REVIEW

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]*†

Although 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 challenges, 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 substantial 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 the 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 to 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 interspecies 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 postmortem processes lead to successive degradation of DNA molecules into increasingly small fragments,

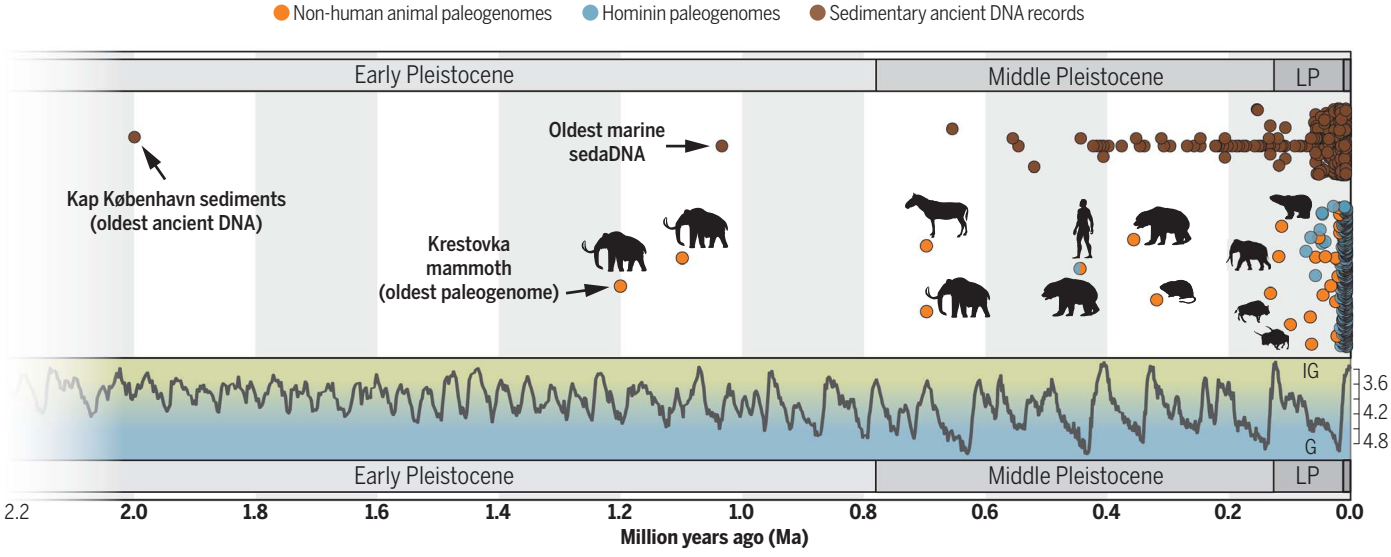

Fig. 1. The temporal distribution of ancient DNA studies to date highlights gaps and opportunities for deep-time paleogenomics and sedimentary ancient DNA. Most ancient DNA studies fall within the last 50 ka and the most recent glacial cycle. The climate curve is based on benthic δ^{18}-oxygen measurements [per mil; LR04 stack from (*42*)]. Sedimentary ancient DNA data are from the AncientMetagenomeDir (v23.06.0) (*57*) and (*58*), with metabarcoding records older than one million years excluded. Paleogenomic data are available from (*59*). 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*, *60*); 700-ka horse (*8*); and 700-ka, 1.1-Ma, and 1.2-Ma mammoths (*10*). LP, Late Pleistocene; IG, Interglacial; G, Glacial.

[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, CA, 95064, USA. [6]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, 95064, USA. [7]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA, 95064, USA.
*Corresponding author. Email: bashapir@ucsc.edu
†These authors contributed equally to this work.

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 considerably longer than predicted by the earliest models. 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 September 2023, the oldest reconstructed paleogenome is from a permafrost-preserved mammoth dating to between 1 and 2 Ma (*10*) and the oldest isolated DNA is from ~2-Ma sediment from northern Greenland (*11*). However, 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 cross-linking (*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 with 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*). Although 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 to 10 nucleotides (*26*). Deep-time DNA molecules are already short—often <35 bases (*15*)—and consequently this may reduce the proportion of useful endogenous DNA.

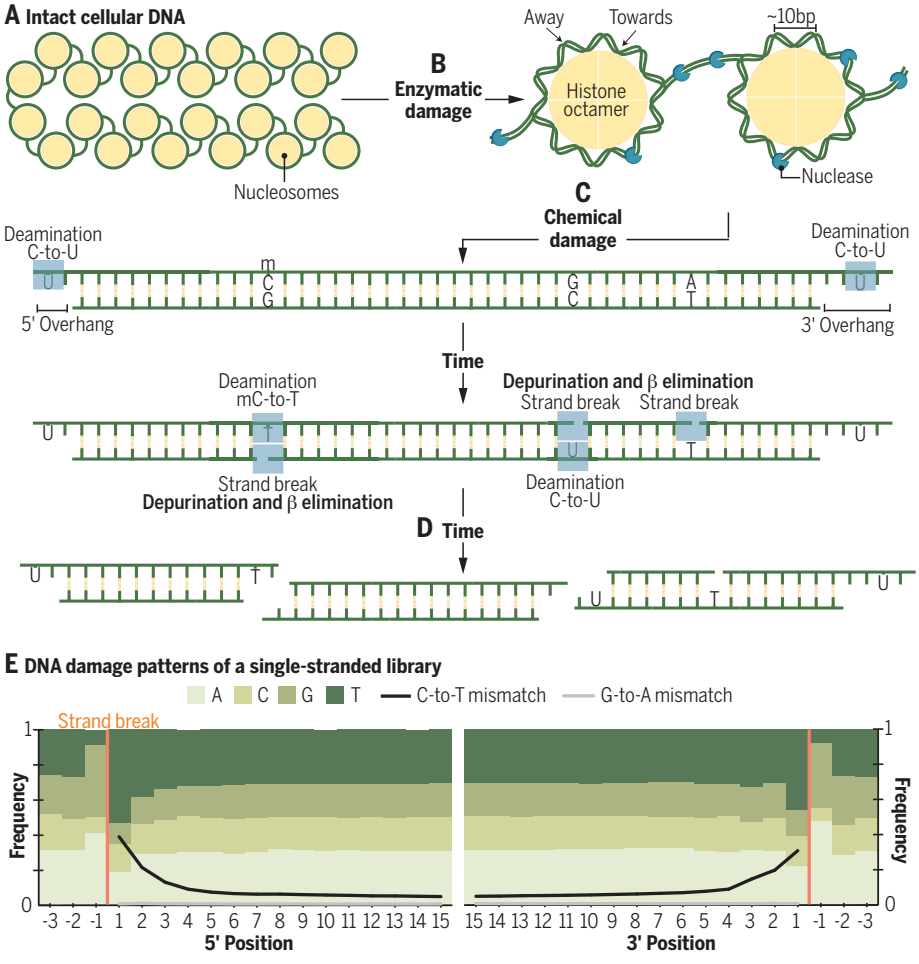

Fig. 2. DNA fragmentation and degradation begins after death and continues until fragments are too short to be useful. (**A**) The integrity of megabase length DNA is maintained by a cell's enzymatic repair machinery and, in eukaryotic genomes, packaged in histone-DNA complexes. (**B**) Following death, 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 [library JKFC14; (*25*)]. 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. Whereas 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.

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 through 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 affected by cytosine deamination. Reference genomes can also be modified to create artificially closer references, such as a "Neanderthalized" version of the human reference genome for reference-guided mapping of Neanderthals 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 hybridized during previous glacial and interglacial periods (*30*, *31*) as well as in the modern era. 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 paleogenomes (*9*, *32*). Similarly, a mammoth paleogenome dating to the Early Pleistocene revealed that Columbian mammoths (*Mammuthus columbi*) 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 and Fig. 3). Deep-time DNA can reveal genomic snapshots of an individual 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, through exploration of 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 Neanderthals 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 that previously hypothesized (*8*).

Paleogenomes from the Early and Middle Pleistocene can also be used to test hypotheses about relationships among species, including 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 from ~130 to 110 ka, for example, showed that two samples exhibiting extreme size dimorphism and representing supposedly distinct species—the longhorn bison

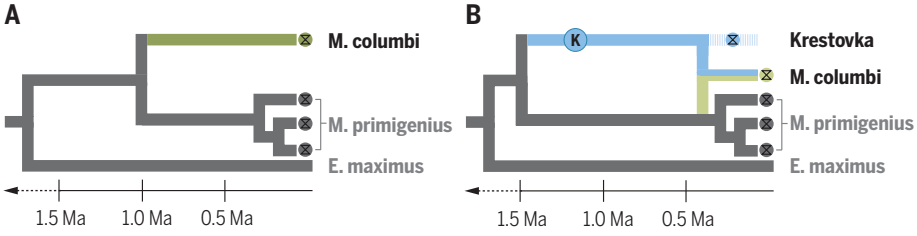

Fig. 3. Deep-time paleogenomes provided new understanding of the evolutionary history of mammoths. Paleontological hypotheses assumed that the *M. columbi* lineage evolved after early divergence from *M. primigenius* (**A**), however, isolation of a deep-time paleogenome from the Krestovka mammoth (blue circle) revealed that *M. columbi* emerged more recently following admixture between the Krestovka and M. primigenius lineages (**B**). Adapted from (*10*).

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 was collected from three Siberian mammoths dated to approximately 700 ka to 1.2 Ma, which made it possible to examine mammoth evolution from multiple genomic angles:

A new lineage: The oldest of the mammoth specimens belonged to a previously unknown and divergent evolutionary lineage, Krestovka. 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 to 1.2 Ma (*61*).

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. They concluded 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 specific to woolly mammoths (*37*).

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 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 the woolly mammoth 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 Early (2.6 Ma to 780 ka) and Middle (780 to 126 ka) Pleistocene sub-epochs relies mostly on more traditional approaches, including morphometrics, stable isotope analysis, and pollen records. Increasing access to genomic data from fossils and sediments dating to these earlier sub-epochs will enable more explicit tests of hypotheses about how glacial cycles affect evolution and biodiversity.

A special attribute of the earlier Pleistocene, for example, is the change in periodicity of glaciations from ~40 ka cycles to ~100 ka cycles that occurred 1.2 to 0.7 Ma (*38*) (Fig. 1). This change resulted in isolation of 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, 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 to 370 ka (Marine Isotope Stage 11) (*42*). Paleogenomes from individuals that lived during this long bottleneck and earlier could be used to test these hypotheses and reveal evolutionary changes that were overwritten by subsequent genetic bottlenecks.

Inference of ancient ecosystems

We have described insights that could be 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 of an Early Pleistocene interglacial ecosystem from sediment 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 through hybridization to synthesized baits designed to target Arctic or Antarctic taxa. In contrast to the PCR-based metabarcoding approach used by the other three studies, hybridization-based targeted enrichment can capture molecules of any length and are therefore powerful even when preserved molecules are short. Although the hybridization-based approach is currently limited to capturing sequences that are genetically similar to other known taxa, methodological improvements in hybridization capture is a ripe area of research that will expand access to deep-time sedimentary DNA.

Deep-time sedimentary DNA research will enable 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 that was part of the Early Pleistocene community (*11*), although no fossil remains from such an animal have been discovered. Deep-time sedimentary DNA can also reveal past connectivity among populations, as indicated by 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—particularly 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 recovery of 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 with nonclay surfaces, particularly 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 that most DNA was inaccessible. Although 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 that DNA preserved in cave sediment is concentrated in microscale 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 ~10 to 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, bio-orthogonal 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 are currently 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. Although 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 multispecies 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.

One considerable challenge for studies of deep-time DNA is knowing the age of samples so that they can be placed into broader evolutionary and geological contexts. As most ancient DNA to date is from organisms that lived within the last several tens of thousands of years, 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. 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 remain 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" are known to 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 such as glass fission-track or argon-argon dating. Tephra beds, which are detectable 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 tephra is present. As volcanic eruptions were common throughout the Pleistocene, improved tephrochronologies spanning 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.

Conclusions

The next decade will bring continued technical advances that will expand the taxonomic and geographic range of deep-time paleogenomes and deep-time ancient sedimentary DNA data sets. These will include 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. These newly assembled deep-time paleogenomes will be placed into chronological context through developments in geochronology and paleoecology, together with more powerful computational approaches to estimate the age of samples using a molecular clock. The resulting deep-time data sets will enable reconstruction of evolutionary histories 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.

REFERENCES AND NOTES

1. J. Ehlers, P. L. Gibbard, P. D. Hughes, *Quaternary Glaciations - Extent and Chronology: A Closer Look* (Elsevier, 2011).
2. C. Posth *et al.*, *Curr. Biol.* **26**, 827–833 (2016).
3. M. Baca *et al.*, *Proc. Biol. Sci.* **290**, 20222238 (2023).
4. L. Loog *et al.*, *Mol. Ecol.* **29**, 1596–1610 (2020).
5. R. E. Green *et al.*, *Science* **328**, 710–722 (2010).
6. A. D. Foote *et al.*, *Nat. Commun.* **4**, 1677 (2013).
7. C. Valdiosera *et al.*, *Biol. Lett.* **2**, 601–603 (2006).
8. L. Orlando *et al.*, *Nature* **499**, 74–78 (2013).
9. A. Barlow *et al.*, *Curr. Biol.* **31**, 1771–1779 (2021).
10. T. van der Valk *et al.*, *Nature* **591**, 265–269 (2021).
11. K. H. Kjær *et al.*, *Nature* **612**, 283–291 (2022).
12. J. Courtin *et al.*, *Environ. DNA* **4**, 1265–1283 (2022).
13. E. Willerslev *et al.*, *Science* **300**, 791–795 (2003).
14. E. Willerslev *et al.*, *Science* **317**, 111–114 (2007).
15. M. Meyer *et al.*, *Nature* **505**, 403–406 (2014).
16. E. I. Zavala *et al.*, *Nature* **595**, 399–403 (2021).
17. T. Lindahl, *Nature* **362**, 709–715 (1993).
18. J. S. Pedersen *et al.*, *Genome Res.* **24**, 454–466 (2014).
19. A. J. Hansen *et al.*, *Genetics* **173**, 1175–1179 (2006).

20. A. W. Briggs *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
21. L. Bokelmann, I. Glocke, M. Meyer, *Genome Res.* **30**, 1449–1457 (2020).
22. J. Dabney, M. Meyer, S. Pääbo, *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
23. E. Hempel *et al*., *Mol. Biol. Evol.* **39**, msac241 (2022).
24. M.-T. Gansauge *et al*., *Nucleic Acids Res.* **45**, e79–e79 (2017).
25. J. D. Kapp, R. E. Green, B. Shapiro, *J. Hered.* **112**, 241–249 (2021).
26. A. W. Briggs *et al*., *Nucleic Acids Res.* **38**, e87 (2010).
27. C. de Filippo, M. Meyer, K. Prüfer, *BMC Biol.* **16**, 121 (2018).
28. K. Prüfer, *Bioinformatics* **34**, 4165–4171 (2018).
29. S. Peyrégne *et al*., *Sci. Adv.* **5**, eaaw5873 (2019).
30. W. Miller *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2382–E2390 (2012).
31. J. A. Cahill *et al*., *Mol. Biol. Evol.* **35**, 1120–1129 (2018).
32. M.-S. Wang *et al*., *Nat. Ecol. Evol.* **6**, 936–944 (2022).
33. J. T. Weir, D. Schluter, *Proc. Biol. Sci.* **271**, 1881–1887 (2004).
34. A. M. Lister, *Philos. Trans. R. Soc. B*, 221–241 (2004).
35. M. Meyer *et al*., *Nature* **531**, 504–507 (2016).
36. D. Froese *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3457–3462 (2017).
37. D. Díez-Del-Molino *et al*., *Curr. Biol.* **33**, 1753–1764.e4 (2023).
38. P. U. Clark *et al*., *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
39. J. R. Stewart, A. M. Lister, I. Barnes, L. Dalén, *Proc. Biol. Sci.* **277**, 661–671 (2010).
40. E. Lord *et al*., *BMC Ecol. Evol.* **22**, 126 (2022).
41. J. Ortego, L. L. Knowles, *Mol. Ecol.* **31**, 296–312 (2022).
42. L. E. Lisiecki, M. E. Raymo, L. E. Lisiecki, M. E. Raymo, *Paleoceanography* **20**, (2005).
43. L. Armbrecht *et al*., *Nat. Commun.* **13**, 5787 (2022).
44. A. J. Monteath *et al*., *Quat. Sci. Rev.* **299**, 107878 (2023).
45. C. L. Clarke *et al*., *Sci. Rep.* **9**, 19613 (2019).
46. M. W. Pedersen *et al*., *Curr. Biol.* **31**, 2728–2736.e8 (2021).
47. D. Massilani *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113666118 (2022).
48. N. Kaplan, J. Dekker, *Nat. Biotechnol.* **31**, 1143–1147 (2013).
49. A. C. Payne *et al*., *Science* **371**, eaay3446 (2021).
50. M. C. Wibowo *et al*., *Nature* **594**, 234–239 (2021).
51. M. Klapper *et al*., *Science* **380**, 619–624 (2023).
52. R. Martiniano, E. Garrison, E. R. Jones, A. Manica, R. Durbin, *Genome Biol.* **21**, 250 (2020).
53. R. E. Green *et al*., *Cell* **134**, 416–426 (2008).
54. K. E. H. Penkman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2109324119 (2022).
55. S. M. Davies, *J. Quat. Sci.* **30**, 114–130 (2015).
56. M. Meyer *et al*., *Science* **338**, 222–226 (2012).
57. J. A. Fellows Yates *et al*., *Sci. Data* **8**, 31 (2021).
58. J. Von Eggers *et al*., Inventory of ancient environmental DNA from sedimentary archives: locations, methods, and target taxa, Version 1.0, Zenodo (2022); https://zenodo.org/record/6847522.
59. D. Diez-del-Molino, P. D. Heintzman, L. Dalén, A list of representative paleogenomic datasets derived from human and faunal remains, Version 1.0, Zenodo (2023); https://zenodo.org/record/8270285.
60. J. Dabney *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
61. A. M. Lister, *Quat. Int.* **443**, 14–31 (2017).

ACKNOWLEDGMENTS

We thank D. Diez-del-Molino for assembling the database of paleogenomes used in Fig. 1. We also thank the editor and two reviewers for helpful comments that improved the manuscript. **Funding:** P.D.H. and L.D. were supported by the Knut and Alice Wallenberg Foundation (KAW 2021.0048 to P.D.H. and K.A.W. 2022.0033 to P.D.H. and L.D.). L.D. also acknowledges support from the Swedish Research Council (2021-00625) and the European Union (ERC, PrimiGenomes, 101054984). B.S. was supported in part by NSF OPP-2131589. J.K. was supported by NIJ 2020-DQ-BX-0014. **Competing interests:** Authors declare that they have no competing interests.

Submitted 31 May 2023; accepted 5 September 2023
10.1126/science.adh7943

Deep-time paleogenomics and the limits of DNA survival

Love Dalén, Peter D. Heintzman, Joshua D. Kapp, and Beth Shapiro

Science **382** (6666), . DOI: 10.1126/science.adh7943