Dear Author

Please use this PDF proof to check the layout of your article. If you would like any changes to be made to the layout, you can leave instructions in the online proofing interface. First, return to the online proofing interface by clicking "Edit" at the top page, then insert a Comment in the relevant location. Making your changes directly in the online proofing interface is the quickest, easiest way to correct and submit your proof.

Please note that changes made to the article in the online proofing interface will be added to the article before publication, but are not reflected in this PDF proof.

If you would prefer to submit your corrections by annotating the PDF proof, please download and submit an annotatable PDF proof by clicking the link below.

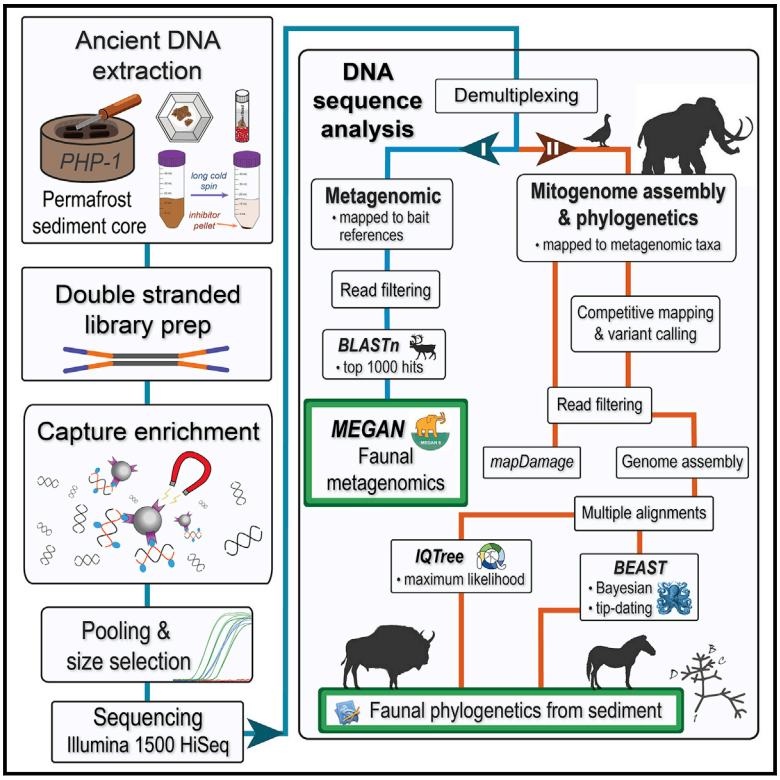
Annotate PDF

Report

Current Biology

Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost

Graphical abstract



Highlights

- Reconstructed 18 mitochondrial genomes solely from <1 g samples of permafrost silts
- Found co-occurrence of distinct mitochondrial clades of mammoth and steppe bison
- Reassembled megafaunal (horse, mammoth, and bison) and avian (ptarmigan) mitogenomes
- Identified a wide breadth of animals that lived 30,000 years ago in central Yukon

Authors

Tyler J. Murchie, Emil Karpinski, Katherine Eaton, ..., Ross D.E. MacPhee, Duane Froese, Hendrik Poinar

Correspondence

murchiet@mcmaster.ca (T.J.M.), duane@ualberta.ca (D.F.), poinarh@mcmaster.ca (H.P.)

In brief

Murchie et al. use sedimentary ancient DNA from Yukon permafrost cores to reassemble 18 mitochondrial genomes of Pleistocene fauna, including woolly mammoth, steppe bison, horse, and ptarmigan. Distinct lineages of woolly mammoth and steppe bison are identified in a core dating to 30,000 years ago, and a wide spectrum of fauna are identified.

Murchie et al., 2022, Current Biology *32*, 1–10
February 28, 2022
https://doi.org/10.1016/j.cub.2021.12.023

CellPress

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

Current Biology

CellPress

Report

Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost

Tyler J. Murchie,[1,2,11,*] Emil Karpinski,[1,3] Katherine Eaton,[1,2] Ana T. Duggan,[1,2] Sina Baleka,[1] Grant Zazula,[4,5] Ross D.E. MacPhee,[6] Duane Froese,[7,*] and Hendrik Poinar[1,2,8,9,10,*]

[1]McMaster Ancient DNA Centre, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[2]Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[3]Department of Biology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[4]Yukon Government, Palaeontology Program, Department of Tourism and Culture, Box 2703, Whitehorse, YT Y1A 2C6, Canada
[5]Collections and Research, Canadian Museum of Nature, PO Box 3443, Station D, Ottawa, ON K1P 6P4, Canada
[6]Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, 200 Central Park West, New York, NY 10024, USA
[7]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB T6G 2E3, Canada
[8]Department of Biochemistry, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[9]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada
[10]CIFAR, Humans and the Microbiome Program, MaRS Centre, West Tower, 661 University Avenue, Suite 505, Toronto, ON M5G 1M1, Canada
[11]Lead contact
*Correspondence: murchiet@mcmaster.ca (T.J.M.), duane@ualberta.ca (D.F.), poinarh@mcmaster.ca (H.P.)
https://doi.org/10.1016/j.cub.2021.12.023

SUMMARY

Traditionally, paleontologists have relied on the morphological features of bones and teeth to reconstruct the evolutionary relationships of extinct animals.[1] In recent decades, the analysis of ancient DNA recovered from macrofossils has provided a powerful means to evaluate these hypotheses and develop novel phylogenetic models.[2] Although a great deal of life history data can be extracted from bones, their scarcity and associated biases limit their information potential. The paleontological record of Beringia[3]—the unglaciated areas and former land bridge between northeast Eurasia and northwest North America—is relatively robust thanks to its perennially frozen ground favoring fossil preservation.[4,5] However, even here, the macrofossil record is significantly lacking in small-bodied fauna (e.g., rodents and birds), whereas questions related to migration and extirpation, even among well-studied taxa, remain crudely resolved. The growing sophistication of ancient environmental DNA (eDNA) methods have allowed for the identification of species within terrestrial/aquatic ecosystems,[6–12] in paleodietary reconstructions,[13–19] and facilitated genomic reconstructions from cave contexts.[8,20–22] Murchie et al.[6,23] used a capture enrichment approach to sequence a diverse range of faunal and floral DNA from permafrost silts deposited during the Pleistocene-Holocene transition.[24] Here, we expand on their work with the mitogenomic assembly and phylogenetic placement of *Equus caballus* (caballine horse), *Bison priscus* (steppe bison), *Mammuthus primigenius* (woolly mammoth), and *Lagopus lagopus* (willow ptarmigan) eDNA from multiple permafrost cores spanning the last 30,000 years. We identify a diverse metagenomic spectra of Pleistocene fauna and identify the eDNA co-occurrence of distinct Eurasian and American mitogenomic lineages.

RESULTS

Sufficient on-target sedimentary ancient DNA (sedaDNA) sequence data ($\gtrsim$3× coverage of at least 80% of the reference genome; Data S1A–S1F) were recovered to reconstruct 18 mitochondrial genomes from four targeted families—Elephantidae, Equidae, Bovidae, and Phasianidae (Figures 1, 2, 3, and 4). These libraries contain distinctive terminal deamination patterns,[25] and the negative controls contain no reads that passed map-filtering and could be *BLASTn* identified or mapped to the reference genomes that make up the PalaeoChip Arctic v1.0 bait set.[6] Combined with the temporally distinctive palaeoecological reconstructions detailed in Murchie et al.[6,23] (where these cores were Bayesian age modeled), the consistent damage patterns and clear negative controls serve as strong support that these sequenced reads originate from the loessal permafrost silts themselves and are characteristically ancient.

This dataset is augmented from Murchie et al.[23] to include a set of uracil-DNA glycosylase (UDG)-treated libraries for sample PHP-1 (Bear Creek, 30,000 calendar/calibrated years before present [cal BP])[26] wherein deamination-derived errors on the ends of ancient DNA (aDNA) fragments are repaired to improve the identification of phylogenetically informative single-nucleotide polymorphisms (SNPs). As expected, UDG treatment was successful in removing deamination from our augmented dataset; these additional libraries retain the same aDNA typical short

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

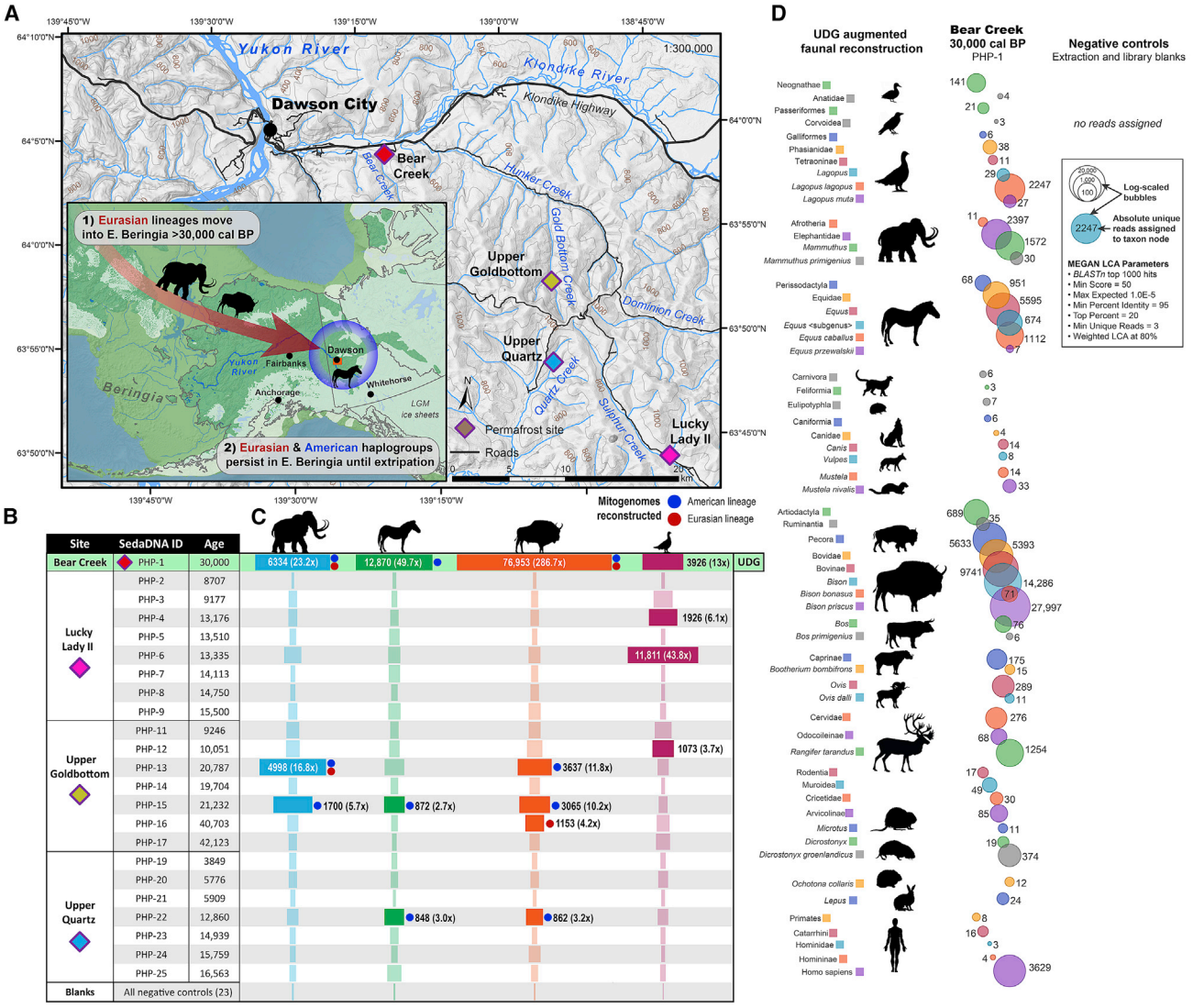


Figure 1. sedaDNA permafrost sites, samples, and metagenomics

(A) Permafrost sites from the Klondike goldfields of Yukon, Canada. Base map data from GeoYukon; contours in meters above sea level. Inset: Beringia during the last glacial maximum (LGM, 26.5–19 ka BP).[40] Ice sheet data from Dyke.[69] Sea level during LGM set to 126 m below sea level based on Clark and Mix.[70] Beringian paleodrainage data from Bond.[71]

(B) Murchie et al.[23] reanalyzed sedaDNA libraries.

(C) Number of filtered reads mapped to target fauna with mean mapped coverages of the mitochondrial genome in parentheses. Bars transformed on a square-root scale by taxon; libraries with sufficient coverage depth for consensus calling highlighted.

(D) Faunal metagenomic reconstruction of PHP-1 with newly sequenced libraries.

See Figure S4 for an expanded analysis of the *Bos* reads, as well as the STAR Methods and Data S1M for analysis of the human reads. See Data S1A and S1B for sequencing summaries.

fragment length distributions (mode ≈ 30–50 bp) as the rest of the dataset.

From these data, we observe a wide metagenomic spectra of paleofauna (Figure 1D) along with the genetic co-occurrence of Eurasian- and American-associated mitochondrial lineages in both *Mammuthus* and *Bison* from the same core samples (Figures 2, 3, and S1–S3). These eDNA-reconstructed mitogenomes likely represent inputs from multiple related individuals. We were able to disentangle inputs from deeply diverged haplogroups with competitive mapping. However, Bayesian tip dating of the sedaDNA genomes was only successful from the deeply sequenced and UDG-treated core extracts (PHP-1) (Figures S1 and S2) likely in part because these sedaDNA genomes do not contain terminal damaged bases and have enough coverage to be carefully reconstructed. In the case of less deeply sequenced libraries, reconstructing multiple haplotypes was impossible as the DNA of many animals was likely deposited over years of loess sedimentation wherein unique SNPs on exceptionally short DNA fragments are generally not represented in the derived consensus sequences (Figure S3).

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

CellPress

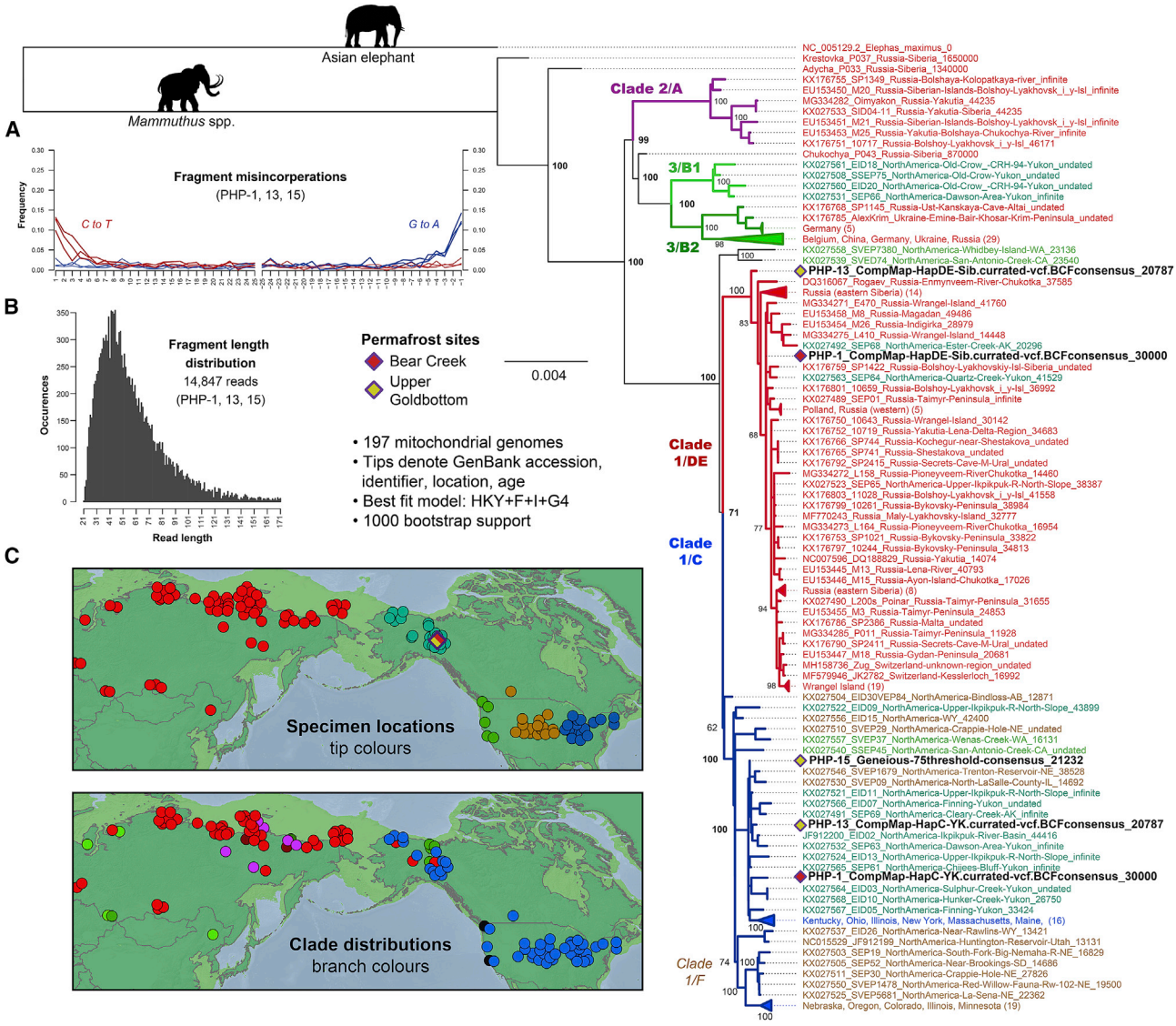


Figure 2. Maximum likelihood tree of *Mammuthus* ssp. with sedaDNA genomes bolded

(A and B) Combined *mapDamage* fragment misincorporation plots (A) and fragment length distributions (B).

(C) Map of sample locations; tree tips are color coded based on region, whereas branches are colored based on clades as per Enk et al.[2] and Chang et al.[27] See Figures S1–S3 for further analyses of allelic diversity, Bayesian phylogenetics, and competitive mapping. See Data S1C for mapping summary and Data S1G–S1L for BEAST tracer logs as related to Figure S2.

DISCUSSION

Mammuthus primigenius

Our Murchie et al.[23] expanded dataset had sufficient *Mammuthus primigenius* (woolly mammoth) reads to reconstruct five mitogenomes from three permafrost samples (Figure 2). These sedaDNA genomes fall primarily within Clade 1, haplogroup C,[2,27] which was distributed from Alaska/Yukon (eastern Beringia) southeast to the Great Lakes and Atlantic Coast along the southern margin of the Laurentide ice sheet. This lineage was originally presumed to be genetically distinct from haplogroup 1-F, composed of *Mammuthus columbi* (Columbian mammoth) and *M. jeffersonii* (likely a hybrid between *M. primigenius* and *columbi*) that occupied the Great Plains and West Coast regions of North America.[2] Nuclear genomic data provide clear evidence of hybridization between North American *M. primigenius* and *columbi*[28] and likely a more panmictic population spread across North America.[27]

Likewise, multiple haplogroups have been observed among eastern Beringian woolly mammoths, presumably because *M. primigenius* genotypes from Eurasia were introduced into American lineages.[2,27] A careful investigation of segregating SNPs in our *Mammuthus* spp. alignment compared with mapped reads and derived consensus sequences shows multiple *M. primigenius* inputs in the same core sample (PHP-1) from both haplogroup 1-C (American) and haplogroups 1-D&E (Eurasian) dating to ~30,000 cal BP (Figures 2, S1, and S2). This Eurasian gene flow may suggest that there was an overlap

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

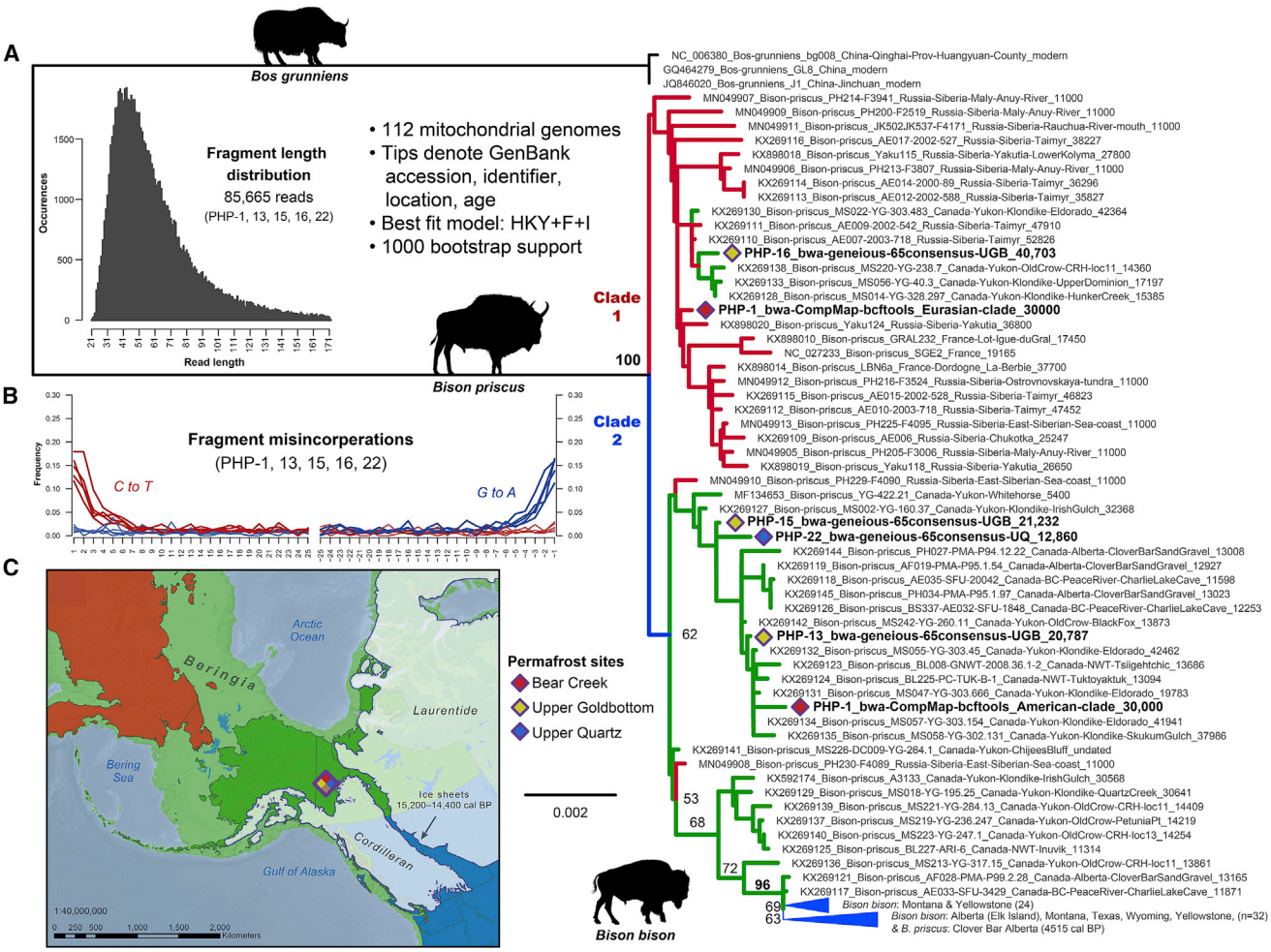


Figure 3. Maximum likelihood tree of *Bison* with emphasis on extinct *Bison priscus*

(A and B) Combined fragment length distributions (A) and *mapDamage* plot depicting aDNA typical damage patterns (B).

(C) Specimen recovery locations split between Siberia (red), eastern Beringia (Green), and areas south of the ice-free corridor (blue).

See Figures S1–S4 for further analyses of allelic diversity, Bayesian phylogenetics, and competitive mapping. See Data S1D for mapping summary and Data S1G–S1L for BEAST tracer logs as related to Figure S2.

of distinct *Mammuthus* mitochondrial lineages in the Klondike region during the time-interval represented by this core sample.

However, it is unclear how much sedaDNA accumulation time is represented by each subsample of the permafrost cores beyond their calibrated radiocarbon age. The cryostratigraphy and age/temperature deamination patterns identified by Murchie et al.[23] suggest that leaching and reworking are unlikely to have significantly contributed to the ecological reconstructions of these core samples, particularly for late glacial cores that predate the early Holocene thaw unconformity identified in the Klondike.[29–32] Even so, it is reasonable to suppose that each ~250 mg core subsample represents the ecological accumulation of at least decades or perhaps even hundreds of years of genetic material and thus likely includes eDNA from multiple individual animals of different time periods. Therefore, haplogroup Q9 1-C and 1-D&E herds may have spatially overlapped at different points in time within the sedaDNA accumulation range of this core sample. It has been observed in modern and ancient brown bears that deeply diverged mitochondrial clades coexisted in eastern Beringia for over 60,000 years despite large-scale phases of dispersal, extirpation, and replacement.[33] Similar patterns of long term population structure have been observed among Asian elephants in Sri Lanka.[34] Determining whether the co-occurrence we observe in *Mammuthus* is the result of gene flow in central Yukon, or if external admixture had resulted in the partial retention of Eurasian mitochondrial DNA in some haplogroup 1-C individuals—such as observed with cave and brown bears[35]—would necessitate nuclear sedaDNA, which was not targeted in this experiment. Nevertheless, the presence of regionally distinct clades of *Mammuthus primigenius* predating the Dawson tephra (29,000 years ago)[36–39] and last glacial maximum (LGM [26,500–19,000 BP])[40] makes phylogeographic sense in eastern Beringia and suggests that range co-occupation and interbreeding among Eurasian and American lineages of mammoths was possible.

Competitive mapping paired with the variant consensus approach of Lammers et al.[41] successfully disentangled much of the haplogroup 1-C and 1-D&E inputs in samples PHP-1, Q10

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

CellPress

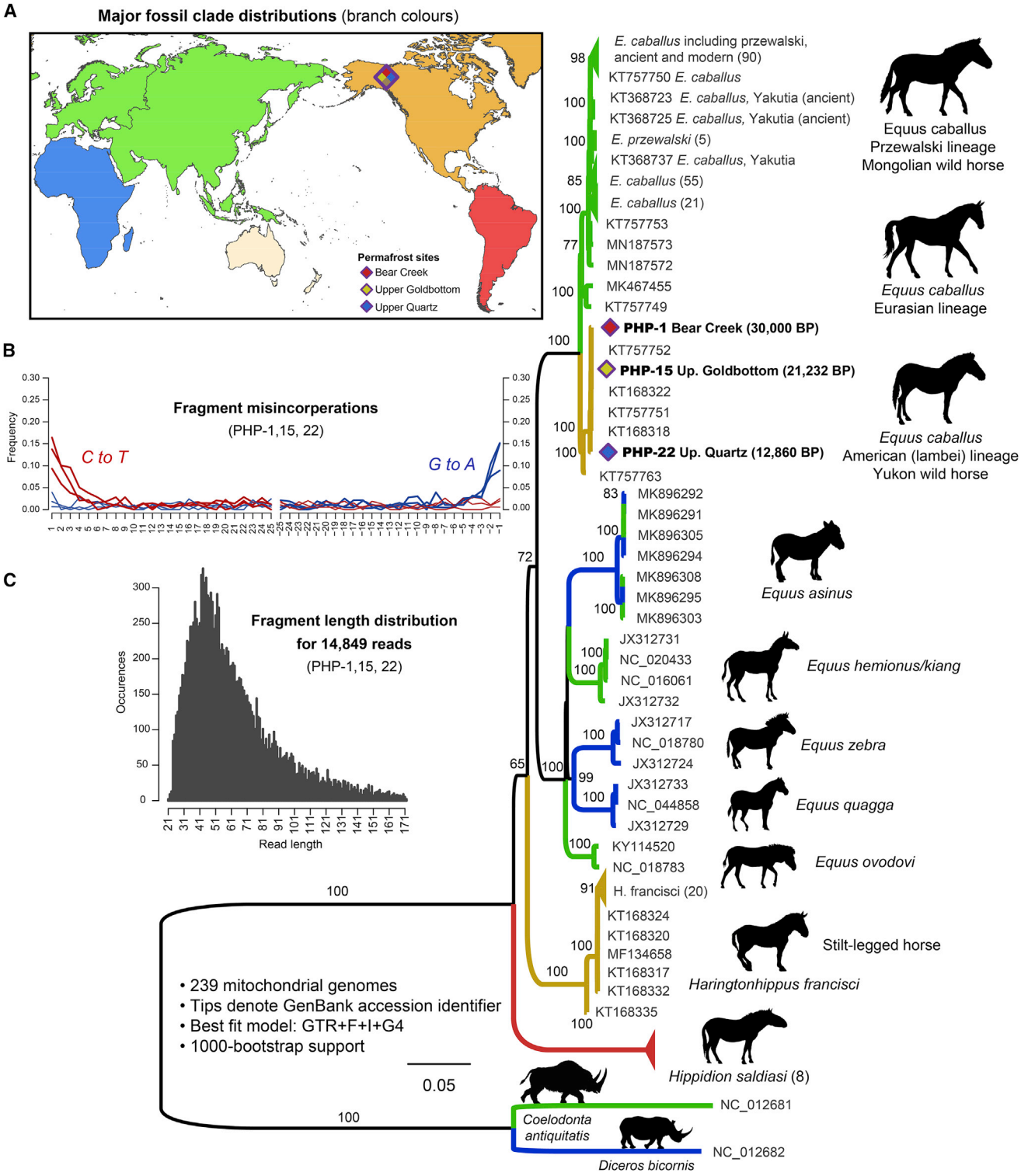


Figure 4. Maximum likelihood tree of Equidae with emphasis on *Equus caballus* (extinct and extant caballine horses)

Regionally color coded branches as per Heintzman et al.[72]

(A) Major equid clade distributions in the fossil record.

(B and C) *mapDamage* plot depicting aDNA typical terminal deamination patterns (B) and a fragment length distribution plot of merged reads (C).

See Data S1D–S1F for mapping summaries.

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

and to a lesser extent with PHP-13 (20,787 cal BP). In the case of sample PHP-13 (haplogroup 1-D&E-variant consensus), low coverage of segregating sites necessitated abundant ambiguous bases, which explains the pseudo-ancestral position of this sequence in both the maximum likelihood and Bayesian trees (Figures 2 and S2). Coverage of PHP-15 was similarly too low to effectively discriminate between these two lineages. By contrast, both haplogroup variants of PHP-1 and the American variants of PHP-13 and PHP-15 fall well within the known diversity of North American and Eurasian lineages of *Mammuthus primigenius*.

Bison priscus

Five sedaDNA libraries had sufficient read coverage to reconstruct mitochondrial genomes of *Bison priscus* (steppe bison) (Figure 3). Four fall within what is a predominantly American clade (referred to here as Clade 2), whereas two fall within the diversity of Eurasian steppe bison (Clade 1).[42–44] As with *Mammuthus*, we see multiple instances of segregating, clade-specific SNPs in the mapped reads wherein both American and Eurasian mitochondrial lineages are represented (Figures 3, S1, and S2). The youngest of these sedaDNA mitogenomes (PHP-22; 12,860 cal BP) is most clearly associated with Clade 2 (American) *B. priscus*, whereas the oldest (PHP-16; 40,703 cal BP) falls within the known diversity of Clade 1 (Eurasian) *B. priscus*. Samples of intermediate age (PHP-1, PHP-13, PHP-15) contain reads clearly associated with both lineages. Competitive mapping was able to separate these clades within sample PHP-1 due to the additional sequencing depth and UDG treatment.

As with mammoths, the overlap of Eurasian and American lineages makes phylogeographic sense for Beringian bison. It is unclear whether this co-occurrence is the result of previous admixture, offset occupations, or direct range co-occupation of Clade 1 and 2 individuals. Following the last interglaciation (MIS 5e, 130,000–115,000 BP)[45] the Bering land bridge between Eurasia and North America re-emerged between 70,000 and 60,000 BP. The bridge intermittently connected the continents from 60,000 to 30,000 BP until reaching its maximum extent during the LGM (~21,000 BP) before the Bering strait was inundated by rising sea level (c. 11,000 BP[46–49]). The ~40,000 cal BP Upper Goldbottom core may correspond to a movement of steppe bison into eastern Beringia during periods when at least a partial land bridge connection existed. This is consistent with previous research that identified *B. priscus* gene flow from Eurasia into Beringia nearer to the LGM[42,44] and corresponds with trends observed in our *Mammuthus* mapped reads. However, the precise mechanism for the presence of these multiple sedaDNA inputs of Eurasian mitochondria observed here remains unknown at this time.

Equus caballus

Three equid mitogenomes had sufficient read coverage to call consensus sequences (Figure 4). Unlike *Mammuthus* and *Bison*, we did not observe mitochondrial allelic diversity within the reads mapped to *Equus caballus*, although the reads are clearly distinct from the Eurasian lineage of *E. caballus* utilized for the reference genome. These sedaDNA genomes contain a unique set of SNPs most closely associated with the lambei clade (Yukon horse) of *E. caballus* (Figure 4). These sedaDNA mitogenomes fall within the known diversity of caballine horses in Eurasia and the Americas, being distinct from non-caballine *Haringtonhippus francisci* (New World stilt-legged horse) and other extinct (*Hippidon* sp., *E. ovodovi*) and extant (*E. zebra*, *E. asinus*, *E. kiang*) equids. The *BLASTn* to *MEGAN* approach reported in Murchie et al.[6,23] was only able to identify reads to the genus *Equus* sp., in part due to the ongoing systematics debate regarding species boundaries in *Equus*.[50] However, our sedaDNA mitogenome assembly and phylogenetics approach confidently identified the lambei lineage of *Equus caballus*, even in libraries bordering the minimum read coverage cut-off (~3×) with ~800 mapped reads (Data S1F).

Lagopus lagopus

Four mitogenomes could be reconstructed for *Lagopus lagopus* (willow ptarmigan) (Figure S3). Only a single publicly available mitogenome exists for *L. lagopus* at the time of writing, which limits investigations of phylogenetic diversity. All four genomes form a monophyletic clade with the *L. Lagopus* reference, being distinct from *L. muta* (rock ptarmigan) and other taxa within Tetraoninae. We observe low genetic diversity between the four sedaDNA mitogenomes, which date to between 30,000 and 10,626 cal BP. The capability of permafrost to preserve avian sedaDNA at a genomic scale is significant, as Beringian paleontological records of Pleistocene birds are very limited. Bluefish Caves (Yukon)[51,52] and Lost Chicken Creek (Alaska)[53] are two notable exceptions with diverse assemblages of birds and other small-bodied fauna. SedaDNA methods by comparison have the potential to greatly expand our understandings of understudied but ecologically significant organisms for whom macrofossils are rarely preserved or recovered.

Faunal metagenomics

The augmented Bear Creek UDG-treated libraries improves the faunal resolution reported by Murchie et al.[6,23] and D'Costa et al.[26] at the 30,000 cal BP time point immediately below the Dawson Tephra[36,38,39] (Figure 1). Although most reads that were binned are still assigned to the same predominant taxa as in Murchie et al.[23] (*Mammuthus*, *Equus*, *Bison*, *Lagopus*, and *Rangifer*), we see an increased species richness in the diversity of lower abundance fauna. Newly identified taxa include: *Ochotona collaris* (Collared pika), more specific hits to *Ovis dalli* (dall sheep), *Bootherium bombifrons* (helmeted muskox), potentially *Bos primigenius* (aurochs), *Mustela nivalis* (least weasel), *Vulpes* sp. (fox), *Canis* sp. (wolf), Feliformia (suborder containing "cat-like" carnivorans), Eulipotyphla (order containing hedgehog/mole/shrew and allies), more specific assignments within Equidae to *Equus caballus/przewalskii*, as well as Corvoidea (superfamily of oscine passerine birds that contains crows, ravens, and magpies), and Anatidae (family of water birds that includes ducks and geese). No faunal DNA reads were identified in the negative controls despite being processed identically. *Bos primigenius* and *Homo sapiens* are both peculiarities of this metagenomic dataset that we believe are the result of false-positive assignment and contamination, respectively (Figure S4; STAR Methods).

Prospects and caveats of environmental genomes

The rapid progress of ancient eDNA methods in the last decade has highlighted the significance of ordinary sediments and soils

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

for complementing paleontological and archaeological research. Very few animals ever become macrofossils,[5] and fewer still are recovered by scientists. Even in areas like Beringia with abundant and well-studied buried records,[4] taphonomy significantly limits our understandings of past ecosystems. Not only can sedaDNA preserve taxonomically informative fragments of past life at all ecological scales, but also these microfossils can be used to help address more nuanced evolutionary questions about iconic species that have been obscured by the rarity of their macrofossils. Identifying the simultaneous presence of such a wide spectrum of ancient animals, while also being able to reconstruct and phylogenetically place multiple ancient genomes, would necessitate exceptional levels of bone preservation that exist at only a handful of Quaternary sites. Although, here and in Murchie et al.,[23] all permafrost cores had ecologically informative sedaDNA. Had we sequenced all cores as deeply as PHP-1, it is likely that we could have reconstructed genomes from multiple ancient animals simultaneously from almost all of the cores, on top of the considerable constituent of plant, fungal, and microbial DNA. Environmental archives of genetic microfossils are poised to broaden Quaternary knowledge in understudied regions, small-bodied and rarely preserved fauna, as well as ecosystem and population dynamics. Environmental genomics is also likely to revolutionize archaeological research related to controversial sites, migration, and the paleodemography of ancient humans. This extends to sites without human remains, which opens a range of ethical concerns.[54]

There are several caveats to keep in mind in this sedaDNA revolution as we increasingly improve our ability to recover and study the smallest and most damaged aDNA fragments. First, assessing to what degree leaching or reworking have mixed stratigraphic components at a site.[55,56] Second, whether the wet-lab methods, reference databases, taxon binning, and assembly algorithms can accurately assign and disentangle mixed eDNA inputs from many closely and distantly related individuals.[6,57] Third, assessing to what degree eDNA abundance correlates with a population's living biomass.[58,59] And fourth, determining which environmental conditions are most amenable to sedaDNA preservation,[60,61] how that shifts through time, and improving techniques for efficiently isolating eDNA from challenging environmental substrates.[6,62,63]

Genomic archives within perennially frozen ground

This study demonstrates the genomic resolution potential of permafrost sedaDNA, allowing population genetic questions to be investigated for multiple organisms simultaneously, even in the total absence of macrofossils. These tiny subsamples of permafrost cores not only contain the genomes of quintessential Pleistocene fauna of Yukon Territory, but they also preserve ecosystem-scale eDNA from a diverse spectrum of late glacial organisms. It is increasingly clear that permafrost can serve as a unique source of ancient genomes that are not reliant on the preservation and recovery of macrofossils.

The growing sophistication of sedaDNA methods and recognition of the immense genetic archives preserved in the Holarctic should be paired with the mounting realization that perennial frozen ground is increasingly poised to undergo widespread, accelerating, and irreversible permafrost thaw.[64–68] As this permafrost thaws, it is likely to lose much of its contextual association as it slumps, as well as undergoing mass degradation of the innumerable biomolecules that have been cold preserved therein for tens, to perhaps hundreds, of thousands of years. These paleogenetic repositories of the north's ecological heritage will not survive indefinitely given current projections of a warming Arctic. This underscores the growing importance of archiving permafrost materials for future sedaDNA research wherein genomes of long extinct organisms can be reassembled solely from submicroscopic biomolecules preserved within the sediments and soils of ancient environments.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Ancient DNA clean lab
  - Subsampling
  - Lysing and purification
  - Library preparation, quantitative PCR, and indexing
  - RNA hybridization targeted capture
  - Total quantification, pooling, size selection, and sequencing
  - Bioinformatics
- COMPETITIVE MAPPING
- PHYLOGENETICS
- BAYESIAN PHYLOGENETICS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.12.023.

ACKNOWLEDGMENTS

Many thanks to the CANA Foundation for their generous support of a PDF and operating costs to T.J.M. and H.P. This work was funded by the Belmont Forum and BiodivERsA grants (to D.F. and H.P.) for the Future ArcTic Ecosystems (FATE) research consortium, as well as NSERC Discovery grants to D.F. and H.P. We wish to also thank the Garfield Weston Foundation for their generous support of T.J.M., as well as McMaster University, The Arctic Institute of North America, Polar Knowledge Canada, and the Social Sciences and Humanities Research Council of Canada for each funding various components of this research. Thank you to the placer gold mining community of the Klondike and the Tr'ondëk Hwëch'in for their continued support of our research and access to study sites in the central Yukon. Our thanks to Beth Shapiro and Jake Enk for providing steppe-bison and mammoth alignments during the early stages of this research that, although we did not end up using further, aided T.J.M. significantly in the completion of the dissertation version of this research. Thanks as well to Love Dalén, David Diez, and Peter Heintzman for sharing their reconstructed million-year-old mammoth mitogenomes. Thanks to Brian Golding for providing access to his computational resources, which were invaluable to the processing of these datasets. Thanks to Shanti Morell-Hart, Brian Golding, and Peter Heintzman for their thoughtful and

Q15

thorough critiques of this work in T.J.M.'s doctoral thesis. Thanks to all members and affiliates of the McMaster Ancient DNA Centre for their ongoing support, as well as the admin and faculty of the Anthropology department at McMaster University. Special thanks to Debi Poinar, Melanie Kuch, Jennifer Klunk, Marissa L. Ledger, and Kévin Roche for their wet-lab assistance during the *Quaternary Research* PalaeoChip paper that laid the groundwork for this study.

AUTHOR CONTRIBUTIONS

Conceptualization, T.J.M., E.K., A.T.D., D.F., and H.P.; methodology, T.J.M., E.K., A.T.D., and H.P.; software, T.J.M., K.E., and A.T.D.; investigation, T.J.M.; resources, S.B., D.F., and H.P.; data curation, T.J.M., E.K., K.E., and S.B.; writing – original draft, T.J.M., G.Z., R.D.E.M., D.F., and H.P.; writing – review & editing, T.J.M., E.K., K.E., A.T.D., S.B., G.Z., R.D.E.M., D.F., and H.P.; visualization, T.J.M.; supervision, A.T.D., D.F., and H.P.; project administration, H.P.; funding acquisition, D.F. and H.P.

DECLARATION OF INTERESTS

T.J.M., H.P., and R.D.E.M. are members of the CANA Foundation science advisory board, a non-profit organization with horse re-wilding initiatives. T.J.M. and H.P. are currently supported by CANA through a PDF to T.J.M. and consumable costs. The research presented in this paper was completed long prior to funding by CANA. The authors declare no other competing interests.

Received: September 7, 2021
Revised: October 20, 2021
Accepted: December 8, 2021
Published: January 10, 2022

REFERENCES

1. Lister, A.M., and Sher, A.V. (2001). The origin and evolution of the woolly mammoth. Science *294*, 1094–1097.
2. Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., Rouillard, J.-M., Shapiro, B., Golding, G.B., Zazula, G., et al. (2016). Mammuthus population dynamics in Late Pleistocene North America: divergence, phylogeogrpaphy, and introgression. Front. Ecol. Evol. *4*, 1–13.
3. D.M. Hopkins, J.V. Matthews, and C.E. Schweger, eds. (1982). Paleoecology of Beringia (Academic Press).
4. Stuart, A.J. (2015). Late Quaternary megafaunal extinctions on the continents: a short review. Geol. J. *50*, 414–433.
5. Guthrie, R.D. (1990). Frozen Fauna of the Mammoth Steppe: The Story of Blue Babe (University of Chicago Press).
6. Murchie, T.J., Kuch, M., Duggan, A.T., Ledger, M.L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., MacPhee, R., et al. (2021). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quat. Res. *99*, 305–328.
7. Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature *506*, 47–51.
8. Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., De La Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., et al. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. Science *356*, 605–608.
9. Sjögren, P., Edwards, M.E., Gielly, L., Langdon, C.T., Croudace, I.W., Merkel, M.K.F., Fonville, T., and Alsos, I.G. (2017). Lake sedimentary DNA accurately records 20th century introductions of exotic conifers in Scotland. New Phytol *213*, 929–941.
10. Graham, R.W., Belmecheri, S., Choy, K., Culleton, B.J., Davies, L.J., Froese, D., Heintzman, P.D., Hritz, C., Kapp, J.D., Newsom, L.A., et al. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. Proc. Natl. Acad. Sci. USA *113*, 9310–9314.
11. Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L.Z., Beaudoin, A.B., Zutter, C., Larsen, N.K., et al. (2016). Postglacial viability and colonization in North America's ice-free corridor. Nature *537*, 45–49.
12. Anderson-Carpenter, L.L., McLachlan, J.S., Jackson, S.T., Kuch, M., Lumibao, C.Y., and Poinar, H.N. (2011). Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. BMC Evol. Biol. *11*, 30.
13. Karpinski, E., Mead, J.I., and Poinar, H.N. (2017). Molecular identification of paleofeces from Bechan Cave, southeastern Utah, USA. Quat. Int. *443*, 140–146.
14. Warinner, C., Speller, C., Collins, M.J., and Lewis, C.M. (2015). Ancient human microbiomes. J. Hum. Evol. *79*, 125–136.
15. Poinar, H.N., Hofreiter, M., Spaulding, W.G., Martin, P.S., Stankiewicz, B.A., Bland, H., Evershed, R.P., Possnert, G., and Pääbo, S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. Science *281*, 402–406.
16. Weyrich, L.S., Dobney, K., and Cooper, A. (2015). Ancient DNA analysis of dental calculus. J. Hum. Evol. *79*, 119–124.
17. Ozga, A.T., Nieves-Colón, M.A., Honap, T.P., Sankaranarayanan, K., Hofman, C.A., Milner, G.R., Lewis, C.M., Stone, A.C., and Warinner, C. (2016). Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. Am. J. Phys. Anthropol. *160*, 220–228.
18. Weyrich, L.S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A.G., Alt, K.W., Caramelli, D., Dresely, V., et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature *544*, 357–361.
19. Søe, M.J., Nejsum, P., Seersholm, F.V., Fredensborg, B.L., Habraken, R., Haase, K., Hald, M.M., Simonsen, R., Højlund, F., Blanke, L., et al. (2018). Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. PLoS One *13*, e0195481.
20. Zavala, E.I., Jacobs, Z., Vernot, B., Shunkov, M.V., Kozlikin, M.B., Derevianko, A.P., Essel, E., de Fillipo, C., Nagel, S., Richter, J., et al. (2021). Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature *595*, 399–403.
21. Gelabert, P., Sawyer, S., Bergström, A., Margaryan, A., Collin, T.C., Meshveliani, T., Belfer-Cohen, A., Lordkipanidze, D., Jakeli, N., Matskevich, Z., et al. (2021). Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. Curr. Biol. *31*, 3564–3574.e9.
22. Pedersen, M.W., De Sanctis, B., Saremi, N.F., Sikora, M., Puckett, E.E., Gu, Z., Moon, K.L., Kapp, J.D., Vinner, L., Vardanyan, Z., et al. (2021). Environmental genomics of Late Pleistocene black bears and giant short-faced bears. Curr. Biol. *31*, 2728–2736.e8.
23. Murchie, T.J., Monteath, A.J., Mahony, M.E., Long, G.S., Cocker, S., Sadoway, T., Karpinski, E., Zazula, G., MacPhee, R.D.E., Froese, D., and Poinar, H.N. (2021). Collapse of the mammoth-steppe in central Yukon as revealed by ancient environmental DNA. Nat. Commun. *12*, 7120.
24. Froese, D.G., Zazula, G.D., Westgate, J.A., Preece, S.J., Sanborn, P.T., Reyes, A.V., and Pearce, N.J.G. (2009). The Klondike goldfields and Pleistocene environments of Beringia. GSA Today *19*, 4–10.
25. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
26. D'Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W.L., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., et al. (2011). Antibiotic resistance is ancient. Nature *477*, 457–461.
27. Chang, D., Knapp, M., Enk, J., Lippold, S., Kircher, M., Lister, A., Macphee, R.D.E., Widga, C., Czechowski, P., Sommer, R., et al. (2017).

The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. Sci. Rep. *7*, 44585.

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. Burn, C.R., Michel, F.A., and Smith, M.W. (1986). Stratigraphic, isotopic, and mineralogical evidence for an early Holocene thaw unconformity at Mayo, Yukon Territory. Can. J. Earth Sci. *23*, 794–803.
30. Kotler, E., and Burn, C.R. (2000). Cryostratigraphy of the Klondike "muck" deposits, west-central Yukon Territory. Can. J. Earth Sci. *37*, 849–861.
31. Fraser, T.A., and Burn, C.R. (1997). On the nature and origin of "muck" deposits in the Klondike area, Yukon Territory. Can. J. Earth Sci. *34*, 1333–1344.
32. Mahony, M.E. (2015). 50,000 years of paleoenvironmental change recorded in meteoric waters and coeval paleoecological and cryostratigraphic indicators from the Klondike Goldfields, Yukon, Canada, Master's thesis (University of Alberta).
33. Barnes, I., Matheus, P., Shapiro, B., Jensen, D., and Cooper, A. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. Science *295*, 2267–2270.
34. Fernando, P., Pfrender, M.E., Encalada, S.E., and Lande, R. (2000). Mitochondrial DNA variation, phylogeography and population structure of the Asian elephant. Heredity (Edinb) *84*, 362–372.
35. Barlow, A., Cahill, J.A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G.G., Paijmans, J.L.A., Rabeder, G., Frischauf, C., Grandal-D'anglade, A., et al. (2018). Partial genomic survival of cave bears in living brown bears. Nat. Ecol. Evol. *2*, 1563–1570.
36. Froese, D., Westgate, J., Preece, S., and Storer, J. (2002). Age and significance of the Late Pleistocene Dawson tephra in eastern Beringia. Quat. Sci. Rev. *21*, 2137–2142.
37. Davies, L.J., Jensen, B.J.L., Froese, D.G., and Wallace, K.L. (2016). Late Pleistocene and Holocene tephrostratigraphy of interior Alaska and Yukon: key beds and chronologies over the past 30,000 years. Quat. Sci. Rev. *146*, 28–53.
38. Westgate, J.A., Preece, S.J., Kotler, E., and Hall, S. (2000). Dawson tephra: a prominent stratigraphic marker of Late Wisconsinan age in west-central Yukon, Canada. Can. J. Earth Sci. *37*, 621–627.
39. Zazula, G.D., Froese, D.G., Elias, S.A., Kuzmina, S., La Farge, C., Reyes, A.V., Sanborn, P.T., Schweger, C.E., Scott Smith, C.A., and Mathewes, R.W. (2006). Vegetation buried under Dawson tephra (25,300 14C years BP) and locally diverse Late Pleistocene paleoenvironments of Goldbottom Creek, Yukon, Canada. Palaeogeogr. Palaeoclimatol. Palaeoecol. *242*, 253–286.
40. Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The last glacial maximum. Science *325*, 710–714.

Q13

41. Lammers, Y., Heintzman, P.D., and Alsos, I.G. (2021). Environmental palaeogenomic reconstruction of an Ice Age algal population. Commun. Biol. *4*, 220.
42. Froese, D., Stiller, M., Heintzman, P.D., Reyes, A.V., Zazula, G.D., Soares, A.E.R., Meyer, M., Hall, E., Jensen, B.J.L., Arnold, L.J., et al. (2017). Fossil and genomic evidence constrains the timing of bison arrival in North America. Proc. Natl. Acad. Sci. USA *114*, 3457–3462.
43. Heintzman, P.D., Froese, D., Ives, J.W., Soares, A.E.R., Zazula, G.D., Letts, B., Andrews, T.D., Driver, J.C., Hall, E., Hare, P.G., et al. (2016). Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. Proc. Natl. Acad. Sci. USA *113*, 8057–8063.
44. Shapiro, B., Drummond, A.J., Rambaut, A., Wilson, M.C., Matheus, P.E., Sher, A.V., Pybus, O.G., Gilbert, M.T.P., Barnes, I., Binladen, J., et al. (2004). Rise and fall of the Beringian steppe bison. Science *306*, 1561–1565.
45. Shackleton, N.J., Sánchez-Goñi, M.F., Pailler, D., and Lancelot, Y. (2003). Marine isotope substage 5e and the Eemian interglacial. Glob. Planet. Change *36*, 151–155.
46. Hu, A., Meehl, G.A., Otto-Bliesner, B.L., Waelbroeck, C., Han, W., Loutre, M.F., Lambeck, K., Mitrovica, J.X., and Rosenbloom, N. (2010). Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. Nat. Geosci. *3*, 118–121.
47. Meiri, M., Lister, A.M., Collins, M.J., Tuross, N., Goebel, T., Blockley, S., Zazula, G.D., van Doorn, N., Guthrie, R.D., Boeskorov, G.G., et al. (2013). Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. Proc. Biol. Soc. *281*, 20132167.
48. Elias, S.A., and Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? Quat. Sci. Rev. *27*, 2473–2483.
49. Vershinina, A.O., Heintzman, P.D., Froese, D.G., Zazula, G., Cassatt-Johnstone, M., Dalén, L., Sarkissian, C.D., Shelly, G.D., Ermini, L., et al. (2021). Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. Mol. Ecol. *30*, 6144–6161.
50. Barrón-Ortiz, C.I., Avilla, L.S., Jass, C.N., Bravo-Cuevas, V.M., Machado, H., and Mothé, D. (2019). What is Equus? Reconciling taxonomy and phylogenetic analyses. Front. Ecol. Evol. *7*, 343.
51. Harington, C.R. (2011). Pleistocene vertebrates of the Yukon Territory. Quat. Sci. Rev. *30*, 2341–2354.
52. Cinq-Mars, J. (1979). Bluefish Cave I: a Late Pleistocene Eastern Beringian cave deposit in the Northern Yukon. Can. J. Archaeol. *3*, 1–32.
53. Porter, L. (1988). Late Pleistocene fauna of Lost Chicken Creek, Alaska. Arctic *41*, 303–313.
54. Handsley-Davis, M., Kowal, E., Russell, L., and Weyrich, L.S. (2021). Researchers using environmental DNA must engage ethically with Indigenous communities. Nat. Ecol. Evol. *5*, 146–148.
55. Arnold, L.J., Roberts, R.G., Macphee, R.D.E., Haile, J.S., Brock, F., Möller, P., Froese, D.G., Tikhonov, A.N., Chivas, A.R., Gilbert, M.T.P., and Willerslev, E. (2011). Paper II – dirt, dates and DNA: OSL and radiocarbon chronologies of perennially frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. Boreas *40*, 417–445.
56. Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M.T.P., Nielsen, R., Munch, K., Ho, S.Y.W., Shapiro, B., and Willerslev, E. (2007). Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? Mol. Biol. Evol. *24*, 982–989.
57. Cribdon, B., Ware, R., Smith, O., Gaffney, V., and Allaby, R.G. (2020). PIA: more accurate taxonomic assignment of metagenomic data demonstrated on sedaDNA From the North Sea. Front. Ecol. Evol. *8*, 1–12.
58. Yoccoz, N.G., Bråthen, K.A., Gielly, L., Haile, J., Edwards, M.E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., et al. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. Mol. Ecol. *21*, 3647–3655.
59. Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., and Minamoto, T. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. Freshw. Biol. *62*, 30–39.
60. 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.
61. Hofreiter, M., Paijmans, J.L.A., Goodchild, H., Speller, C.F., Barlow, A., Fortes, G.G., Thomas, J.A., Ludwig, A., and Collins, M.J. (2015). The future of ancient DNA: technical advances and conceptual shifts. BioEssays *37*, 284–293.
62. Epp, L.S., Zimmermann, H.H., and Stoof-Leichsenring, K.R. (2019). Sampling and extraction of ancient DNA from sediments. In Ancient DNA: Methods and Protocols, B. Shapiro, A. Barlow, P.D. Heintzman, M. Hofreiter, J.L.A. Paijmans, and A.E.R. Soares, eds. (Humana Press), pp. 31–44.
63. Rawlence, N.J., Lowe, D.J., Wood, J.R., Young, J.M., Churchman, G.J., Huang, Y.T., and Cooper, A. (2014). Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. J. Quat. Sci. *29*, 610–626.

64. Kokelj, S.V., Tunnicliffe, J., Lacelle, D., Lantz, T.C., Chin, K.S., and Fraser, R. (2015). Increased precipitation drives mega slump development and destabilization of ice-rich permafrost terrain, northwestern Canada. Glob. Planet. Change *129*, 56–68.

65. Kokelj, S.V., Lantz, T.C., Tunnicliffe, J., Segal, R., and Lacelle, D. (2017). Climate-driven thaw of permafrost preserved glacial landscapes, northwestern Canada. Geology *45*, 371–374.

66. Nitze, I., Grosse, G., Jones, B.M., Romanovsky, V.E., and Boike, J. (2018). Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and subarctic. Nat. Commun. *9*, 5423.

67. Lewkowicz, A.G., and Way, R.G. (2019). Extremes of summer climate trigger thousands of thermokarst landslides in a High Arctic environment. Nat. Commun. *10*, 1329.

68. Ward Jones, M.K., Pollard, W.H., and Jones, B.M. (2019). Rapid initialization of retrogressive thaw slumps in the Canadian high Arctic and their response to climate and terrain factors. Environ. Res. Lett. *14*.

69. Dyke, A.S. (2004). An outline of the deglaciation of North America with emphasis on central and northern Canada. Dev. Quat. Sci. *2*, 373–424.

70. Clark, P.U., and Mix, A.C. (2002). Ice sheets and sea level of the Last Glacial Maximum. Quat. Sci. Rev. *21*, 1–7.

71. Bond, J.D. (2019). Paleodrainage map of Beringia. Yukon Geological Survey, Open File 2019–2. https://data.geology.gov.yk.ca/Reference/81642#InfoTab.

72. Heintzman, P.D., Zazula, G.D., MacPhee, R.D.E., Scott, E., Cahill, J.A., McHorse, B.K., Kapp, J.D., Stiller, M., Wooller, M.J., Orlando, L., et al. (2017). A new genus of horse from Pleistocene North America. eLife *6*, 1–43.

73. Sadoway, T.R. (2014). A metagenomic analysis of ancient sedimentary DNA across the Pleistocene-Holocene transition (PhD thesis (McMaster University)).

74. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

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

76. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res *40*, e3.

77. Klunk, J., Duggan, A.T., Redfern, R., Gamble, J., Boldsen, J.L., Golding, G.B., Walter, B.S., Eaton, K., Stangroom, J., Rouillard, J.-M., et al. (2019). Genetic resiliency and the Black Death: no apparent loss of mitogenomic diversity due to the Black Death in medieval London and Denmark. Am. J. Phys. Anthropol. *169*, 240–252.

78. Renaud, G., Stenzel, U., and Kelso, J. (2014). LeeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res *42*, e141.

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

80. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

81. Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., and Tappu, R. (2016). MEGAN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol. *12*, e1004957.

82. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res *17*, 377–386.

83. Zeyland, J., Wolko, L., Lipiński, D., Woźniak, A., Nowak, A., Szalata, M., Bocianowski, J., and Słomski, R. (2012). Tracking of wisent–bison–yak mitochondrial evolution. J. Appl. Genet. *53*, 317–322. Q14

84. Lott, M.T., Leipzig, J.N., Derbeneva, O., Michael Xie, H., Chalkia, D., Sarmady, M., Procaccio, V., and Wallace, D.C. (2013). MtDNA variation and analysis using Mitomap and Mitomaster. Curr. Protoc. Bioinformatics *44*, 1.23.1–1.23.26.

85. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res *44*, W58–W63.

86. Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Hooshiar Kashani, B.H., et al. (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am. J. Hum. Genet. *90*, 915–924.

87. 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.

88. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792–1797.

89. Eaton, K. (2020). NCBImeta: efficient and comprehensive metadata retrieval from NCBI databases. J. Open Source Softw. *5*, 1990.

90. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

91. 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.

92. 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.

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Yukon permafrost cores	This study	N/A
Chemicals, peptides, and recombinant proteins		
Tris-Cl (1M)	Fisher Scientific	BP1758-500
20% SDS	Fisher Scientific	AM9820
Proteinase K	Roche Molecular Systems	05114403001
Calcium chloride ($CaCl_{2)}$	VWR	CA80501-994
Dithiothreitol (DTT)	VWR	CAJTF780-1
Polyvinylpyrrolidone (PVP)	Fisher Scientific	BP431-100
N-Phenacylthiazolium bromide (97%) (PTB)	Oakwood Chemical	080244
Ultrapure water	VWR	75832-010
5 M Guanidine Hydrochloride	Fisher Scientific	BP178-1
Isopropanol (100%), molecular grade	Fisher Scientific	BP26181
Tween 20	Sigma-Aldrich	P9416-100ML
3 M Sodium Acetate (pH 5.2)	VWR	97062-834
Sodium chloride (NaCl)	VWR	CASX0420-1
EDTA 0.5M (pH 8.0)	VWR	97062-656
NEBuffer 2.1	New England BioLabs	NEB #B7202
dNTP Set [100 mM]	VWR	CA95057-688
ATP Solution (100 mM)	Fisher Scientific	FERR0441
T4 Polynucleotide Kinase	New England BioLabs	M0201L
Uracil-DNA glycosylase	New England BioLabs	M0280S
Endonuclease VIII	New England BioLabs	M0299S
T4 DNA polymerase	New England BioLabs	M0203L
T4 DNA Ligase	New England BioLabs	M0202L
50% w/v Polyethylene glycol 4,000 (PEG 4000)	NeXtal BioTech	133085
T4 DNA Ligase Buffer	New England BioLabs	M0202L
ThermoPol Reaction Buffer	New England BioLabs	M0275L
Bst DNA Polymerase, Large Fragment	New England BioLabs	M0275L
KAPA SYBR FAST Bio-Rad iCycler 2X	Sigma-Aldrich	KK4608
Ethyl Alcohol Anhydrous (Ethanol)	Greenfield	P006EAAN
Invitrogen UltraPure Agarose	Fisher Scientific	16-500-100
GeneRuler 50 bp DNA Ladder	Thermo Fisher	SM0373
SYBR Safe DNA Gel Stain	Thermo Fisher	S33102
TAE Buffer	Thermo Fisher	B49
TriTrack DNA Loading Dye	Thermo Fisher	R1161
Critical commercial assays		
High Pure Viral Nucleic Acid Large Volume Kit	Roche Molecular Systems	05114403001
PowerBead Tubes, Garnet	Qiagen	12888-100-PBT
myBaits Custom DNA-Seq Kit (baits, hybridization, block, and washing reagents)	Daicel Arbor Biosciences	N/A
QiaQuick Nucleotide Removal Kit	Qiagen	28306
Minelute PCR Purification Kit	Qiagen	28006
MinElute Gel Extraction Kit	Qiagen	28606

(*Continued on next page*)

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
NCBI SRA sequence data	This study	BioProject: PRJNA752360. Accessions: SRR15356337– SRR15356342.
Oligonucleotides		
Illumina library adapter sequences, indices, primers	Meyer et al.[75] Kircher et al.[76]	N/A
PalaeoChip Arctic v1.0 bait set	Murchie et al.[6]	N/A
Software and algorithms		
bcl2fastq	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
fastq2bam	Gabriel Renaud	https://github.com/grenaud/BCL2BAM2FASTQ
BLASTn	Renaud et al.[78]	https://blast.ncbi.nlm.nih.gov/
network-aware-BWA	Zeyland et al.[84]	https://github.com/mpieva/network-aware-bwa
Libbam	Gabriel Renaud	https://github.com/grenaud/libbam
Biohazard	Udo Stenzel	https://bitbucket.org/ustenzel/biohazard
NGSXRemoveDuplicates	Katherine Eaton	https://github.com/ktmeaton/NGSeXplore
MEGAN	Huson et al.[81,82]	https://github.com/husonlab/megan-ce
Samtools	Genome Research Limited	https://github.com/samtools/samtools
Geneious Prime 2021.2.2	Geneious	https://www.geneious.com/
mapDamage	Jónsson et al.[25]	http://https//ginolhac.github.io/mapDamage/
Bcftools	Samtools	https://samtools.github.io/bcftools/
NCBImeta	Eaton[89]	https://github.com/ktmeaton/NCBImeta
IQ-Tree v1.6.12	Weissensteiner et al.[85]	http://www.iqtree.org
BEAUti & The BEAST, TreeAnnotator, LogCombiner,	Pala et al.[86]	https://beast.community/
Tracer v1.7.1	Nguyen et al.[90]	http://beast.community/tracer
FigTree v1.4.4	Andrew Rambaut	http://tree.bio.ed.ac.uk/software/figtree/
Muscle	Lott et al.[85]	https://github.com/rcedgar/muscle
GenBank	Li and Durbin[79]	https://www.ncbi.nlm.nih.gov/genbank/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Tyler J. Murchie (murchiet@mcmaster.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

SedaDNA sequence data mapped to the PalaeoChip reference sequences have been deposited at the NCBI SRA under BioProject PRJNA752360 and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. PalaeoChip reference and bait sequences used for capture enrichment are available at https://doi.org/10.5281/zenodo.5643845. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The central Yukon permafrost cores used in this research were previously collected, dated, and analyzed by D'Costa et al.,[26] Mahony,[32] Sadoway,[73] and Murchie et al.,[6,23] and are currently in cold storage at the McMaster Ancient DNA Centre and the Permafrost Archives Laboratory at the University of Alberta. These cores were collected between June and August of 2010, 2012, and 2013 with research permits issued to DF from the Yukon Heritage Branch. The cores were sampled at placer gold mining exposures chosen for

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

the quality of the exposure and expected age of the sediments. Prior to sample collection by all three original research teams, the sampling area was cleared of eroded materials back to frozen sediments to create a fresh coring surface for a ~10 cm diameter coring tube ~30 cm in length. Horizontal core samples were drilled with a small portable gas-powered drill (Echo), recovered frozen, stored individually in plastic bags, immediately placed in a -20°C chest freezer, and transported in the freezer to the University of Alberta or McMaster University for subsampling. Core locations were recorded with a GPS in the field. Stratigraphic information was recorded in a field notebook at the time of sampling. Horizontal permafrost cores were collected from Bear Creek, Upper Quartz, and Upper Goldbottom. Vertical cores were taken from Lucky Lady II. Murchie et al.[23] redated and Bayesian age-modelled the cores. See Data S1A for a breakdown of core identifiers and ages, and see Murchie et al.[23] for further stratigraphic, aging, palynological, and metagenomic data.

METHOD DETAILS

Ancient DNA clean lab

Laboratory work was conducted in clean rooms at the McMaster Ancient DNA Centre, which are subdivided into dedicated facilities for sample preparation (with separate facilities for subsampling eDNA versus discrete materials like bone), stock solution setup, and DNA extraction through library preparation. The post-indexing and capture enrichment clean room is in a physically isolated facility from the Centre's standard aDNA labs, while the subsequent high-copy post-PCR workspace is in a separate building. The Centre has a unidirectional workflow progressing from low-copy to high-copy facilities to reduce the chance of cross-contamination. Each dedicated workspace is physically separated with air pressure gradients between rooms to reduce exogenous airborne contamination. Prior to all phases of laboratory work, dead air hoods and workspaces were cleaned using a 6% solution of sodium hypochlorite (commercial bleach) followed by a wash with Nanopure purified water (Barnstead) and 30 minutes of UV irradiation at >100 mJ/cm^2. Gloves are changed frequently throughout, and negative controls (extraction, library, and PCR blanks) accompany each sample batch.

Subsampling

Most samples were lysed and extracted by Murchie et al.,[23] although 4 new extracts (0.25 g raw sediment input each) from core BC 4-2B (PHP-1) were processed for UDG treatment. These additional samples were treated identically but were each subdivided into multiple replicates to increase sequencing coverage. They were subsampled as follows.

Metal sampling tools were cleaned with commercial bleach, rinsed with Nanopure water immediately thereafter, UV irritated on both sides for >30 min, then heated overnight in an oven at ~130°C. Once the tools had cooled the next day, work surfaces were Q12 cleaned with bleach and Nanopure water and covered with sterile lab-grade tin foil. Sediment cores previously split into disks[105,108] and stored at −20°C had the upper ~1 mm of external sediment chiseled off to create a fresh sampling area free of exogenous contaminants. A small (~1/4 inch) decontaminated chisel was used to carefully remove interior sediment from the core, which was collected in a weigh boat. After enough material was acquired for multiple extractions (~2–5 g), the core was covered in sterile tin foil and re-frozen. The subsampled material in the weigh boat was homogenized by manually stirring using a small metal chisel as the sediment thawed. This sediment was transferred to a 50 mL falcon tube and refrozen. Gloves were changed frequently throughout subsampling (multiple times per core) to minimize contamination. The homogenized sediments were later subsampled for subsequent DNA extractions.

Lysing and purification

Homogenized subsamples (0.25 g each) were loaded into DNeasy PowerBead Garnet tubes (preloaded with 750 μL PowerBead Solution) with a digest buffer containing 0.02 M Tris-Cl (pH 8.0), 0.5% SDS, 0.01 M $CaCl_2$, 100 mM DTT, 5 mM PTB, and 2.5% PVP. The PowerBead tubes were then vortexed for 20 minutes using a TissueLyser II (Qiagen). Thereafter, the tubes were briefly centrifuged to remove liquid from the lids, and proteinase K (0.25 mg/mL) was pipetted into each tube individually. The tubes were then briefly finger vortexed to disturb the sediment-bead pellets that had formed at the base and the tubes and were loaded in an incubator to oscillate overnight at 35°C. The next day, the PowerBead tubes were centrifuged at 10,000 x g for 5 minutes and the supernatant was transferred to a 2 mL MAXYMum Recovery tube and stored at −20°C for later purifications.

For sedaDNA purification, the digestion supernatant (≈1.25 mL) was thawed, briefly centrifuged, and added to ≳16.25 mL (13 volumes) of high-volume guanidinium binding buffer (5 M guanidinium hydrochloride, 40% isopropanol, 0.05% Tween-20, 0.1 Sodium acetate [pH 5.2]) in a 50 mL falcon tube and mixed by repeated inversion. The 50 mL tubes were loaded into a refrigerated centrifuge for the Murchie et al.[6] long cold spin, where they were centrifuged at 2500 x g at 4°C for ~20 hours overnight. Thereafter, the falcon tubes were carefully removed from the centrifuge buckets, and the supernatant was decanted, taking care to not disturb the darkly coloured pellet that had formed during the cold spin. The binding buffer was passed through a high-volume silica-column (High Pure Extender Assembly, Roche Molecular Systems) over multiple rounds of centrifugation and extraction proceeded as per Dabney et al.[74] with binding and wash centrifugation at 3300 x g, two rounds of PE wash, followed by two 30 second dry spins at 16,000 x g with the tubes rotated 180° between spins to minimize the chance of ethanol retention. Purified DNA was eluted off the silica columns with two volumes of 25 μL EBT (each while waiting 5 minutes after EBT loading to maximize elution, then centrifuging at 16,000 x g for 1 minute). Prior to all subsequent experiments the extracts were centrifuged at 16,000 x g for ≳5 minutes to pellet any remaining co-eluted inhibitors. Care was taken when subsampling these extracts to avoid disturbing any pellet precipitates.

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

Library preparation, quantitative PCR, and indexing

Doubled stranded libraries were prepared as described in Meyer and Kircher[75] with modifications from Kircher et al.[76] This augmented set of subsamples from core BC 4-2B (PHP-1) was library prepared with UDG treatment for improved mitogenomic assembly, which was split into multiple library replicates for downstream processing. UDG Blunt-end repair (step 1: 1X NE Buffer 2.1, 1 mM DTT, 100 μM dNTP mix, 1 mM ATP, 0.4 U/μL T4 polynucleotide kinase, 0.1 U/μL Uracil-DNA glycosylase, and 0.4 U/μL Endonuclease VIII; step 2: 0.2 U/μL T4 DNA polymerase, see Data S1N) was followed with a QIAquick PCR Purification Kit (Qiagen) to maximally retain small fragments. After adapter ligation (1X T4 DNA Ligase Buffer, 5% PEG-4000, 0.25 μM library adapter mix (P5/P7), 0.125 U/μL T4 DNA Ligase) a MinElute PCR Purification Kit (Qiagen) was used as the fragments were sufficiently long. Heat deactivation was used after adapter fill-in (1X ThermoPol Rxn Buffer, 250 μM dNTP mix, 0.4 U/μL BST Polymerase) to avoid further purification loss. Quantitative PCR (qPCR) on a BioRad CFX96 was used to assess library conversion success (1X KAPA SYBR FAST qPCR Master Mix [2X], 0.2 μM Meyer IS7 primer, 0.2 μM Meyer IS8 primer). Thereafter, libraries were dual indexed with qPCR (1X KAPA SYBR FAST qPCR Master Mix [2X], 750 nM Meyer unique forward index primer, 750 nM Meyer unique reverse index primer) on a BioRad CFX96 for 8-12 cycles (libraries that reached plateau earlier than cycle 12 were removed early and placed at 60°C in a secondary thermocycler, then returned to the BioRad CFX96 for the final extension). See Murchie et al.[6,23] for further details on master mix concentrations and thermocycling parameters.

RNA hybridization targeted capture

Probe design

In-solution enrichments were carried out using the previously designed PalaeoChip Arctic v1.0 bait set.[6] This bait set targets whole mitochondrial genomes from approximately 180 extinct and extant Holarctic fauna, and the chloroplast barcoding loci (*trn*L, *rbc*L, and *mat*K) from approximately 2100 species of plants. See Murchie et al.[6] for further details on the design of PalaeoChip Arctic v1.0. The bison bait set contains probes designed by myBaits at Daicel Arbor Biosciences for the capture of American bison (*Bison bison*) and Steppe bison (*Bison priscus*) mitochondrial genomes in equal proportions.

Enrichment wet lab

In solution enrichments were carried out using a modified version of the myBaits v4.1 protocol (Daicel Arbor Biosciences). For each library, 7 μL of template was combined with 5 μL of the library block master mix (using IDT xGen blocking oligos [0.04 μg/μL], Human COt-1 DNA [0.19 μg/μL], and salmon sperm DNA [0.19 μg/μL]). The myBaits hybridization mix (Hyb N [9X SSPE, 6.25 mM EDTA], 8.75 X Hyb D, 0.25% Hyb S, 1.56 X Hyb R, 200 ng/rxn of PalaeoChip Arctic plant/animal baits) was pre-warmed to 60°C before being combined with the library-block mixture. The first batch of samples (PHP) was incubated for 48 hours at 55°C for bait-library hybridization. The subsequent round of libraries (PHP_{ii} and the augmented Bear Creek batches) were enriched with a hybridization temperature of 60°C over ~72 hours to improve off-target exclusion. Only a single round of hybridization was performed on each library. See Murchie et al.[6,23] and the myBaits v4.1 for further details.

After the hybridization, beads were dispensed (20 μL per reaction), washed with 200 μL of binding buffer per reaction, then resuspended in 20 μL binding buffer per reaction and aliquoted into PCR strips. Baits were captured using 20 μL of the bead binding buffer suspension per library, incubated at 55°C for 2.5 minutes (60°C for the second round), finger vortexed and spun down, then incubated for another 2.5 minutes. Beads were pelleted and the supernatant (the non-captured library fraction) was removed and stored at −20°C as per Klunk et al.[77] The beads were resuspended in 180 μL of 60°C Wash Buffer X per tube and washed four times following the myBaits v4.1 protocol. Beads were resuspended in 18.8 μL EBT, PCR reamplified for 12 cycles, then purified with MinElute columns following manufacturer's protocols and eluted in 15 μL EBT.

Total quantification, pooling, size selection, and sequencing

Enriched and indexed libraries were quantified using the long-amplification total library qPCR assay and pooled to equimolar concentrations.[23] Pools were size-selected with gel excision following electrophoresis for molecules ranging between 150–500 bp. Gel plugs were purified using the QIAquick Gel Extraction Kit (QIAGEN), according to the manufacturer's protocol, then sequenced on an Illumina HiSeq 1500 with a 2x75 bp paired-end protocol at the Farncombe Metagenomics Facility (McMaster University, ON).

Bioinformatics

Demultiplexing and preliminary metagenomics

Reads were demultiplexed with *bcl2fastq* (v 1.8.4), converted to bam files with *fastq2bam* (https://github.com/grenaud/BCL2BAM2FASTQ), then trimmed and merged with *leeHom*[78] using ancient DNA specific parameters (–ancientdna). Initially for *BLASTn* identification, reads were mapped (map-filtered) to a concatenation of the PalaeoChip Arctic v1.0 plant and animal probe references with *network-aware-BWA*[79] with a maximum edit distance of 0.01 (-n 0.01), allowing for a maximum two gap openings (-o 2), and with seeding effectively disabled (-l 16500). Mapped reads that were merged or unmerged but properly paired were extracted with *libbam* (https://github.com/grenaud/libbam), collapsed based on unique 5' and 3' positions with *biohazard* (https://bitbucket.org/ustenzel/biohazard) (for PCR deduplication), and converted to FASTA files and restricted to a minimum length of 24 bp. Subsequent FASTA files were additionally filtered to remove any reads with lingering sequence similarity to the Illumina adapter sequences (sequences with an edit distance of 1 of the common adapter sequence AGATCGGAA and its reverse complement) and were string deduplicated using the *NGSXRemoveDuplicates* module of *NGSeXplore* (https://github.com/ktmeaton/NGSeXplore).

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

These filtered FASTAs were used as the input for *BLASTn*,[80] which were aligned against a July 2019 local copy of the GenBank NCBI nucleotide database set to return the top 1000 alignments (unique accession hits) per read with e-values less than 1.0E-5 (flags: -num_alignments 1000 -max_hsps 1 -evalue 0.00001). The *BLASTn* outputs were then passed to *MEGAN* (Community Edition, v.6.21.7)[81,82] where the *BLASTn* results were filtered through a lowest common ancestor (LCA) algorithm using the following parameters: Min-score = 50, Max expected (e-value) = 1.0E-5, Minimum percent identity = 95%, Top percent consideration of hits based on bit-score = 20%, Minimum read support = 3, Minimum complexity = 0.3, LCA weighted algorithm at 80%. Replicate libraries were merged in *MEGAN* by core (Data S1A and S1B).

Extended faunal metagenomics from the Bear Creek site (PHP-1)

There are two notable peculiarities in this enhanced faunal dataset (Figure 1). First, D'Costa et al.[26] and Murchie et al.[6,23] both observed hits to the genus *Bos*, with Murchie and collegues identifying *Bos mutus/grunniens* (yak) in their *BLASTn* data initially, which return to *Bos* sp. after increasing the LCA stringency. Here, 76 reads are retained to *Bos* sp. (aligning across the mitogenome) with 6 being specifically called as *Bos primigenius* (aurochs). This consistent *Bos* signal may either suggest A) that Bos is a common false positive or contaminant within Bovinae, or B) that one of the Eurasian *Bos* species (*Bos primigenius* or *Bos mutus/grunniens*) was present in eastern Beringia at some low population density during the Pleistocene. These hits retain their *Bos* binning when percent identity is increased to 100%,[6] however *Bos* was also one of the few taxa of macro-ecological interest identified in the negative controls of Murchie et al.[23] This sample blank *Bos* contamination was entirely absent in the augmented negative control dataset here with increased *BLASTn* hits to 1000. However, there is reason to doubt this taxonomic identification as no Pleistocene *Bos* spp. macrofossils have been found in eastern Beringia to date.[4] Our UDG treated dataset was initially intended in part to more closely investigate this anomalous *Bos* signal as we suspected that some number of these *Bos* binned reads were false positives within Bovinae due to database bias, familial mitogenomic conservation, and previously inferred cross-breeding between *Bos* and *Bison priscus*.[84] *Bos* is highly over-represented on GenBank, which raises the question as to whether 1000 top hits is sufficient to classify these reads when there are 1000+ *Bos* spp. mitochondrial GenBank entries capable of swamping the *BLAST* results for conserved regions of the Bovinae mitogenome. When mapping to *Bos mutus, Bison* reads overwhelm the dataset, returning a *Bison priscus* consensus sequence, challenging attempts to isolate this *Bos* signal from other bovid DNA.

To investigate this *Bos* signal further, we competitively *network-aware-bwa* mapped the BC 4-2B core libraries against the RefSeq entries for *Bison priscus* (NC_027233.1) and *Bos grunniens* (NC_006380.3). Thereafter, reads with a map quality of >30 (and minimum size of 24 bp) were extracted. Reads with a map quality less than 30 in this case mapped equally well to either reference genome and were not useful in discriminating taxonomic identity. These competitively mapped reads were visualized in *Geneious*, edit distances (counts of nucleotide mismatches relative to the reference) were computed and plotted, and reads mapped to *B. grunniens* were extracted then run with *BLASTn* to retrieve the top 5000 hits. These *B. grunniens* MQ30 reads were then taxonomic binned with *MEGAN*[81] using both the naïve and weighted lowest common ancestor algorithm (LCA).

Based on these analyses (Figure S4) it would appear as if these *Bos* reads are likely the result of false positive binning within conserved regions Bovinae mitogenomes, with their genus/species assignment within *BLASTn* and *MEGAN* being driven by an over-abundance of *Bos* spp. genomes on GenBank NCBI relative to other bovids, as well ancient cross-breeding between the two genera.[84] The edit distance plots show that there are significant mismatches for reads mapped to the B. *grunniens* reference genome (relative to B. *priscus*) despite UDG treatment, and that database over-representation of domestic *Bos* species drove false positive assignments when using a weighted LCA.

The second peculiarity is that human is also identified in this metagenomic dataset with >3000 *MEGAN* binned reads despite neither bait-set targeting human DNA. When mapping to the rCRS (NC_012920.1) we can assemble a 56.2X mitochondrial genome with 91.8% reference coverage. After curation, *Mitomaster*[84] was used to identify this consensuses genome as haplogroup T2b5 with 86.99% quality as per *HaploGrep2*[85] (Data S1M). These libraries were UDG treated, so we cannot use *mapDamage* to evaluate deamination. However, haplogroup T2b is only considered to have diverged around 10,000 years BP and is predominantly a Near Eastern and European lineage.[86] In this case, we can conclude that this human signal is very likely the result of contamination. Approximately 9500 mapped reads came from a single indexed replicate of a library that otherwise had minimal human reads mapping in other sequenced index combinations, suggesting that one of the index primer aliquots may be contaminated. The negative controls run with this batch only contain 6 reads that map to the rCRS, suggesting that this contamination is specific to this indexed library. Only the lead author (TJM) worked with these materials in the lab, and his mitochondrial haplogroup is H1. This indicates that this human contamination originates from a third-party source, illustrating why typing everyone in a lab setting cannot simply be used to 'subtract' all possible sources of human (or other) contamination. In terms of the palaeo-faunal reconstruction of Bear Creek some 30,000 years ago, we can dismiss the human signal observed in Figure 1.

Mitogenome reconstruction

Metagenomically identified organisms (bison, mammoth, horse, ptarmigan) were used as mapping targets with the same filtering parameters in addition to a >30 map quality filter and minimum 24 bp fragment length with *samtools* (https://github.com/samtools/samtools). These mapped bam files were then imported into Geneious Prime 2021.2.2 (https://www.geneious.com/) for visualization where the contigs were manually curated using pre-assembled alignments and NCBI references as guides. Spurious-indels were removed, nucleotides with a read coverage less than 3 were 'N' masked, and regions with high coverage (>3 standard deviations from the coverage mean) were carefully inspected for non-specific mapping—all polymorphisms relative to the reference in these high coverage regions were N-masked. Consensus sequences were called at either a 65% (*Bison* and *Equus*) or 75% (*Mammuthus*) nucleotide identity threshold depending on the abundance of allelic diversity, otherwise ambiguous bases were called in those

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

positions. Libraries from Murchie et al.[6,23] that had not been UDG treated were processed with *mapDamage*[25] (v 2.0.3) to assess fragment length distributions (FLDs) and terminal deamination patterns. Mapping summary statistics of cores by taxon: woolly mammoth (*Mammuthus primigenius*), Data S1C; steppe bison (*Bison priscus*), Data S1D; caballine horse (*Equus caballus*), Data S1E; willow ptarmigan (*Lagopus lagopus*), Data S1F.

COMPETITIVE MAPPING

A second set of mitogenomes were reconstructed for *Mammuthus* and *Bison* to disentangle the intermixed lineages observed in the mapped data for cores dating prior to ca. 20,000 cal BP (Figures 2, 3, S1, and S2). Two *Mammuthus* and *Bison* references were selected for competitive mapping (**Data S1O**) and concatenated into two index references. Libraries were competitively mapped with *network-aware-BWA* using the same parameters as described previously, filtered to a minimum length of 24 bp, a minimum map quality (MQ) of 30, and core replicates were merged with *samtools*. Most reads in non-segregating regions have an MQ of 0 as these could map equally well to either reference, whereas reads with a >MQ30 have segregating polymorphisms that discriminate between the two references (Figures S1 and S2). A list of MQ30 North American and Eurasian specific reads were extracted using *samtools view* and *grep*. This list was then used to remove reads associated with the Eurasian lineage from the bam files originally mapped to either the *Mammuthus or Bison* NCBI RefSeq to create a competitively mapped, eastern Beringian specific bam file, and vice versa to create a Eurasian specific bam file.

This approach worked well for disentangling regions with a dense set of segregating SNPs, but singletons were still only assigned as ambiguous bases (as these are analogous to deamination derived polymorphisms or random sequencing errors). To further separate these mixed eDNA inputs we utilized the approach described by Lammers et al.[41] Our competitively mapped bam files were then used to generate variant call files (VCF) with *bcftools* using a combination of *mpileup* (-Ou –skip-indels –max-depth 1000), *call* (-Ou –variants-only –multiallelic-caller), *view* (-i '%QUAL>=30'), and *norm* (-m -). Each variant listed was manually inspected and compared to the competitively mapped bam files, the original mammoth/bison alignment, and the RefSeq entry to confirm whether the variant position existed within the known diversity of that organism, and whether there was sufficient coverage to call the variant. As with Lammers et al.[41] we observed that relative frequency could be used to differentiate between the two lineages, although off-target mapping in conserved regions such as 12S and 16S necessitated manual curation to remove erroneous SNPs. This approach is conservative insofar as we attempted to remove the influence of non-specific mapped reads (an expectation for metagenomic libraries), but at the cost of only accepting variants observed within the mitogenomic diversity of these organisms to date. Both the *Mammuthus* and *Bison* RefSeq entries are from Eurasian lineages. As such, variants from the RefSeqs were accepted at each position in the American lineage sedaDNA mitogenomes if the variants were present in known mammoth/bison diversity based on the alignment and if the variant had >3 coverage. For the alternative Eurasian lineage mitogenome, singleton variation from the RefSeq was only accepted if the variant was observed in other Eurasian mammoths/bison, the position had >3 coverage, or if all reads from that position were for the same SNP. Therein, in the American lineage VCF, the emphasis was on calling all variant alleles that differed from the RefSeq, whereas in the Eurasian lineage VCF, alternate alleles were only called when the RefSeq alelle was not present in our data. After curating the VCFs, they were compressed with *gzip*, indexed with *bcftools index*, and consensus sequences were called with *bcftools consensus*. Positions with insufficient total coverage (<3) were replaced with 'N's.

PHYLOGENETICS

Comparative mitochondrial genomes were acquired from GenBank NCBI for entries within Equidae, Bovidae, Elephantidae, and Phasianidae. Additional comparative genomes unavailable on GenBank (but available on the SRA) were acquired through contacting the corresponding authors from those publications; in particularly by contacting the authors of van der Valk et al.[87] for their million-year-old mammoth sequences. Comparative genomes were aligned independently by family using *muscle* v3.8.425,[88] curated for only sequences with $\gtrsim$ 80% completeness, and sequences with numerous highly divergent indels were removed (under suspicion that these genomes contained assembly errors). The control region was masked in the Bovidae and Elephantidae alignments. Geographic, age, and other associated metadata were retrieved using *NCBImeta*,[89] which were used to rename the entries for all retrieved sequences. Our sedaDNA reconstructed mitochondrial genomes were then *muscle* re-aligned with these curated multiple alignments and gaps at the ends of the alignment were replaced with Ns following a final visual inspection. Model testing and maximum likelihood analyses were conducted using *IQ-Tree* v1.6.12[90] with outgroup rooting and 1000 bootstrap replicates. The Bovidae and Elephantidae alignments were also assessed with Bayesian phylogenetic inference using the BEAST v1.10.5pre[91] software package and visualized using Tracer v1.7.1.[92] Priors and other details are elaborated on in the next subsection. Both the maximum likelihood and Bayesian trees were analyzed in *FigTree* v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

BAYESIAN PHYLOGENETICS

We ran an iterative series of Bayesian analyses with *BEAUti*/*BEAST* v1.10.5pre[91] to A) confirm the phylogenetic placement of our sedaDNA reconstructed mitogenomes relative to the maximum likelihood trees ran in parallel, and B) to assess the potential for tip dating with environmentally intermixed organelle genomes. Only sequences with associated dating information were used for the final Bayesian analyses. Tip dates were used as years before present. Nucleotide substitution models were derived from the

Please cite this article in press as: Murchie et al., Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.12.023

parallel maximum likelihood analyses processed with *IQ-Tree*[90] model test. A strict clock was used for all instances except Bison run 3. All priors and tracer summary stats are listed in Data S1G–S1L. Triplicate *BEAUti* files were generated for each, which were ran simultaneously with a chain length of 100,000,000. Each replicate was manually inspected in *Tracer v1.7.1*[92] to monitor for convergence (both within and between files) and violin plots of tip dating estimates were generated. Thereafter log and tree files were combined using log combiner for final analysis. *LogCombiner* and *TreeAnnotator*, as part of the *BEAST*v1.10.5pre package, were used to remove 10% chain burn-in as well as combine and annotate the replicates, which were then imported into *FigTree* v1.4.4 for visualization.

The Bayesian phylogenetic trees for both mammoth and bison (Figure S2) recapitulate all major clades identified in the corresponding maximum likelihood trees (Figures 2 and 3), as well as the differential placement of our sedaDNA mitogenomes within both Eurasian and American clades.

In our UDG augmented library (PHP-1; 30,000 cal BP) with significantly deeper sequencing depth (Data S1A and S1B), tip dating was consistently successful in estimating the age of the *Mammuthus primigenius* Clade 1/C (American) lineage genome across all runs (Figure S2; Data S1G, S1I, and S1K). This approach also estimated the same age ranges as van der Valk et al.[87] for their million-year-old mammoths (Figure S2). However, most other libraries across *Mammuthus* and *Bison* were not age estimated accurately. In many cases, samples were aged nonsensically as being some 100,000 years older than the core age estimates from Murchie et al.[23] While competitive mapping and the VCF approach of Lammers et al.[41] were successful in disentangling phylogenetically significant segregating SNPs (especially in regions with multiple nearby polymorphisms [<100 bp]), this approach is too conservative to fully disentangle the inputs of multiple individuals from each major lineage.

A Bayesian tip dating approach uses the number of mutations and prior estimated evolutionary rates to determine the probable age of a given sequence relative to samples with known dates. In cases of evolutionary insignificant, novel, or individual specific mutations, these were generally assigned ambiguous bases as there was insufficient depth to determine whether these polymorphisms were real or the result of non-specific mapping from closely related organisms, damage, or sequencing errors. Only in our deeply sequenced and UDG treated libraries (PHP-1) were we able achieve a sufficient sequencing depth to confidently accept these SNPs. In the case of the Eurasian lineage VCF variant consensus, most mutations were rejected to disentangle the Clade 1/DE haplogroup from the American Clade 1/C eDNA inputs. As such, this genome has fewer unique mutations relative to the reference, making it appear far older than it is. So, while our competitive mapping approach was largely successful in disentangling inputs from different lineages of *Mammuthus* and *Bison* for their phylogenetic placement with both maximum likelihood and Bayesian inference, these environmental genomes remain imperfectly disentangled for more sensitive tip dating analyses at this time. UDG treatment and deeper sequencing aids here significantly, but to some degree this problem may be inherent with environmentally mixed DNA. Individual specific SNPs are unlikely to ever be called in a consensus sequence if there are inputs from multiple diverse individuals (Figure S3). Despite this imperfect disentangling of multiple lineages (a problem in need of further bioinformatic research and development), these analyses aid in confirming the genetic diversity of eDNA inputs from multiple mitochondrial lineages of woolly mammoths and steppe bison. In cases where samples are sequenced deeply and UDG treated with high target coverage, Bayesian tip dating can be used to age estimate genomes reconstructed solely from sediment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Maximum likelihood and Bayesian inference was used to phylogenetically place the reconstructed sedaDNA sequences with reference to other publicly available mitogenomes on GenBank NCBI. See Method Details for further information on the phylogenetic approach used here.