Deep-time paleogenomics and the limits of DNA survival

Love Dalén[#1,2,3], **Peter D. Heintzman**[#1,4], **Joshua D. Kapp**[#5,6], **Beth Shapiro**[#6,7,*]

[1]Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, Sweden

[2]Department of Zoology, Stockholm University, SE-10691, Stockholm, Sweden

[3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE- 10405 Stockholm, Sweden

[4]Department of Geological Sciences, Stockholm University, SE-10691, Stockholm, Sweden

[5]Department of Biomolecular Engineering, University of California Santa Cruz; Santa Cruz, California, 95064, USA

[6]Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, California, 95064, USA

[7]Howard Hughes Medical Institute, University of California Santa Cruz; Santa Cruz, California, 95064, USA

[#] These authors contributed equally to this work.

Abstract

While most ancient DNA studies have focused on the last 50,000 years, paleogenomic approaches can now reach into the early Pleistocene, an epoch of repeated environmental changes that shaped present-day biodiversity. Emerging deep-time genomic transects, including from DNA preserved in sediments, will enable inference of adaptive evolution, discovery of unrecognized species, and exploration of how glaciations, volcanism, and paleomagnetic reversals shaped demography and community composition. In this review, we explore the state-of-the-art in paleogenomics and discuss key bottlenecks, including technical limitations, evolutionary divergence and associated biases, and the need for more precise dating of remains and sediments. We conclude that with improvements in laboratory and computational methods the emerging field of deep-time paleogenomics will expand the range of questions addressable using ancient DNA.

The Pleistocene epoch (approximately 2.6 million years ago; Ma, to 10 thousand years ago; ka) was a time of considerable environmental upheaval that shaped the present worldwide distribution of biodiversity. Environmental changes during the Pleistocene included cyclical fluctuations in global temperatures and precipitation patterns, advances and recessions of high-latitude ice sheets, and dramatic changes in sea-level, together with large-scale volcanism, paleomagnetic reversals, and the global spread of humans (1). These events

altered habitats around the world, driving changes in resource availability and ecological community composition.

The rich fossil record of the Pleistocene has been instrumental for testing hypotheses about correlation between these environmental changes and biodiversity dynamics, especially at high latitudes where the cold climate favors fossil preservation. This is particularly true for the Late Pleistocene (126-11.7 ka), thanks to fine-scale inferences enabled by ancient DNA preserved in fossils dating to this period. Such inferences have allowed insights into population turnover (2–4) and inter-species gene flow (5)–processes that are invisible to traditional paleontological techniques–and shown that demographic trends in large mammals closely track available habitat (6).

Technical advances in DNA recovery have extended the ability to make these inferences deeper into the Pleistocene. DNA from bones and teeth that are several hundreds of thousands of years old (7–9) and beyond one million years old (10) has now been recovered and analyzed (Fig. 1). Such deep-time paleogenomes, which we consider here to refer to genomes assembled from organisms that lived during or earlier than the Middle Pleistocene, i.e. >126 ka, are still rare because post-mortem processes lead to successive degradation of DNA molecules into increasingly small fragments, making DNA recovery more difficult with age. Early and Middle Pleistocene DNA has, however, been recovered from remains and sediments in high-latitude permafrost (10–14) and lower latitude caves (15, 16), suggesting that deep-time genomics is feasible in ideal preservation environments. Here, we explore the current state-of-the-art in deep-time paleogenomics research, the key obstacles preventing wider adoption, and scientific questions that deep-time paleogenomics can address.

DNA persistence into deep time

DNA does not survive indefinitely, but it does survive for significantly longer than the earliest models predicted. In 1993, Lindahl estimated that hydrolytic depurination would lead to complete degradation of DNA molecules within several tens of thousands of years (17). This limit has since been exceeded, and DNA is regularly recovered from remains and sediments that date to within the last 100 ka. As of May 2023, the oldest reconstructed paleogenome is from a permafrost-preserved mammoth dating to 1-2 Ma (10) and the oldest isolated DNA is from ~2 Ma sediment from northern Greenland (11), but the maximum age of recoverable and useful DNA molecules–those that are long enough to retain information– remains uncertain.

DNA begins to degrade immediately following organismal death, initially through microbial and endogenous nuclease activity (Fig. 2). In nuclear DNA, strands are cleaved in labile regions of histone-DNA complexes, resulting in a ~10-base periodicity in the distribution of the lengths of recovered molecules (18). The primary chemical mechanism of DNA fragmentation is hydrolytic depurination. This process removes adenine or guanine bases, creating abasic sites that can be cleaved by β elimination (19; Fig. 2C), and leading to purine overrepresentation adjacent to strand breaks (20; Fig. 2E) and interior gaps (21). Hydrolytic deamination, another common form of chemical damage, converts cytosine to

uracil and is observed as thymine in sequencing data, or "C-to-T transitions" (Fig. 2C). Deamination occurs primarily near strand ends and in single-stranded DNA (17, 21, 22; Fig. 2E). DNA crosslinking (19, 22) and oxidative damage (20, 23) also occur but are observed less frequently than depurination and deamination. These typical damage patterns can be used to bioinformatically corroborate the authenticity of recovered ancient sequences and, to reduce their impact on sequence accuracy, can be identified and removed from ancient DNA data sets using standard bioinformatic approaches.

Recovery of increasingly old and damaged DNA is possible in part due to technical advances in the laboratory. Ancient DNA isolation methods are optimized to recover both short DNA molecules and molecules containing nicks and gaps. Extracted molecules are prepared for sequencing by ligating platform-specific adapters to either double-stranded or single-stranded DNA. Single-stranded approaches to genomic library preparation (24, 25) convert natively single-stranded DNA as well as double stranded DNA and more effectively convert molecules containing nicks and gaps compared to double-stranded approaches. DNA extracts are also often treated with uracil DNA glycosylase and endonuclease VIII to reduce deamination damage by removing uracil bases (26). While this approach reduces damage-induced errors in the resulting sequencing data, it also cuts the DNA backbone at abasic sites and shortens the recovered molecules by 5-10 nucleotides (26). As deep-time DNA molecules are short, often <35 bases (15), this may reduce the proportion of useful endogenous DNA.

The short nature of deep-time DNA molecules makes them prone to spurious alignment and reference bias (27), complicating genome assembly and analysis. For example, ancient DNA data sets comprise both endogenous DNA from the target organism(s) and introduced exogenous DNA. These categories of molecules can be separated by identifying each read via taxonomic assignment, which can be problematic if the ancient organism has no close living relative to act as a genomic reference. Lack of a close reference, reference bias, and errors introduced by damage will also impede variant and consensus calling. Bioinformatic approaches mitigate these challenges by directly modeling DNA damage and/or bias as part of genotyping (28), or considering only substitutions that are not impacted by cytosine deamination. Reference genomes can also be modified to create artificially closer references, such as a "Neandertalized" version of the human reference genome for reference-guided mapping of Neandertal reads (29). Genotype likelihoods rather than strictly called genotypes can also be used during downstream analysis, although imputation-based analytical methods may be inappropriate for deep-time data sets if ancient genomic diversity is not represented in existing reference panels.

Research opportunities arising from deep-time DNA

Speciation and evolution

Speciation is not always a simple process of cladogenesis followed by reproductive isolation. Instead, modern and paleogenomic data have shown that interspecific hybridization is surprisingly common and perhaps driven in part by repeated habitat redistribution associated with glacial cycles (5, 9, 10). For example, brown bears and polar bears hybridize today and also hybridized during previous glacial and interglacial periods

(30, 31). Recently, polar bear and cave bear paleogenomes dating to up to 360 ka revealed that all living brown bears derive a portion of their ancestry from admixture with these other bear lineages–evolutionary events that were invisible without these paleogenome (9, 32). Similarly, a mammoth paleogenome dating to the Early Pleistocene revealed that Columbian mammoths originated after hybridization between two distinct ancient mammoth lineages (10; Fig. 3). Taxonomically diverse deep-time paleogenomes could clarify the timing, rate, and extent of genomic introgression episodes and their role in evolution. Paleogenomic data from species that went extinct during the Early and Middle Pleistocene, such as short-faced hyenas, European jaguars, and the enigmatic *Xenocyon* canids, could shed light on whether these taxa contributed to the genetic make-up of living carnivores. Deep-time paleogenomes could also identify unknown "ghost" lineages that contributed to species' ancestries, as exemplified in the paleogenomic characterization of the Krestovka mammoth (10; Box 1, Fig. 3).

Deep-time DNA can also reveal genomic snapshots of a species' entire evolutionary story (Box 1). As many temperate and cold-adapted birds and mammals trace their origin to the Early and Middle Pleistocene (33, 34), paleogenomes from these species could correlate evolutionary changes to specific environmental perturbations, such as transitions between climate regimes or community reshuffling. The process of speciation can be investigated as it happens, exploring founder event bottlenecks and testing whether speciation occurred through strict allopatry or gradually with post-divergence gene flow. As deep-time paleogenomes tend to occupy basal phylogenetic positions within their clades, they can also provide important calibrations for estimating rates of molecular evolution. For example, paleogenomic data from a Middle Pleistocene hominin from Sima de los Huesos in present-day Spain confirmed hypotheses from Late Pleistocene genomes that Neandertals and Denisovans diverged during the early Middle Pleistocene (35), whereas the inclusion of a ~700 ka horse paleogenome in the equid phylogeny pushed the estimated time for the origin of living equids to more than twice as old as previously hypothesized (8).

Deep-time paleogenomes can also be used to test hypotheses about relationships between species and how derived forms are related to earlier forms. An outstanding question in paleontology is whether fossil morphospecies are true species, synchronous ecomorphs, or chronospecies that were direct ancestors of succeeding species. A paleogenomic study of ancient North American bison dating to ~130-110 ka, for example, showed that two samples exhibiting extreme size dimorphism and representing supposedly distinct species– the longhorn bison and the steppe bison–actually belong to the same lineage that dispersed into North America only a few tens of thousands of years earlier (36). Conversely, deep-time paleogenomics can also give context to species for which we have only limited remains, such as Denisovans (35).

Finally, paleogenomes across deep time-scales will also make it possible to explore aspects of adaptive evolution. At the most basic level, deep-time genomes can help identify when adaptive mutations arose. For example, comparative analysis of mammoth paleogenomes ranging from a few thousand to more than a million years old identified genes associated with hair and skin development, fat storage and metabolism, immune system function, and body size that evolved in that lineage within the last 700 ka (37). Paleogenomes will also

allow exploration of how the rate of protein-coding changes varies over time, such as in conjunction with past changes in climate, as well as to assess when genomic deletions arose and the rate of positive and purifying selection in introgressed genomic regions.

The impact of glacial cycles on biodiversity

Nearly all ancient DNA studies to date have for practical reasons focused on Late Pleistocene or more recent materials (Fig. 1). Thus, our current understanding of evolutionary processes during the Pleistocene mostly relies on more traditional approaches, including morphometrics, stable isotope analysis, and pollen records. This is despite the fact that the majority of the Pleistocene glacial oscillations occurred during the Early (2.6 Ma - 780 ka) and Middle (780 - 126 ka) Pleistocene sub-epochs, which are now accessible with deep-time paleogenomes.

A special attribute of the Pleistocene is the change in periodicity of glaciations from ~40 ka cycles to ~100 ka cycles that occurred 1.2-0.7 Ma (38) (Fig. 1). This change isolated temperate species in glacial refugia for longer periods, providing more time for local adaptation and increasing the rate of population divergence. Biological communities may also have been reshuffled following this periodicity change, as the longer and higher amplitude glaciations allowed sufficient ice sheet accumulation for the Bering Land Bridge to form, making land dispersal between Eurasia and North America possible.

Since the change in glacial periodicity, the dominant pattern has been cycles of long glaciations separated by short warm interglacials. This pattern is believed to have driven the demography and range dynamics of many species (39). Long interglacials, for example, have been correlated with bottlenecks in cold-adapted taxa (40) and expansion and speciation in warm-adapted taxa (41). Of particular interest is the unusually long interglacial that occurred 420-370 ka (Marine Isotope Stage 11) (42). Paleogenomes from individuals that lived during and earlier than this long bottleneck could test these hypotheses and reveal evolutionary changes that may have been overwritten by subsequent genetic bottlenecks.

Inference of ancient ecosystems

Above, we describe insights potentially derived from DNA extracted from remains of individuals that lived during the Middle Pleistocene and earlier. However, the advances that enable deep-time paleogenomics also make it possible to reconstruct entire deep-time ecological communities. To date, only five studies have attempted to use sedimentary ancient DNA to reconstruct plant and/or animal communities dating to the Middle Pleistocene or older: Kjær et al (11) reconstructed components an Early Pleistocene interglacial ecosystem from a sediment core extracted from the present day polar desert in northern Greenland, Armbrecht et al (43) reconstructed an Early to Middle Pleistocene marine ecosystem from Iceberg Alley in the Southern Ocean, Courtin et al (12) reconstructed a Middle Pleistocene interglacial ecosystem from a permafrost megaslump in Eastern Siberia, and Willerslev et al reconstructed Middle Pleistocene plant communities from sediments collected below the Greenland ice sheet (14) and from coastal Siberian permafrost (13). Among these, Kjær et al and Armbrecht et al enriched libraries for sequences of interest via hybridization to synthesized baits designed to target Arctic or

Antarctic taxa. In contrast to metabarcoding methods, which use targeted PCR amplification, hybridization-based targeted enrichment can capture molecules of any length and are therefore powerful even when preserved molecules are short. While this approach is limited today to capturing sequences that are genetically similar to other known taxa, methodological improvements in hybridization capture is a ripe area of research that will no doubt expand access to deep-time sedimentary DNA.

Deep-time sedimentary DNA research will allow better understanding of the effect of glacial-interglacial transitions on community composition. Reconstructions of communities spanning the transition into the present Holocene, for example, have revealed rapid biological turnover that closely tracked abiotic changes (44, 45). Comparison with older transitions will test whether patterns are predictable or idiosyncratic, whether some species or communities are more resilient to environmental upheaval than others, and whether some transitions or events leave lasting signatures on community biodiversity.

Reconstructions of communities that thrived in past warm interglacials may provide insight into the potential composition of communities in a future, warmer world (11), and improve our understanding of how ecosystem-level interactions among species evolve and are maintained. They also enrich our understanding of these extinct ecosystems beyond what is knowable from the fossil record. Deep-time sedimentary DNA from northern Greenland, for example, revealed a mastodon or mastodon-like animal was part of the Early Pleistocene community (11) despite that no fossil remains from such an animal have been discovered. Deep-time sedimentary DNA can also reveal past connectivity among populations, as in a recent study of Late Pleistocene sedimentary DNA from a cave in Mexico that linked an extinct population of black bears to living populations in eastern North America (46). As technologies improve, in particular those that allow increasingly sensitive targeted enrichment, we envisage deep-time sedimentary DNA as a powerful tool to explore the ecological and evolutionary consequences of environmental change on community-level biodiversity.

Future research to enable deep-time DNA

It has been shown that DNA can survive in ideal preservation conditions into at least the Early Pleistocene. The next phase of deep-time DNA research is to expand the taxonomic, geographic, and temporal range of recovered and authenticated deep-time DNA. This challenge presents new research opportunities in the field, at the bench, and bioinformatically.

Deep-time genomics is today mostly conducted on substrates with optimal DNA preservation such as those derived from permafrost or caves. However, more efficient approaches to recover ancient DNA molecules will continue to expand the range of samples and substrates suitable for analysis. Today, methods for DNA extraction and library conversion do not recover all potentially preserved DNA molecules. For example, Kjaer et al (11) found that DNA adsorbed preferentially to clay mineral surfaces compared to non-clay surfaces, and in particular to the clay mineral smectite, which can bind 200 times more DNA than quartz and is a common mineral in terrestrial samples. Their best

performing extraction protocol recovered 40% of DNA bound to quartz and only 5% of DNA bound to smectite, suggesting the majority of DNA was inaccessible. While anecdotal, this observation points to several opportunities for improving deep-time DNA research, including using mineralogical characterization to identify the most promising sites for deep-time sedimentary DNA recovery and refining experimental approaches to recover DNA bound to all mineral surfaces. In the absence of improved methods to release bound DNA, microscopic evaluation of sedimentary samples will improve the efficiency of DNA recovery. Massilani et al (47), for example, showed DNA preserved in cave sediment is concentrated in micro-scale particles, especially fragments of bone and feces preserved within the substrate.

Library conversion protocols could also be made more efficient. Optimized library conversion protocols use enzymatic ligation and polymerization, but ancient DNA extracts contain inhibitors as well as molecules with uncharacterized DNA damage. Although we can convert as little as 100 picograms of DNA into libraries using the Santa Cruz method (25), library preparation has been shown to typically convert only around 10-50% of extracted DNA (21), suggesting that most recovered molecules are lost at this experimental step. Improvements in library preparation may include engineering more robust enzymes to combat inhibitors or developing protocols that incorporate enzymatic repair during library conversion. Additionally, reducing reliance on ligase and polymerase steps through alternative enzymatic strategies, bioorthogonal chemistry, or native DNA sequencing may offer new approaches to convert currently unsequenceable DNA molecules.

Many species that are obvious targets for deep-time DNA research are extinct, and some, such as *Xenocyon* canids and basal members of the elephant and horse families, have no evolutionarily close living relative for which an ideal reference genome can be produced. This presents challenges to ancient DNA authentication and identification as well as to reference-guided genome assembly. Although the average fragment length of deep-time DNA sequences is short, it may be possible to generate de novo assemblies from ancient extracts by capitalizing on methods that use chromosome conformation capture to retain proximity information useful to link short reads within a chromosome (48). Approaches that sequence DNA in situ (49) are also promising, but in early stages of development. Improvements in bioinformatic processing will also benefit eukaryotic paleogenomic reconstruction and variant calling. Recently, microbial genomes were assembled from DNA recovered from relatively recent paleofecal samples (50) and from archaeological dental calculus dating to as old as 100 ka (51), suggesting a bioinformatic path toward de novo assembly of some small paleogenomes. While this approach is not likely to apply to complex eukaryotic genomes, other bioinformatic approaches can improve the accuracy of these assemblies from short read data. Replacing linear single-species reference genomes with multi-species variation graphs that incorporate variants from several genomes (52), for example, can increase the number of reads that map to a reference genome. This approach has the additional benefit of allowing variation among indel lengths as well as among nucleotides. Iterative assembly approaches, such as the mapping-iterative-assembler used to generate the first Neanderthal mitochondrial genome (53) may improve mapping to more complex genomes. Finally, as reference-based taxonomic assignment is always limited to sequences deposited in public databases, the ongoing population of these databases

will continue to improve robust identification of DNA recovered from Early and Middle Pleistocene remains and sediments.

A considerable challenge for studies of deep-time DNA is to know how old samples are so that they can be placed into broader evolutionary and geological contexts. As most ancient DNA to date is from organisms that lived relatively recently, it is usually possible to estimate their age directly using radiocarbon dating. However, the short radioactive half-life of carbon-14 means that age estimates are often unreliable if organisms lived more than ~50 ka. For samples older than ~50 ka, alternative methods are necessary. Trapped charge dating methods, such as electron spin resonance (ESR) for tooth enamel or luminescence approaches for minerals such as quartz and feldspar, can provide age estimates for samples dating throughout the Pleistocene, but require that sediments have remained undisturbed since burial (for a review see 54). When proteins are preserved, the extent of amino acid racemization, hydrolysis, and decay can also estimate time since death, although amino acid "clocks" vary among species and localities (54).

In some cases, paleoenvironmental, geological, and geophysical markers can provide clues about a sample's age. A fossil might be found in the Arctic with other paleoecological proxies that suggest a warm and wet environment, for example, indicating that the animal lived during a previous interglacial, or in sediments with reversed polarity, suggesting that it lived prior to the last paleomagnetic reversal some 780 ka. In some environments, tephra beds–layers of fine, settled, volcanic ash–can be dated by methods including glass fission-track and argon-argon dating. Tephra beds, which can be detected even when present in only microscopic amounts (55), have been particularly important in dating sediment cores, but can also provide contextual clues about the age of samples found *in situ* at sites where the tephra is present. As volcanic eruptions were common throughout the Pleistocene, improved tephrochronology for the Early and Middle Pleistocene will help place deep-time DNA into a chronological context.

Other approaches to dating deep-time genomes might rely on the predictable nature of evolutionary change in organisms. Molecular clock methods infer the age of paleogenomes by estimating the amount of "missing" evolution along a phylogenetic branch leading to the paleogenome, often called "branch shortening" (56). Because the accumulation of mutations is approximately constant over time, the differences between these branch lengths should correspond to the number of generations that separate the represented paleogenome from extant or more recent individuals. To translate missing generations into calendar time, however, the branch shortening approach requires either an independent fossil calibration or an estimate of generation length. For many lineages that lived during the Early and Middle Pleistocene, dated ancestral fossils are few and, with no close living relatives, estimates of generation time would be imprecise. Variation among evolutionary rates between distantly related lineages may also reduce the power of a comparative molecular dating approach. Nonetheless, development of approaches that use genomic information to estimate the age of paleogenomes and their evolutionary relationships to other species is a rich area for future research.

Conclusion

The next decade will bring continued technical advances that will increase the taxonomic and geographic range of deep-time paleogenomes and deep-time ancient sedimentary DNA data sets. Most crucially, new insights into what substrates are likely to preserve deep-time ancient DNA as well as refined approaches to release DNA bound to biological or mineralogical matrices will increase the number and taxonomic range of recoverable deep-time paleogenomes. These will need to be placed into chronological context, which will be addressed with developments in geochronology and paleoecology together with increasingly powerful computational tools to estimate the age of samples using a molecular clock. The resulting deep-time DNA will enable increasingly detailed reconstruction of evolutionary history across repeated environmental perturbations, refining understanding of adaptive evolution, community organization, and ecosystem resilience. Moreover, as the past by its nature is different from anything that exists today, access to deep-time DNA provides ample yet unpredictable opportunities for scientific discovery.

Acknowledgements

References and Notes

1. Ehlers, J, Gibbard, PL, Hughes, PD. Quaternary Glaciations-Extent and Chronology: A Closer Look. Elsevier; 2011.

2. Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, Valentin F, Thevenet C, Furtwängler A, Wißing C, Francken M, et al. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. Curr Biol. 2016; 26: 827–833. [PubMed: 26853362]

3. Baca M, Popović D, Agadzhanyan AK, Baca K, Conard NJ, Fewlass H, Filek T, Golubiński M, Horáček I, Knul MV, Krajcarz M, et al. Ancient DNA of narrow-headed vole reveal common features of the Late Pleistocene population dynamics in cold-adapted small mammals. Proc Biol Sci. 2023; 290: 2022–2238.

4. Loog L, Thalmann O, Sinding M-HS, Schuenemann VJ, Perri A, Germonpré M, Bocherens H, Witt KE, Samaniego JA, Castruita Velasco MS, Lundstrøm IKC, et al. Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. Mol Ecol. 2020; 29: 1596–1610. [PubMed: 31840921]

5. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010; 328: 710–722. [PubMed: 20448178]

6. Foote AD, Kaschner K, Schultze SE, Garilao C, Ho SYW, Post K, Higham TFG, Stokowska C, van der Es H, Embling CB, Gregersen K, et al. Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. Nat Commun. 2013; 4 1677 [PubMed: 23575681]

7. Valdiosera C, García N, Dalén L, Smith C, Kahlke R-D, Lidén K, Angerbjörn A, Arsuaga JL, Götherström A. Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. Biol Lett. 2006; 2: 601–603. [PubMed: 17148299]

8. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PLF, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013; 499: 74–78. [PubMed: 23803765]

9. Barlow A, Paijmans JLA, Alberti F, Gasparyan B, Bar-Oz G, Pinhasi R, Foronova I, Puzachenko AY, Pacher M, Dalén L, Baryshnikov G, et al. Middle Pleistocene genome calibrates a revised evolutionary history of extinct cave bears. Curr Biol. 2021. [PubMed: 33592193]

10. van der Valk T, Pečnerová P, Díez-Del-Molino D, Bergström A, Oppenheimer J, Hartmann S, Xenikoudakis G, Thomas JA, Dehasque M, Sağlican E, Fidan FR, et al. Million-year-old DNA sheds light on the genomic history of mammoths. Nature. 2021; 591: 265–269. [PubMed: 33597750]

11. Kjær KH, Winther M, De Pedersen B, De Sanctis B, Cahsan Korneliussen TS, Michelsen CS, Sand KK, Jelavić S, Ruter AH, Schmidt AMA, Kjeldsen KK, et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature. 2022; 612: 283–291. [PubMed: 36477129]

12. Courtin J, Perfumo A, Andreev AA, Opel T, Stoof-Leichsenring KR, Edwards ME, Murton JB, Herzschuh U. Pleistocene glacial and interglacial ecosystems inferred from ancient DNA analyses of permafrost sediments from Batagay megaslump, East Siberia. Environmental DNA. 2022; 4: 1265–1283.

13. Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C, Gilichinsky DA, Cooper A. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science. 2003; 300: 791–795. [PubMed: 12702808]

14. Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB, Hofreiter M, Bunce M, Poinar HN, Dahl-Jensen D, Johnsen S, et al. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science. 2007; 317: 111–114. [PubMed: 17615355]

15. Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J-L, Martínez I, Gracia A, de Castro JMB, Carbonell E, Pääbo S. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature. 2014; 505: 403–406. [PubMed: 24305051]

16. Zavala EI, Jacobs Z, Vernot B, Shunkov MV, Kozlikin MB, Derevianko AP, Essel E, de Fillipo C, Nagel S, Richter J, Romagné F, et al. Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature. 2021; 595: 399–403. [PubMed: 34163072]

17. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993; 362: 709–715. [PubMed: 8469282]

18. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, Andersson R, et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. 2014; 24: 454–466. [PubMed: 24299735]

19. Hansen AJ, Mitchell DL, Wiuf C, Paniker L, Brand TB, Binladen J, Gilichinsky DA, Rønn R, Willerslev E. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. Genetics. 2006; 173: 1175–1179. [PubMed: 16582426]

20. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A. 2007; 104: 14616–14621. [PubMed: 17715061]

21. Bokelmann L, Glocke I, Meyer M. Reconstructing double-stranded DNA fragments on a single-molecule level reveals patterns of degradation in ancient samples. Genome Res. 2020; 30: 1449–1457. [PubMed: 32963029]

22. Dabney J, Meyer M, Pääbo S. Ancient DNA Damage. Cold Spring Harb Perspect Biol. 2013; 5 a012567 [PubMed: 23729639]

23. Hempel E, Bibi F, Faith JT, Koepfli K-P, Klittich AM, Duchêne DA, Brink JS, Kalthoff DC, Dalén L, Hofreiter M, Westbury MV. Blue Turns to Gray: Paleogenomic Insights into the Evolutionary History and Extinction of the Blue Antelope (Hippotragus leucophaeus). Mol Biol Evol. 2022; 39 doi: 10.1093/molbev/msac241

24. Gansauge M-T, Gerber T, Glocke I, Korlević P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. 2017; 45: e79. [PubMed: 28119419]

25. Kapp JD, Green RE, Shapiro B. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. J Hered. 2021; 112: 241–249. [PubMed: 33768239]

26. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. 2010; 38: e87. [PubMed: 20028723]

27. de Filippo C, Meyer M, Prüfer K. Quantifying and reducing spurious alignments for the analysis of ultra-short ancient DNA sequences. BMC Biol. 2018; 16: 121. [PubMed: 30359256]

28. Prüfer K. snpAD: an ancient DNA genotype caller. Bioinformatics. 2018; 34: 4165–4171. [PubMed: 29931305]

29. Peyrëgne S, Slon V, Mafessoni F, de Filippo C, Hajdinjak M, Nagel S, Nickel B, Essel E, Le Cabec A, Wehrberger K, Conard NJ, et al. Nuclear DNA from two early Neandertals reveals 80,000 years of genetic continuity in Europe. Sci Adv. 2019; 5 eaaw5873 [PubMed: 31249872]

30. Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE, Tomsho LP, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A. 2012; 109: E2382–90. [PubMed: 22826254]

31. Cahill JA, Heintzman PD, Harris K, Teasdale MD, Kapp J, Soares AER, Stirling I, Bradley D, Edwards CJ, Graim K, Kisleika AA, et al. Genomic Evidence of Widespread Admixture from Polar Bears into Brown Bears during the Last Ice Age. Mol Biol Evol. 2018; 35: 1120–1129. [PubMed: 29471451]

32. Wang M-S, Murray GGR, Mann D, Groves P, Vershinina AO, Supple MA, Kapp JD, Corbett-Detig R, Crump SE, Stirling I, Laidre KL, et al. A polar bear paleogenome reveals extensive ancient gene flow from polar bears into brown bears. Nat Ecol Evol. 2022; 6: 936–944. [PubMed: 35711062]

33. Weir JT, Schluter D. Ice sheets promote speciation in boreal birds. Proc Biol Sci. 2004; 271: 1881–1887. [PubMed: 15347509]

34. Lister AM. The impact of Quaternary Ice Ages on mammalian evolution. Philos Trans R Soc Lond B Biol Sci. 2004; 359: 221–241. [PubMed: 15101579]

35. Meyer M, Arsuaga J-L, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, Martinez I, Gracia A, Bermúdez de Castro JM, Carbonell E, Viola B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016; 531: 504–507. [PubMed: 26976447]

36. Froese D, Stiller M, Heintzman PD, Reyes AV, Zazula GD, Soares AER, Meyer M, Hall E, Jensen BJL, Arnold LJ, MacPhee RDE, et al. Fossil and genomic evidence constrains the timing of bison arrival in North America. Proc Natl Acad Sci U S A. 2017; 114: 3457–3462. [PubMed: 28289222]

37. Díez-Del-Molino D, Dehasque M, Chacón-Duque JC, Pečnerová P, Tikhonov A, Protopopov A, Plotnikov V, Kanellidou F, Nikolskiy P, Mortensen P, Danilov GK, et al. Genomics of adaptive evolution in the woolly mammoth. Curr Biol. 2023; doi: 10.1016/j.cub.2023.03.084

38. Clark PU, Archer D, Pollard D, Blum JD, Rial JA, Brovkin V, Mix AC, Pisias NG, Roy M. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO2. Quat Sci Rev. 2006; 25: 3150–3184.

39. Stewart JR, Lister AM, Barnes I, Dalén L. Refugia revisited: individualistic responses of species in space and time. Proc Biol Sci. 2010; 277: 661–671. [PubMed: 19864280]

40. Lord E, Marangoni A, Baca M, Popović D, Goropashnaya AV, Stewart JR, Knul MV, Noiret P, Germonpré M, Jimenez E-L, Abramson NI, et al. Population dynamics and demographic history of Eurasian collared lemmings. BMC Ecol Evol. 2022; 22: 126. [PubMed: 36329382]

41. Ortego J, Lacey Knowles L. Geographical isolation versus dispersal: Relictual alpine grasshoppers support a model of interglacial diversification with limited hybridization. Mol Ecol. 2022; 31: 296–312. [PubMed: 34651368]

42. Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography. 2005; 20 doi: 10.1029/2004pa001071

43. Armbrecht L, Weber ME, Raymo ME, Peck VL, Williams T, Warnock J, Kato Y, Hernández-Almeida I, Hoem F, Reilly B, Hemming S, et al. Ancient marine sediment DNA reveals diatom transition in Antarctica. Nat Commun. 2022; 13 5787 [PubMed: 36184671]

44. Monteath AJ, Kuzmina S, Mahony M, Calmels F, Porter T, Mathewes R, Sanborn P, Zazula G, Shapiro B, Murchie TJ, Poinar HN, et al. Relict permafrost preserves megafauna, insects, pollen,

soils and pore-ice isotopes of the mammoth steppe and its collapse in central Yukon. Quaternary Science Reviews. 2023; 299 107878

45. Clarke CL, Edwards ME, Gielly L, Ehrich D, Hughes PDM, Morozova LM, Haflidason H, Mangerud J, Svendsen JI, Alsos IG. Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. Sci Rep. 2019; 9 19613 [PubMed: 31873100]

46. Pedersen MW, De Sanctis B, Saremi NF, Sikora M, Puckett EE, Gu Z, Moon KL, Kapp JD, Vinner L, Vardanyan Z, Ardelean CF, et al. Environmental genomics of Late Pleistocene black bears and giant short-faced bears. Curr Biol. 2021; 31: 2728–2736. e8 [PubMed: 33878301]

47. Massilani D, Morley MW, Mentzer SM, Aldeias V, Vernot B, Miller C, Stahlschmidt M, Kozlikin MB, Shunkov MV, Derevianko AP, Conard NJ, et al. Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. Proc Natl Acad Sci U S A. 2022; 119 doi: 10.1073/pnas.2113666118

48. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. Nat Biotechnol. 2013; 31: 1143–1147. [PubMed: 24270850]

49. Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao C-C, Markoulaki S, Earl AS, Labade AS, Jaenisch R, Church GM, et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. Science. 2021; 371 doi: 10.1126/science.aay3446

50. Wibowo MC, Yang Z, Borry M, Hübner A, Huang KD, Tierney BT, Zimmerman S, Barajas-Olmos F, Contreras-Cubas C, García-Ortiz H, Martínez-Hernández A, et al. Reconstruction of ancient microbial genomes from the human gut. Nature. 2021; 594: 234–239. [PubMed: 33981035]

51. Klapper M, Hübner A, Ibrahim A, Wasmuth I, Borry M, Haensch VG, Zhang S, Al-Jammal WK, Suma H, Fellows Yates JA, Frangenberg J, Velsko IM, et al. Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. Science. 2023; 380: 619–624. [PubMed: 37141315]

52. Martiniano R, Garrison E, Jones ER, Manica A, Durbin R. Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. Genome Biol. 2020; 21: 250. [PubMed: 32943086]

53. Green RE, Malaspinas A-S, Krause J, Briggs AW, Johnson PLF, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, Prüfer K, et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell. 2008; 134: 416–426. [PubMed: 18692465]

54. Penkman KEH, Duller GAT, Roberts HM, Colarossi D, Dickinson MR, White D. Dating the Paleolithic: Trapped charge methods and amino acid geochronology. Proc Natl Acad Sci U S A. 2022; 119 e2109324119 [PubMed: 36252044]

55. Davies SM. Cryptotephras: the revolution in correlation and precision dating. J Quat Sci. 2015; 30: 114–130. [PubMed: 27512240]

56. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012; 338: 222–226. [PubMed: 22936568]

57. Lister AM. On the type material and evolution of North American mammoths. Quat Int. 2017; 443: 14–31.

58. Fellows Yates JA, Andrades Valtueña A, Vågene ÅJ, Cribdon B, Velsko IM, Borry M, Bravo-Lopez MJ, Fernandez-Guerra A, Green EJ, Ramachandran SL, Heintzman PD, et al. Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. Sci Data. 2021; 8: 31. [PubMed: 33500403]

59. Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, Meyer M. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A. 2013; 110: 15758–15763. [PubMed: 24019490]

Box 1

Deep-time mammoth DNA and the inference of a lineage's entire evolutionary story

The power of a deep-time genomic approach was showcased in a study by van der Valk and colleagues (10) in which genome-wide data collected from three Siberian mammoths, dated to approximately 700 ka to 1.2 Ma, made it possible to examine mammoth evolution from multiple genomic angles:

A new lineage

The oldest of the mammoth specimens, Krestovka, belonged to a previously unknown and divergent evolutionary lineage. This implied that two distinct lineages of mammoth, Krestovka and the ancestors of woolly mammoths, lived in Siberia during the later stages of the Early Pleistocene. The analysis also revealed that mammoths belonging to the Krestovka lineage were the first mammoths to colonize North America 1.5-1.2 Ma (57).

Hybridization

Multiple lines of evidence suggested that the Columbian mammoth originated as a result of hybridization between the Krestovka lineage and early woolly mammoths. This hybridization took place as woolly mammoths expanded into North America during the Middle Pleistocene, after the Krestovka lineage was already established on the continent (Fig. 3). Columbian mammoths derive approximately 50% ancestry from each of these two lineages.

Adaptive evolution

The deep-time nature of the mammoth data set allowed van der Valk et al to estimate the rate of adaptive evolution in mammoths. Their results suggested that the evolutionary origin of the woolly mammoth lineage did not coincide with an increased rate of protein-coding changes and therefore higher rates of positive selection across the genome (10). Subsequent analyses identified a suite of genes that underwent protein-coding changes during the last 700 ka and were thus unique to woolly mammoths (37).

Teaser

A review of research and discovery opportunities by leveraging paleogenomes from the Early and Middle Pleistocene.
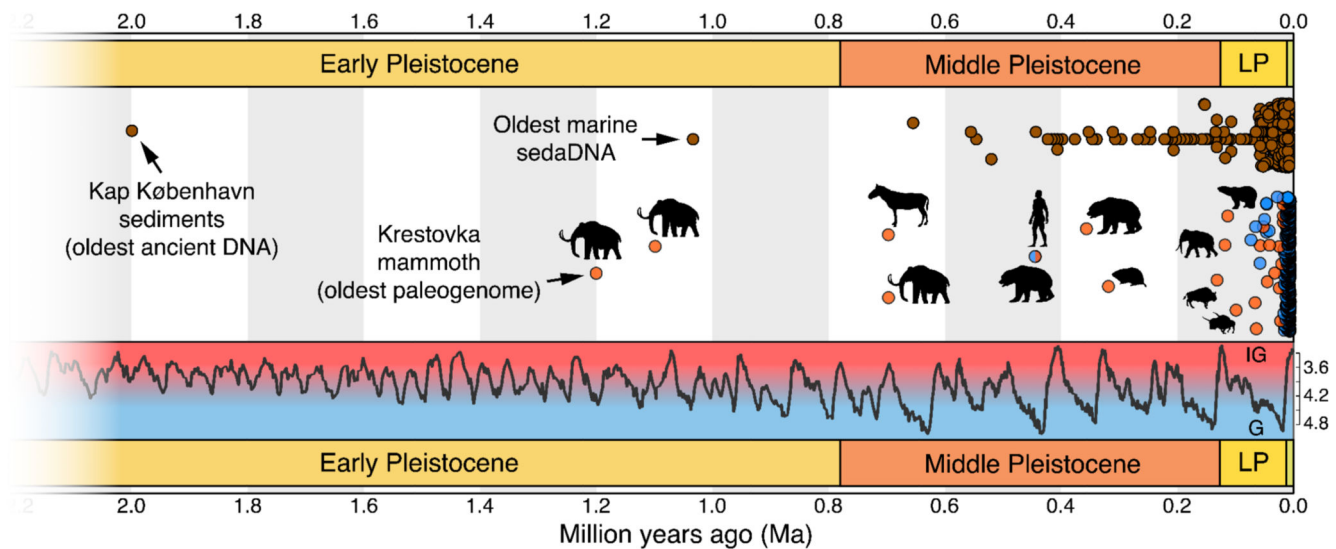
Fig. 1. The temporal distribution of ancient DNA studies to date highlights gaps and opportunities for deep-time paleogenomics and sedimentary ancient DNA.

Circles in orange are non-human animal paleogenomes, in blue are hominin paleogenomes, and in brown are sedimentary ancient DNA records. Most ancient DNA studies fall within the last 50 ka and the most recent glacial cycle. The climate curve is based on benthic δ^{18}-Oxygen measurements (per mil, ‰o, LR04 stack from (42). Sedimentary ancient DNA data are from the AncientMetagenomeDir (v23.06.0, 58) and von Eggers et al. (v1, https://doi.org/10.5281/zenodo.6847522), with metabarcoding records older than one million years excluded. Paleogenomes older than 100 ka are annotated with a silhouette of the study taxon, with the deep-time paleogenomes including a 130 ka steppe bison (36); 330 ka collared lemming (40); 360 ka cave bear (9); 430 ka cave bear and hominin (35, 59); 700 ka horse (8); and 700 ka, 1.1 Ma, and 1.2 Ma mammoths (10). Silhouettes are from PhyloPic (https://beta.phylopic.org/) and are in the public domain with credits to Zimices (mammoth, two bison) and Robert Bruce Horsfall (horse). LP: Late Pleistocene; IG: Interglacial; G: Glacial.
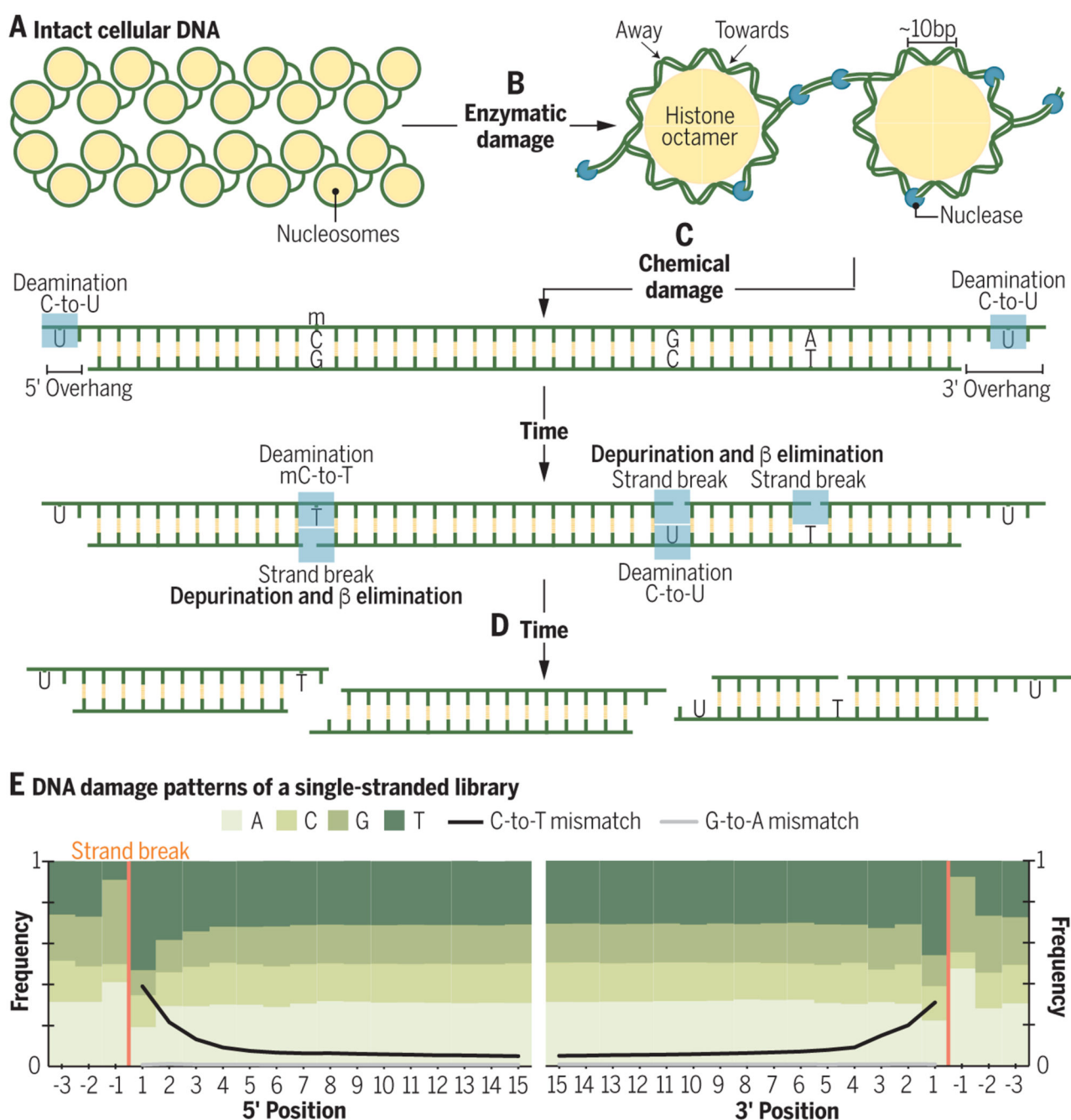
Fig. 2. DNA fragmentation and degradation begins after death and continues until fragments are too short to be useful.

(**A**) The integrity of megabase length DNA is maintained by a cell's enzymatic repair machinery and, in eukaryotic genomes, packaged in histone-DNA complexes. (**B**) Following death, repair stops and DNA damage begins to accumulate. Nucleases and microorganisms cleave DNA in labile regions between nucleosomes and when the DNA backbone faces away from histones. (**C**) Over time, chemical damage also accumulates. Cytosine bases are converted to uracil and methylated cytosines are converted to thymines (by deamination).

Cytosines are particularly vulnerable to deamination in single-stranded regions such as in overhanging regions at DNA termini, but deamination is possible in some double-stranded contexts. Fragmentation occurs after the loss of purine bases (depurination), creating abasic sites that can be cleaved by β elimination. Depurination and β elimination create a region of single-stranded DNA, which leaves cytosines vulnerable to deamination.. (**D**) Given enough time, DNA molecules will become too short to be identifiable. (**E**) A summary of base and mismatch frequencies along the initial 15 5' and 3' bases of reads generated using a single-stranded DNA library protocol. Depurination leads to overrepresentation of adenine and guanine bases adjacent to strand breaks. C-to-T mismatches are elevated near read ends and observed throughout damaged reads. While 3' G-to-A mismatches are observed in double-stranded libraries, single-stranded libraries show a C-to-T signal at both ends by retaining the native termini of the molecules.
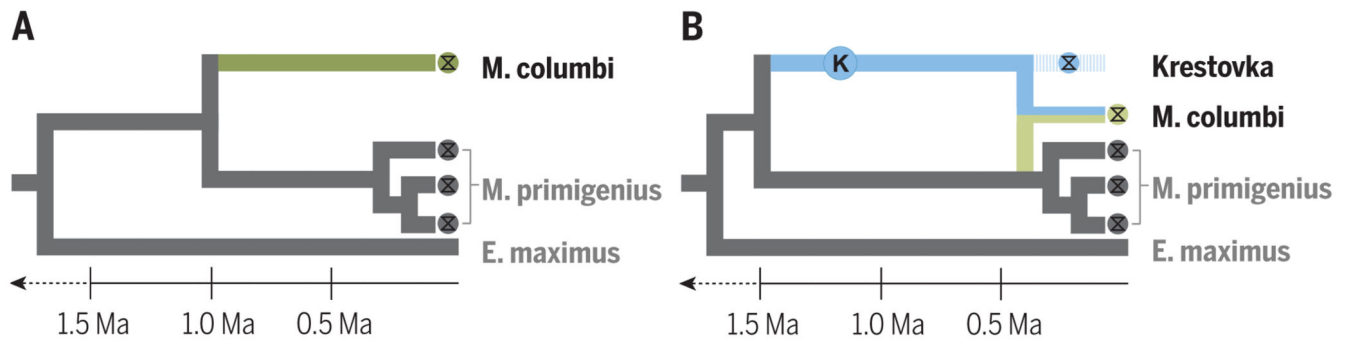
Fig. 3. Deep-time paleogenomes provided new understanding of the evolutionary history of mammoths.

Paleontological hypotheses assumed that the *M. columbi* lineage evolved after early divergence from *M. primigenius* (**A**), however, isolation of a deep-time paleogenome from the Krestova mammoth (blue circle) revealed that *M. columbi* emerged more recently and following admixture with the Krestova mammoth lineage (**B**).