quality greater than or equal to 30 for both sequences in the read pair) to the largest scaffold in Loxafr3.0 (GL010027.1) were extracted for the contact probability analysis. Logarithmic binning was used. Values have been normalized by the total number of contacts within GL010027.1 in each corresponding experiment.

All other things equal, it would carry greater significance to compare the PaleoHi-C data (not treated with the USER cocktail) to the non-USER-treated aDNA-Seq library for the same sample since exposure to the USER enzyme cocktail has the capacity to reduce the DNA fragment length.[127] In practice, the final size distribution depends on other experimental factors such as, e.g., DNA quality variation, amplification conditions and cleanup procedure details. In view of this, we examined the size distribution of the treated and untreated aDNA-Seq datasets and chose the USER-treated dataset from Díez-del-Molino et al.[24] for the most conservative comparison with PaleoHi-C 1D separation probability distribution (see Figure 1E; Figure S1G). Only a subset (ERR10173641) of a full dataset was used for Figure 1E. The full dataset is included in Figure S1G (top).

The Illumina-sequenced PaleoHi-C dataset has been used when comparing to the Illumina shotgun aDNA-Seq data in Figure 1E to avoid possible bias associated with a particular sequencing platform. Figure S1G (top) includes a version of the same plot with PaleoHi-C data curve generated from data collated from across the BGI and Illumina sequencing platforms.

Figure S1G (middle) compares the mammoth skin PaleoHi-C dataset with the modern Asian elephant skin sample across a wider range of distances compared to Figure 1E. The log-log read separation probability curves are almost identical, up to a vertical shift.

We include an additional supplementary figure (Figure S1G, bottom) in order to demonstrate the robustness of these results with respect to the exact input material used for the PaleoHi-C protocol (see additional details in the sample preparation step in the PaleoHi-C protocol section and in the description of IN18-032 PaleoHi-C datasets). Figure S1G (bottom) presents a breakdown of contact probability curves across the four dataset groups listed in description of IN18-032 PaleoHi-C datasets (EtOH precipitate, EtOH supernatant, no EtOH precipitate and no EtOH supernatant). All datasets show the characteristic contact probability decline, with approximately the same slope. Setting aside the EtOH-p dataset, one can see elevated statistical noise in the other curves on account of relatively little data generated from the corresponding sample handling procedure variants. The analysis of the protocol variants is done using BGI sequencing data.

It is worth noting that while the contact probability curves for the mammoth and the Asian elephant skin samples demonstrate strong consistency in terms of the slope (Figure S1G, middle), there is still an overall downward shift in terms of the fraction of long-range contacts in the mammoth vs. elephant data at the same genomic separation. (For this comparison, we analyzed the same number of reads aligning to GL010027.1 from both species.) This effect is due to the fact that we have been much less aggressive in removing short sequences from the PaleoHi-C library when compared to standard experimental procedures for a modern *in situ* library on account of small amounts of input material. Short sequences are often enriched for non-restricted/non-ligated reads (i.e., DNA-Seq reads): the non-restricted/non-ligated read fraction in PaleoHi-C can be readily seen in Figure S1G (top). The more aggressive cleanup step in the modern Hi-C library preparation led to the removal of more such reads from the modern elephant library, resulting in a higher long-range fraction per total sequenced read count. Filtering out all reads with genomic separation under 1 kb removes some of the associated bias, allowing for a fairer comparison of the quality of the 3D signal between the modern elephant and the ancient woolly mammoth datasets (Table S1, #8).

Another way to ensure that aDNA-Seq component does not skew the comparison between the paleo and modern datasets is to filter Hi-C data based on contact type. Following Juicer classification,[23,77] all alignable Hi-C read pairs can be classified into one of four types based on their alignment to the reference. In a ''left'' pair, both ends map to the reverse strand. In a ''right'' pair, both reads map to the forward strand. In an ''inner'' pair, the ends map to different strands and point (5′ to 3′) toward each other. In an ''outer'' pair, reads land on opposite strands but point away from one another. Importantly, read pairs in the ''right'' and ''left'' configurations are rarely seen outside of Hi-C experiments. By contrast, ''inner'' pairs are overwhelmingly expected in a DNA-Seq experiment; circularization of non-chimeric DNA fragments will sometimes produce ''outer'' pairs during library construction. Thus, excluding the

"inner" and "outer" pairs from the analysis will (at the expense of losing some real Hi-C contacts) reliably filter out the non-Hi-C component, and allow for comparison of the contacts without the aDNA-Seq bias. This is the approach we take when plotting the contact probability curves *p*(*s*) in Figure 6C of the main text.

It is important for those seeking to use PaleoHi-C in their own laboratories to mention that alignment noise is a factor that can influence the appearance of read separation probability curves, especially when analyzing low-coverage data. For example, a DNA-Seq read pair from a repeat locus could be mistaken for a long-range contact if the two reads from a read pair get erroneously aligned to two different copies of the repeat sequence. The variation of the associated read separation probability signal with distance reflects the distribution of loci with high sequence identity in a particular genome. It is typically different from the dependency observed during successful Hi-C experiments. Thus, the presence of putative long-range contacts does not on its own establish that a PaleoHi-C experiment was successful. Rather, it is the distance distribution of such contacts, along with the presence of other known features of 3D genome architecture, that should be used to assess the success or failure of a PaleoHi-C experiment.

Proof-of-principle scaffolding using PaleoHi-C data

Before attempting to assemble a chromosome-length genome of a woolly mammoth, we ran a proof-of-concept test. In this test we: 1) scaffolded the African elephant draft Loxafr3.0 from Palkopoulou et al.[28] using PaleoHi-C data; and 2) scaffolded Loxafr3.0 using 5 million African elephant *in situ* Hi-C read pairs. The latter dataset is comparable to PaleoHi-C in terms of putative long-range contacts (e.g., once aligned to Loxafr3.0_HiC, a total of 1,771,672 read pairs are classified as intrachromosomal long-range (>20 kb), comparable to 1,763,225 PaleoHi-C contacts in the same category, see Figure 1F). The scaffolding procedure followed the one described in the genome assembly of modern elephants section. We then compared the resulting assemblies using GMASS.[109] GMASS measures structural similarity between two assemblies by calculating a score, measuring from 0 (no similarity) to 1 (identical), based on the length and number of syntenic genomic regions as identified by alignment of the corresponding sequences.

First, we calculated the GMASS similarity score between the subsampled African elephant assembly (Loxafr3.0_sub_HiC) and Loxafr3.0_HiC. The similarity score was based on comparing assemblies across the following resolutions: 100 kb, 250 kb, 500 kb, 1 Mb, 2.5 Mb, 5 Mb, 10 Mb, 25 Mb, 50 Mb, 100 Mb. The alignment strictness parameter was set to "-s near", and the following parameters were added to LASTZ calls within GMASS to allow for alignment of longer sequences with ambiguous characters: "--allocate:traceback = 1.99G --ambiguous = iupac". The comparison yielded a GMASS score of 0.996 (compare this to GMASS score of 0.813 when running the same comparison for Loxafr3.0 and Loxafr3.0_HiC). A high similarity score suggests that high quality genome assembly is possible with low amounts of contact data, which is essential if we are to successfully assemble genomes using PaleoHi-C.

We then compared Loxafr3.0, Loxafr3.0_sub_HiC and Loxafr3.0 scaffolded using PaleoHi-C data (Loxafr3.0_mock_HiC). We expected one of the following. If all ancient contacts were effaced by subsequent movement, then PaleoHi-C essentially comprises random ligations. Random contacts would have no capacity to order and orient sequences in Loxafr3.0, and the resulting assembly, when compared to Loxafr3.0_sub_HiC, would have the same GMASS score as the one between Loxafr3.0 and Loxaf3.0_sub_HiC. By contrast, if PaleoHi-C reflects ancient chromatin conformation, then, just like modern Hi-C data, the data will have some capacity to order and orient sequences, and the resulting assembly will be more similar to Loxaf3.0_sub_HiC than Loxafr3.0. (This assumes that a typical synteny block between the two species is larger than the typical fragment length in Loxafr3.0.)

The results of the comparison are summarized in Table S2, #1. A high GMASS score of 0.983 between Loxafr3.0_mock_HiC and Loxafr3.0_sub_HiC is consistent with the conclusion that ancient chromosome conformation persists in the woolly mammoth sample, and that PaleoHi-C is able to assay the resulting contacts.

Reference-assisted 3D genome assembly of the woolly mammoth

Design of a reference-assisted 3D genome assembler

Here, we introduce the 3D Comparative Assembler using Reference To Assist (3D-CARTA), a program for performing reference-assisted 3D genome assembly. 3D-CARTA can generate a chromosome-length genome assembly fasta from a set of reads, DNA-Seq and/or Hi-C, from an organism by mapping them to an assembly (fragmentary or chromosome-length) of a closely related organism. Similarly to other tools of this type (see, e.g., AMOScmp[128] that has served as a primary inspiration for the tool) 3D-CARTA attempts to substitute the traditional overlap-layout-consensus approach to assembly with alignment-layout-consensus. As part of the procedure, the read alignments to the assisting reference (provided in the form of *.sam* files, typically produced as part of the Juicer workflow) are examined in a base-by-base fashion to detect small-scale differences (single nucleotide polymorphisms, as well as deletions and insertions that are isolated and small as compared to the read length) between the two species. The identified differences are then incorporated into the assisting reference to create a consensus approximation of the organism represented by the read set.

Rearrangements between the two genomes pose the most difficult challenge to comparative assembly using the alignment-layout-consensus strategy. For example, the analysis of read alignment signatures associated with putative rearrangements constitutes the most complex part of the AMOScmp pipeline.[128] Unlike other comparative assembly tools, 3D-CARTA does not attempt to extract the information necessary to resolve rearrangements from the alignment data. Instead, it relies on Hi-C for this purpose. In particular, it relies on the 3D-DNA pipeline, whose results it takes as input in the form of a *.assembly* file.

The pipeline is available on GitHub (https://github.com/aidenlab/3d-carta) and includes the following steps.

(1) In the preliminary step, the workspace is organized and the assisting reference is preprocessed to create a reference *.map* file. A *.map* file is a text file format in which, similar to *.mpileup*, each line represents a single genomic position.
(2) During the next step (-S sort) the SAM file(s) are merge-sorted into a master sorted SAM file. The master SAM is then split into manageable chunks. The split size is adjusted dynamically to result in a predefined number of files determined by jobcount parameter (unprompted).
(3) In the mapping step (-S map) a *.map* file is constructed from split alignment data to record bases observed at each genomic position. Indels and read break positions with respect to the reference are also recorded.
(4) The results from individual chunks are merged during the merge step (-S merge). Reference map data is also added during this step. (Including the assisting reference when generating the consensus enhances the accuracy of the assembly in regions of low coverage, or regions covered only by a single read, helping distinguish the true base from contaminants and sequencing errors.) Optionally, read clips that originate from restriction sequence-based cutting in the Hi-C protocol can be filtered during this step.
(5) The next step (-S reduce) generates a preliminary consensus from the map.
(6) The pipeline includes an optional step to reexamine indel and clip features that were not incorporated during preliminary data consolidation and reassemble the corresponding regions (-S reassemble).
(7) In the last step (-S finalize) the rearrangement file (provided via -a|--assembly flag) is incorporated and the final FASTA is constructed that reflects both the local consensus and the rearrangement data.

The 3D-CARTA pipeline has a number of limitations, the most obvious of which is its dependence on the assembly of a closely related species. Even with a closely related species some highly divergent regions (arguably the most interesting regions!) might not be assembled due to issues with read alignment. Also, the current pipeline relies on majority voting when generating consensus, which has limited ability to handle indels and does not address several important evolutionary events such as large insertions in the target (long segments of DNA present in the target genome that do not have a counterpart in the assisting reference genome) or segmental duplication events. Note that the latter appear as coverage anomalies at the corresponding sequences. (In the elephantids, for example, the affected sequences include the following genes: *Rnasel*, *Rgs16* and *Lipa*.) Often, these end up in the unanchored portion of the assembly on account of them being identified as problematic by the 3D-DNA rearrangement detection workflow.

It is also worth noting the following. The pipeline examines the differences between the assisting reference and the data in a 'bifocal' fashion. Specifically, 3D-CARTA is well equipped to search for small polymorphisms (those shorter than a read). At the same time, the larger the rearrangement, the easier it is to detect using Hi-C. Thus, the most readily resolved rearrangements are those at opposite extremes of the size spectrum. (Note that, for the purposes of detection by Hi-C, the scale associated with a translocation-type rearrangement is a function of the length of the fragment being translocated and the 1D distance between the original and the translocated position. Small sequences that are translocated further away along the chromosome may be easier to detect than larger sequences that remain close to their original position, since the signal associated with the latter will be obscured by all sequences in the 1D neighborhood interacting in 3D.) Previously, we showed that 3D-DNA can use Hi-C at ~10× coverage to reliably identify contact frequency differences between sequences separated by over ~10 kb.[22] While additional coverage can be used to push the limit of detection to smaller scales, some rearrangements corresponding to the scales between ~100 bp and ~10 kb may remain unresolved when using the experimental and computational methods described in the present manuscript.

Note that the comparative assembly strategy described above is somewhat different from another powerful assisted assembly approach when contigs in a species of interest are scaffolded according to synteny blocks reconstructed by analyzing the genomes of multiple related species (see, e.g., Kim et al., Kolmogorov et al., and Alonge et al.[129–131]). 3D-CARTA essentially assumes perfect synteny with the assisting reference at the outset, an assumption that is corrected based on Hi-C data from the target species. The Hi-C data is thus used to identify the synteny breaks *de novo* rather than to infer them via multiple genome comparison. Importantly, 3D-CARTA also has the capacity to scaffold sequences beyond synteny blocks, and typically yields chromosome-length scaffolds. Of course, when Hi-C data from a species of interest is not available, synteny block reconstruction can serve as a compelling alternative.

Reference-assisted 3D genome assembly of the domestic donkey

The donkey *in situ* Hi-C sequencing data from two replicate libraries was aligned to the EquCab2.0 (RefSeq: GCF_000002305.2[78]) reference genome assembly and deduplicated using Juicer version 2.0/BWA 0.7.17-r1188.[77] The pipeline was run with "-s MboI" flag. The resulting statistics for both libraries as generated by the Juicer pipeline are listed in Table S2, #2.

The resulting alignments were used as input to the 3D-DNA pipeline.[22] The pipeline was run with default parameters to identify chromosomal breaks, fusions and intrachromosomal rearrangements that separate the donkey and the horse and edit the horse reference to reflect these events. The resulting candidate assembly representing a tentative donkey karyotype was examined and polished using Juicebox Assembly Tools.[33] The Hi-C data aligned to the original EquCab2.0 reference and the "donkified" EquCab2.0 are shown in Figure S2A.

The core steps of the 3D-DNA pipeline, as applied in reference-assisted 3D genome assembly, are: 1) misjoin detection; and 2) scaffolding (anchoring, ordering, and orientation) of the resulting reference fragments. When performing an assisted assembly from a relatively accurate reference such as EquCab2.0, step 1 primarily detects chromosome breaks that took place over the course of evolution and which, just as with misjoins due to an assembly error in the assisting reference, appear as a disagreement between the Hi-C data and the underlying reference. Specifically, this disagreement manifests in the form of a near-diagonal depletion of Hi-C

proximity signal not typically observed in contact maps of truly contiguous fragments. This is because sequences that are nearby in 1D tend to form contacts in 3D, whereas sequences that are not, do not. Similarly, step 2 primarily detects chromosomal fusion events that took place over the course of evolution and that appear as increased Hi-C contacts between different fragments of the assisting reference genome assembly. In the case of a less accurate reference genome assembly, these same steps detect misjoins and scaffold draft sequences that are not associated with particular evolutionary events (chromosome breaks and fusions), just in the same way that they can function in the context of *de novo* 3D genome assembly workflows.

The output of the 3D-DNA/Juicebox Assembly Tools pipeline is a *.assembly* file that encodes the large-scale differences between the EquCab2.0 reference (2*n* = 64) and the donkey (2*n* = 62). The file is passed as input to 3D-CARTA, a custom comparative assembly pipeline that examines per-base read alignment statistics to edit the reference at base-pair resolution while at the same time incorporating larger rearrangements as annotated by the assembly file.

The 3D-CARTA pipeline has the capacity to examine DNA-Seq and Hi-C alignment data (see the design of a reference-assisted 3D genome assembler section). In order to simplify the comparison with the *de novo* donkey genome assembly (see *de novo* chromosome-length genome assembly of the domestic donkey) we used reads from the same donkey individual, Willy (*Equus asinus asinus*), for both the *de novo* and assisted genome assemblies. Specifically, the short-read DNA-Seq data from Orlando et al.[6] (SRA: SRR873443, SRR873444, SRR873445) aligned to EquCab2.0 (RefSeq: GCF_000002305.2) were used as input to 3D-CARTA along with the *.assembly* file. The DNA-Seq reads were aligned using BWA-MEM (v 0.7.17-r1198-dirty[98,99]) and deduplicated using SAMtools.[101] The resulting EquCab2.0 sequence coverage was 15.6×. The alignment and deduplication statistics are included in Table S2, #3.

De novo *chromosome-length genome assembly of the domestic donkey*

For *de novo* donkey genome assembly we combined the donkey PBMC *in situ* Hi-C data with the draft genome assembly ASM303372v1 (GenBank: GCA_003033725.1) from Renaud et al.[79] The ASM303372v1 draft genome assembly was generated using Chicago HiRise (a technology that relies on high molecular weight DNA) providing scaffolds of subchromosomal size (scaffold N50 = 15 Mb). The two Hi-C datasets were aligned to the ASM303372v1 reference using Juicer (version 1.5.6, BWA 0.7.15-r1140[77]) with '-s MboI' flag. The alignment statistics generated by the Juicer pipeline can be found in Table S2, #4. The resulting merged_nodups.txt file was passed as input to 3D-DNA[22] which was run with default parameters to perform misjoin correction of the draft sequences as well as anchor them to chromosomes, as well as to order and orient them. The resulting candidate chromosome-length genome assembly was further examined using Juicebox Assembly Tools.[33] Hi-C data aligned to the ASM303372v1 draft and the final chromosome-length genome assembly (ASM303372v1_HiC) is shown in Figure S2B. For convenience of comparison with the assisted assembly, the final chromosome-length scaffolds are ordered and oriented to match the order and orientation in the assisted donkey genome assembly shown in Figure S2A.

Comparison of the donkey genome assemblies

Basic assembly statistics for the two donkey assemblies as well as the EquCab2.0 reference and ASM303372v1 can be found in Table S2, #5. The stats were generated using the BBMap (version 38.79) stats.sh script with default parameters.[132]

We further compared the original horse assembly, the 3D reference-assisted donkey assembly, and the *de novo* donkey assembly by doing a whole-genome alignment using LASTZ.[80] The code was run with "--masking = 3 --notransition --step = 20 --nogapped --format = maf --ambiguous = iupac --hspthresh = 50000" command options, and EquCab2.0 was used as a target and ASM303372v1 was used as query sequence. The resulting alignments were lifted over to ASM303372v1_HiC and to EquAsi_EquCab2.0_assisted_HiC to create the dotplots shown in Figure S2C.

We also calculated pairwise GMASS scores[109] between the assisted donkey and the *de novo* donkey references as well as between the two versions of the donkey and the assisting horse reference (see Table S2, #6). GMASS was run as described in the proof-of-principal scaffolding using PaleoHi-C data section. As one would expect from viewing the dotplots, the GMASS similarity score between the assisted and the *de novo* donkey references is very high (0.995), much higher than, e.g., between the donkey reference and the assisting horse genome assembly (0.863). The comparison supports the notion that reference-assisted 3D genome assembly is accurate and not overly biased by the assisting genome assembly.

Reference-assisted 3D genome assembly of the woolly mammoth

The reference-assisted 3D genome assembly pipeline was run as described in the previous section for the donkey. Specifically, the Hi-C data was aligned to the Loxafr3.0 reference using Juicer (version 1.6; BWA 0.7.17-r1188[77]), see Table S2, #7. The resulting deduplicated alignments were analyzed using 3D-DNA[22] and Juicebox Assembly Tools[33] to produce a *.assembly* file. The file was used as input to 3D-CARTA, along with SAMtools deduplicated BWA-MEM alignments of adapter-trimmed aDNA-Seq reads from the same individual published in Díez-del-Molino et al.[24] The assisted assembly procedure resulted in 28 chromosome-length scaffolds (Figure 2C). The chromosome-length sequences were ordered, oriented, and labeled (chr1 through chr27 and chrX) to match the ordering and orientation of the African elephant genome assembly (see Hi-C data analysis and *de novo* chromosome-length assembly of the African elephant and conservation of synteny in the family Elephantidae sections).

Basic assembly statistics for MamPri_ MamPri_Loxafr3.0_assisted_HiC can be found in Table S2, #8. Also listed are the statistics for Loxafr3.0, Loxafr3.0_HiC, ASM1433276v1, ASM1433276v1_HiC and MamPri_ASM1433276v1_assisted_HiC (see reference-assisted 3D genome assembly of the woolly mammoth is robust to the choice of assisting assembly). The stats were generated using BBMap (version 38.79) stats.sh script with default parameters.[132]

Conservation of synteny in the family Elephantidae

We compared the three elephantid genome assemblies by doing a whole-genome alignment using LASTZ.[80] The code was run with "--masking = 3 --notransition --step = 20 --nogapped --format = maf --ambiguous = iupac --hspthresh = 50000" command options, and Loxafr3.0 was used as target and ASM1433276v1 was used as query. The resulting alignments were lifted over to ASM1433276v1_HiC, Loxafr3.0_HiC, and MamPri_Loxafr3.0_assisted_HiC to create dotplots shown in Figure S2D. The dotplot suggests high conservation of synteny in elephantids, consistent with the published karyotyping data.[76]

Elephantidae chromosome assignment

The karyotypes of African and Asian elephants were determined in Houck et al.[76] In this study, autosomes were enumerated in order of apparent (i.e., cytogenetic) size, with acrocentric/telocentric chromosomes placed first, and were oriented so that the centromere is closer to the chromosome start. The sex chromosome was labeled by comparing male and female samples. To minimize confusion when comparing the results of this study to prior studies in this clade, we sought to enumerate, orient, and label the chromosomes for all three Elephantidae species consistently with Houck et al.[76]

This is not entirely straightforward, since Houck et al.[76] presents exclusively cytogenetic data, and does not contain sequence information.

To address this issue, we used the comparative genome maps of human and the African elephant published in Frönicke et al.[81] These maps, generated via Zoo-FISH, outline the regions of homology between GTG banded elephant karyotype presented by Houck et al.[76] and human chromosomes. For comparison, we identified the regions of homology between Loxafr3.0_HiC, our chromosome-length genome assembly, and human reference hg38[84] using SegAlign,[133] a scalable GPU system for pairwise whole genome alignments based on LASTZ's seed-filter-extend paradigm. Prior to alignment Loxafr3.0_HiC was repeat-masked using MAVR (v0.1), a wrapper script to deploy three independent repeat annotation tools: RepeatMasker, version open-4.0.7,[110] Tandem Repeats Finder trf, version 4.04,[134] and Windowmasker 1.0.0.[111] SegAlign was run with the following parameters: "--wga_chunk_size = 2500000 --lastz_interval_size = 4000000 --seq_block_size = 40000000 --hspthresh = 2400 --gappedthresh = 3000 --ydrop = 9400".

The results are shown in Figure S2E (left). The alignment results were fully consistent with Zoo-FISH data allowing us to assign labels (chr1 through chr27 and chrX) to all 28 Loxafr3.0_HiC chromosome-length scaffolds. Of these 28 chromosomes, 14, syntenic to more than one human chromosome, were oriented using comparative data from Frönicke et al.[81] The metacentric X chromosome was oriented to match human chrX orientation. The remaining 13 chromosomes were oriented based on the interaction analysis of Loxafr3.0_HiC repeat sequences. The chromosome-length scaffolds in the Asian elephant and the woolly mammoth were ordered, oriented, and labeled to match those of the African elephant.

Chromosome assignment was done after the main analyses were completed for the paper. The deposited data files and figures were then updated throughout the manuscript to match chromosome assignments as described above. We note that the deposited CRUSH tracks for certain chromosomes (e.g., 5) have bins that start and stop at numbers which are not round (e.g., 77,444) on account of chromosome orientation changes to the reference genome after the analysis was performed.

(It is perhaps worth noting that another human/elephant comparative genome map is available in Yang et al.,[135] also based on reciprocal chromosome painting. This latter map is identical to that published in Frönicke et al.[81] with the exception of two pairs of chromosomes: chromosome labeled as #5 in Yang et al.[135] is labeled as #6 in Frönicke et al.[81] and, vice versa, chromosome labeled as #6 in Yang et al.[135] is labeled as #5 in Frönicke et al.[81] Similarly chromosome #19 and #22 are switched between the two studies. We chose Frönicke et al.[81] as a basis for our assignment since the latter study had a higher number of resolved conserved chromosomal segments.)

Woolly mammoth gene annotation

We used TOGA, a homology-based annotation pipeline that takes as input a gene annotation of a well-annotated reference species such as human and a whole-genome alignment between the reference and a query genome.[34] From the alignment data, TOGA infers orthologous gene loci in the query genome, annotates them, and classifies them.

Prior to running TOGA, we masked repeats in the mammoth FASTA using RepeatMasker, version open-4.0.7,[110] Tandem Repeats Finder trf, version 4.04,[134] and Windowmasker 1.0.0,[111] all deployed using MAVR (v0.1) (https://github.com/mahajrod/MAVR). We aligned the masked assisted mammoth assembly to the human reference hg38[84] using SegAlign.[133] The alignment was run with the following parameters: "--wga_chunk_size = 2500000 --lastz_interval_size = 4000000 --seq_block_size = 40000000 --hspthresh = 2400 --gappedthresh = 3000 --ydrop = 9400". We then proceeded with creating alignment chains including identifying and incorporating repeat-overlapping alignments with (custom parallelized using GNU Parallel[117]) RepeatFiller[113] and removing chain-breaking alignments with chainCleaner.[114] The steps followed the recommendations outlined in https://github.com/hillerlab/GenomeAlignmentTools and https://github.com/ucscGenomeBrowser/kent/.[115]

The resulting chains were used as input to run TOGA (v1.1) with a "--mask_stops" flag using human annotations, U12 introns and isoform data shared at https://github.com/hillerlab/TOGA/tree/master/TOGAInput/human_hg38. We examined the resulting protein annotations using BUSCO v5.4.6,[35] which was run with a "-m proteins" flag and "-l eutheria_odb10" (Eutheria odb10.2021-02-19).

We ran the same annotation pipeline for the African elephant chromosome-length genome assembly Loxafr3.0_HiC. The resulting annotations agreed well with the RefSeq gene annotations for Loxafr3.0 lifted to the chromosome-length genome assembly (data not shown). The comparison of completeness of the resulting gene annotations, as a percentage of 11,366 eutherian genes from BUSCO, is included in Table S2, #9.

Reference-assisted 3D genome assembly of the woolly mammoth is robust to the choice of assisting assembly

To demonstrate robustness of the assembly procedure with respect to the assisting reference we generated another woolly mammoth chromosome-length genome assembly, MamPri_ASM1433276v1_assisted_HiC, this time using ASM1433276v1, the Asian elephant draft genome assembly from Tollis et al.[29] as an assisting reference.

The assisted pipeline was run as described above. Specifically, PaleoHi-C data was aligned to the ASM1433276v1 reference using Juicer (version 1.6; BWA 0.7.17-r1188[77]); see Table S2, #10. The resulting deduplicated alignments were analyzed using 3D-DNA[22] and Juicebox Assembly Tools[33] to produce a *.assembly* file. The file was used as input to 3D-CARTA, along with SAMtools deduplicated BWA-MEM alignments of aDNA-Seq data from the same individual published in Díez-del-Molino et al.[24] (The reads were subject to an adapter removal step prior to alignment.) The assisted assembly procedure resulted in 28 chromosome-length scaffolds (Data S1 Figure M1 and Table S2, #8). The chromosome-length scaffolds in the final assembly were ordered, oriented, and labeled to match the chromosome-length scaffolds in the Elephantidae assemblies described above.

The two mammoth genome assemblies, MamPri_Loxafr3.0_assisted_HiC and MamPri_ASM1433276v1_assisted_HiC, are highly concordant (Figure S2E, right panel), despite the fact that the respective assisting references, Loxafr3.0[28] and ASM1433276v1,[29] not only represent two different elephantid species but also differ significantly with respect to their underlying sequencing platform, data models, assembly statistics, etc.

We further quantified the similarity between the two assisted assemblies and their relationship to their respective assisting references by calculating six pairwise GMASS scores[109] between the following sequences: Loxafr3.0, MamPri_Loxafr3.0_assisted_HiC, ASM1433276v1 and MamPri_ASM1433276v1_assisted_HiC (see Table S2, #11). The scores were calculated as described in the proof-of-principal scaffolding using PaleoHi-C section. For reference, we used the same parameters to calculate GMASS scores across several independent versions of the human genome: hg38,[84] hg19,[85] CHM1_1.1[87] and Ash1_v2.2,[86]; see Table S2, #12. The highest GMASS score in the elephantid reference set (0.984) was observed when comparing the two versions of the mammoth reference, confirming that the chromosome-length assemblies are structurally similar across all examined resolutions. The associated score is comparable to that observed for different human references (average: 0.935).

In the section 3D architecture results are robust with respect to the choice of assisting reference, we use MamPri_ASM1433276v1_assisted_HiC alongside MamPri_Loxafr3.0_assisted_HiC to reproduce the main findings of the study, namely, the preservation of cell-type-specific segregation between open and closed chromatin, point-to-point chromatin loops, and the tetradic structure of mammoth chromosome X.

3D genome architecture persists in IN18-032

Mammoth DNA is spatially segregated from contaminant DNA and at least partially encapsulated in the original nuclei

We explored the organization of the mammoth DNA in the context of the sample as a whole. Taken together with histological observations (Figures 1C, S1B, and S1C), the data are consistent with a model where the mammoth DNA in IN18-032 is spatially segregated from the contaminant DNA, and at least some of it remains in the original nuclei.

Metagenomic PaleoHi-C maps. We constructed an extended reference comprising the woolly mammoth fasta, MamPri_Loxafr3.0_assisted_HiC, as well as three reference genomes corresponding to the most abundant contaminants we detected with FCS, one for each of the three top scoring divisions (see metagenomic analysis of PaleoHi-C sequencing data section). Specifically, we used ASM651725v1 (*Janthinobacterium tructae*, from Jung et al.[91]) for the "prok:b-proteobacteria" division; ASM3043605v1 (*Pseudomonas veronii*, from Shideler et al.[92]) for "prok:g-proteobacteria"; and ASM331443v2 (*Flavobacterium petrolei*, from Chaudhary et al.[93]) for "prok:CFB group bacteria". We then used Juicer (version 1.6; BWA 0.7.17-r1188) to realign BGI-sequenced PaleoHi-C data to this extended reference. The corresponding alignment and deduplication statistics are listed in Table S2, #13.

We used the resulting alignments to build an extended PaleoHi-C map. In order to better visualize the contact pattern between the smaller contaminant chromosomes and the mammoth sequences we show a chromosome-resolution contact map, i.e., a contact map in which each bin is a chromosome, with the three bins at the top corresponding to three contaminant references, and the 28 bins that follow corresponding to the mammoth chromosomes. We included only long-range intrachromosomal data (read separation $\geq$20 kb) and normalized all counts by the respective probabilities of observing a contact in question assuming a uniform contact distribution (i.e., the total count between chromosomes a and b, $a \neq b$, was divided by the product of the corresponding chromosome lengths $L_a \times L_b$).

The resulting contact map is presented in Figure S2F. We note that nearly all chromatin contacts were between loci in the same species, and that inter-species contacts were extremely rare. This is consistent with the notion that the mammoth DNA occupies distinct territories within the sample, spatially segregated from the DNA of other species. The result is in agreement with the nucleus-like features observed in histological preparations (see Figures 1C and S1B). Refer also to Figure S1C for histological staining of contaminants in the mammoth sample.

Isolation of putative nuclei. We wanted to check if it might be possible to recover the nuclei from the sample. First, we used a nucleus isolation protocol from Trifonov et al.[136] originally designed for high molecular weight DNA extraction. We stained the extract with ethidium bromide, a DNA intercalator, alongside a nuclear preparation from a modern elephant cell line as a control (see method details). The mammoth and elephant preparations looked similar, with clusters of brightly stained nucleus-like structures of similar size and shape (Figure S2G, top two rows).

We also performed a nuclei isolation experiment from crosslinked tissue similar to the method described in Nohara et al.[137] We stained the resulting preparation using propidium iodide (PI) as described in the method details section. Although the efficacy of

the recovery was low, the resulting preparations revealed structures resembling mammalian nuclei (Figure S2G, bottom row). The findings are in line with prior studies that were able to extract nucleus-like structures from woolly mammoth samples.[26,138]

Woolly mammoth exhibits nuclear architecture marked by strong chromosome territories

Eukaryotic genomes tend to fold into one of two architectural types,[37] shown in Figure S2H. Type I exhibits features reminiscent of Rabl chromosome configuration[139] in which centromeres cluster, often accompanied by telomere clustering and chromosomes folding along a telomere-to-centromere axis. Type II lacks the above features and is instead marked by prominent chromosome territories, when a chromosome occupies a discrete subvolume of the nucleus, excluding other chromosomes.

The PaleoHi-C mammoth data indicates a territorial, or Type II, architecture, commonly observed in mammals.[27,37]

Compartment analysis at 1-Mb resolution

Benefiting from the high degree of synteny conservation in the Elephantidae and in order to take advantage of the ''gold standard'' RefSeq annotations available for the African elephant (see apparent differences in gene activity in mammoth skin vs. Asian elephant skin section) we performed the compartment analysis in the main text with respect to the Loxafr3.0_HiC, the chromosome-length African elephant genome assembly. Compartment analysis is repeated with respect to the genome assemblies MamPri_Loxafr3.0_assisted_HiC and MamPri_ASM1433276v1_assisted_HiC in the section 3D architecture results are robust with respect to the choice of assisting reference.

The compartmentalization analysis at 1-Mb resolution was largely performed as previously described.[17] We observed the characteristic plaid pattern, with the resultant bright squares along the diagonal, consistent with contact domains that form due to compartmentalization. The first order Pearson correlation analysis sharpens the signal, producing the characteristic plaid pattern that suggests that each chromosome can be decomposed into two sets of loci (traditionally labeled as A and B) such that contacts within each set are enriched, and contacts between sets are depleted. The signal can be sharpened even further using second order Pearson autocorrelation analysis (Figure 3A).

The entries for the first autocorrelation matrix p_{ij} were computed from KR-normalized observed/expected values (c_{ij}) by taking the Pearson correlation coefficient of the ith row (c_{ix}) and the jth column (c_{xj}). Note that we skip two entries in each vector c_{ii}, c_{ij}, and c_{ji}, c_{jj} in order to avoid entries associated with the diagonal of the contact matrix in either vector. The KR-normalized observed/expected values for the purpose were extracted from the corresponding *.hic* files using Juicer Tools.[77] For second order autocorrelation matrices, Pearson correlation coefficients between the two vectors representing all first-order Pearson correlation coefficients for loci *i* (p_{ix}) and *j* (p_{xj}) are calculated in the same fashion.

For the mammoth data, the first principal component of the KR-normalized contact matrix clearly corresponded to the plaid pattern for all but one chromosome (positive values defining one set, negative values the other) and was used to partition each chromosome into A and B. For chr8, the second principal component corresponded to the plaid pattern. (This type of behavior for the principal components is well-known.[17]) For elephant data the first principal component corresponded to the plaid pattern for all chromosomes.

We examined the distribution of values in the eigenvectors calculated for the mammoth and elephant skin contact matrices at 1-Mb resolution (Data S1 Figure M2). While the elephant histogram reflected a typical bimodal distribution, the mammoth histogram did not show a binary A/B association reflecting, presumably, reduced confidence in assigning the relevant bins due to low data coverage in the mammoth. In view of this, we imposed a threshold of 0.05 as the minimal absolute value for the assignment in the mammoth dataset to be considered reliable for 1-Mb resolution analysis.

As such, the cross-tissue comparison described in the text was done as follows. We first filtered out all bins for which the eigenvector value in the mammoth data was either positive and <0.05 or negative and >-0.05. We then applied the sign function to the remaining values and calculated corresponding Pearson coefficients. Chromosome 8 was excluded from the analysis. Note that including the second principal component values for chr8 does not change the overall result. When doing so, the Pearson *r* values are as follows: elephant skin vs. mammoth skin: 0.920; vs. elephant ovary: 0.849; vs. elephant liver: 0.819, vs. elephant brain: 0.796; vs. elephant PBMCs: 0.758.

It is worth noting that the compartmentalization signal is robust enough that it can be observed when analyzing the data generated solely from the suboptimal non-ethanol preserved samples (Figure S3A). This signal, while noisier on account of the lower contact counts, yields a highly concordant segregation pattern at 1-Mb resolution with that calculated for the ethanol preserved sample.

In addition to chromosome-specific Pearson-correlation maps shown in Figures 3 and S3A, we calculated a genome-wide second-order Pearson correlation map for the woolly mammoth skin and the Asian elephant skin datasets at 1-Mb resolution. The matrix was calculated as described above, with Observed/Expected values for interchromosomal data extracted from the corresponding *.hic* files using Juicer Tools[77] along with intrachromosomal Observed/Expected. The *.hic* files were generated to include genome-wide normalization on the entire input matrix, and this is the normalization we used. The resulting matrices demonstrate excellent agreement between the two datasets, for intra- and interchromosomal signals alike (Figure S3B).

PaleoHi-C contacts near isolated CTCF motifs tend to point in the same direction as the motif

It is well known that, in human Hi-C datasets, loops often occur between CTCF sites in the convergent orientation.[23] This phenomenon manifests, among other things, as a directionality bias among contacts. Thus, right-facing CTCF sites, on average, preferentially form contacts with genomic sequences to the right along a chromosome, toward the q terminus, and left-facing CTCF sites form contacts to the left, toward the p-terminus.[49] This bias forms due to the extrusion mechanism by which loops form (and is effaced by cohesin degradation, data not shown). Importantly, contact directionality bias analysis only requires the position of CTCF motifs in the genome to be known, and thus, unlike the methods described below, can detect architectural evidence for loop extrusion without requiring a loop annotation set.

To calculate CTCF motif directionality bias we used FIMO[116] with the MA0139.1 PWM matrix from https://jaspar.genereg.net/[140] to annotate CTCF motifs in the Loxafr3.0_HiC reference. We also used bedtools shuffle[112] to randomly permute the genomic locations of the annotated CTCF motifs in order to serve as a negative control. We then analyzed the positions of the resulting 58,958 hits to find 25 kb bins in the genome that harbored motifs in just one orientation, either positive (forward) or negative (reverse). We then calculated interaction directionality bias for these bins for three contact datasets: the Asian elephant skin *in situ* Hi-C, the subsampled Asian elephant skin Hi-C, and mammoth PaleoHi-C, all aligned to Loxafr3.0_HiC and filtered for mapping quality (mapq $\geq$30). (To construct the subsampled Asian elephant skin dataset, we matched the intrachromosomal long-range interaction count [unique interactions between loci on the same chromosome separated by distances $\geq$20 kb] in the mammoth PaleoHi-C data.)

The calculation was done following the approach described in Rowley et al.[49] Briefly, we summed the KR-normalized interaction counts at 25-kb resolution between the selected motif bins and sequences to the left and to the right of the motif at each distance. We then plotted the ratio of right/left summed signals (see Figure S4A). Similar to the modern Asian elephant dataset, mammoth PaleoHi-C demonstrates directional bias associated with CTCF motif orientation, whereas the negative control (scrambled motif positions) does not.

Calling point-to-point loops in the African elephant using HiCCUPS

We generated loop calls using HiCCUPS (Juicer Tools version 1.9.9)[23,77] and published data from an African elephant fibroblast cell line[27] analyzed with respect to our chromosome-length African elephant genome assembly Loxafr3.0_HiC. The Loxafr3.0_HiC contact map was generated by 3D-DNA from alignments to Loxafr3.0 (Table S1, #2) as part of the scaffolding procedure. HiCCUPS was run with default parameters using the contact map consisting of high mapping quality reads (mapq $\geq$30) as input.

The resulting 3,723 calls were further examined and rigorously filtered, manually, to remove 145 putative false positive calls. This was done in order to minimize the chance of false signal accumulation during aggregate peak analysis. Note that the loop count is lower than what is typically expected in a mammalian *in situ* Hi-C dataset[23] on account of the African elephant dataset being of relatively low depth. (The African elephant map comprised 511,627,023 raw read pairs and 345,636,788 Hi-C contacts, as compared to e.g., the 4.9 billion contacts that were used to generate loop calls for the GM12878 human cell line in Rao et al.[23] Consequently, only the strongest loops can be detected in the African elephant map.)

Aggregate Peak Analysis of PaleoHi-C data in the woolly mammoth

The Juicer Tools (version 3.25.21) APA script was used to run aggregate peak analysis with KR normalization (-k "KR").[77] We ran APA for three elephantid datasets: the African elephant fibroblast cell line data used to generate the loop calls, mammoth skin PaleoHi-C data, and a contact map generated from the modern Asian elephant skin sample. All the analyses were performed using the same filtered loop list (see calling point-to-point loops in the African elephant using HiCCUPS). Note that the script filters the input loop list to exclude loops that are too close to the diagonal to be included at a given resolution.

All datasets were analyzed against the Loxafr3.0_HiC genome assembly to avoid the need to liftover loop positions, and only reads with mapping quality score mapq $\geq$30 were used.

APA plots in Figure 4 of the main text are shown at 10-kb resolution with 1,883 of 3,582 loops included in the calculation. Two additional resolutions (5 kb and 25 kb) are included in Figure S4B.

In order to rule out the possibility that the mammoth enrichment is brought about by an artifact we repeated the APA analysis using three random non-overlapping subsets of the African elephant loop calls, each containing 1,000 loop calls. The idea behind such an analysis is that if, for example, the observed enrichment is associated with a local "jackpot" effect where the number of contacts in the vicinity of one of the annotated loops is markedly increased in the mammoth with respect to the background model due to an alignment artifact or other non-looping feature, we would expect that the enrichment in the central bin would only be observed in the subset that included the corresponding loop position and not in the other two subsets. Instead, the analysis shows comparable enrichment across all three subsets (see Figure S4D), consistent with the notion that the mammoth APA enrichments reflect real loop signatures.

Loop persistence results are robust to choice of loop caller and aggregation strategy

In order to demonstrate the robustness of the point-to-point chromatin loop preservation result with respect to the loop annotation tool and aggregation strategy, we repeated our elephantid loop analyses using the Significant Interaction Peak (SIP) caller and the SIPMeta aggregation tool.[102] We used SIP (https://github.com/PouletAxel/SIP) with default parameters to call loops at 10 kb and 25 kb bin sizes in the published data from an African elephant fibroblast cell line[27] analyzed with respect to our chromosome-length African elephant genome assembly Loxafr3.0_HiC. Only reads that aligned with mapping quality $\geq$30 were used for the analysis. We then merged these loop calls at 25-kb resolution. The resulting 6,340 loop calls can be viewed interactively[100] via the following link: https://t.3dg.io/3d-mammoth-SIP.

We then plotted the SIPMeta (https://github.com/JRowleyLab/SIPMeta) bullseye average metaplots of the contact signal in the corresponding positions across three contact maps: the African elephant fibroblast cell line used to call the loops, the Asian elephant skin contact map subsampled to match the mammoth for long-range contact count, and the mammoth skin contact map at 25-kb resolution. (All datasets were aligned to Loxafr3.0_HiC assembly and thresholded for mapping quality. Loops with anchors less than 250 kb apart were excluded from the analysis.) The resulting average metaplots are shown in Figure S4E (rightmost column).

Loops are preserved from human to mammoth

We repeated the aggregate loop analysis using loops annotated in human skin by the ENCODE project (https://www.encodeproject.org/files/ENCFF531FEY). We lifted the corresponding annotations to the African elephant genome assembly Loxafr3.0_HiC as follows.

We started with the SegAlign-generated alignment file between Loxafr3.0_HiC and hg38 described in the Elephantidae chromosome assignment section. We then proceeded with creating alignment chains, running RepeatFiller[113] and removing chain-breaking alignments with chainCleaner.[114] The resulting chain file was used to liftover the loop anchor coordinates from hg38 to Loxafr3.0_HiC using the command-line version of the LiftOver tool from the Kent Utilities suite https://github.com/ucscGenomeBrowser/kent/[115] with the minimum ratio of bases that must remap set to 10% (-minMatch = 0.1).

Of the 14,339 loops annotated in the human skin Hi-C dataset, 3,439 loops had at least one anchor that was either missing from the alignment chains, partially deleted, split, or duplicated. Anchors from another 217 loops lifted over to genomic positions in Loxafr3.0_HiC such that the corresponding sequences for the two anchors were no longer on the same chromosome. The remaining 9,683 loops, when analyzed in aggregate, were associated with pronounced enrichment at the center of APA plots, at multiple resolutions, for both the Asian elephant and mammoth skin data, with similar levels of enrichment (see Figures 4B and S4C). Thus PaleoHi-C data preserves signatures of loops that persist over vast evolutionary distances as well as millennia in permafrost.

3D architecture results are robust with respect to the choice of assisting reference

We redid the analysis of 3D genome architecture in the woolly mammoth using MamPri_Loxafr3.0_assisted_HiC and MamPri_ASM1433276v1_assisted_HiC in order to demonstrate that the main findings of the study, namely, the preservation of cell-type-specific segregation between open and closed chromatin, the persistence of point-to-point chromatin loops, and superdomains on mammoth chromosome X, are robust with respect to the choice of assisting reference.

Detailed alignment, duplication rate, and signal distribution statistics calculated by the Juicer pipeline[77] while specifying the respective assemblies as a genome reference using a ''-z'' flag, can be found in Table S3, #1–2. Juicer (Juicer version 1.6; BWA 0.7.17-r1188; 8 threads; openjdk version "1.8.0_222-ea"; Juicer Tools Version 1.9.9) was run with ''-s none'' flag to account for non-enzymatic fragmentation of the sample. Table S3, #3–4, lists the corresponding numbers for the Asian elephant datasets (skin, ovary, peripheral blood monocytes, liver, and brain tissues), and the African elephant datasets (primary fibroblast cultures derived from two individuals, a male and a female).

Following the procedure described in the main text and in the section compartment analysis at 1-Mb resolution, we used Juicer Tools Version 1.9.9[77] to calculate the eigenvectors for all 28 chromosomes in MamPri_ASM1433276v1_assisted_HiC for the combined woolly mammoth skin contact map and the Asian elephant tissue panel at 1-Mb resolution. Only contacts with mapq $\geq$30 were included in the analysis. The same analysis was repeated for MamPri_Loxafr3.0_assisted_HiC. The results are summarized in Figure S3C, in the form of similarity matrices and accompanying dendrograms. Similarity score for a pair of eigenvectors was calculated as a fraction of 1-Mb bins such that the eigenvalues corresponding to that bin had the same sign in both maps, i.e., either were both positive or both negative. (Because the sign of the eigenvector is arbitrary, for each chromosome we calculated the fraction of bins with the same sign for both the original pair of eigenvectors, and with one of the eigenvectors in the pair inverted. If the fraction of bins with eigenvalues of the same sign in the latter case was higher than in the former, all eigenvalues for that particular chromosome were multiplied by -1 prior to genome-wide calculation of the similarity score.) Bins in which the absolute value of the eigenvector was less than 0.05 in at least one of the contact maps were excluded from the analysis. The dendrogram shown above the similarity matrix is based on Pearson distances between the eigenvectors. The results are consistent with those described in the main text, and suggest that the woolly mammoth skin is a closer match to modern Asian elephant skin than any other Asian elephant tissue.

We then used HiCCUPS to identify point-to-point loops in the African elephant fibroblast contact maps generated using MamPri_Loxafr3.0_assisted_HiC and MamPri_ASM1433276v1_assisted_HiC (see calling point to point loops in the African elephant using HiCCUPS). The number of loops annotated was 3,495 and 3,575, respectively. Following the strategy outlined in the aggregate peak analysis of PaleoHi-C data in the woolly mammoth section, we generated APA plots. The plots show enrichment of aggregate signal associated with the African elephant loop calls in PaleoHi-C, consistent with the persistence of loops in the woolly mammoth. The enrichment is comparable to that calculated for the Asian elephant skin contact maps (Figure S4E).

Finally, we examined the chromosome X contact maps generated by aligning the mammoth PaleoHi-C data to the MamPri_ASM1433276v1_assisted_HiC and to the MamPri_Loxafr3.0_assisted_HiC genome assemblies (Figure S5B). Consistent with the results shown in Figure 5, the maps exhibit a tetradic structure. We clearly observe a superdomain boundary at DXZ4 and two additional boundaries (ICCE and FROST, see tetradic structure of woolly mammoth and modern elephant X chromosomes).

Apparent differences in gene activity in mammoth skin vs. Asian elephant skin

High-resolution compartment analysis in sparse Hi-C datasets using CRUSH

Previous studies have shown that transcriptional state and compartment assignment organization are highly correlated.[17,23,82] Hence, it would be desirable to explore the compartment assignment obtained from PaleoHi-C to explore the transcriptional state of individual mammoth genes.

However, the eigenvector method described above, and which is typically used in Hi-C studies, is inadequate for this purpose. This is for two reasons. First, it cannot be efficiently calculated at high resolutions. As such, even using a deeply sequenced Hi-C dataset, the method cannot be used to identify compartments at the scale of individual gene bodies (median length $\sim$30 kb[141]). To address the problem of algorithmic efficiency, we recently introduced a much more efficient numerical procedure for calculating the compartment eigenvector via principal component analysis, dubbed POSSUMM.[82]

But even this method fails when the Hi-C data is sparse, as in the case of PaleoHi-C. For example, we generated 50 kb POSSUMM eigenvector calls in the elephant skin dataset, and compared them with those generated using the same procedure on a

downsampled version of the same dataset. The downsampled dataset comprised only 7 million read pairs from the *in situ* Hi-C elephant library in order to roughly match the mammoth PaleoHi-C contact count (1,793,040 contacts in the inrachromosomal long-range category when aligned to Loxafr3.0_HiC). The eigenvector-based compartment assignments at 50-kb resolution in the downsampled dataset were not meaningful on many chromosomes, disagreeing with eigenvector-based compartment assignments for the full dataset as often as 49% of the time (since eigenvectors are determined up to sign, higher than 50% disagreement is not possible). Thus, the eigenvector did not reliably calculate compartment assignments at 50-kb resolution in a Hi-C dataset whose sparsity is comparable to the depth of PaleoHi-C. This suggested that using a more robust methodology for compartment assignment would be helpful for the analysis of PaleoHi-C data.

To address this challenge, we generated A/B annotations in both woolly mammoth and Asian elephant at 50 kb using CRUSH (Compartment Refinement for the Ultraprecise Stratification of Hi-C, https://github.com/JRowleyLab/CRUSH), a method specifically designed for analyzing sparse Hi-C datasets.

CRUSH is based on the A/B index, a non-principal component analysis-based method for compartment identification.[49] In CRUSH, a subset of loci is initially labeled as pseudo-A or pseudo-B, based on characteristics that correlate with and have been previously used to predict A and B compartments.[142,143] Because the A compartment has high GC content and is gene-rich,[144] for each chromosome we initialize bins with GC content one standard deviation below the average as pseudo-B. Pseudo-A bins are defined as bins that contain an annotated gene. These initialized bins represent only a small fraction of the genome, and are subsequently filtered and refined using data from Hi-C experiments.

Next, CRUSH examines each row within the Hi-C matrix and measures the relative interaction intensities of that row with pseudo-A and pseudo-B bins, calculating a score similar to the A/B index.[49] The calculation includes distance normalization (by calculating (observed +1)/(expected +1)) followed by independently z-scoring each row in the Hi-C matrix. Then the row's A/B index is calculated as the average *Z* score for interactions with pseudo-A bins minus the average *Z* score for interactions within pseudo-B bins. This initial calculation is used to refine the initialization state by removing pseudo-A bins that interact preferentially with pseudo-B, and vice versa. Then the A-B index calculation is rerun. The resultant 1D track can be thought of as the row's relative interaction strength with features that correlate with A or B compartments.

The analysis is repeated across multiple resolutions using a "resolution-walking" technique, from the coarsest resolution down to the target resolution (e.g., 2.5 Mb, 1 Mb, 500 kb, 250 kb, 100 kb, and 50 kb). For each subsequent (higher) resolution in the walk, the CRUSH calculation is informed by the tracks from a preceding lower resolution step. E.g., as the pseudo-A and pseudo-B are defined at 500-kb resolution, bins whose assignment based on the GC and gene content is in conflict with the 1-Mb resolution track are filtered out from the initialization set for the 500 kb CRUSH calculation.

An example of such a resolution walk for one of the mammoth chromosomes is shown in Figure S3D. Both the mammoth and the Asian elephant skin datasets were aligned to Loxafr3.0_HiC for this analysis. The B compartments were initialized using Loxafr3.0_HiC.fasta. The A compartments were initialized using RefSeq transcript annotations[94] for Loxafr3.0 lifted to the Loxafr3.0_HiC chromosome-length genome assembly via the 3D-DNA[22] liftover script (https://github.com/aidenlab/3d-dna/blob/phasing/lift/lift-input-bed-to-HiC-bed.awk).

We validated the CRUSH results by comparing them to the eigenvector-based A/B calls.

First, we sought to check whether the results of CRUSH were consistent with the results of low-resolution (1Mb) compartment eigenvector calculations. We compared the compartment assignments for the elephant skin dataset at 1Mb resolution. We coarse-grained the 50 kb CRUSH calls to 1-Mb resolution by calculating the median score across 20 bins. The resulting calls and the 1-Mb eigenvector track were in excellent agreement: among the 1,874 bins that were confidently assigned to either A or B compartment by both methods, we found only 5 bins (0.27%) with discordant assignments genome-wide. (For the purpose of this analysis we ignored ~30% of the bins in which absolute CRUSH score or eigenvalue were closest to zero, suggesting ambiguous compartment assignment.) The concordance was still high, albeit somewhat lower, for the mammoth skin dataset (2.72% discordant assignment genome-wide).

Next, we checked whether CRUSH results were consistent with the results of higher resolution (50 kb) compartment eigenvector calculations, in Hi-C datasets where sufficient data is available to generate the latter. We therefore compared CRUSH calls to eigenvector-based calls at 50-kb resolution in the elephant skin *in situ* Hi-C dataset. The 50-kb eigenvector was calculated using POSSUMM.[82] Once again, the eigenvector and CRUSH tracks were highly concordant, with only 24 (0.06%) contradictory assignments among the 39,095 confidently assigned bins.

Taken together, the above results confirmed that CRUSH agrees with the compartment eigenvector so long as there is sufficient data available to calculate both.

Finally, we revisited the analysis which we performed at the outset of this section, and compared the 50 kb POSSUMM eigenvector calls for deeply sequenced elephant skin Hi-C with those generated by CRUSH in a downsampled elephant dataset. As before, for this test, we used only 7 million read pairs from the *in situ* Hi-C elephant library to roughly match the mammoth PaleoHi-C contact count. We observe excellent agreement, with 0.73% discordant bins among the 36,820 bins with high-confidence compartment assignments, validating high-resolution CRUSH calls in a sparse dataset. Thus, CRUSH can be used to annotate compartments reliably at 50 kb, in maps that are too sparse to be annotated using the compartment eigenvector.

Differential compartment calls in the mammoth and Asian elephant skin Hi-C datasets

We examined CRUSH annotations at 50-kb resolution for the mammoth and Asian elephant skin datasets to identify significant changes in compartmentalization between the two. We dubbed such regions MARs, for Mammoth Altered Regions.

Specifically, we annotated MARs as described below. First, we calculated the difference between the mammoth and the elephant CRUSH score in every bin, genome-wide. For the resulting distribution, we calculated the 5th and 95th percentile values. Then, we marked ~5% of the bins where the absolute value of the CRUSH score was low as ambiguous bins, in which a compartment assignment could not be made with certainty. Finally, we defined two kinds of MARs, one for sequences where mammoth PaleoHi-C data suggested a stronger association with the B compartment (closed chromatin) compared to the elephant, and another for sequences where mammoth was more strongly associated with the open A compartment than elephant.

The first kind, dubbed "M < E", was defined as a bin (or a stretch of consecutive bins) such that: 1) its differential CRUSH score was below the 5th percentile value, and 2) the bin was a part of a stretch at least 3 bins wide (150 kb) in which the mammoth was consistently closed (CRUSH score <0) or ambiguous, and the elephant was consistently open (CRUSH score >0) or ambiguous. Similarly, MARs demarcating regions more open in the mammoth dataset compared to the elephant ("M > E"), were defined as bins where: 1) the differential score was above 95th percentile value, and 2) the bin was a part of a stretch at least 3 bins wide (150 kb) in which the mammoth was consistently open (CRUSH score >0) or ambiguous, and the elephant was consistently closed (CRUSH score <0) or ambiguous.

The procedure identified 731 "M < E" and 697 "M > E" MARs, respectively, totaling 106.9 Mb of sequence with significant changes in compartmentalization, or ~3.4% of chromosomal sequences (Table S3, #5).

We associated the MARs with individual genes as follows. First, we lifted RefSeq[94] protein-coding gene annotations for Loxafr3.0 to the Loxafr3.0_HiC chromosome-length genome assembly using the 3D-DNA lift-input-bed-to-HiC-bed.awk script (https://github.com/aidenlab/3d-dna/blob/phasing/lift/lift-input-bed-to-HiC-bed.awk). We then used bedtools intersect[112] with default parameters to screen for overlaps between the lifted gene annotations and MARs features. We identified 395 and 425 genes overlapping with "M < E" and "M > E" MARs, respectively.

It is worth noting some caveats to the annotation of differences in gene activity in woolly mammoth skin vs. Asian elephant skin. First, our analysis is too coarse to reliably see a potential compartmentalization effect in small genes, which may bias the differential lists. Second, compartmentalization of neighboring and overlapping genes is not independent, and as such, neighboring bystander genes may end up being annotated as highly differential in our analysis.

Examples of apparent differences in gene activity in mammoth skin vs. Asian elephant skin

There are many more genes that change in mammoth vs. elephant than what was highlighted in the main text.

We include CRUSH tracks for several other genes of interest in Figure S3E. These genes are known to be highly expressed in the skin and participate in a wide array of crucial skin processes. Each panel shows a 1-Mb stretch of the Loxafr3.0_HiC genome centered on a gene of interest. The gene track is shown at the top of each panel. Below the genes are two CRUSH tracks at 50-kb resolution, those calculated for mammoth skin based on PaleoHi-C (top), and those for the Asian elephant skin dataset (bottom).

The following genes are included in the figure:

(1) *Barx2*, whose deletion in mice leads to short hair (Olson et al.[184]), is thought to have a role in controlling adhesive processes in keratinizing epithelia[145];
(2) *Apobec1* ("apolipoprotein B mRNA editing enzyme catalytic subunit 1") and *Syk* ("Spleen Tyrosine Kinase") are hair follicle epithelial cell signature genes[146,147];
(3) *Gja3* ("Gap Junction Protein Alpha 3) and *Enpp6* ("Ectonucleotide Pyrophosphatase/Phosphodiesterase 6") are melanocyte signature genes[146,147];
(4) *Ovol2* ("ovo like zinc finger 2") and *Grhl2* ("Grainyhead-like 2") modulate cell differentiation and proliferation in human keratinocytes and hair follicles[148,149];
(5) *Dusp10* ("dual specificity phosphatase 10") is a regulator of brown adipocyte differentiation[150];
(6) *Tdo2* has been shown to play an important role in the initiation of hair growth in rats.[151]

Gene ontology (GO) analysis

In order to test whether the list of genes overlapping MARs is enriched for particular functional categories, we followed the example of a recent woolly mammoth study[24] and performed gene ontology (GO) term enrichment analysis using GOrilla.[57]

Briefly, we performed a simple overrepresentation test using RefSeq protein gene annotations for Loxafr3.0 as a background gene set, specifically restricting the gene set to those protein-coding genes successfully lifting to chromosome-length scaffolds in Loxafr3.0_HiC (20,246 genes). Following the recommendations in Hong et al.,[152] we performed the analysis separately for M < E genes, i.e., those that, based on MARs analysis, might have been associated with the B compartment, and therefore with a lower expression level in the mammoth sample compared to the elephant, and for M > E genes, i.e., those that are more strongly associated with the A compartment in mammoth and might have been associated with higher expression levels. The gene sets were analyzed using the database for *Mus musculus* (GOrilla database update from Mar 6, 2021).

We found several GO terms overrepresented among the M < E genes with $p \leq 0.0001$ (Table S3, #6) including those related to hair (GO:0001942: hair follicle development) and to cell fate in the epidermis more generally (GO:0045605: negative regulation of epidermal cell differentiation; GO:0008544: epidermis development; GO:0045683: negative regulation of epidermis development; GO:0045617: negative regulation of keratinocyte differentiation; GO:0010837: regulation of keratinocyte proliferation; note that keratinocytes are the most common cell type in the epidermis). The same analysis of M > E overrepresented genes highlights GO terms associated with cell-cell signaling and transmembrane transport (Table S3, #7).

In addition to the MARs gene overrepresentation analysis described above, we ranked all annotated protein-coding genes according to the difference between their CRUSH scores in the elephant and mammoth skin samples. When multiple 50 kb bins overlapped a gene, we examined the difference between the CRUSH scores for each of the overlapping bins, and used the median of these differences for ranking purposes. The scores for each gene can be found in Table S3, #8.

Comparison of elephant dermis vs. epidermis

The differences in gene expression between mammoth and elephant that are observed in this study may be interpreted in one of three ways: a) they reflect species-specific changes in gene expression in a particular cell type in mammoth vs. elephant; b) they are due to species-specific differences in cell type composition of the comparable skin samples; and c) they are not due to species-specific differences at all, but instead are due to differences in the anatomical location from which the samples were derived.

To explore whether variation in the anatomical sampling that are not in fact species-specific could lead to the kind of differences we observed between woolly mammoth skin and elephant skin, we compared two tissue samples derived from elephant skin. These samples were selected so that one overlapped the epidermis more extensively (as described in the main text), and one preferentially overlapped the dermis (see Asian elephant samples and library preparation). We performed CRUSH on both datasets.

We then checked whether skin-layer-specific differences could account for the apparent differences in gene expression seen in our comparisons of woolly mammoth and Asian elephant skin. We focused on the genes responsible for the enrichment of the GO term ''hair follicle development'' in the overrepresentation analysis. These are:

(1) *Egfr* (''epidermal growth factor receptor''), described in the main text, plays a crucial role in skin development and physiology, influencing processes such as keratinocyte survival, differentiation, and migration. *Egfr* inhibition has been connected to epidermal thickening and various hair phenotypes.[54,55,153] Increased exposure to *Egf* (''epidermal growth factor''), the canonical ligand for *Egfr*, has been shown to result in epidermal thickening and inhibition of hair growth[56];
(2) *Edaradd* (''*Edar*-associated via death domain protein''), described in the main text, is a member of the *Eda*/*Edar* pathway involved in the induction of primary hair follicles and associated with hypohidrotic ectodermal dysplasia, a rare disease characterized by sparse hair and absence of sweat glands.[154] Population studies in Arctic Indigenous people[52] suggested that *EDAR*/*EDARADD* pathway may be relevant for adaptation to cold environments;
(3) *Lgr5* (''leucine rich repeat containing G protein-coupled receptor 5'') is a conserved marker of hair follicle stem cells, and is present early and throughout follicle morphogenesis[155];
(4) *Tgfb2* (''transforming growth factor β 2'') plays an important role in the human hair cycle. Research has highlighted the potential of *TGF-β* antagonists in preventing male pattern baldness and promoting hair elongation[156];
(5) *Lgr4* (''leucine rich repeat containing G protein-coupled receptor 4'') has been researched in connection to epidermal hyperplasia and keratinocyte proliferation[157];
(6) *Trp63* (''transformation related protein 63'') transcription factor is considered a master gene for the development of skin and hair follicles. Overexpression of *Trp63* is known to cause dramatic defects in hair follicle development and hair cycling[158];
(7) *Tnfrsf19* (''tumor necrosis factor receptor superfamily, member 19'', also known as *TROY* in humans) marks epidermal cells that govern interfollicular epidermal renewal and cornification.[159]

We plotted CRUSH tracks in the vicinity of these seven genes (Figure S3F). We exhibit CRUSH-generated compartmentalization tracks from the derma-rich Asian elephant *in situ* Hi-C library alongside those generated from the woolly mammoth and the epidermis-rich Asian elephant dataset. In all but one case, the difference between mammoth CRUSH and the elephant epidermis-enriched sample was also seen when comparing to the derma-enriched sample. The sole exception was *Tnfrsf19*.

We then repeated the GO overrepresentation analysis described above, this time comparing the Asian elephant epidermis-enriched sample and the Asian elephant dermis-enriched samples. Consistent with the similarity of the CRUSH tracks at the associated genes between the two datasets, ''hair follicle development'' is not detected as an enriched GO term in this analysis (Table S3, #9–10).

Taken together, these findings illustrate that the differential gene expression analyses reported in the main text are robust to the choice of Asian elephant skin layer that is used as a comparator. Of course, such conclusions could be drawn with even greater certainty by constructing an atlas of Hi-C maps derived from skin samples from all across elephant anatomy, but this was beyond the scope of the present study.

Tetradic structure of woolly mammoth and modern elephant X chromosomes

Estimating CTCF binding capacity along the X chromosome

We estimated CTCF-binding capacity along the X chromosome as follows. We extracted all PaleoHi-C reads where one end of the read aligned to the motif ''CC.C[TC].G[AC]TGGCA.T''. This is a consensus motif for the DXZ4/ICCE class of repeats that was manually curated based on a review of the literature.[48,95] We then plotted the position of both the sequence matching read and its pair, on the logic that: a) if parts of the repeat sequence were assembled into the chromosome-length scaffolds, the procedure would produce a hit in approximately the correct position; and b) if the repeat sequence was not assembled or was not anchored, the second read might nonetheless produce a hit to a nearby assembled portion of the chromosome-length scaffold. The procedure was designed to yield some robustness with respect to genome assembly quality, but can in principle lead to false positive hits, e.g., when

a large number of repeats occur interspersed with a CTCF binding motif in one genomic position and without CTCF in another. The resulting chromosome-wide tracks for the X chromosome are shown in Figure 5. Peaks are visible at DXZ4, ICCE, and FROST. Peaks visible downstream of DXZ4 correspond to another CTCF-binding element, FIRRE.[41]

The ICCE and DXZ4 boundaries disappear in males

We asked whether, like the DXZ4 boundary in humans, the superdomain boundaries identified on the mammoth X chromosome are specific to the inactive X chromosome. If this is indeed the case, we would predict that these boundaries would be observed in unphased contact maps only if the maps derive from females, since such maps will reflect a combination of data from both the active, Xa chromosome, and the inactive, Xi chromosome. By contrast, the boundaries should disappear in contact maps generated from male samples, which only contain the Xa chromosome.

Figures 5 and S5A show the male vs. female contact maps for the Asian (PBMCs, see Table S1, #1 and Table S3, #11) and African elephants (fibroblast data from Álvarez-González et al.[27]), as well as for human (lymphoblastoid data from Rao et al. and Harris et al.[23,82]). The images show the disappearance of the FROST, DXZ4 and ICCE boundaries in male elephants. See also the interactive version of Figure 5 shared at https://t.3dg.io/3d-mammoth-Fig-5.

FROST boundary lies at the edge of elephantid pseudoautosomal region

The analysis of male and female data near FROST can be used to show that the element flanks the PAR boundary. We observed that Hi-C coverage tracks for the X chromosome in both male and female Asian elephants are similar to one another from the p-terminus of the X chromosome until the position of this boundary element. On the opposite side of FROST, coverage is roughly 2-fold lower in males than in females (Figure S5A). This is consistent with the hypothesis that the pseudoautosomal region (PAR) in mammoth and other elephantids extends from the p terminus to FROST, and with earlier studies suggesting that the elephantid pseudoautosomal region is larger than its human counterpart.[58]

Modeling the diffusion of DNA fragments in samples of ancient woolly mammoth skin

Theoretical analysis of a random coil polymer after fragmentation and simple diffusion of the monomers

Consider a sequence of points in three-dimensional space:

$$\overrightarrow{r_i} = (x_1, y_1, z_1), (x_2, y_2, z_2), \ldots (x_N, y_N, z_N).$$

When using such points to represent a polymer or genome, it is natural to think of $\boldsymbol{r}_i$ as corresponding to the spatial position of the i^{th} monomer in the chain. We can say that two monomers, $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, are in contact if the Euclidean distance between the points $\|\boldsymbol{r}_i, \boldsymbol{r}_j\|$ is less than a fixed threshold k. It is possible to define a contact probability function as follows:

$$p(s) = P(\|\overrightarrow{r_i}, \overrightarrow{r_j}\| < k,\ s = |i - j|).$$

This function describes the likelihood that two monomers are in contact as a function of their distance, in monomers, along the 1D contour of the polymer. The behavior of $p(s)$ has been carefully studied for many types of polymers, as it frequently exhibits specific power laws that are characteristic of various polymer states. It has also been studied for other trajectories like random walks and fractal curves.[160] As described in the body of the paper, the behavior of $p(s)$ is also examined in many studies of genome architecture, as it can be readily measured using Hi-C experiments.[17]

For an ancient biosample, the chromosomal polymer fragments, in 1D, into numerous oligonucleotides that are free, in principle, to diffuse independently through space. It is natural to ask how this process of diffusion affects the behavior of $p(s, t)$. Here we have made the dependence on t explicit in order to highlight that the contact probability function – which reflects the architecture of the polymer or genome – will change over time, as a result of the effacement of the architecture via the ongoing diffusion of the monomers. The exact relationship between the extent of the diffusion and the extent of the effacement is not immediately obvious, but certain aspects of this relationship are apparent. For instance, as diffusion goes on and the positions of the monomers become entirely decorrelated with their initial position, $p(s)$ becomes independent of s.

This process can be modeled by considering an initial polymer trajectory $\boldsymbol{r}$, and then allowing the individual monomers $\boldsymbol{r}_i$ to diffuse through space independently of one another. (Note that, in this model, each monomer might correspond to, e.g., a short oligo, rather than a single basepair, in an ancient biosample.)

Such a model is amenable to theoretical analysis. For instance, let us, for the sake of an example, model the structure of the ancient woolly mammoth genome as a polymer in a neutral solvent. Such polymers adopt a random coil conformation, equivalent to the monomers performing a random walk in three-dimensional space. The trajectories for random walks in general can be described by the Gaussian probability distribution:

$$G(\overrightarrow{r}; \overrightarrow{\mu}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} e^{-\frac{r^2}{2\sigma^2}},$$

where

(1) $\boldsymbol{\mu}$ is the expected value of the end-to-end vector,
(2) σ^2 is the variance,

(3) $G(\boldsymbol{r};\boldsymbol{\mu},\sigma^2)$ represents the probability of obtaining the trajectory $\boldsymbol{r}$ given $\boldsymbol{\mu}$ and σ^2.

In particular, the trajectories for a random coil correspond to:

$$G\left(\vec{r};0,sb^2\right),$$

where

(1) 0 is the all zero vector (the origin, since the expected displacement for an isotropic random walk is 0 in all directions),
(2) b is the Kuhn length of the polymer,
(3) s is the dimensionless total length of the polymer, measured in Kuhn lengths (i.e., monomers).

For such trajectories, it is well known that

$$R(s) = b\sqrt{s},$$

where

(1) b is the Kuhn length of the polymer,
(2) s is the total length of the polymer, measured in Kuhn lengths,
(3) $R(s)$ is Flory radius, the root-mean-square distance between two monomers that are s Kuhn lengths apart.

One can calculate the probability of contact as follows:

$$p(s) = \int_{CV} G\left(\vec{r};0,sb^2\right)d\vec{r},$$

where CV can be thought of as a "contact volume" in which contacts take place. Considering the contact volume to be a cube of volume αb^3 where α is a dimensionless constant we obtain:

$$p(s) = \left(\frac{3}{2\pi R^2}\right)^{\frac{3}{2}}\left(\int_{-\frac{\sqrt[3]{\alpha b}}{2}}^{\frac{\sqrt[3]{\alpha b}}{2}} e^{\frac{-x^2}{\left(\frac{2R^2}{3}\right)}}dx\right)^3 = \left(erf\left(\frac{\frac{\sqrt[3]{\alpha b}}{2}}{\sqrt{\frac{2R^2}{3}}}\right)\right)^3 \approx \left(\frac{2}{\sqrt{\pi}}\frac{\frac{\sqrt[3]{\alpha b}}{2}}{\sqrt{\frac{2R^2}{3}}}\right)^3 = \left(\frac{3}{2\pi}\right)^{\frac{3}{2}}\frac{\alpha b^3}{R^3} = \left(\frac{3}{2\pi}\right)^{\frac{3}{2}}\alpha s^{-\frac{3}{2}} = \alpha' s^{-\frac{3}{2}},$$

i.e., that the scaling of the contact probability for a 3D random coil has a power of $-3/2$.

The trajectories for the diffusion of each monomer after the dissolution of the chain are given by

$$G\left(\vec{r};0,6Dt\right),$$

where

(1) D is the diffusivity of the ancient sample, and
(2) t is the age of the sample (here we make the simplifying assumption that the decay of the phosphate backbone takes place rapidly after the death of the mammoth).

The relative displacement of two monomers in the present sample, conditioned on the fact that they were s monomers apart in the original genome polymer, is the result of (i) their relative positioning along the original chain, $G(\boldsymbol{r};0,sb^2)$; and (ii) the subsequent diffusion of the two monomers, each of which can be described by $G(\boldsymbol{r};0,6Dt)$. The total displacement can therefore be represented as the convolution of the corresponding gaussians, which yields still another Gaussian:

$$G\left(\vec{r};0,6Dt\right)\cdot G\left(\vec{r};0,sb^2\right)\cdot G\left(\vec{r};0,6Dt\right) = G\left(\vec{r};0,sb^2+12Dt\right).$$

This gives an expected inter-monomer distance of

$$R(s,t) = \sqrt{sb^2+12Dt},$$

and a contact probability given by:

$$p\left(s,t\right) \approx \left(\frac{3}{2\pi}\right)^{\frac{3}{2}}\frac{\alpha b^3}{R^3} = \left(\frac{3}{2\pi}\right)^{\frac{3}{2}}\frac{\alpha b^3}{\left(sb^2+12Dt\right)^{\frac{3}{2}}} = \alpha' s^{-\frac{3}{2}}\left(1+\frac{12Dt}{sb^2}\right)^{-\frac{3}{2}}.$$

The diffusion has little effect on the original power law (exponent = −3/2) so long as the second term in parentheses is dominated by the first, i.e., so long as:

$$b\sqrt{s} \gg \sqrt{12Dt}.$$

Thus, the contact probability's power law scaling will be preserved at all scales where the distance between two monomers in the original polymer greatly exceeds the distance diffused by the two monomers after fragmentation. Of course, in the $sb^2 \ll 12Dt$ limit we find that

$$p(s,t) = \alpha'\left(\frac{12Dt}{b^2}\right)^{-\frac{3}{2}},$$

which, as expected, no longer depends on s at all.

The results can be generalized to random walks of arbitrary dimension n to yield:

$$p(s,t) = \alpha\left(\frac{n}{2\pi}\right)^{\frac{n}{2}} s^{-\frac{n}{2}}\left(1+\frac{2nDt}{sb^2}\right)^{-\frac{n}{2}}.$$

Simulations of a random coil polymer after fragmentation and simple diffusion of the monomers illustrate theoretical analysis

To illustrate the above analysis, we constructed 2D and 3D random walks of unit step length comprising 10^5 steps each. We simulated the effects of diffusion by displacing the coordinates of each point in the random walk by a value distributed according to normal distribution with variance q ($q^2 = 2Dt$). We studied the following q values: 0, 4, 8, 12, and 16. We then examined the effects of diffusion on contact probability. At each q the contact probability curves are constructed by examining 10 random walks. The results are summarized in Figure S6A. There is a very close correspondence between the theoretical predictions and the simulated values.

Simulations of other fractal conformations after fragmentation and simple diffusion exhibit similar behavior

In the preceding sections, we explored how diffusion influences contact probability $p(s, t)$ in random coil polymers. The analysis of this model revealed a persistent power-law relationship at larger scales when the distance between two monomers in the original polymer greatly exceeds the distance diffused by the two monomers after fragmentation. At smaller scales, $p(s, t)$ is independent of s, reflecting that at distances smaller than the range covered by diffusion after fragmentation, the diffusion process effectively obliterates the memory of the monomers' initial positions.

Intuitively it seems that we should be able to extend this result beyond simple random coils. To explore further, in this section we will demonstrate that space-filling curves, a class of structures whose contact probability exhibits a power-law scaling and that are sometimes used to model chromatin,[17,38] demonstrate the same basic behavior in response to diffusion.

To illustrate, we constructed 2D and 3D Hilbert curves of unit step length comprising 2^{20} and 2^{21} steps, respectively. Similar to the simple random walk case, we simulated the effects of diffusion by displacing the coordinates of each point in the Hilbert curve by a value distributed according to normal distribution with variance q ($q^2 = 2Dt$). We studied the following q values: 0, 4, 8, 12, and 16. We then examined the effects of diffusion on contact probability, see Figure S6B. The qualitative behavior of the power law is the same as in the case of a random coil, with similar parameter regimes for scaling effacement and preservation. There is also an overall multiplicative increase in the contact frequency, which is not seen in the case of a random coil. This overall increase does not affect the presence or absence of the power laws themselves. Similar behavior can be seen when analyzing the effects of diffusion on contact probability in polymers modeled after other fractal curves (Figures S6C–S6F).

PaleoHi-C contact probability scaling analysis using the simple diffusion model indicates that DNA has diffused less than 50 nm in the woolly mammoth sample

In the preceding section, we argued that diffusion of monomers (or short oligonucleotides) ought to be detectable by looking for changes in $p(s)$. In this section, we put this strategy to use using PaleoHi-C data.

Comparing $p(s)$ for ancient mammoth and modern elephant, we find that they are identical in woolly mammoth and modern elephant at scales of 500 bp or larger, see Figure 6B. This observation is inconsistent with diffusion of mammoth DNA in the sample over distances that are larger than the spatial extent of 500 bp in the original polymer. Of course, the length of 500 bp of B-DNA is 170 nm, but inside the nucleus, DNA is not fully stretched out, and is packaged into nucleosomes.

We therefore want to calculate the end-to-end distance of 500 bp of chromatinized DNA *in vivo*. However, we note that 500 bp is not a multiple of the nucleosome repeat length (200 bp), making it somewhat trickier to consider. Instead, we will consider 550 bp of chromatinized DNA, which turns out to be a little simpler, for reasons that will become apparent below.

To estimate the length of 550 bp of chromatinized DNA *in vivo*, we use the wormlike chain (WLC) model, which can be used to estimate the average end-to-end distance (R) of a polymer segment, as a function of its contour length, L, and the persistence length, P:

$$R = \sqrt{2P \times \left(L - P + P \times e^{-\frac{L}{P}}\right)}.$$

For this calculation, we will use persistence lengths of 30 nm for chromatin and 50 nm for naked DNA.[161]

One way to perform the calculation is to use the WLC model to estimate the length of the full 550 bp of chromatinized DNA. We calculate the contour length as follows. (1) Assuming a nucleosome repeat length of 200 bp, there are 3 nucleosomes and two segments of linker DNA. (2) Nucleosomes have a diameter of 10 nm and a height of 5.7 nm. The height is the relevant parameter for calculating contour length. (3) The height of a single base pair is 0.34 nm. (4) We assume 147 bp wound around each of three nucleosomes, leaving a total of 109 bp of linker DNA. (5) Thus the total contour length L is 3 × 5.7 nm + 109 × 0.34 nm = 54.16 nm, combining DNA wrapped around nucleosomes and linker DNA.

Using P = 30 nm, this yields an end-to-end distance of 41.78 nm for chromatinized DNA.

A second way to get insight into the size of 550 bp of chromatinized DNA is to note that since the contour length is mostly due to the naked DNA, the linker DNA should account for most of the end-to-end distance of the folded configuration. For 109 bp of naked DNA, one obtains a contour length of 37.06 nm (= 109 bp x 0.34 nm). Combined with P = 50 nm, the WLC model yields an end-to-end distance of approximately 33.0 nm.

Both approaches yield similar values. To be as conservative as possible, we round up to 50 nm in the main text.

We note that the deviations between contact probability curves generated for the mammoth vs. the modern elephant data seen for $s < 500$ bp may be unrelated to diffusion, as such deviations may arise as an artifact of the Hi-C protocol itself. Regardless, using arguments based on simple diffusion, we cannot rule out the possibility of diffusion at scales smaller than 50 nm. See the next section, however, for a discussion of how more sophisticated models that reflect the excluded volume of the monomers can influence these findings.

Incorporating excluded monomer volume into simulations of genome polymer fragmentation and diffusion suggests a smaller RMSD than the simple diffusion model

In prior sections, we considered extremely simplified models of diffusion after chromosome fragmentation. The models were simplified in two key respects. First, the initial chromosome configuration was an idealized curve, such as a random walk, rather than a realistic model of chromatin conformation. Second, the dynamics after breakdown of the polymer were governed by simple diffusion. In this section we consider a more realistic physical model of chromosomal fragment diffusion. This model uses realistic chromatin conformations to provide the initial conditions. It also considers the excluded volume of the monomers during the diffusion process, i.e., the fact that two monomers cannot be in the same place at the same time. We then study the consequences of such a model for the dynamics of the contact probability scaling as well as for the Hi-C contact map as a whole. Since such models are more sophisticated, theoretical analysis is less tractable, and we instead study this process exclusively using simulations.

We perform diffusion simulations using molecular dynamics on an ensemble of 5,000 native structures representative of a 1.1-Mb region of human chromosome 7 (corresponding to genomic positions 95.4 to 96.5 Mb) at nucleosome resolution (200 bp). Each bead in the structures corresponds to a nucleosome (~150 bp) and its corresponding linker DNA (~50 bp), so that each structure is composed of 5,500 nucleosomes in total. The initial ensemble of structures can be found in Harris et al.[82] Analysis of contact frequency within the initial ensemble (two nucleosomes are considered to be in contact if their centers are within 15 nm of one another) yields a contact map for the ensemble that closely matches the contact map seen for this locus in human lymphoblastoid cells, as assayed by Hi-C (Figure S7A).

The ensemble of these structures is subjected to fragmentation, which we simulate by instantaneously eliminating all bonds between successive nucleosomes along the polymer chains. Effectively, the bond removal represents the degradation of the DNA backbone in the samples, producing a 'gas' of nucleosomes. Nucleosomes diffuse according to Langevin dynamics, and the only interactions among nucleosomes are those due to excluded volume effects modeled through truncated Lennard-Jones potentials.

The results of these simulations are summarized in Figure S7A. We calculate how contact probability between nucleosomes changes as the structures are allowed to diffuse, as well as present example structures and contact maps. We also calculate the average root-mean-squared displacement (RMSD) of nucleosomes throughout the diffusion process, measured in relation to the position of the nucleosomes of the native, undegraded structures. This allows us to link the changes of the contact probabilities scaling and Hi-C contact maps to the displacement of nucleosomes.

The ensemble of initial, native structures corresponds to collapsed, globular polymers. The undegraded ensemble captures salient features of chromosome architecture, as seen through DNA-DNA proximity ligation experiments. Indeed, as is apparent from the contact map generated using the undegraded ensemble (Figure S7A, initial structure), the native *in silico* structures recapitulate typical features of Hi-C contact maps. These include short-range interactions between successive nucleosomes along the polymer chain, which manifest as a bright diagonal on the contact map, and long-range compartment interactions manifesting as a characteristic plaid pattern.

Unlike in the simple diffusion models of the preceding section, in the present model we take into account the fact that two monomers cannot occupy the same place at the same time. Once diffusion begins, the effect of the increasing RMSD of the nucleosomes becomes important and has a measurable effect on the contact maps and the contact probability scalings. Indeed, even a small amount of diffusion can be sufficient to account for a sharp decrease of the contact probability scaling exponent, accompanied by a significant loss of discernible structure in the average contact map (Figure S7A, "diffusion" panels). Progressive loss of structure is accompanied by the domains and compartments becoming noticeably fainter on the map and by the decay of the scaling exponent of the contact probability. In fact, an ensemble with RMSD of ~80 nm, or mere 8 nucleosome diameters, corresponds to a scaling exponent that is less than 50% of its original value in the undegraded ensemble.

By the time the nucleosomes have diffused by an average RMSD of ~200 nm, about 20 nucleosome diameters in length, the scaling exponent has fallen to less than one-fifth of its original value. At this stage, deterioration is visually apparent in example structures. As can be seen in the structure rendering, the non-mixed, sequential, color ordering according to genomic distance from blue to green to yellow to red that was strikingly evident in the native, globular structure, is lost at this point as the colors of the diffused

nucleosomes in the deteriorated sample have become mixed. This is suggestive of a rather dramatic loss of structure in chromatin and is consistent with the fact that the average contact map is, at this stage of diffusion, almost devoid of features.

With further increase of RMSD the contact map tends to a uniform, constant matrix, depleted of any of the interesting features of the original, native, undegraded ensemble. The scaling exponent of the contact probability tends to zero, in accordance with said uniformity and with the complete effacement of any short-range and long-range structure.

The analysis suggests that the relative decay of the scaling exponent can be used to assess the degree of diffusion in a nucleosome gas. The rapid character of this decay with respect to the average RMSD of the nucleosomes suggests that nucleosome diffusion over a distance of only a dozen nucleosome diameters or so is enough to produce significant loss of structure in Hi-C maps.

Estimates of diffusivity and viscosity of ancient woolly mammoth skin are in line with the behavior of an amorphous solid

The arguments in this paper provide a variety of bounds on the degree of diffusion in our sample. For instance, microscopy data suggests that a significant portion of the woolly mammoth DNA oligonucleotides must have diffused less than 5 microns (5 μm = 5×10^{-6} m, the diameter of a typical mammalian cell nucleus) in 52,000 years. We can use such facts to derive an upper bound on the diffusion coefficient for short DNA oligonucleotides in the sample, and thereby a lower bound for the sample's viscosity. The initial set of calculations will serve as an illustration. We will then construct a table of the various bounds that can be obtained based on different lines of evidence from the present study.

We begin by calculating an upper bound on the diffusion coefficient using the formula for mean squared distance:

$$\langle x^2 \rangle = 2nDt,$$

where

(1) D is the diffusivity,
(2) x is the distance traveled by the particle,
(3) n is the number of dimensions in which the particle is moving, and
(4) t is the time taken.

The formula can be rearranged to obtain:

$$D = \frac{\langle x^2 \rangle}{2nt}.$$

In our case, the diffusing particles are DNA oligonucleotides. Taking the following values:

(1) $x = 5$ μm $= 5 \times 10^{-6}$ m (the maximum distance that the oligos could have diffused),
(2) n = 3 (diffusion is occurring in three dimensions),
(3) $t = 52{,}000$ years $= 52{,}000 \times 365.25 \times 24 \times 60 \times 60$ s $\approx 1.64 \times 10^{12}$ s,

and substituting them into the equation for D, we get:

$$D \leq \frac{(5 \times 10^{-6}\ \mathrm{m})^2}{2 \times 3 \times 1.64 \times 10^{12}\ \mathrm{s}} = 2.54 \times 10^{-24}\ \mathrm{m^2/s}.$$

We can use the upper bound on diffusivity to get a lower bound on the viscosity of the sample, by using the Stokes-Einstein-Sutherland[162–164] equation:

$$D = \frac{k_b T}{6\pi\eta r},$$

where

(1) k_b is the Boltzmann constant 1.38×10^{-23} m^2 kg/($s^2 \cdot$K),
(2) T is the absolute temperature (in Kelvin), and
(3) r is the radius of the diffusing particle (in m).

This equation can be rearranged to get:

$$\eta = \frac{k_b T}{6\pi D r}.$$

The appropriate values are

(1) $D = 2.54 \times 10^{-24}$ m^2/s (the upper bound obtained above),
(2) $T = 273.15$ K (0°C, an upper estimate for the permafrost's temperature),

(3) $r = 0.5 \times 3.4 \times 10^{-10}$ m/bp × 100 bp = 1.7×10^{-8} m (assuming 100 bp oligonucleotides).

Substituting these values into the equation for η, we get:
$\eta \geq (1.38 \times 10^{-23}\ m^2\ kg/(s^2 \cdot K) \times 273.15\ K)/(6\pi \times 2.54 \times 10^{-24}\ m^2/s \times 1.7 \times 10^{-8}\ m)$
or
$\eta \geq 4.6 \times 10^9\ kg/(m \cdot s)$.

In the table below, we use various bounds on RMSD that can be derived from the presence of particular architectural features or from theoretical and simulation-based analyses of contact probability scaling in order to derive bounds on the diffusivity and viscosity in mammoth samples. We note that viscosities in excess of 10^{12} kg/(m·s) are associated with the behavior of amorphous, or glassy, solids.[165] Thus the presence of features such as point-to-point chromatin loops suggests glassy behavior in the woolly mammoth sample.

Diffusivity and viscosity estimates for woolly mammoth chromatin based on the preservation of various architectural features

Feature	RMSD (nm)	Diffusivity (m^2/s)	Viscosity (kg/m·s)
Nucleus	≤5000	$\leq 2.5 \times 10^{-24}$	$\geq 4.6 \times 10^{9}$
Chromosome territories	≤2000	$\leq 4.1 \times 10^{-25}$	$\geq 2.9 \times 10^{10}$
Barr body	≤1000	$\leq 1.0 \times 10^{-25}$	$\geq 1.2 \times 10^{11}$
Point-to-point loops	≤50	$\leq 2.5 \times 10^{-28}$	$\geq 4.6 \times 10^{13}$
Contact probability …			
… using simple diffusion model	≤50	$\leq 2.5 \times 10^{-28}$	$\geq 4.6 \times 10^{13}$
… using excluded volume model	≤20	$\leq 4.1 \times 10^{-29}$	$\geq 2.9 \times 10^{14}$

After drying, the 3D genome architecture of a tissue can remain stable for months, even at room temperature

The history of the sample IN18-032 is complex, including long-term burial in permafrost followed by a thawing process during sample retrieval when the permafrost was crumbled with pressurized water from a nearby waterway. To gain a better understanding of the factors that might have contributed to the preservation of 3D architecture in such samples, we performed Hi-C experiments on cow tissue to test whether drying, and in particular freeze-drying (lyophilization), has the capacity to conserve genome architecture, even at elevated temperatures.

The sequencing data from all experiments described below was aligned to a cattle genome bosTau9[88] using Juicer 2.0/BWA 0.7.17-r1188, run with the "-s MboI" flag.[77] The alignment and deduplication statistics for all libraries are listed in Table S4, #1.

The following architectural features were analyzed in the resulting bovine contact maps: chromosome territories, compartment segregation, and looping.

Juicer-generated contact maps were used to assess chromosome territories. Contact data for bovine chromosome #1 was used across all datasets to calculate the second-order Pearson correlation maps and eigenvector, both at 1 Mb, to assess compartment segregation. Finally, the presence of loops was assessed using APA. These were generated using a list of 7,821 conserved loops obtained by lifting over 9,448 human loops annotated in Rao et al.[23] to bosTau9, analogous to what was described in the loops are preserved from human to mammoth section.

Without any intervention, the 3D genome architecture degrades within days

First, we performed an experiment to establish how genome architecture degrades if left to its own devices. For the purpose, we obtained a beef liver sample 2 h postmortem (hpm). We used some of the sample to perform a "fresh sample control" *in situ* Hi-C experiment. The resulting contact maps displayed typical 3D chromosome morphology including chromosome territories, compartments and loops (see Figures 6D and S7B, top row).

We split the remaining liver into 9 chunks, each large enough to fill a 50 mL conical. (The cutting was done in a biosafety cabinet to reduce exposure to potential contaminants.) Each conical was then sealed with Parafilm and kept at room temperature, in the dark, for 12 h, 24 h, 36 h, 48 h, 60 h, 72 h, 84 h, 96 h, and 120 h, respectively. After incubation each tube was sampled twice (targeting different areas of the original tissue chunk), and the resulting material used as input for *in situ* Hi-C library preparation. After sampling the respective tube was discarded.

The results of this experiment are summarized in detail in Data S1 Figure M3. (One data point, corresponding to 96 h incubation, is also shown in Figures 6D and S7B, second row.) The following conclusions can be drawn from the experiment. 1) Loops disappear within 48h of degradation. 2) Compartments disappear at around the same time. 3) Contact patterns consistent with chromosome territories (bright squares along the diagonal, corresponding to each chromosome) tend to persist for longer, up until about 96h of room temperature incubation.

It is worth noting that analysis of contacts between nuclear and mitochondrial sequences (data not shown) suggest that the nucleus disintegrates prior to the disappearance of chromosome territory-like map features. One possible interpretation is that the bright, whole-chromosome squares stem from free-floating monochromosomal chromatin bodies (remnants of the original chromosome territories), i.e., chromatin chunks in which the majority of sequences originate on the same chromosome and which persist after the nucleus has been disrupted. The exact time for the disappearance of these whole-chromosome squares is hard to establish

with precision as we observed considerable variability between biological replicates for later incubation time points. Interestingly, DNA fragmentation does not appear to align well with the disappearance of the loops/compartments or chromosome territories (see Data S1, Figure M4). Instead, the average DNA length appears to be relatively unchanged up to about 60 h of incubation, after which it drops to mononucleosome-scale fragments.

We repeated the degradation experiment with another tissue, the beef cheek muscle (see Table S4, #1, and Data S1, Figure M5). The muscle tissue was left to degrade for 2 weeks (336 h). The results were largely consistent with those observed in the liver degradation experiment, with somewhat delayed decay for all observed architectural features. For example, it took at least 60 h of incubation for beef cheek loops to disappear. The compartment pattern again disappeared at around the same time. Surprisingly, the whole-chromosome squares could be observed in some timepoints and some samples, until about 300 h. Examination of the samples suggested that parts of the muscle have air-dried, which could explain the added resilience. It is worth noting that this is despite the fact that after ~200 h we started seeing considerable bacterial contamination in the sequencing data (percent of unaligned reads consistently above 25%, with some libraries showing as high as 90% of sequences not aligning to the cow reference genome assembly, see Table S4, #1), despite our efforts to reduce contamination at the beginning of the experiment by starting with vacuum-sealed beef and splitting the sample in a biosafety cabinet. Without such precautions a typical tissue sample is lost to contamination within 68 h, as shown in Data S1, Figure M6A.

Drying stabilizes 3D genome architecture

We then tested if drying helps stabilize 3D genome architecture. We dried the 2 hpm liver sample described above in a PRESTO DehydroTM electric food dehydrator for 22 h. (See Video S1 for a visual representation of the sample's appearance after drying.) A Hi-C experiment performed immediately afterward showed little to no difference in genome architecture in this hot-air dried liver sample when compared to the original sample (Figures 6D and S7B; Data S1, Figure M7), showing chromosome territories, compartments and loops. The results were unchanged even after the dehydrated sample was stored for one month at room temperature (Data S1, Figure M7). As noted above, chromatin architecture in a sample that had not been dehydrated was fully degraded by this time (Figures 6D and S7B; Data S1, Figure M3).

To assess the robustness of dehydrated samples at even longer time scales, we performed an experiment on a store-bought freeze-dried raw liver sample (see Video S1). Similar to the one month-old sample, the freeze-dried sample's contact map contained chromosome territories, compartments and loops at all tested time points, even after 1 year of storage at room temperature (Figures 6D and S7B, Data S1 Figure M7).

We performed similar experiments in two other bovine tissue samples, sirloin and heart.

We dried a sirloin sample for up to 6 h in a toaster oven and compared the resulting 3D genome architecture with that of a fresh sirloin sample. (Refer to Video S1 for a visual representation of the sample's appearance after drying.) The resulting contact maps and architectural features were virtually identical. The same results were observed when dehydrating the sample in a dehydrator (Data S1 Figure M8).

Finally, we performed experiments in several store-bought dried beef heart samples, one hot-air dried (Beef Heart Jerky, single ingredient dog treats from Natural Rapport, moisture content 12%, experiment performed 5 months after packaging date) and two freeze-dried heart samples, (Stella&Chewy's Freeze-dried Raw Beef Heart Treats, moisture content 8%, experiment performed 5 months after packaging date; and smallbatch Pets Freeze-Dried Beef Heart treats for dogs and cats, moisture content 7%, experiment performed 8 months after the listed packaging date), see Data S1 Figure M6B. The results were compared to those obtained in a fresh heart sample (Data S1, Figure M6A). Hi-C data across all experiments showed comparable architectural features, with chromosome territories, compartment segregation, and loops.

Genome architecture is extremely resilient in modern dehydrated samples

After confirming that genome architecture in dehydrated samples is shelf-stable for extended periods, even at room temperature, we proceeded to investigate whether dried chromatin exhibits robustness against the types of mechanical, thermal, and other disturbances that ancient dehydrated samples might plausibly encounter through millennia. We included the following perturbations: run over by a car, hit with a fastball, and pulverized with a shotgun, for varying degrees of mechanical impact; baking in a toaster oven for 1 h, microwaving for 6 min, dropping into liquid nitrogen, for thermal disturbance; and soaking, in either plain water or with added lemon juice, for 1 h, to assess the effects of water damage under neutral or acidic conditions (see Video S1). None of the perturbations resulted in the destruction of architectural features, see Figures 6D and S7B; Data S1, Figure M9.

Additional notes

We note that several store-bought dehydrated samples did not display the architectural features as expected and were in fact partially or completely degraded (Data S1, Figure M10). We hypothesize that in these cases chromatin features were already absent at the time of dehydration. As such, sample dehydration is a necessary, but not a sufficient condition for a successful PaleoHi-C experiment. Nevertheless, the results in beef tissue broadly indicate that the 3D genome architecture can be stabilized and preserved by dehydration, whether via hot-air or freeze-drying. In view of these experiments, we hypothesize that the mammoth sample underwent a glass transition that was induced by spontaneous freeze-drying of the woolly mammoth tissue, shortly after death, in the Pleistocene Siberian climate.

The physics behind tissue dehydration has been widely studied by food scientists. Food science attributes the prolonged shelf life of dehydrated foods to glass transition, or vitrification, a process in which the viscosity of a sample is raised to levels that effectively arrest diffusion. There are several ways to achieve vitrification (see Data S1, Figure M11, adapted from Devireddy et al.[166]). One

approach is to rapidly cool the sample to below the glass transition temperature (T_g), e.g., by plunging the sample into liquid nitrogen. This approach is used, for example, for cryoprotectant-free spermatozoa preservation.[167] Another approach is desiccation at temperatures above freezing. Anhydrobiosis, based on this type of desiccation, is a strategy for survival adopted by numerous organisms across the tree of life.[67] As the water evaporates, the remaining solution becomes more and more concentrated until it reaches its vitrification point. The glass transition temperatures of desiccated substances frequently fall within the range of ambient temperatures.

Freeze-drying is a process whereby excess water is removed by sublimation at sub-zero temperatures. In this process, the point of vitrification is reached at subzero temperatures. However, as the sample continues to desiccate, T_g can increase substantially. In the food industry this results in products that are stable at room temperature for long periods of time.[61] In the case of the mammoth, it may have contributed to making the sample robust against thawing during its extraction from permafrost. Note that gradual freezing, as might be expected when a sample spontaneously freeze-dries, in-and-of-itself contributes to desiccation: as the ice crystals initially form in the free water present in the extracellular space, they create an osmotic gradient that draws water out of the cells. Repeated freezing whereby the resulting interstitial ice crystals thaw and drain out of the tissue, and then the process of freezing-out of cellular water repeats when the temperatures drop below zero once again, is a known mechanism of natural vitrification, alongside spontaneous freeze-drying.[168] Freeze-thaw mummies however are typically associated with substantial tissue damage including hair loss. As such, intact hair on the mammoth skin sample described in this paper appears to favor spontaneous freeze-drying rather than repeated freezing as the primary mechanism behind the glass transition. Future research may help clarify whether vitrification through repeated freeze-thaw is compatible with PaleoHi-C.

While the results observed in modern dried samples are encouraging and suggest that vitrified chromatin is extremely robust, it must be noted that the tested samples are an imperfect model for ancient samples because, among other factors, the DNA in these samples is considerably less fragmented than ancient DNA. For example, the freeze-dried liver sample had an average fragment size of 7,200 bp; the freeze-dried heart sample was ~2,850 bp and the heart jerky was ~4,850 bp (DNA extracted using Qiagen DNeasy Blood and Tissue kit (69504); size distribution measured using Agilent 2100 Bioanalyzer, DNA 12000 assay). Longer DNA fragments are not free to diffuse, even when the sample is not in a glassy state, due to entanglement,[169] which may contribute to sample stability in ways that may not apply to ancient samples.

3D modeling of ancient chromosomes

In this section, we describe the approaches we used for modeling the 3D structural ensembles of woolly mammoth chromosomes.

Maximum entropy approach

Energy landscape theory paired with the maximum entropy approach to derive the effective energy landscape from experimental Hi-C data has proved useful for understanding genome folding in contemporary species.[63,170] Here we explore the application of the maximum entropy approach for modeling the 3D structural ensembles of woolly mammoth chromosomes based on PaleoHi-C data. We employ two distinct modeling methods: the Direct Inversion method, which focuses on a detailed reconstruction using every individual contact in the input dataset, and the Minimal Chromatin Model (MiChroM) which leverages PaleoHi-C compartment annotations to infer chromosomal architecture.[64]

There is an important difference between the two approaches. The Direct Inversion method models chromosomes as they are at the moment when the Hi-C experiment is conducted, i.e., in our case, the conformation present 52,000 years after the mammoth died. As such, this modeling approach will reflect aspects of the sample's present condition, e.g., the consequences of ancient DNA diffusion. MiChroM, on the other hand, is less susceptible to the influence of diffusion because it makes explicit assumptions about chromatin compartment interactions in living nuclei. (Simplifying somewhat, this is because, although diffusion can create contacts that are not ancient, i.e., contacts that were not present in the mammoth at the time of its death, diffusion is less likely to affect the compartment assignment of an entire locus, which reflects a large number of contacts.) As such, the MiChrom approach is oriented toward reconstructing the ancient conformation. We explore the convergence of these two methods in the comparison of direct inversion and MiChroM structural ensembles section. For now, we will simply point out that, because the simulation bead size we employed, 250 kb, was selected to be far larger than the scale of the diffusion in the sample, the differences between the ancient and modern conformations at this scale should be minimal.

Building an energy function. A simple way to think about a chromosome is as a homopolymer, which can be described using a bead-spring model. This model, however, will not reproduce the experimental Hi-C contact frequencies. In order to create a least-biased polymer model that would reproduce the experimentally determined contact frequencies, we add biases to a homopolymer model using a maximum entropy approach.

For the homopolymer potential $U_{HP}(\vec{r})$ and for a generic phase space observable function $\Phi(\vec{r})$, a simulation in the canonical ensemble yields the following value:

$$\Phi_{HP} = \langle \Phi(\vec{r}) \rangle_{HP} = \frac{\int \Phi(\vec{r}) e^{-\beta U_{HP}(\vec{r})} d\vec{r}}{\int e^{-\beta U_{HP}(\vec{r})} d\vec{r}} = \int \Phi(\vec{r}) \pi^{HP}(\vec{r}) d\vec{r}, \quad \text{(Equation 1)}$$

where $\boldsymbol{r}$ is the vector of positions (Cartesian space) of all beads in the polymeric chain, $\pi^{HP}(\vec{r})$ is the probability density of the homopolymer model, and $\beta = 1/k_bT$. As mentioned, we do not expect that the homopolymer sampling Φ_{HP} will match the experimental values Φ^{exp}.

Let us use $\pi^{ME}(\vec{r})$ to denote the probability density that matches the following constraints:

$$\begin{aligned} c_0 &= \int \pi^{ME}(\vec{r})d\vec{r} - 1 \\ c_1 &= \int U_{HP}(\vec{r})d\vec{r} - \frac{3}{2}Nk_bT \\ c_i^{DATA} &= \int \Phi(\vec{r})\pi^{ME}(\vec{r})d\vec{r} - \Phi_i^{exp}, \quad i = 1, \ldots, n. \end{aligned} \tag{Equation 2}$$

All constraint equations must be equal to zero. The first constraint c_0 ensures that probability density $\pi^{ME}(\vec{r})$ is normalized, i.e., the summation of the probabilities must be equal to one. The second constraint c_1 determines the average potential energy equal to the thermal energy $3Nk_bT/2$. The last set of constraints is c_i^{DATA} where i can be related to n observables, and the equations ensure that the theoretical and experimental values coincide.

To determine the probability density $\pi^{ME}(\vec{r})$ (that now includes the experimental biases), we maximize the information entropy

$$S = -\int \pi^{ME}(\vec{r})\ln\left(\pi^{ME}(\vec{r})\right)d\vec{r} \tag{Equation 3}$$

subject to the constraints. The constrained maximization of the entropy is equivalent to minimizing the amount of additional information between the distributions with and without the observable biases which leads to the extremum condition:

$$\frac{\partial S}{\partial \pi^{ME}} - \lambda_0\frac{\partial c_0}{\partial \pi^{ME}} - \lambda_1\frac{\partial c_1}{\partial \pi^{ME}} - \sum_{i=1}^{n}\lambda_i^{DATA}\frac{\partial c_i^{DATA}}{\partial \pi^{ME}} = 0, \tag{Equation 4}$$

where λs are the Lagrange multipliers. Rearrangement of the expressions leads to the probability distribution

$$\pi^{ME}(\vec{r}) = \frac{e^{-\lambda_1 U_{HP}(\vec{r}) - \sum_{i=1}^{n}\lambda_i^{DATA}\Phi(\vec{r})}}{\int d\vec{r}\, e^{-\lambda_1 U_{HP}(\vec{r}) - \sum_{i=1}^{n}\lambda_i^{DATA}\Phi(\vec{r})}}. \tag{Equation 5}$$

Observing that λ_1 coincides with β, we can think of the maximum entropy probability distribution $\pi^{ME}(\vec{r})$ as the distribution sampled from the maximum entropy potential energy:

$$U_{ME}(\vec{r}) = U_{HP}(\vec{r}) + \frac{1}{\beta}\sum_{i=1}^{n}\lambda_i^{DATA}\Phi_i(\vec{r}). \tag{Equation 6}$$

Equation 6 has two terms. The first is related to the homopolymer chain where the bead interactions are non-specific. The second term is associated with the information provided by the experimental data.

Now that we have a general form for the potential energy we still need a procedure to calculate the λ coefficients. This requires an optimization procedure that minimizes the difference between the simulation ensemble observables and experimental data.

Crosslinking probability function. Following Di Pierro et al.,[64] we are going to postulate that the probability of observing a contact between a pair of loci i and j in a Hi-C experiment $\Phi(i, j)$ corresponds to a simulation observable that can be calculated using the following scalar function:

$$f(r_{i,j}) = \frac{1}{2}(1 + \tanh[\mu(r_c - r_{i,j})]). \tag{Equation 7}$$

This function assumes that the likelihood of a locus pair i and j forming a contact in 3D space decreases as a function of distance between loci $r_{i,j}$, and it is close to one when the distance is small compared to a characteristic contact distance r_c, and to zero when the distance is large. Similar to Di Pierro et al.,[64] the parameters in 7 were calibrated based on near-diagonal interactions in the PaleoHi-C experiment and set to $\mu = 5.32\sigma^{-1}$ and $r_c = 2.31\sigma$, where σ is the diameter of the polymer bead.

Homopolymer potential. The potential energy $U_{HP}(\vec{r})$ describes a spatially self-avoiding polymer. Following Zhang et al.,[63] we use this formula for the homopolymer potential:

$$U_{HP}(\vec{r}) = \sum_{i\in\{Loci\}}U_{FENE}(r_{i,i+1}) + \sum_{i\in\{Loci\}}U_{HC}(r_{i,i+1}) + \sum_{i\in\{Angles\}}U_{Angle}(\theta_i) + \sum_{\substack{i,j\in\{Loci\}\\ j>i+2}}U_{SC}(r_{i,j}). \tag{Equation 8}$$

The potential consists of the following four terms, U_{FENE}, U_{HC}, U_{Angle} and U_{SC}.

U_{FENE} (Finite Extensible Nonlinear Elastic potential) is the bonding term applied between two consecutive monomers, connecting a sequence of beads with nonlinear springs:

$$U_{FENE}(r_{i,j}) = \begin{cases} -\frac{1}{2}K_B R_0^2 \ln\left[1 - \left(\frac{r_{i,j}}{R_0}\right)^2\right], & r_{i,j} \le R_0 \\ 0, & r_{i,j} > R_0 \end{cases}, \quad \text{(Equation 9)}$$

where K_B is the spring constant and R_0 is the equilibrium bond length.

The hard-core repulsive potential U_{HC} is included to avoid overlap between bonded monomers:

$$U_{HC}(r_{i,j}) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r_{i,j}}\right)^{12} - \left(\frac{\sigma}{r_{i,j}}\right)^{6} + \frac{1}{4}\right], & r_{i,j} \le \sigma \times 2^{\frac{1}{6}} \\ 0, & r_{i,j} > \sigma \times 2^{\frac{1}{6}} \end{cases}, \quad \text{(Equation 10)}$$

where $\varepsilon = k_b T$ is the energy scale-factor.

The angular potential that regulates the chain flexibility is given by:

$$U_{Angle}(\theta_i) = K_A[1 - \cos(\theta_i - \theta_0)], \quad \text{(Equation 11)}$$

where θ_i is the angle defined by two vectors $\boldsymbol{r}_{i,i+1}$ and $\boldsymbol{r}_{i,i-1}$, $\theta_0 = \pi$ and K_A is a bending coefficient. This three-body interaction is included between all connected three consecutive monomers to model a nonzero bending stiffness.

All non-bonded pair interactions are described by a soft-core repulsive potential of the following form:

$$U_{SC}(r_{i,j}) = \begin{cases} \frac{1}{2}E_{cut}\left[1 + \tanh\left(\frac{2U_{LJ}(r_{i,j})}{E_{cut}} - 1\right)\right], & r_{i,j} < r_0 \\ U_{LJ}(r_{i,j}), & r_0 \le r_{i,j} \le \sigma \times 2^{\frac{1}{6}} \\ 0, & r_{i,j} > \sigma \times 2^{\frac{1}{6}} \end{cases}. \quad \text{(Equation 12)}$$

The expression U_{LJ} corresponds to the Lennard-Jones potential

$$U_{LJ}(r_{i,j}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{i,j}}\right)^{12} - \left(\frac{\sigma}{r_{i,j}}\right)^{6} + \frac{1}{4}\right] \quad \text{(Equation 13)}$$

capped off at short distances, allowing for chain crossing at finite energetic cost. r_0 is chosen as the distance at which $U_{LJ}(r_0) = 1/2E_{cut}$.

The simulations were performed using reduced units with energy (ε) and length (σ). The model parameters and potential constants are described below:

$$\varepsilon = k_b T,\ \sigma = 1,\ K_A = 2\varepsilon,\ K_B = \frac{30\varepsilon}{\sigma^2},\ E_{cut} = 4\varepsilon,\ R_0 = 1.5\sigma.$$

Direct inversion approach

There are several different ways in which one can incorporate information provided by the Hi-C data into the maximum entropy potential energy. In the Direct Inversion approach, the observable signal is the contact frequency between all pairs of loci along the chromosome. (Simply put, the higher the contact frequency for a locus pair, the stronger the interaction is expected between the corresponding beads in the polymer model.) The parameters in Equation 6 are calculated by minimizing the difference between the simulation data and experimental results.

In order to simulate woolly mammoth chr10 using this approach, we used PaleoHi-C contact data at 250-kb resolution. We employed the OpenMiChroM chromatin dynamics software's minimization protocol to calculate the lambdas.[65] The energy function calculated with the optimized lambdas was then used to generate an ensemble of 200,000 3D chromosomal structures using OpenMiChroM.[65] The structures can be explored interactively using the Spacewalk browser (https://t.3dg.io/3d-mammoth-Fig-7A). An *in silico* contact map generated for the 200,000 structures in the ensemble closely matches the PaleoHi-C experiments (see Figures 7A and S8A).

Minimal Chromatin Model (MiChroM)

Another approach for incorporating experimental data to build the potential energy function is known as the Minimal Chromatin Model (MiChroM), which has been successfully used to study the organization of genomes in different organisms.[64,65,171,172] Unlike Direct Inversion, which makes no prior assumption about the nature of interactions between individual beads, MiChroM makes several assumptions about the process of chromatin folding. Most importantly, it assumes that chromatin can be classified into a handful of different chromatin types, typically referred to as compartments and subcompartments. In the simplest implementation

this is just two compartments, or types, namely open (A) and closed (B) chromatin. When two segments of chromatin come into contact, the effective free energy change due to this contact depends, to first order, only on the chromatin type identity of each segment. MiChroM also assumes that every time a pair of loci comes into contact, there is a gain/loss of effective free energy that depends only on the genomic distance. This "ideal chromosome" term models the local structure of chromatin in the absence of A/B compartmentalization.

The Direct Inversion model tries to model the exact interactions observed in an experiment. This makes it potentially prone to error when the signal to noise ratio is poor in the observable function that one tries to match, or to yielding different results when the observables may be affected by non-biological processes such as postmortem chromatin diffusion. Because it makes explicit assumptions about the nature of chromatin folding, MiChroM has a potential advantage in such cases on account of the fact that it can aggregate interactions when approximating the energy function. It has a much lower number of free parameters, and so the training and calibration of the MiChroM model is also more computationally scalable than that for Direct Inversion. For these same reasons MiChroM is especially suitable for reconstructing the ancient conformation for many chromosomes.

The set of constraints generated by all the observables in the MiChroM model is as follows[64]:

$$
\begin{aligned}
c_0 &= \int \pi^{MiChroM}(\vec{r})d\vec{r} - 1 \\
c_1 &= \int U_{HP}(\vec{r})d\vec{r} - \frac{3}{2}Nk_bT \\
c_T^{kl} &= \int T_{kl}(\vec{r})\pi^{MiChroM}(\vec{r})d\vec{r} - T_{kl}^{exp}, \quad \forall k, l \in Set\ of\ Types: l \geq k \\
c_G^d &= \int G(\vec{r})\pi^{MiChroM}(\vec{r})d\vec{r} - G_{kl}^{exp}, \quad 3 \leq d \leq d_{cutoff}
\end{aligned}
\qquad \text{(Equation 14)}
$$

Compared to (2) we reformulate the observables. First, we constrain the average number of contacts happening between chromatin types (c_T). Second, we constrain the number of contacts that happen at a certain fixed genomic distance d along the chromosome (c_G).

These constraints define the MiChroM probability distribution, and the information theoretic energy function is accordingly[64]:

$$
U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k \geq l \\ k,l \in Types}} \alpha_{kl} \sum_{\substack{i \in \{Loci\ of\ Type\ k\} \\ j \in \{Loci\ of\ Type\ l\}}} f(r_{i,j}) + \sum_{d=3}^{d_{cutoff}} \gamma(d) \sum_{i \in \{Loci\}} f(r_{i,i+d}), \qquad \text{(Equation 15)}
$$

where $U_{HP}(\boldsymbol{r})$ is the potential energy of a generic homopolymer; α_{kl} are parameters that govern interactions between chromatin types k and l; $f(r_{i,j})$ represents the probability of crosslinking of loci i and j; and $\gamma(d)$ is the ideal chromosome energy interaction as a function of the genomic distance d. Similar to the Direct Inversion approach, αs and γs need to be determined by minimizing the difference between simulation and experimental results.

As in the Direct Inversion case, we performed this minimization for PaleoHi-C data for mammoth chromosome 10 at 250-kb resolution using OpenMiChroM chromatin dynamics software's minimization protocol[65] (see Table S4, #2). Taking advantage of the fact that the MiChroM energy function approximated using one chromosome can be applied to all other chromosomes of the same organism,[64,172] we did not limit ourselves to simulating just one chromosome, but did the simulations for all 28 mammoth chromosomes to produce an ensemble of 5,600,000 structures (200,000 structures for each chromosome). (The first eigenvector was used to assign the chromatin A/B types for all chromosomes except chromosome 8. A custom script was used to extract the second eigenvector for chromosome 8 following the procedure described in Lieberman-Aiden et al.[17]) The structures can be explored interactively using the Spacewalk browser (https://t.3dg.io/3d-mammoth-Fig-S8B). Similar to the Full Inversion case, the contact maps generated from the simulated ensembles closely resemble experimental data (see Figures S8B and S8E).

Comparison of direct inversion and MiChroM structural ensembles

We compare the two modeling approaches, Direct Inversion and MiChroM, by comparing the 3D structure ensembles generated for chromosome 10 by both methods. Critically, given that the monomers employed in both methods are 250 kb, far larger than the scale of diffusion in the sample, the two ensembles - representing the modern state of the sample vs. its ancient state, respectively - ought to be extremely similar. We include a homopolymer model (Equation 8) of the same chromosome in this comparison as a baseline.

The comparison shows that the ensembles resulting from Direct Inversion and the MiChroM approaches differ from the homopolymer ensembles and are similar to each other, demonstrating the robustness of the two methods. We further expand the comparison between the two methods by calculating the fold enrichment of contacts between A-A, B-B and fold depletion of A-B compartment interactions compared to those in the homopolymer model for the Direct Inversion structures and the MiChroM structures. The resulting contact enrichments, which can be thought of as a measure of compartment segregation, are very similar despite the fact that the Direct Inversion does not explicitly assume the existence of chromatin types. This further strengthens the confidence in the results from the two modeling methods.

Q-metric similarity for chromosomal ensemble comparison. The Q-metric is a widely used measure for quantifying the structural similarity between two polymer chains.[173] Usually it is employed as a protein folding reaction coordinate, giving a quantitative measure of conformational differences between any given structure and its final folded state as the reaction proceeds through a funneled energy landscape.[174] Following Cheng et al.,[175] we extend the application of the Q-metric and use it to quantify conformational differences between any two structures, not just between a structure and its native folded state.

We thus define the Q-value between two structures α and β (see Figure S8C) as:

$$Q^{\alpha,\beta} = \frac{1}{N}\sum_{i<j} exp\left(\frac{-\left(r_{ij}^{(\alpha)} - r_{ij}^{(\beta)}\right)^2}{2\rho^2}\right), \quad \text{(Equation 16)}$$

where $r_{ij}^{(\alpha)}$ and $r_{ij}^{(\beta)}$ are the distances between chromatin loci i and j in structures α and β, respectively, N denotes the number of loci pairs being considered in the summation, and ρ is a scaling factor that helps tune the sensitivity of the Q-value to structural differences of relevant scale, not too small so that all structures look different on account of minor shifts of individual beads, and not too large so that all structures look the same on account of being viewed from a "bird's eye" perspective. A smaller Q-value ($Q \simeq 0$) corresponds to highly dissimilar structures. Two identical structures would give a value of $Q = 1$.

In order to compare the Direct Inversion and the MiChroM ensembles we randomly selected 1,000 structures from each of the ensembles and calculated the pairwise Q-values for all structures in this pool of 2,000 structures, both for structure pairs from the same ensemble, and for pairs across the two ensembles. The calculation was done using the CNDBTools analysis library implemented as part of the OpenMiChroM chromatin dynamics package.[64,65,171] We also generated 200,000 structures using the Homopolymer model (Equation 8), randomly selected 1,000 homopolymer structures, and calculated the pairwise Q-values for all pairs within the ensemble and between the Homopolymer, the Direct Inversion, and the MiChroM 3D structures.

The results of the analysis are summarized in Figure S8C, right panel. The figure is a 3,000 x 3,000 similarity matrix, with the first 1,000 rows and columns corresponding to the 1,000 randomly selected MiChroM structures included in the comparison, the second 1,000 rows and columns to those from the Direct Inversion model, and the last 1,000 rows and columns to the homopolymer model. The darker colors correspond to higher pairwise Q-values, i.e., more similar structures. The close resemblance between structures from the MiChroM and Direct Inversion models, as compared to those modeled as a homopolymer, is readily apparent.

The distribution of values within submatrices of interest, specifically, Q-values calculated between all Homopolymer structures (HP), those between Homopolymer and Direct Inversion (HP x Direct), Homopolymer and MiChroM (HP x MiChroM), MiChroM and Direct Inversion (MiChroM x Direct), intra-MiChroM and intra-Direct Inversion ensemble Q-values (MiChroM and Direct Inv., respectively), are plotted to the right from the similarity matrix. As expected, the Homopolymer ensemble demonstrates the highest degree of intra-ensemble structural variability of the three models, while the Direct Inversion has the least amount. This is because all interactions in the Homopolymer are non-specific. On the other hand, the Direct Inversion has the most constraints of the three models. Consistent with the similarity matrix on the left, the distributions demonstrate a high degree of similarity between the MiChroM and Full Inversion structures, with Q-values between the two ensembles comparable to those calculated for pairs of MiChroM structures.

Quantifying A/B compartment segregation in direct inversion and MiChroM models. We further extend the scope of comparison between the Direct Inversion and the MiChroM modeling approaches by calculating the extent of A/B compartment segregation in structures from the corresponding ensembles.

First, we used the Hi-C eigenvector to assign beads to either A (open) or B (closed) type for the 3,000 randomly chosen structures, i.e., 1,000 from each of the three ensembles analyzed in the previous section, the Direct Inversion, MiChroM and Homopolymer ensembles. We computed the 3D Voronoi tessellation[176] for each structure using the Freud analysis library.[177] In this tessellation, each bead of the polymer chain serves as a generating point for a Voronoi cell, and the cell is assigned the type of the bead, i.e., type A or type B (Figure S8D). We then calculate the number of cell interfaces of various kinds, A-to-A, B-to-B, and A-to-B for all structures from the three ensembles, while excluding the immediate 1D chromosomal neighborhood from the calculation (i.e., ignoring interfaces that correspond to beads i and j such that $|i-j|<3$). We then convert the MiChroM and Direct Inversion counts to fold enrichment values with respect to counts calculated for the Homopolymer model.

The resulting fold enrichment values are listed in Figure S8D. The values are very similar between the two models, with both showing that beads of the same type are more likely to come in contact than those of two different types (i.e., the structures are consistent with the notion of compartment segregation), even though the Direct Inversion model is agnostic to the concept of chromatin type. The consistency between the results generated by the Direct Inversion and MiChroM models supports the robustness of chromatin modeling of ancient chromosomes from PaleoHi-C data.

3D genome architecture persists in Yuka

The Yuka PaleoHi-C sequencing data was aligned to Loxafr3.0_HiC and deduplicated using the Juicer Hi-C processing pipeline, Juicer version 1.6; BWA 0.7.17-r1188; 8 threads; openjdk version "1.8.0_222-ea"; Juicer Tools Version 1.9.9.[77] The pipeline was run with a "-s none" flag to account for non-enzymatic fragmentation of the DNA in the sample. The statistics associated with the Juicer run can be found in Table S4, #3.

The resulting map was too shallow to perform all the analysis presented for IN18-032. However, we were able to perform several analyses. First, just as with IN18-032, we confirmed that Yuka PaleoHi-C library derives from authentic ancient mammoth DNA.

Replicating the analysis described in PaleoHi-C read pairs aligning to elephant derive from ancient DNA, we established that Yuka PaleoHi-C reads reflected ancient DNA using PMDTools.[30] Damage (~7.3% reads) was significantly elevated in Yuka PaleoHi-C mammoth library as compared to modern elephant libraries (Table S4, #4), and was comparable to the damage rates calculated for aDNA-Seq reads from the same individual in Maschenko et al.[83] We also repeated the SNP analysis described in PaleoHi-C read pairs aligning to elephant derive from mammoth DNA. The Yuka PaleoHi-C reads overwhelmingly matched mammoth rather than elephant (93.3% vs. 6.7%, Table S4, #4).

Next, we examined the relationship between 1D distance along the contour of the elephant chromosome, *s*, and the contact probability *p*(*s*) (see Figure 6C). A power law almost identical to that seen in modern elephant *in situ* Hi-C libraries was visible. Finally, we used Aggregate Chromosome Analysis (ACA) to assess the presence or absence of chromosome territories in the data. ACA involves scaling the chromosomes to a uniform length and aggregating the signal of all intra- and interchromosomal contacts.[37] The resulting map (Figure S6G) looks like a typical Type II ACA map, consistent with the persistence of chromosome territories in the Yuka sample.

Theoretical predictions suggest that DNA architecture in chromoglass could survive for far longer than the oldest known ancient DNA samples

In earlier sections, we derived empirical bounds on diffusion and viscosity for the woolly mammoth sample and concluded that these values were consistent with a glassy state. In this section, we assume, based on the above analysis, that ancient DNA samples could be preserved indefinitely in a glassy state. We use this assumption, combined with analytic estimates for glassy behavior in general, to estimate the viscosity of such samples purely analytically, as well as to estimate the diffusion coefficient for 100 bp oligonucleotides. This allows us to calculate the length of time for which DNA architecture could remain intact in chromoglass.

Of course, as we shall see, our analysis suggests that DNA architecture could survive for an extraordinarily long time - longer than the oldest samples of ancient DNA that have been sequenced to date. As such, implicit in the below is the strong assumption that other factors, such as hydrolysis, oxidation, and bacterial digestion, have not degraded the DNA to the point that a discussion of its architecture is irrelevant. Put simply, when preserved in the glassy state, the architecture of DNA oligonucleotides is likely to survive for nearly as long as the oligonucleotides themselves.

Our strategy will be to begin with the diffusivity of water in amorphous ice, and introduce several corrections, to adjust for the viscosity of the nucleoplasm, the age of ancient glasses, and the relative size of DNA as compared to water molecules.

We begin with the observation that water diffuses on the scale of hours in amorphous ice, with a diffusivity of roughly 0.1 Å^2/hour.[178]

$$D_{\text{water,glassy-ice}} = 0.1\ \text{Å}^2/\text{hour} \times \left(10^{-10}\text{m}/\text{Å}^2\right)^2 \times (1\ \text{hour}/3600\ \text{s}) = 2.8 \times 10^{-25}\text{m}^2/\text{s}.$$

Next, we adjust for the fact that the nucleoplasm is ~100 times more viscous than water,[179] yielding:

$$D_{\text{water,nucleoplasm}} = D_{\text{water,glassy-ice}}/100 = 2.8 \times 10^{-27}\text{m}^2/\text{s}.$$

Although the glass transition itself is associated with a massive decline in diffusivity, diffusion coefficients continue to decline in glassy materials as they age. We therefore adjust for reductions in diffusivity due to the age of the glass, using a factor of 1000 as a conservative estimate.[180]

$$D_{\text{water,ancient-nucleoplasm}} = D_{\text{water,nucleoplasm}}/1000 = 2.8 \times 10^{-30}\text{m}^2/\text{s}.$$

Finally, we adjust for the size of a 100 bp oligonucleotide (34 nm) relative to the size of a water molecule (0.34 nm). The diffusivity varies inversely with size, so:

$$D_{\text{DNA,ancient-nucleoplasm}} = D_{\text{water,ancient-nucleoplasm}} \times (0.3\ \text{nm}/34\ \text{nm}),\ \text{or}$$

$$D_{\text{DNA,ancient-nucleoplasm}} = 2.5 \times 10^{-32}\ \text{m}^2/\text{s}$$

Using the Stokes-Einstein-Sutherland equation, we can calculate the corresponding viscosity:

$$\eta_{\text{DNA,ancient-nucleoplasm}} = k_b T/6\pi D_{\text{DNA,ancient-nucleoplasm}} r$$

$$= \left(1.38 \times 10^{-23}\text{J/K} \times 273.15\ \text{K}\right)/\left(6\pi \times 2.5 \times 10^{-32}\ \text{m}^2/\text{s} \times 1.7 \times 10^{-8}\ \text{m}\right),\ \text{or}$$

$$\eta_{\text{DNA,ancient-nucleoplasm}} \cong 5 \times 10^{17}\ \text{kg}/(\text{m}\cdot\text{s})$$

As expected, the above values are consistent with glassy dynamics and with the bounds calculated in the preceding section.

The value obtained for $D_{DNA,ancient\text{-}nucleoplasm}$ allows us to predict the absolute displacement for 100 bp oligos in ancient samples, such as our 52,000 year old woolly mammoth sample, by taking the square root of the mean squared displacement:

$$x_{mammoth} = \sqrt{2 \times 3 \times D_{DNA,ancient\text{-}nucleoplasm} \times t_{mammoth}}$$

$$= \sqrt{\left(2 \times 3 \times 2.5 \times 10^{-32}\ m^2/s \times 52,000 \times 365.25 \times 24 \times 60 \times 60\ s\right)}$$

$$\cong 5 \times 10^{-10}\ m$$

or 5 Å.

This is a little larger than the size of a single DNA base along the double helix (3.4 Å). Interestingly, this small value suggests that other aspects of the sample's history, rather than its time as a glass in permafrost, are responsible for the degree of noise observed in the PaleoHi-C map as compared to a modern elephantid map. These include the period of time after the mammoth's death, and before it became freeze-dried, as well as the handling of the sample after it was unearthed.

Of course, we can also ask about the degree of diffusion to be expected in other samples. For example, roughly 66 million years ago, the KT extinction led to the disappearance of the dinosaurs. It is interesting to ask whether, given a sample of flesh containing intact ancient DNA from a *Tyrannosaurus rex* (one of the species that perished around the time of the KT extinction), the DNA architecture would be preserved. Of course, here we are assuming not only that the DNA is still present, but that the sample has been preserved in a glassy state, akin to that of our mammoth. Using the same equation for mean squared displacement, but with t = 66,000,000 years, we obtain:

$$x_{trex} = \sqrt{2 \times 3 \times D_{DNA,ancient\text{-}nucleoplasm} \times t_{trex}}$$

$$= \sqrt{2 \times 3 \times 2.5 \times 10^{-32} m^2/s \times 66,000,000 \times 365.25 \times 24 \times 60 \times 60s}$$

$$= 1.7 \times 10^{-8} m$$

or ≈ 20 nm.

This is about half the size of a 100 bp oligonucleotide (see above). As such, it seems likely that, assuming the DNA itself could survive the period between the Cretaceous–Tertiary extinction and the present, its architecture would likely also survive well enough to enable PaleoHi-C experiments.

More generally, we might ask, how long could 3D DNA architecture survive sufficiently well to enable PaleoHi-C experiments? Here we assume that an absolute displacement of $x^* \geq 50$ nm (>1% of the typical nuclear diameter) is incompatible with high-quality contact mapping experiments. (See the section incorporating excluded monomer volume into simulations of genome polymer fragmentation and diffusion suggests a smaller RMSD than the simple diffusion model for a discussion of how much diffusion a contact mapping experiment can tolerate.) Rearranging the equation for mean squared displacement, we obtain:

$$t^* = x^{*2} / (2nD)$$

$$= \left(5 \times 10^{-8} m\right)^2 / \left(2 \times 3 \times 2.5 \times 10^{-32} m^2/s\right)$$

$$\approx 1.7 \times 10^{16} s$$

or 530,000,000 years.

Thus, DNA architecture could survive for a very long time. 530 million years ago corresponds to the time of the Cambrian explosion which gave rise to nearly all modern animal phyla.[181] Interestingly, a sample might also travel quite far in that time. The first confirmed interstellar comet, 2I/Borisov, is moving at roughly 32 km/s.[182] At this rate, it could reach Proxima Centauri, the closest star outside the solar system, in 40,000 years, less than the age of our woolly mammoth sample; and have reached most of the Milky Way galaxy before t^* had elapsed. This is highly speculative, as it is unknown whether either DNA or chromoglass can survive in extraterrestrial environments, except insofar as the Philae space probe landed on comet 67P/Churyumov–Gerasimenko in 2014, and detected

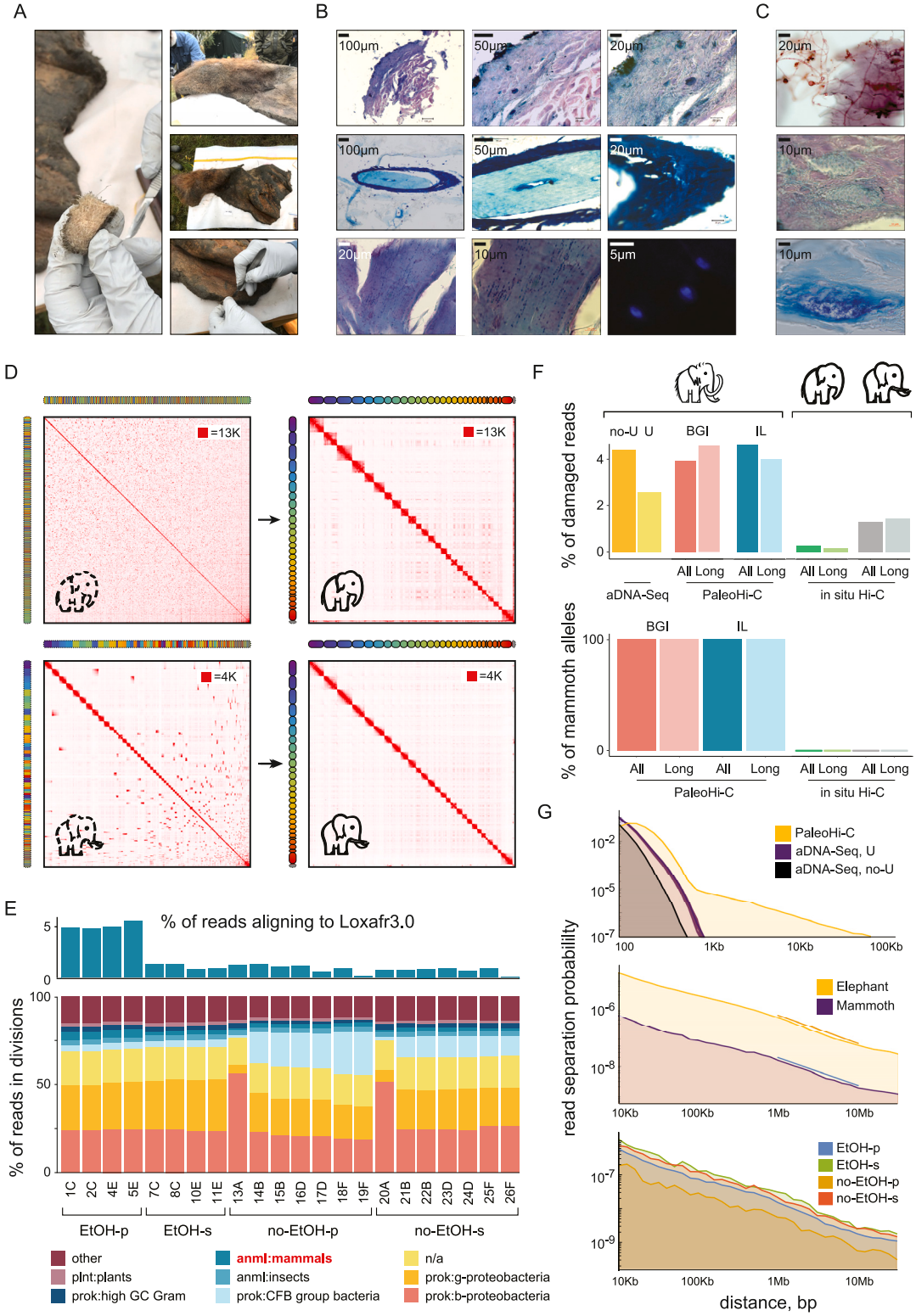
organic compounds there; see Bardyn et al.[183] Of course, t^* also greatly exceeds the age of the oldest successful ancient DNA sequencing experiments to date, and linear DNA molecules are not known to survive for such lengths of time.

Nevertheless, ancient DNA studies have demonstrated that, in an appropriate context, DNA can survive for at least 2 million years.[75] The above estimates suggest that, if preserved in a glassy state, the architecture of ancient DNA might still be well-preserved in samples of this age. Experiments to interrogate this ancient DNA architecture, such as those described in the present study, could therefore facilitate genome assembly, epigenetic studies, and architectural analysis for a broad array of extinct species, and could be useful in reconstructing ecosystems across geologic time.

Of course, it is unlikely that most samples accessible to ancient DNA studies were in a glassy state throughout their history. As such, it is difficult, at present, to determine the fraction of ancient DNA samples in which PaleoHi-C experiments will be possible. Prior to the success of PaleoHi-C in the woolly mammoth sample described herein, several dozen other samples in which ancient DNA was detected were analyzed unsuccessfully using early variants of PaleoHi-C. We cannot be certain as to whether this was because of the disruption of DNA architecture in these samples (e.g., because they had not undergone a glass transition), or simply because of the relatively embryonic state of our PaleoHi-C protocol at the time.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. PaleoHi-C read pairs aligning to elephant derive from authentic ancient mammoth DNA and reflect the ancient chromosome conformation, related to Figure 1 and STAR Methods

(A) Photographs of the mammoth sample used in this study (specimen ID IN18-032) showing intact hair and ear. A close-up (left) shows one of the two skin fragments that have been collected for research purposes. The edge of the skin, from which the two fragments were cut side-by-side, is shown in the lower-right.

(B) Light microscopy, Van Gieson/Giemsa and DAPI staining of the mammoth sample, and successive zoom-ins of the regions shown in Figure 1C. Top row: Derma (Van Gieson/Giemsa staining): general view (left), successive zoom-ins (center and right panel). Middle row: Hair follicle (Giemsa staining): general view (left), and zoom-in on the hair shaft and outer root sheath cells (center, right). Bottom row: Skeletal muscle (Van Gieson/Giemsa staining and DAPI): general view (left), zoom-in (center), and DAPI staining of nuclei-like structures (laser scanning confocal microscopy). More images are available at https://doi.org/10.5281/zenodo.11193545.

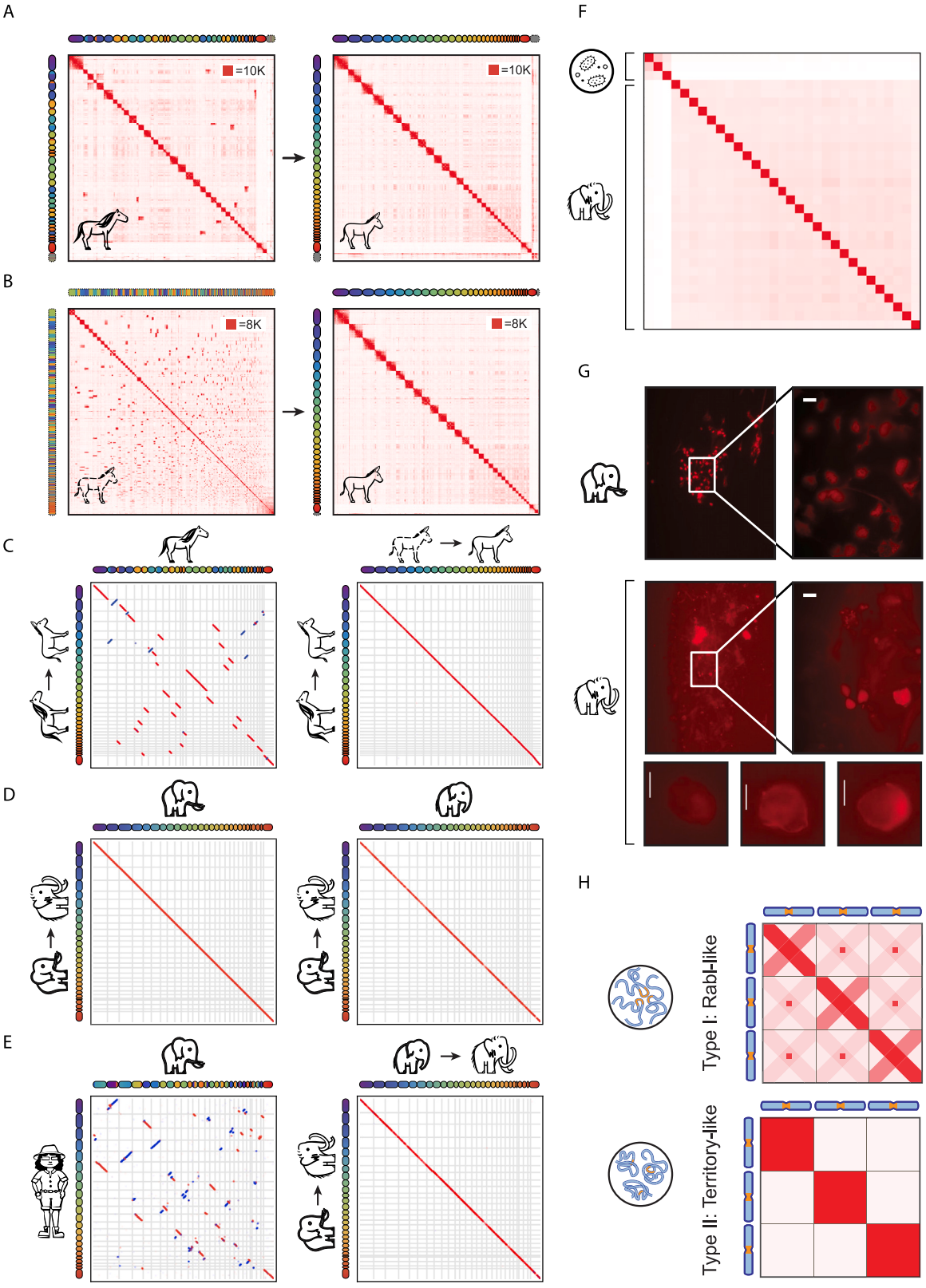
(C) Contaminants in the mammoth sample: fungi (top, Van Gieson/Eosin), unidentified (possibly parasites) (middle, Van Gieson/Giemsa), bacteria (bottom, Giemsa). More images are available at https://doi.org/10.5281/zenodo.11193545.

(D) Top: Starting with the Asian elephant draft genome assembly ASM1433276v1 from Tollis et al.,[29] we use *in situ* Hi-C data (left) to error-correct, anchor, order and orient the draft sequences to produce a chromosome-length *de novo* assembly for the Asian elephant, ASM1433276v1_HiC (right). Rainbow tracks on top and to the left of the contact maps are used to highlight corresponding loci between the two assemblies: the same color is used to show matching sequences. The draft sequences on the left are ordered by size, from largest to smallest. The chromosome-length sequences on the right are ordered according to the published Asian elephant karyotype.[76] The ovals shaping the rainbow track on the right outline the boundaries of the 28 chromosomes in ASM1433276v1_HiC. The dashed oval after the 28th chromosome-length scaffold represents unanchored sequences. The dashed oval around the draft assembly rainbow track highlights that all sequences in the draft are unanchored. Bottom: Starting with the African elephant draft genome assembly Loxafr3.0 from Palkopoulou et al.,[28] we use *in situ* Hi-C data (left) from Álvarez-González et al.[27] to error-correct, anchor, order and orient the draft sequences to produce a chromosome-length *de novo* assembly for the African elephant, Loxafr3.0_HiC, on the right. Rainbow tracks on top and to the left of the contact maps are used to highlight corresponding loci between the two assemblies: the same color is used to show matching sequences. The draft sequences on the left are ordered by size. The chromosome-length sequences on the right are ordered according to the published African elephant karyotype.[76] The ovals shaping the rainbow track on the right outline the boundaries of the 28 chromosomes in Loxafr3.0_HiC. The dashed oval after the 28 chromosome-length scaffolds represents unanchored sequences. The dashed oval around the draft assembly rainbow track highlights that all sequences in the draft are unanchored. An interactive version of the contact maps shown in this figure can be found at https://t.3dg.io/3d-mammoth-Fig-S1D.

(E) Bottom: As is often the case with ancient DNA (aDNA) sequencing, many of the read pairs in PaleoHi-C align to a variety of contaminants. The stacked bar chart shows the taxonomic composition of 22 BGI libraries spanning six biological tissue replicates, labeled alphabetically (see Table S1, #3). The replica ID is listed along the x axis. The y axis shows the number of reads in each library assigned to one of the NCBI Foreign Contamination Screen's Genome Cross-species aligner (FCS-GX) taxonomic divisions shown below the bar chart. These include seven top-scoring known taxonomic divisions across all datasets (b-proteobacteria, g-proteobacteria, CFB group bacteria, insects, mammals, high GC Gram+, and plants) as well as unidentified sequences ("n/a"). All other identified taxonomic categories are bundled together and labeled as "other". Endogenous mammoth sequences are expected to fall into the "mammals" taxonomic division. The libraries are bundled into four groups: EtOH-p for "EtOH precipitate," the main PaleoHi-C dataset; EtOH-s for "EtOH supernatant," a supplementary dataset prepared from the tissue lysis supernatant leftovers; no-EtOH-p for "No EtOH precipitate," a supplementary dataset generated from the sample handled without ethanol; and no-EtOH-s for "No EtOH supernatant," a supplementary dataset generated from the supernatant collected during no-ethanol tissue lysis. The total number of reads analyzed from each library is 20 million (10 million paired-end reads). Top: Immediate on-site preservation of samples in ethanol is compatible with PaleoHi-C and, along with mechanical removal of exterior contaminants, can contribute to curbing microbial activity and improve isolation of endogenous mammoth Hi-C sequences. Although with lower efficiency, endogenous contact data can be collected from non-ethanol samples as well. The percent of total alignable read pairs (calculated with respect to the total raw read pair count) are listed as reported by the Juicer pipeline.[23,77] Alignment statistics are shown for the same 22 BGI libraries as are analyzed in the bottom panel.

(F) Top: Barplot showing the percentage of reads that display damage related to cytosine deamination in the woolly mammoth sample and the modern elephant samples, as determined by PMDtools.[30] As expected, the mammoth-derived PaleoHi-C reads have an elevated degradation signature, on par with the non-USER (no-U) treated aDNA-Seq data. The damage is higher than in any of the modern elephant datasets. Note the elevated damage signature in the skin necropsy for the Asian elephant as compared to the fibroblast-derived African elephant dataset. The relative increase in percent of damaged reads is consistent with the necropsy sample being stored under suboptimal conditions for ~20 years. Bottom: Barplot showing the percentage of mammoth alleles in reads overlapping mammoth-specific fixed genetic variants from Díez-del-Molino et al.[24] in PaleoHi-C data vs. data derived from modern elephants. Whenever such diagnostic reads are examined in the mammoth, they overwhelmingly show the mammoth allele, confirming that PaleoHi-C data derives from ancient mammoth DNA rather than from modern elephantid contaminants. Of course, the diagnostic reads from modern elephants overwhelmingly show the modern elephant allele. The results of both the read damage and allele analyses remain unchanged if only reads representing candidate long-range contacts are analyzed (Long) as opposed to all reads in the library (All) and are robust to the choice of sequencing platform (BGI or Illumina [IL]).

(G) Supplementary details on the probability of observing a read pair in which the two reads map to positions in the genome separated by a given distance in base pairs. Top: A small fraction of PaleoHi-C data reflects contacts between loci that lie far away in 1D. No such fraction exists in aDNA-Seq datasets. The image is analogous to that shown in Figure 1E, with the following modifications: 1) the PaleoHi-C curve is based on data collated from both the BGI and Illumina sequencing platforms; 2) two aDNA-Seq datasets are included: the USER-treated (U) datasets (from across 32 replicates) from Díez-del-Molino et al.[24] and the non-USER (no-U) treated dataset generated for the same sample as part of this study. To be maximally conservative, the dataset with the heaviest tail from Díez-del-Molino et al.[24] was chosen to be plotted alongside PaleoHi-C data in Figure 1E. Middle: Contact probability curves for PaleoHi-C of woolly mammoth skin and *in situ* Hi-C of Asian elephant skin. The slopes between the two datasets are highly consistent across a wide range of distances. The curve is based on a combined dataset from both the BGI and Illumina platforms. Bottom: Contact probability curves plotted separately for the four library groups: EtOH-p (EtOH precipitate), EtOH-s (EtOH supernatant), no-EtOH-p (no EtOH precipitate) and no-EtOH-s (no EtOH supernatant).



(legend on next page)

Figure S2. Reference-assisted 3D genome assembly of a woolly mammoth yields a robust karyotype consistent with modern elephantids, related to Figure 2 and STAR Methods

(A) Reference-assisted 3D genome assembly of the donkey, *Equus asinus*. Starting with the EquCab2.0 domestic horse genome assembly (2*n* = 64[78]), left, we used Hi-C data for the donkey *Equus asinus* to identify the large-scale evolutionary events such as chromosome breaks, fusions, and intrachromosomal rearrangements that differentiate the two species. The events manifest as near-diagonal depletions and off-diagonal enrichments that suggest a disagreement between the 1D distance separating a pair of genomic loci as suggested by the Hi-C data and by the underlying reference. (Simplifying somewhat, off-diagonal enrichments suggest that the sequences are further away on the reference than they should be according to Hi-C data, and conversely, a depletion near diagonal that is not associated with a reference sequence boundary suggests that the sequences flanking the depletion should be moved further apart in the 1D reference.) On this basis, we edited the horse genome assembly's chromosome-length scaffolds, introducing breaks and fusing the horse sequences to reflect the identified evolutionary events. This process continued until the Hi-C data and the underlying edited reference were in agreement, with the resulting map showing no off-diagonal enrichments or near-diagonal depletions not associated with chromosome boundaries, on the right. The resulting genome assembly (after correcting local changes on the basis of SNVs and indels in individual reads) closely approximates the donkey genome (2*n* = 62). The contact maps show the same donkey Hi-C data, aligned to the original EquCab2.0 and lifted over to the rearranged reference. Rainbow tracks on top and to the left of the contact maps are used to highlight syntenic sequences between the two references: the same color is used to show corresponding loci. The ovals shaping the rainbow tracks outline the boundaries of the 32 chromosomes in the EquCab2.0 horse genome assembly on the left, and the boundaries of the 31 chromosomes in the assisted donkey genome assembly on the right. The horse chromosomes are ordered from #1 to #31, followed by chromosome X. The donkey chromosomes are ordered from largest (#1) to smallest (#30) autosome, followed by chromosome X. The dashed ovals shown in the two chromograms after the sex chromosome represent the unanchored portions of the two genome assemblies. For ease of comparison, the orientation of donkey chromosomes was adjusted to match the orientation of the corresponding horse chromosomes. For "composite" chromosomes, the orientation of the horse chromosome that provided the largest amount of sequence was used when assigning orientation. An interactive version of the contact maps shown in the figure can be found at https://t.3dg.io/3d-mammoth-Fig-S2AB.

(B) *De novo* 3D genome assembly of the donkey, *Equus asinus*. Starting with the ASM303372v1 domestic donkey genome assembly draft from Renaud et al.,[79] left, we used Hi-C data to error-correct, anchor, order and orient the draft sequences to produce a chromosome-length *de novo* assembly for the donkey, ASM303372v1_HiC, on the right. The contact maps show the same donkey Hi-C data, aligned to the draft on the left and lifted over to the chromosome-length reference on the right. Rainbow tracks on top and to the left of the contact maps are used to highlight corresponding loci between the two assemblies: the same color is used to show matching sequences. The draft sequences on the left are ordered by size, from largest to smallest. The chromosome-length sequences are ordered to match the ordering and orientation of the assisted donkey assembly. The ovals shaping the rainbow track on the right outline the boundaries of the 31 chromosomes in ASM303372v1_HiC. The dashed oval after the 31 chromosome-length scaffolds represents unanchored sequences. (The dashed oval around the draft assembly rainbow track highlights that all sequences in the draft are unanchored.) An interactive version of the contact maps shown in this figure can be found at https://t.3dg.io/3d-mammoth-Fig-S2AB.

(C) Dotplots showing the correspondence between sequences in the chromosome-length assisted donkey genome assembly EquAsi_EquCab2.0_assisted_HiC and EquCab2.0, the assisting reference genome assembly (left); and a *de novo* chromosome-length assembly of the donkey, ASM303372v1_HiC (right). The left image was generated with liftover using the *.assembly* file describing a set of rearrangements between the donkey and the domestic horse identified from analyzing the donkey Hi-C data alignments in comparison to the horse reference. For the image on the right, the horse reference and *de novo* donkey sequences were aligned using LastZ alignment algorithm[80] using "--masking = 3 --notransition --step = 20 --nogapped --format = maf --ambiguous = iupac --hspthresh = 50000" command options, and the resulting alignments lifted over to the assisted donkey reference. The 15,000 highest-scoring alignment blocks are represented, with direct synteny blocks colored red, and inverted blocks colored blue. The chromosome order and orientation in both panels matches that in panels A and B. The chromosome borders are marked with vertical and horizontal gray lines. The dotplots illustrate excellent correspondence between the assisted and *de novo* assemblies despite many rearrangements occurring in the donkey and horse lineages since their most recent common ancestor.

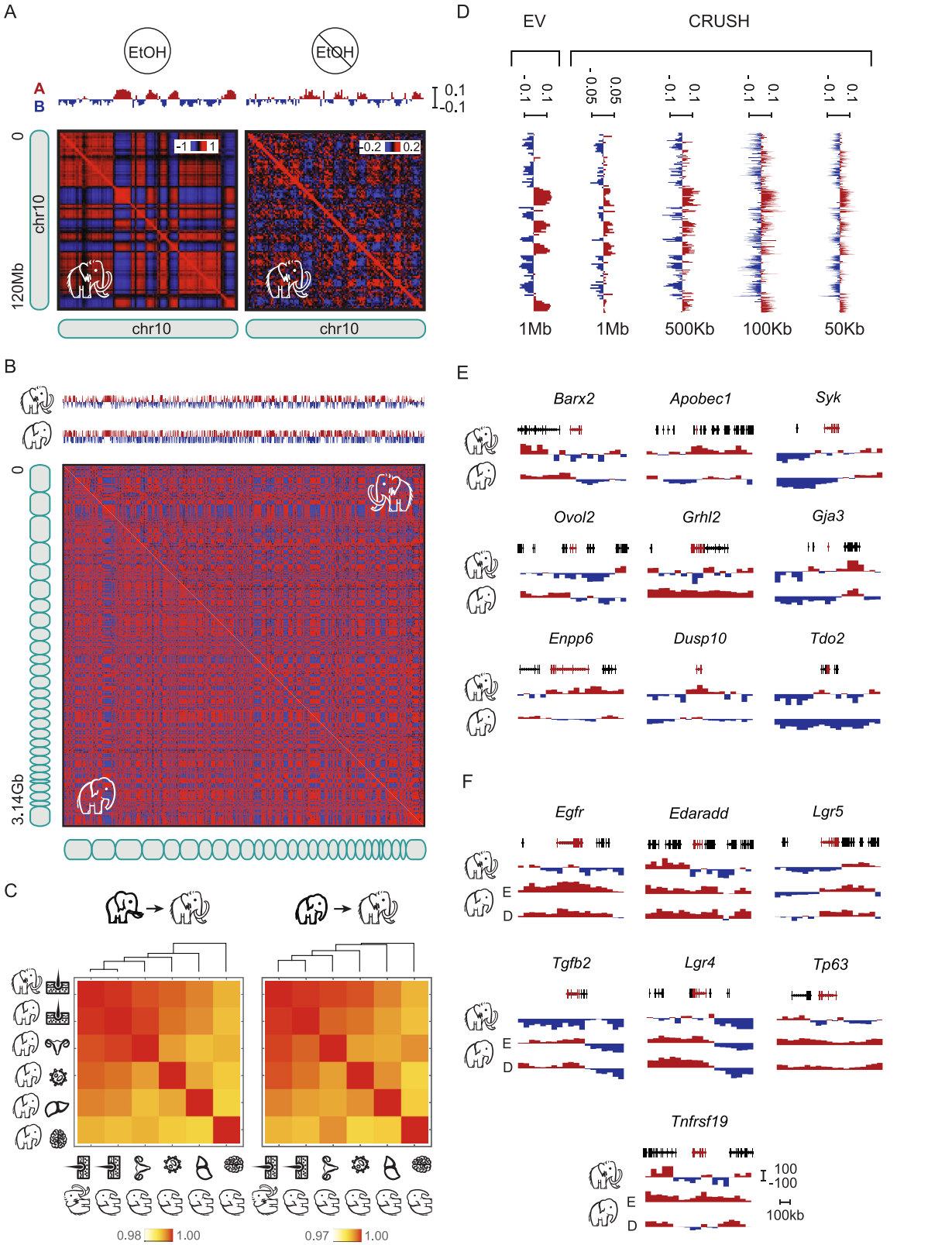
(D) Dotplots showing a conservation of synteny between the woolly mammoth and the African (left) and Asian (right) elephants. Each dot marks syntenic sequences between two genomes, with direct synteny blocks colored red, and inverted blocks colored blue. The syntenic blocks were identified using LASTZ alignment algorithm.[80] The pipeline was run with "--masking = 3 --notransition --step = 20 --nogapped --format = maf --ambiguous = iupac --hspthresh = 50000" command options to align ASM1433276v1 to Loxaf3.0, and the resulting alignments were lifted over to ASM1433276v1_HiC, Loxafr3.0_HiC and MamPri_Loxafr3.0_assisted_HiC. The 15,000 highest-scoring alignment blocks are represented. Chromosome borders are marked with vertical and horizontal gray lines.

(E) Left: A whole-genome alignment between the African elephant chromosome-length genome assembly and human genome reference hg38. African elephant chromosome-length scaffolds were ordered and oriented to match the comparative genome map generated using Zoo-FISH,[81] as were the syntenic chromosomes in Asian elephant and woolly mammoth assemblies. Right: Whole-genome alignment plot between two woolly mammoth genome assemblies, one assisted using the African elephant genome assembly Loxafr3.0, and another one assisted using the Asian elephant genome assembly ASM1433276v1. The plot was generated by lifting over the top scoring 15,000 LastZ alignments between Loxafr3.0 and ASM1433276v1 to the respective references. The analysis suggests excellent agreement, demonstrating that the woolly mammoth assembly is robust with respect to the choice of assisting reference.

(F) An extended PaleoHi-C map that includes references for the three top contaminating species alongside the 28 assembled mammoth chromosomes. Each assembled molecule (one for each contaminant and 28 for the mammoth) is represented as a separate bin, and the resulting contact map is balanced using the Juicer software's VC_SQRT (square root of vanilla coverage) normalization option. The map shows that mammoth chromosomes are more likely to come into contact with other mammoth chromosomes than with chromosomes of the contaminating organisms, suggesting that mammoth DNA occupies a distinct territory within the sample, spatially segregated from the DNA of other species. A slightly elevated contact frequency between *J. tructae*, *P. veronii* and *F. petrolei* bins likely reflects alignment noise, which might relate to the evolutionary distance between the species included in the extended reference (e.g., *J. tructae* and *P. veronii* are the most closely related of the species included, since both are Proteobacteria).

(G) Ethidium bromide, a DNA intercalating dye, reveals brightly staining structures in an African elephant nuclear preparation (top) and in a comparable mammoth nuclear preparation (middle). The structures in both samples are similar in both size and shape. Images are shown at 20× (left) and 100× (right) zoom. The scale bar marks 10 μm. Bottom row: Propidium Iodide (which stains DNA specifically) reveals structures resembling mammalian nuclei in a crosslinked nuclear suspension from the mammoth. Scale bar: 10 μm.

(H) Schematic representation of type I and type II genome architectures and their corresponding Hi-C contact maps. Type I (Rabl-like) nuclei shown on top tend to have centromere clustering, often accompanied by telomere clustering and alignment of chromosome arms. Hi-C contact maps generated from nuclei in type I architecture typically have bright foci corresponding to contacts between centromeres (marked with orange) of different chromosomes. Interactions of centromeric foci with their respective chromosome is often depleted. Type II (chromosome territory-like) nuclei shown on the bottom have very little inter-chromosomal interactions. Instead, there is a strong preference for intrachromosomal contacts, consistent with individual chromosomes occupying distinct subvolumes in the nucleus.



(legend on next page)

Figure S3. The segregation of active (A) and inactive (B) chromatin in the woolly mammoth sample preserves the cell-type specific compartmentalization of ancient woolly mammoth chromosomes, related to Figure 3 and STAR Methods

(A) PaleoHi-C contact maps from woolly mammoth skin exhibit a plaid pattern, characteristic of the spatial segregation between the A and B genome compartments. The 2nd order autocorrelation analysis can be performed even on suboptimal non-ethanol-preserved sample data (right), which independently matches the result of the ethanol-preserved sample data (left). The data is shown for chr10. All maps and tracks are shown at 1-Mb resolution.

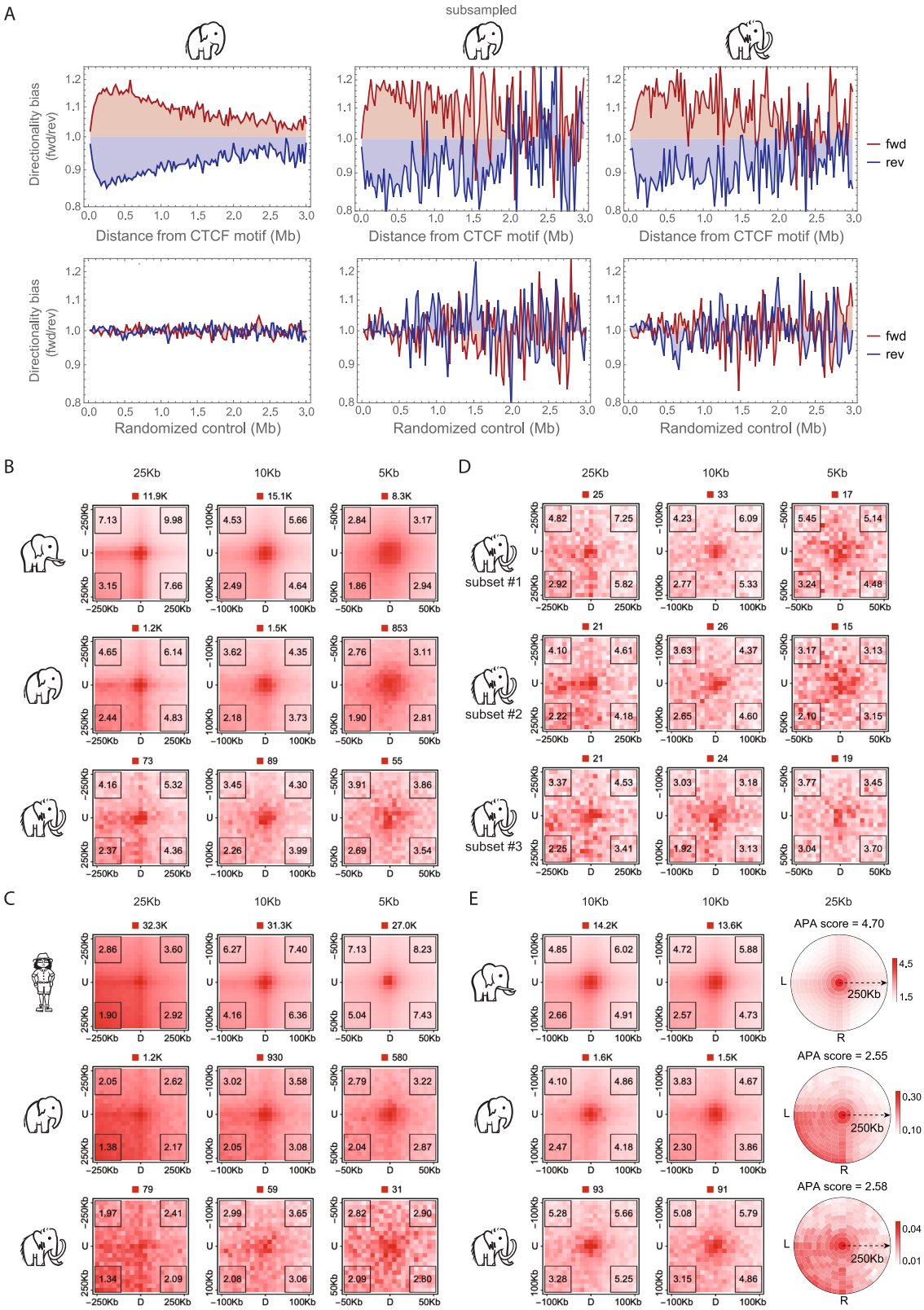
(B) Segregation into the active (A) and inactive (B) genome compartments is evident in the mammoth skin PaleoHi-C data genome-wide (upper right), and agrees extremely well with results of the same analysis for the modern Asian elephant skin contact dataset (lower left). The tracks above the Pearson maps show the first principal component of the elephant and the mammoth genome-wide contact matrices, with the woolly mammoth plotted on top, and the Asian elephant at the bottom. The 1D tracks also closely correspond to one another. Pearson coefficient range shown is from −0.01 to 0.01 for both matrices. Genome-wide eigenvector values range from −0.02 to 0.02 in both tracks. Both the correlation matrices and the 1D tracks closely correspond to one another.

(C) Compartment analyses at 1-Mb resolution using two different mammoth genome assemblies, MamPri_Loxafr3.0_assisted_HiC (left) and MamPri_ASM1433276v1_assisted_HiC (right). Both analyses suggest that the mammoth skin is a closer match to elephant skin than to any other elephant tissue, in agreement with the analysis presented in the main text. The images show, for each of the underlying references, the eigenvector similarity matrix for the mammoth and the elephant tissue datasets, and an associated dendrogram. The fraction of bins with congruent eigenvector compartment calls (i.e., either both positive or both negative) between any two given contact matrices is used as a similarity measure. Bins with small eigenvalues (absolute value < 0.05) were excluded from the analysis. The dendrogram is based on Pearson distance between the binarized eigenvectors.

(D) Mammoth CRUSH annotations at high resolution (right) are highly concordant with eigenvector-based (EV) A/B annotations at low resolution (left). The data is shown for mammoth chromosome 10, aligned with the Pearson correlation maps shown in Panel A. The following resolutions are included in the CRUSH resolution walk: 1 Mb, 500 kb, 100 kb and the target resolution, 50 kb.

(E) Examples of CRUSH compartment tracks in several genes of interest. Each panel shows a 1-Mb stretch of the Loxafr3.0_HiC genome centered on a gene of interest. The gene track is shown at the top of each panel, with the gene of interest highlighted in red. Below the gene tracks are two CRUSH tracks at 50-kb resolution, those calculated in the mammoth skin dataset (top), and the Asian elephant skin dataset (bottom). See text for a brief description of the genes included in the figure.

(F) CRUSH compartment tracks for genes highlighted in association with the "hair follicle development" GO term. Each panel shows a 1-Mb stretch of the Loxafr3.0_HiC genome centered on a gene of interest. The gene track is shown at the top of each panel, with the gene of interest highlighted in red. Below the gene tracks are three CRUSH tracks at 50-kb resolution. These are, from top to bottom, CRUSH scores in the mammoth skin PaleoHi-C dataset, the Asian elephant skin dataset described in the main text (marked with an (E) for Epidermis-enriched), and an *in situ* Hi-C dataset from the Asian elephant in which we targeted deeper layers of the skin (marked with (D) for Dermis-enriched). The first two tracks can be used to assess the difference in compartmentalization signal associated with the genes, as calculated by CRUSH, between the woolly mammoth and Asian elephant skin. The derma tracks give an opportunity to assess the degree to which these differences might be attributed to differences in anatomical sampling location between the two datasets. The tracks for all genes but one are consistent with species-specific differences. *Tnfrsf19* could be explained by differences in anatomical position of the sample.



(legend on next page)

Figure S4. Remains of point-to-point looping and ancient loop extrusion persist in woolly mammoth skin, related to Figure 4 and STAR Methods

(A) Like modern elephant Hi-C contacts (top row, left [full dataset] and center [subsampled to match mammoth dataset]), mammoth PaleoHi-C contacts (top right) exhibit an orientation-dependent directionality bias in the vicinity of CTCF motifs. No directionality bias is seen for any of the three datasets at randomly selected loci (bottom row). The bias is consistent with activity of loop extrusion in the woolly mammoth, arrested by CTCF in an orientation-specific fashion while the mammoth was alive. The analysis does not depend on the use of a loop annotation. The analysis uses the Loxafr3.0_HiC reference for elephantid species.

(B) APA plots for three elephantid datasets, the African elephant fibroblast *in situ* Hi-C, modern Asian elephant skin *in situ* Hi-C, and mammoth skin PaleoHi-C, show enrichment of aggregate signal associated with the 3,578 *Loxodonta africana* loop calls (central pixels) across multiple resolutions (25 kb, 10 kb, 5 kb) relative to the local background (non-central pixels). Distance-filtered loop counts for the three resolutions are as follows: 492, 1,880 and 3,069 for 25 kb, 10 kb and 5 kb, respectively.

(C) Aggregate peak analysis of 9,683 loops with conserved anchors in human, Asian elephant and mammoth skin data reveals loop signatures that persist over vast evolutionary distances as well as 52,000 years in permafrost. Distance-filtered loop counts for 25-kb, 10-kb and 5-kb resolution APA plots are, respectively, 2,832, 4,766 and 6,543 for the human skin dataset, and 3,054, 5,015 and 6,679 for the elephantid skin datasets. The human skin Hi-C data and loop list were generated as part of the ENCODE project (https://www.encodeproject.org/files/ENCFF531FEY). The analysis uses the Loxafr3.0_HiC reference for elephantid species.

(D) Mammoth APA plots, across multiple resolutions, show an enrichment of the central pixel with respect to the background for three non-overlapping loop subsets, each comprising a random selection of 1,000 entries from the 3,578 loops annotated in the African elephant fibroblast data. The analysis is consistent with the notion that the mammoth APA enrichments reflect real loop signatures rather than enrichment in a small number of artifactual loci. Distance-filtered loop counts for subsets 1-2-3 are as follows: 137/135/133, 546/510/519 and 873/850/854 for resolutions 25 kb, 10 kb and 5 kb, respectively. The analysis uses the Loxafr3.0_HiC reference for elephantid species.

(E) Additional analyses support the robustness of the loop preservation result in mammoth PaleoHi-C. Left: APA reanalyzed using MamPri_Loxafr3.0_assisted_HiC as underlying reference for the African elephant (top row), Asian elephant (center) and mammoth (bottom row). The APA plots were constructed as described in the main text for Figure 4A using loop calls generated with HiCCUPS[23] from the African elephant data. The plots, shown here at 10-kb resolution, show a characteristic enrichment of the central pixel compared to the background, a signature of loops in the contact maps. Center column: Same APA analysis, done using MamPri_ASM1433276v1_assisted_HiC as a reference. Right: Similar to Juicer APA plots generated for HiCCUPS loop calls, SIPMeta average bullseye metaplots for loops independently called by SIP show an enrichment of the central pixel with respect to the background in the African elephant fibroblast data (top), Asian elephant skin (middle) and the mammoth skin dataset (bottom).

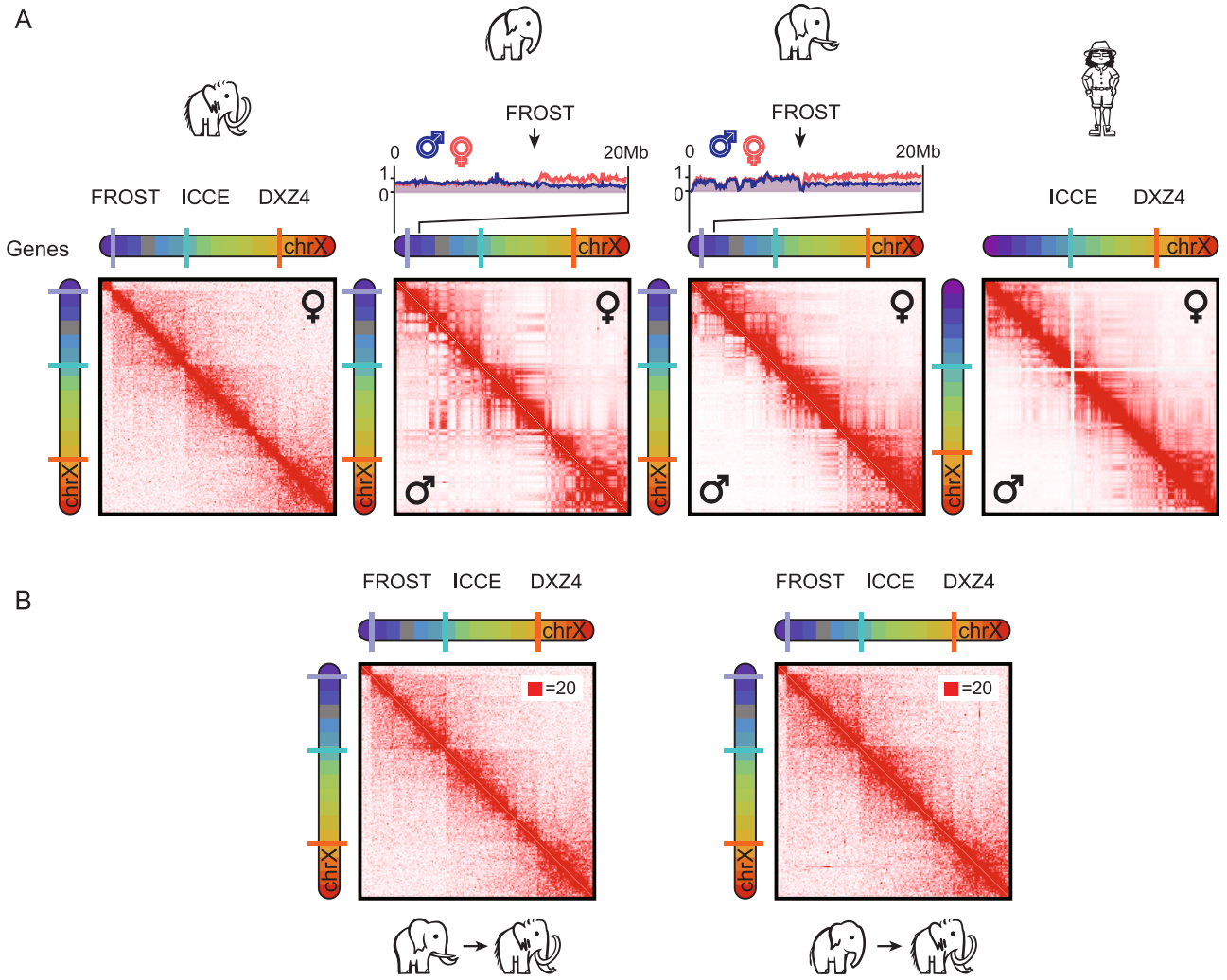
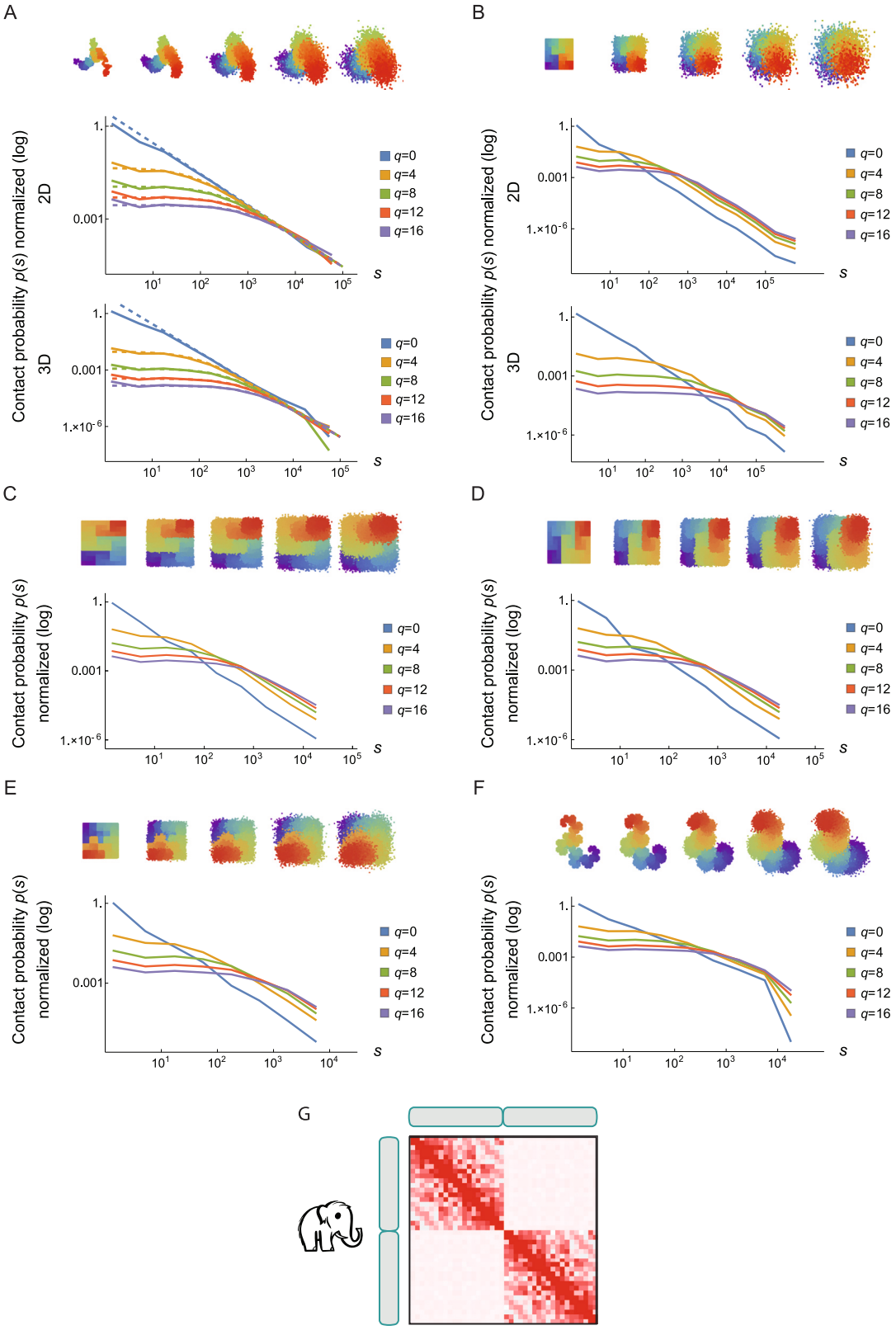


Figure S5. The FROST superdomain boundary element flanks the elephantid pseudoautosomal region (PAR), related to Figure 5 and STAR Methods

(A) FROST flanks the pseudoautosomal region (PAR). The following chrX contact maps are shown: the woolly mammoth (PaleoHi-C skin data for a female individual), the Asian elephant (PBMC data for a male and female animal, in the lower left and the upper right part of the image, respectively), the African elephant (fibroblast data for a male and female animal[27] in the lower left and the upper right part of the image, respectively), and human (male AK1 from Harris et al.[82] in the lower left, and female GM12878 lymphoblastoid data from Rao et al.[23] on the right). Data is aligned to the following genome assemblies: MamPri_Loxafr3.0_assisted_HiC, ASM1433276v1_HiC, Loxafr3.0_HiC and hg19. The chromograms above the contact maps indicate the ordering of the loci. The color scheme is based on the human chrX, purple (p-terminus) to red (q-terminus); corresponding loci across species are shown using the same color. Purple, cyan, and orange ticks on the chromograms indicate the positions of the FROST, ICCE and DXZ4 repeat elements, respectively. The FROST, ICCE, and DXZ4 boundaries disappear in male datasets. The Hi-C coverage tracks centered at FROST are shown at the top for male and female Asian and African elephant datasets. At FROST the coverage ratio for the male vs. female samples changes 2-fold, consistent with FROST lying at the PAR boundary. See also https://t.3dg.io/3d-mammoth-Fig-5.

(B) Contact maps for chromosome X, obtained by reanalyzing PaleoHi-C data using either MamPri_Loxafr3.0_assisted_HiC (left) or MamPri_ASM1433276v1_assisted_HiC (right), display a tetradic structure, consistent with results described in the main text of the paper and confirming the robustness of the result with respect to the assisting reference. Note that the contact map on the left of panel B is the same contact map shown to the left of panel A; it is reproduced to make it easier to compare the contact maps generated using the two mammoth reference genome assemblies.



(legend on next page)

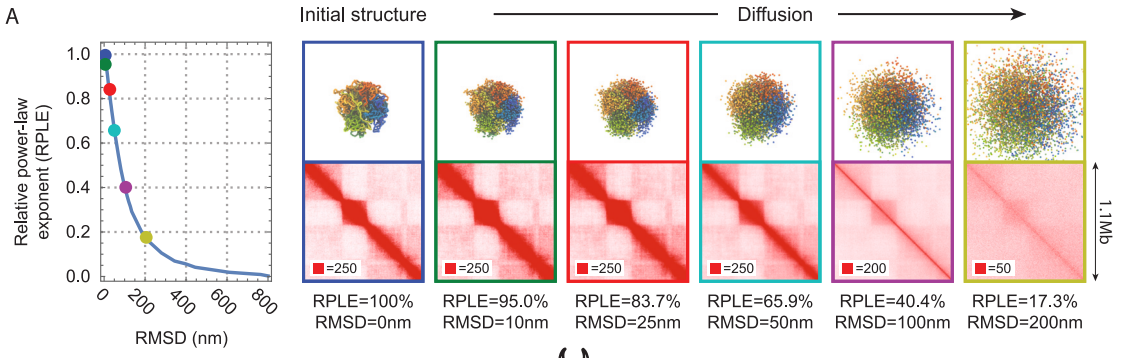
Figure S6. Nuclear architecture is preserved in Yuka, a juvenile female woolly mammoth from 39,000 years ago, related to Figure 6 and STAR Methods

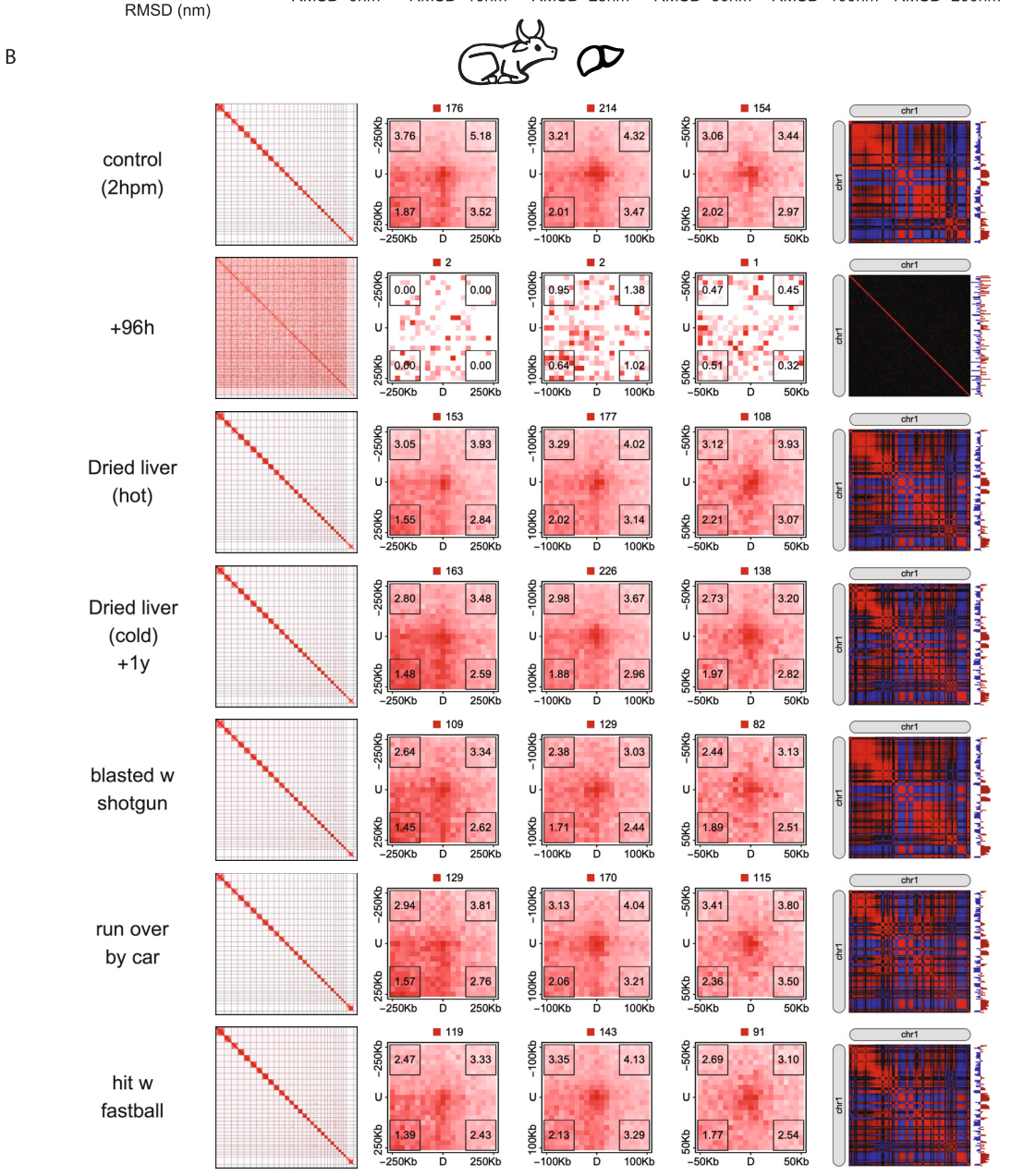
(A) Diffusion of monomers has a pronounced effect on contact probability scalings. Top row: Images illustrating the effect of monomer diffusion on a short (length = 10^3) 2D random walk. As the diffusion level (q) increases, the local structure is lost, but the long-range structure highlighted with a color gradient indicating the initial 1D position is, to a large extent, preserved. Middle row: Plots of $p(s)$ on log-log axes for the 2D random walks for varying diffusion intensities. Each curve has been obtained by simulating 10 random walks of length 10^5. Without diffusion (blue), the plot is linear, with a slope of -1. With diffusion, the slope is altered as it flattens out on the left side, in accordance with theoretical predictions described in the section on theoretical analysis of a random coil polymer after fragmentation and simple diffusion of the monomers. Theoretical predictions for varying levels of q are shown as dashed lines colored to match the simulation curves for the same q. Bottom row: Contact probability curves $p(s)$ for 3D random walks. Similar to the 2D case, all curves have been obtained by aggregating data from 10 random walks of length 10^5. Theoretical curves are shown in dashed lines. Again, there is a close correspondence between theory and simulation.

(B) Simulation of monomer diffusion in a Hilbert curve shows a similar effect on contact probability scaling as was described for the random walk polymers: as the diffusion level increases the local structure is lost while the long-range structure is preserved. Top: Images illustrating the effect of monomer diffusion on a short (length = 2^{12}) 2D Hilbert curve. As the diffusion level (q) increases, the local structure is lost, but the long-range structure reflected by a color gradient is, to a large extent, preserved. Middle: Plots of $p(s)$ on log-log axes for the 2^{20}-long 2D Hilbert curve for varying diffusion intensities. Without diffusion (blue), the plot is linear, with a slope of $-3/2$. With diffusion, the slope is altered as it flattens out on the left side. The transition happens when the initial distance between the monomers ($\sqrt{s}$ in 2D, $\sqrt[3]{s}$ in 3D) becomes comparable to RMSD (defined by q). Bottom: Contact probability curves $p(s)$ for a 3D Hilbert curve 2^{21} monomer-long. Similar to the 2D case and the random walk chains, as diffusion increases the slope in the left part of the plot flattens out while the power law is preserved on the far right.

(C–F) Monomer diffusion and contact probability curves for the 2D Peano curve (length = 3^{10}), the 2D Symmetric Peano curve (length = 3^{10}), the 2D Quadratic Gosper curve (length = 5^6), and the 2D Heighway Dragon curve (length = 2^{15}), respectively. Just as with the random walk and Hilbert curves, as the diffusion level (q) increases, the local structure of these fractal curves is lost, but the long-range structure is largely preserved. This can be seen by means of the 1D color gradient. Contact probability curves at varying diffusion levels are shown in the respective bottom panels. Other than an overall increase in contact frequency at higher q values, the theoretical description of $p(s)$ that we derive for the special case of a random walk appears to hold for other types of fractal curves as well. This includes the gradual effacement of the power law, starting from the left-hand side, as diffusion increases, as well as the preservation of the power-law on the right-hand side, and the positioning of the $p(s)$ shoulder.

(G) Yuka PaleoHi-C Aggregate Chromosome Analysis (ACA). ACA involves scaling the chromosomes to a uniform length and aggregating the signal of all intra- and interchromosomal contacts.[37] The resulting map looks like a typical Type II ACA map, consistent with the persistence of chromosome territories in the Yuka sample.



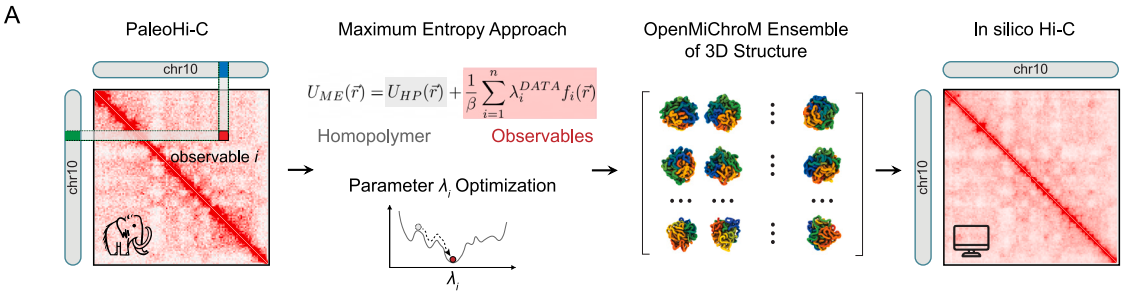


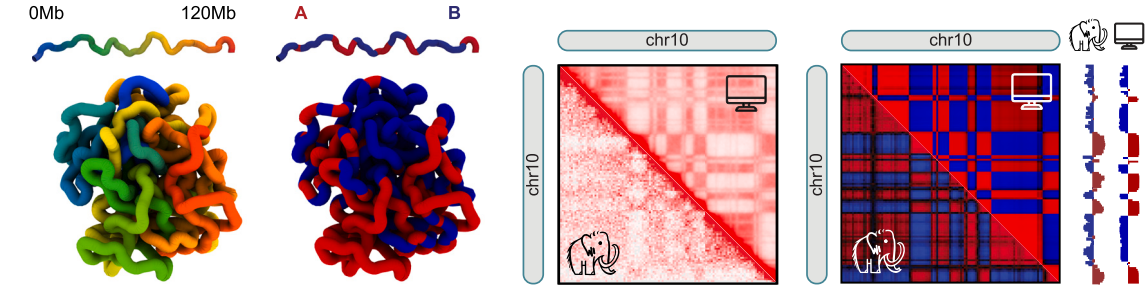
(legend on next page)

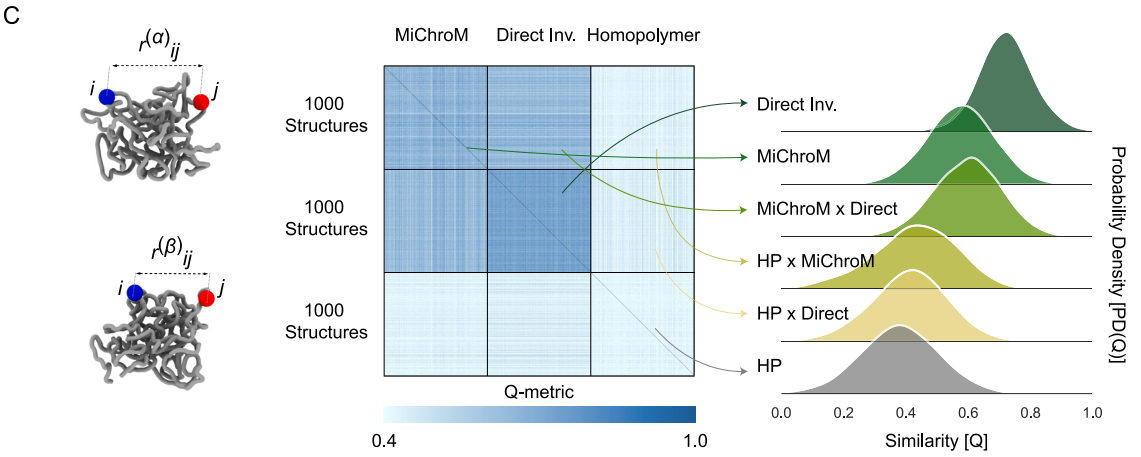
Figure S7. In dehydrated tissue samples exposed to room temperature, nuclear architecture is preserved for a year, related to Figure 6 and STAR Methods

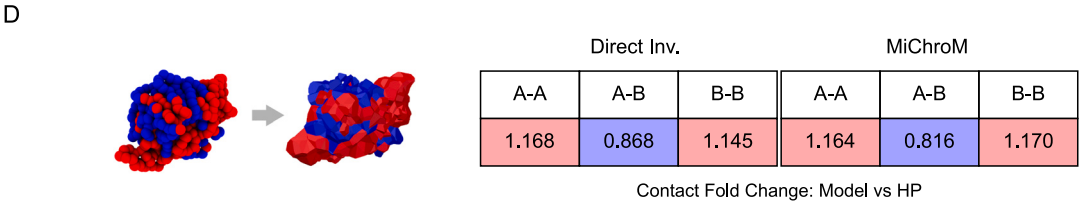
(A) The effects of nucleosome diffusion on the contact probability scaling exponent and Hi-C maps, modeled using a realistic chromatin conformation as initial, non-degraded structure. Leftmost panel: The decay of the contact probability exponent, measured as relative power-law exponent (RPLE), is shown as a function of the average root-mean-square-displacement (RMSD) of the nucleosomes. Included to the right of the contact probability decay plot are panels with structures and contact maps corresponding to six different RMSDs, outlined in blue, green, red, cyan, violet and yellow. (The same colors are used to mark the corresponding points in the exponent decay curve.) Each panel is labeled with the relative power law exponent (RPLE) and RMSD. The blue box to the left refers to the initial, undegraded, native state of the chromosome structures. The original ensemble comprises collapsed, globular structures modeling a 1.1 Mb region of chr7 of human lymphoblastoid cells. The corresponding native contact map shows typical features of chromosome organization such as a high number of counts along the diagonal (corresponding to local interactions between nucleosomes separated by short genomic distances), as well as evidence of compartments. Shown to the right of the native panel is a sequence of increasingly deteriorated states. Throughout the panels, the nucleosomes are shown as beads, with the beads colored according to genomic position from blue to green to yellow to red. The loss of the original structure is clearly visible as initially adjacent nucleosomes become more distant from each other with increasing RMSD, and the colors mix together. Contact map features are similarly lost, with the diagonal becoming less populated with time, and the checkerboard pattern becoming less defined. Contact probability is particularly sensitive to the effects of diffusion, with the scaling exponent falling to about 50% of its value in the undegraded ensemble when nucleosomes are allowed to diffuse about 80 nm, mere 8 nucleosome diameters. Once the nucleosome gas has diffused to a RMSD ~200 nm, or ~20 nucleosome diameters, the scaling exponent falls to less than 20% of the original value. The results suggest that even small amounts of diffusion can result in appreciable changes in the contact probability power law exponent, and that contact probability scaling can be used to estimate the amount of diffusion in a degrading sample. Interactive maps: https://t.3dg.io/3d-mammoth-Fig-S7A.

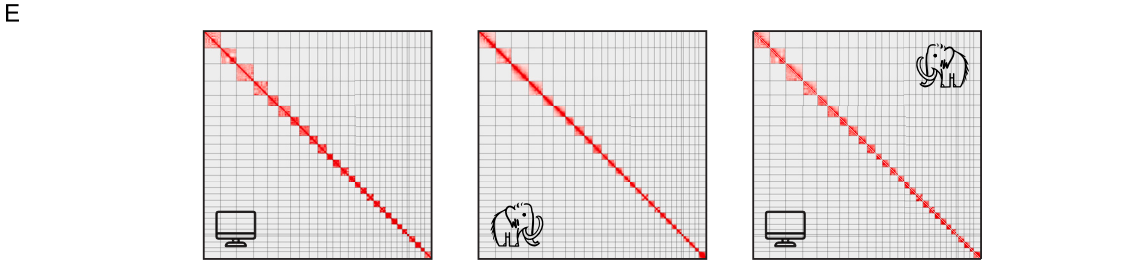
(B) A detailed analysis of 3D genome architecture across samples shown in Figure 6D. Top row: Contact maps obtained from a fresh liver sample (2 hpm) displayed all the expected architectural features: chromosome territories (leftmost panel), compartment segregation (rightmost panel), visualized via a second order Pearson correlation map and a corresponding eigenvector, and loops (3 panels at the center showing APA plots calculated at 25 kb, 10 kb, and 5 kb resolutions). Later time points however showed progressive deterioration of chromosome architecture, with all features gone after 96 h at room temperature (second row). In contrast, the 3D architecture in dehydrated samples is preserved and remains virtually unchanged, both immediately after dehydration (as shown in the 3rd row, liver from the top row after 22 h in a PRESTO DehydroTM electric food dehydrator), and after one year of room temperature storage (4th row, store-bought freeze-dried liver). In this dehydrated state the nuclear architecture is extremely resilient, with chromosome territories, compartment segregation and loops persisting even after the sample is blasted with a shotgun, run over by a car and hit with a fastball (rows #5–7). Interactive maps: https://t.3dg.io/3d-mammoth-Fig-S7B. More data are available in Data S1.











(legend on next page)

Figure S8. The 3D structural ensemble inferred from PaleoHi-C data is robust to the choice of modeling strategy, related to Figure 7 and STAR Methods

(A) Direct Inversion workflow for modeling a woolly mammoth chromosome. We use PaleoHi-C contact data to build a polymer energy function via an optimization procedure implemented in the OpenMiChroM chromatin dynamics software package. We use the resulting empirical energy function to generate an ensemble of 200,000 three-dimensional structures. A contact map generated *in silico* from this ensemble closely matches the experimental PaleoHi-C data, validating the modeling process.

(B) Ancient chromosome reconstruction using MiChroM, a method that uses compartment assignment to help infer the structure of chromosomes in 3D. Shown are a representative structure from the simulated MiChroM ensemble generated for mammoth chr10, colored by genomic position (1st panel) and A/B type (2nd), and a contact map (3rd) and Pearson's autocorrelation matrix (4th) from the simulated structures (above diagonal), compared to mammoth PaleoHi-C (below diagonal). Resolution: 1 Mb. The structures can be explored using Spacewalk at https://t.3dg.io/3d-mammoth-Fig-S8B.

(C) Using Q-metric to compare the two modeling approaches, Direct Inversion and MiChroM. Left: A pair of loci, *i* and *j*, are highlighted in two structures. Q-metric is computed by calculating the differences between the distances $r_{i,j}$ in the two structures, for all pairs of loci *i* and *j*, except for immediate neighbors. A smaller Q-value ($Q \simeq 0$) corresponds to highly dissimilar structures. Two identical structures would give a value of $Q = 1$. Right: A 3,000 x 3,000 similarity matrix that shows, on the scale from white to blue, the pairwise Q-values calculated for structures generated using the MiChroM model, the Direct Inversion, and Homopolymer model, each represented with an equal number of structures. The submatrix corresponding to Q-values calculated between the MiChroM and Direct Inversion structures displays high Q-values, indicating a high degree of similarity between the corresponding ensembles compared to that generated using a Homopolymer (HP) model. Probability density PD(Q) graphs corresponding to various intra- and inter-ensemble value distributions are shown to the right of the similarity matrix and further support this conclusion. Specifically, the Q-values between MiChroM and Direct Inversion ensembles (MiChrom x Direct submatrix) are comparable to those calculated for structures from the MiChroM ensemble (MiChroM submatrix).

(D) Analysis of compartment segregation across 3D structural ensembles generated by MiChroM and Direct Inversion models shows that results across both modeling methods are highly consistent. We computed the 3D Voronoi tessellations of structures from both ensembles where each bead of the polymer chain serves as a generating point for a Voronoi cell, and the cell is assigned the type, A or B, as suggested by the eigenvector analysis of PaleoHi-C data. We then analyzed the interfaces between the Voronoi cells and counted the A-to-A, B-to-B and A-to-B interfaces. When normalized against similar counts in the Homopolymer model (equivalent to random assignment of types), the counts suggest an enrichment of intra-type contacts (A-to-A and B-to-B counts are >1) and the depletion of inter-type contacts (A-to-B counts are <1). The results are consistent across both models, underscoring the robustness of the two modeling approaches and their applicability to model ancient chromosomes from PaleoHi-C data.

(E) Contact map generated from MiChroM ensembles of simulated 3D structures for all 28 mammoth chromosomes (left panel) compared to the raw mammoth data (central panel). The right panel shows the same two contact maps with the simulated data in the lower left, and PaleoHi-C in the upper right corner. An interactive version of this figure can be found at https://t.3dg.io/3d-mammoth-Fig-S8E.