nature communications

Article

https://doi.org/10.1038/s41467-026-72977-6

Ground squirrel coprolites preserve complex archives of ancient environmental DNA over 700,000 years

Received: 17 October 2025

Accepted: 24 April 2026

Published online: 09 June 2026

Check for updates

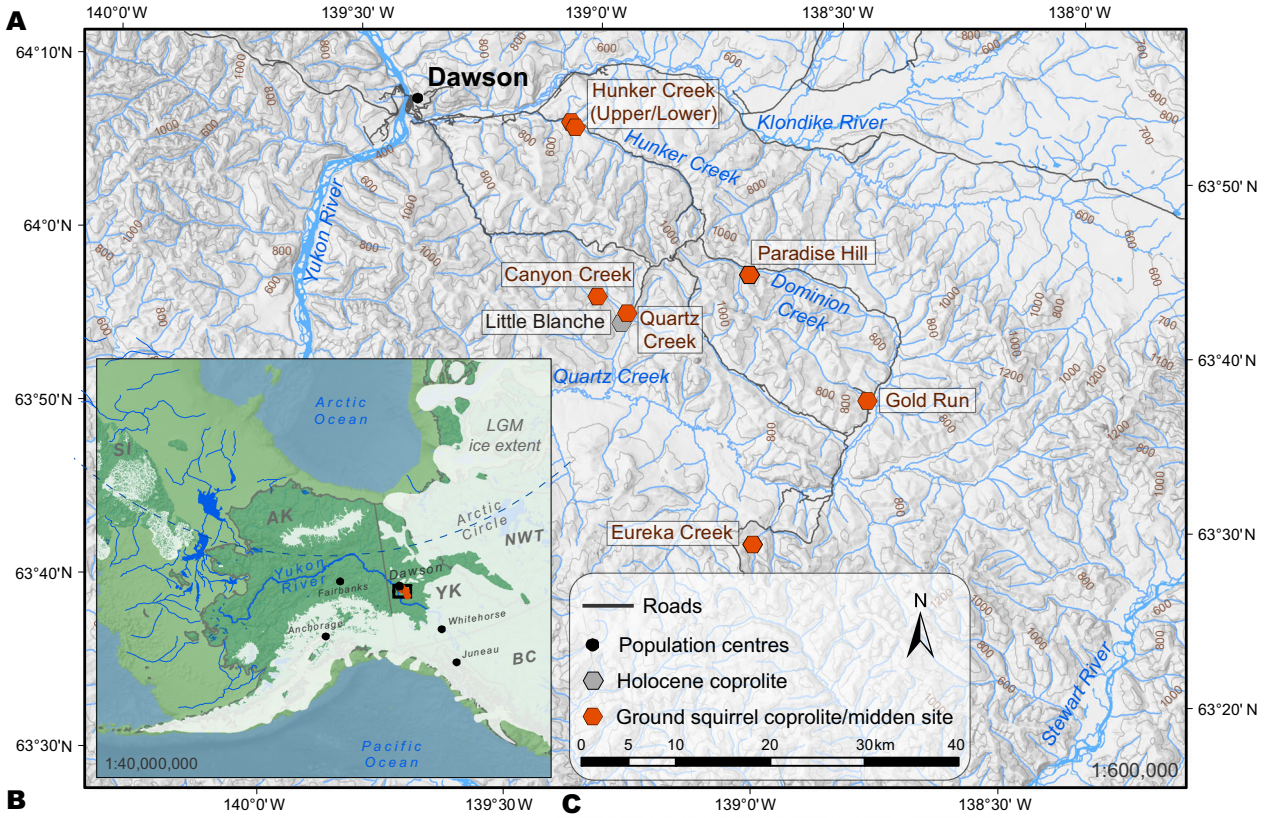
Tyler J. Murchie[1,2] ✉, Scott L. Cocker[3,4], Sina Baleka[2], Nicola Alexandra Vogel[2,5], Libby Natola[1], Emil Karpinski[2,6], Diana Tirlea[7], McIntyre A. Barrera[1], Danielle M. Grant[1], Evan Morien[1], George S. Long[2], Linda Y. Rutledge[1,8], Grant D. Zazula[9,10], Britta J. Jensen[3], Duane G. Froese[3] ✉ & Hendrik N. Poinar[2,11] ✉

Permafrost-preserved ground squirrel (*Urocitellus*) burrows in Yukon, Canada contain coprolites (palaeofaeces) that span from the Holocene to at least the Middle Pleistocene (~700 kya). Using shotgun metagenomics and targeted enrichment, we recover a rich, multi-taxon spectrum of ancient environmental DNA from these pellets, including: plants, insects, microbes, and megafauna consistent with eastern Beringian ecosystems. These coprolites consistently preserve an abundance of eukaryotic DNA, enabling the assembly of >18 mitochondrial genomes (ground squirrel, snowshoe hare, steppe bison, horse, and mammoth), and revealing previously unrecognized diversity within Arctic *Urocitellus*, including a ~700 kya lineage that predates divergence among several extant clades. Characteristic damage patterns, positive/negative controls, and in silico taxon validations strongly support aDNA authenticity, and comparisons with regional permafrost datasets indicate minimal post-depositional leaching. These results show that permafrost coprolites can yield high-resolution records of Quaternary ecosystems and multi-organism population histories, providing a powerful complement to sedimentary and skeletal ancient DNA.

Coprolites (ancient faeces) can serve as a time-restricted biological snapshot of the past, preserving a variety of biomolecules[1,2] and even ancient DNA (aDNA)[3–12]. This includes the aDNA of the defecator, their diet[13], their gut microbiome, and the aDNA of other organisms in their local environment. When preserved in permafrost over immense timescales, faecal pellets can retain ancient environmental DNA (aeDNA) that helps reconstruct past communities and track ecological change across glacial–interglacial cycles. In unglaciated regions of eastern Beringia (Yukon and Alaska) (Fig. 1), ground squirrels construct burrows that can remain frozen and sealed for tens to hundreds of

[1]Biodiversity Genomics, Hakai Institute, Heriot Bay, BC, Canada. [2]McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, ON, Canada. [3]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB, Canada. [4]Department of Geological Sciences and Centre for Palaeogenetics, Stockholm University, Stockholm, Sweden. [5]Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark. [6]Harvard Medical School, Harvard University, Cambridge, MA, USA. [7]Quaternary Environments, Royal Alberta Museum, Edmonton, AB, Canada. [8]Forestry and Conservation, University of British Columbia, Vancouver, BC, Canada. [9]Department of Tourism and Culture, Yukon Government, Palaeontology Program, Whitehorse, YT, Canada. [10]Collections and Research, Canadian Museum of Nature, Ottawa, ON, Canada. [11]McMaster Ancient DNA Centre, Departments of Biochemistry and Biology, McMaster University, Hamilton, ON, Canada. ✉e-mail: tyler.murchie@hakai.org; duane@ualberta.ca; poinarh@mcmaster.ca

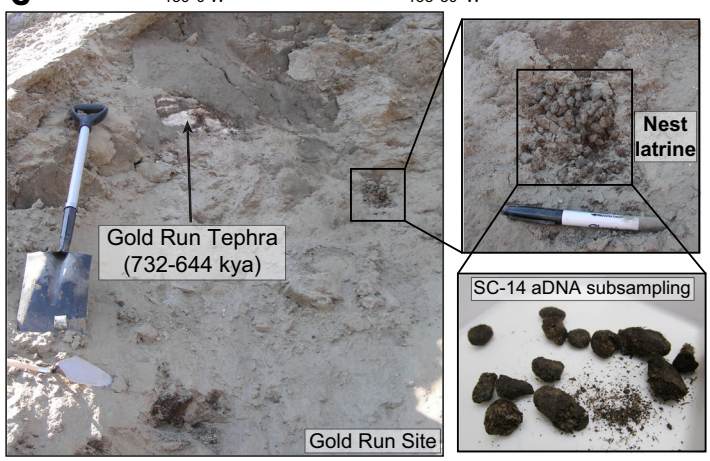
https://doi.org/10.1038/s41467-026-72977-6



B

aDNA ID	Site	Core ID	Age (cal. yr BP)	Shotgun (total reads)	Enriched (total reads)
SC-2	Hunker Creek	DF09-HC-30	17,175	15,920,534	8,619,666
SC-3	Hunker Creek	DF09-HC-29a	17,175		1,867,770
SC-5	Lower Hunker	SLC21-KL-1D	30,000^{DT}		10,531,629
SC-6	Quartz Creek	SC-AGS-01	80,000^{SCT}		2,571,597
SC-8	Quartz Creek	SC-AGS-04	80,000^{SCT}		2,131,519
SC-9	Canyon Creek	SC-AGS-12	80,000^{SCT}		19,883,053
SC-10	Eureka Creek	SC-AGS-15	31,035	16,287,382	19,659,670
SC-11	Eureka Creek	SC-AGS-18	28,660		1,875,061
SC-12	Little Blanche	SC-AGS-21	682		2,777,265
SC-13	Quartz Creek	SC-AGS-22	35,850	12,734,374	15,852,759
SC-14	Gold Run	GZ.07.07	700,000^{GRT}	242,855,809	290,600,543
SC-15	Lower Hunker	SLC21-KL-4A	30,000^{DT}		12,678,828
SC-16	Treadstone Mine	SLC21-TM-1C	30,000^{DT}		1,956,241
SC-17	Lower Hunker	SLC21-KL-8	30,000^{DT}		2,208,612
Negative controls (air, extraction, library) x11				975,495	4,675,363
•Arctic ground squirrel (Urocitellus sp.)			• Positive Holocene control		

Tephrochronology ages: ^{GRT}Gold Run Tephra; ^{SCT}Sheep Creek Klondike Tephra, ^{DT}Dawson Tephra. Other ages based on ^{14}C (median calibrated years BP).



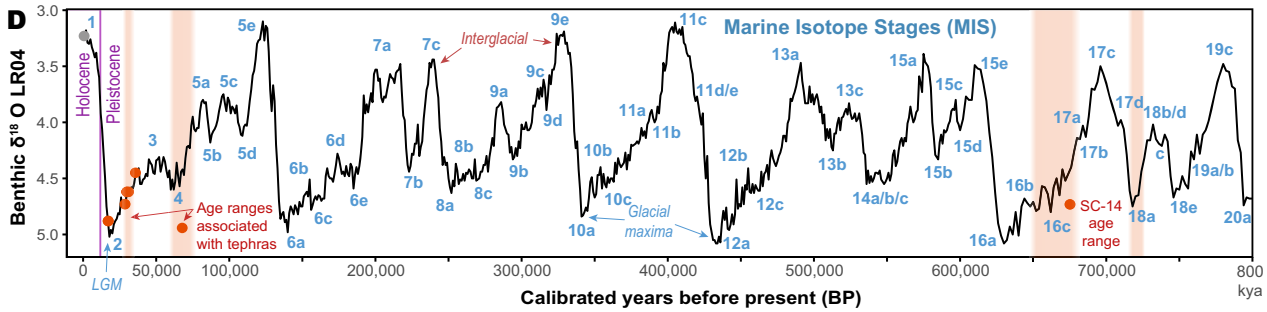


Fig. 1 | Coprolites processed for ancient environmental DNA with their midden locations. A Nest midden sites in the Klondike region of Yukon, Canada. Base map data from GeoYukon (Government of Yukon); contours in metres above sea level. Inset: Beringia during the last glacial maximum (LGM, 26.5–19 thousand years before present [ka BP])[67]. Ice sheet data from Dyke[193]. Sea level during LGM set to 126 m below sea level based on Clark and Mix[194]. Beringian palaeo-drainage data from Bond[195]. **B** Sample metadata and summary of aDNA batches. See Fig. S1 and Tables S1-S4 for additional metadata, and see Fig. S2 for photo examples of faecal pellets. **C** Photos during field recovery from Gold Run (2007) and subsampling (2023). **D** Benthic δ^{18}O record[196] with MIS substages[197]. Source data are provided as a Source Data file.

thousands of years[14–19]. These burrows commonly contain abundant faecal pellets and food caches, providing an opportunity to recover both host genetic material and multi-taxon time capsules of Quaternary ecosystems.

Coprolites were among some of the earliest trace fossils studied with aDNA[1,4,20–25] investigating shasta ground sloths (*Nothrotheriops shastensis*), humans[3,4,20,21,26–28], and parasitic nematodes[29]. Despite repeated observations of their biomolecular potential, coprolites continue to constitute a small proportion of typical aDNA sample types[23]. One reason for this is that coprolites occupy a biochemically intermediary space between discrete tissues (e.g. bone, teeth, skin, hair) and environmentally disseminated DNA substrates (e.g. sediments, soils, ice) that are not always immediately identifiable nor easily linked to their defecating source [22,23].

An immediate challenge when working with coprolites is that faeces are rich in substances like humic acids, phenolics, and carbohydrates (partially degraded organics), which can inhibit the functioning of necessary enzymes for downstream aDNA processing such as polymerization and ligation[30–33]. This is in addition to prolific DNA cross-linking[4] and other forms of DNA damage[5,12,34–37]. Coprolite aDNA processing thus requires an approach cognizant of this biomolecular complexity akin to other genetically mixed aeDNA sample types[22,37]. When enzymatic inhibition can be overcome while retaining a viable aeDNA signal, the resulting data would be expected to be rich biomolecular archives of Quaternary evolution, which is where the behavioural adaptations of ground squirrels become highly fortuitous.

Ground squirrels (family Sciuridae, tribe Marmotini) are omnivorous rodents that live predominantly on the ground or in burrows and include genera such as marmots (*Marmota*), chipmunks (*Tamias*), susliks (*Spermophilus*), ground squirrels (*Urocitellus*), and prairie dogs (*Cynomys*). Species of the genus *Urocitellus* once formed part of the much larger clade *Spermophilus*, but following recent phylogenetic work by Herron et al[38]. who found paraphyletic clades within *Spermophilus*, the large taxonomic clade was split into eight new genera[39]. Within *Urocitellus*, there are 13 species distributed primarily in northwestern North America and across the Bering Strait into Siberia (Fig. 2). Fossil records of ground squirrels as old as 2.5 and 1.8 million years ago (mya) have been found in Alaska, with Yukon records extending to at least 700 thousand years ago (kya)[18,40,41]. The most abundant records come from deposits dating to MIS 2 (~30 kya) and MIS 4 (~80 kya), which are found across Siberia, Alaska, and Yukon Territory [15,18,40,41].

Arctic ground squirrels (*U. parryii*) are the extant circumpolar species found within Beringia today (Fig. 2)[42]. They live in areas where the active layer is sufficiently thick (~1 metre) as permafrost impedes their burrowing[43]. Although they have traditionally been considered to be generalist herbivores, Arctic ground squirrels are opportunistic feeders that rely on a variety of plants (grasses, sedges, leaves, buds, flowers, and seeds), fungi[44], insects, and carrion[45,46]. These meat-eating tendencies are reported in a variety of forms, including: predation[45,47], scavenging[48,49], and infanticide/cannibalism[50,51]. Geist[48] and Cade[49] both report on carnivorous ground squirrel feeding on St. Lawrence Island (Bering Sea, Alaska) with scavenging walrus and whale meat, in addition to killing and trapping other rodents. Additional reports from Alaskan islands identified the impact of Arctic ground squirrels on bird populations by feeding on passerine eggs and on the eggs and chicks of seabirds[48,49,52]. Ebbert and Byrd[52] observed that storm petrels and small burrow-nesting species are often absent from islands that are home to Arctic ground squirrel populations. Active predation has been recorded on lemmings in Northwest Territories[45] along with suspected predation on juvenile snowshoe hares (*Lepus americanus)*[47]. Although fewer reports of predation exist compared to scavenging, predation appears to be a behaviour common in ground squirrel populations. A recent study of California ground squirrels (*Otospermophilus beechyi*) records active predation on voles in the Briones Regional Park, California[53]. Arctic ground squirrels likely adopted such opportunistic feeding strategies due to the considerable physiological demands of hibernation that requires sufficient fat stores to enter into a state of torpor. Hibernation occurs from early October to mid-April, during which time their core body temperature drops below freezing and their heart rate slows to ~1 beat per minute[54]. These many layers of unique behavioural, dietary, and physiological adaptations are what make Arctic ground squirrel burrows and their coprolites so appealing from a Quaternary science perspective for archiving rich, holistic snapshots of past ecosystems.

In this study, we metagenomically leverage the behavioural and dietary adaptations of ground squirrels by analysing 13 permafrost-preserved coprolite samples recovered from burrows (also referred to as middens) in central Yukon, Canada spanning marine isotope stages (MIS) 1–4 and one Middle Pleistocene-aged sample (MIS 16/18) associated with the Gold Run tephra (~700 ka) (Fig. 1). We combine wet-lab extraction optimizations, multi-pipeline metagenomic classifications and validations, multi-species genome assembly and phylogenetic analyses, and a comprehensive array of authentication controls to evaluate (i) how reliably coprolites preserve multi-taxon aeDNA through time and (ii) what these archives reveal about Quaternary ecosystem composition and population history in eastern Beringia. We find that coprolite metagenomics reconstruct diverse snapshots of palaeoecosystems and evolutionary history, highlighting the exceptional biomolecular archives of aeDNA that can be found in the unsuspecting coprolites of ground squirrels.

Results and discussion

Summary of results

Initial morphological classifications, based on descriptions from Cocker et al[41]. placed the Pleistocene coprolites as originating from Arctic ground squirrels (*U. parryii*, $n = 13$)—the extant *Urocitellus* species of Arctic/Subarctic Canada. As a control we processed an additional Holocene-aged coprolite of an unknown, but clearly non-squirrel source. The first batch of extracts on these samples were significantly challenged by inhibitor co-elution with high inhibition indexes (Fig. S1)[36]. This hindered enzymatic functioning while attempting to library adapt the aeDNA for sequencing, prompting a phase of iterative extraction optimization experiments testing the use of ultrasonication for improving inhibitor removal. These experiments are detailed in the supplementary materials (Supplementary Notes 2 and 4, Figs. S3-S17). Inhibitor co-elution could be reduced to the degree that library and PCR enzymes regained functionality simply by reducing raw sample input to 0.1 g when paired with the cold spin and high-volume binding buffer procedure[36]. Libraries were processed through previously published shotgun sequencing and targeted enrichment workflows [36,37,55–62].

We identified diverse and abundant metagenomics signals from the scat producers, their diets, and the local ecosystems of expected Pleistocene/Holocene flora and fauna. This data allowed for whole ecosystem taxonomic cataloguing, genome reassembly, multi-species phylogenetics, taxonomic reassignment of the scat producers, and we found evidence for two genetically undescribed populations of ground squirrels. Our negative controls contain none of the DNA signals we observe in any of the ancient samples, and our aeDNA libraries have deamination rates and fragment length distributions indicative of these molecules being damaged and thus ancient and originating from the coprolites themselves. To avoid over-interpreting weak taxonomic signals of local ecologies, we also integrated our conservative map-filtered *BLASTn/MEGAN* classifications with *Holi/metaDMG* damage evidence, the negative controls, macrofossil corroborations, and ecological plausibility into an ensemble support framework. We find that eukaryotic aDNA preservation from these coprolites exceeds that of other locally preserved sample types (including bone), highlighting coprolites as exceptional biomolecular archives of Quaternary ecosystems.

https://doi.org/10.1038/s41467-026-72977-6

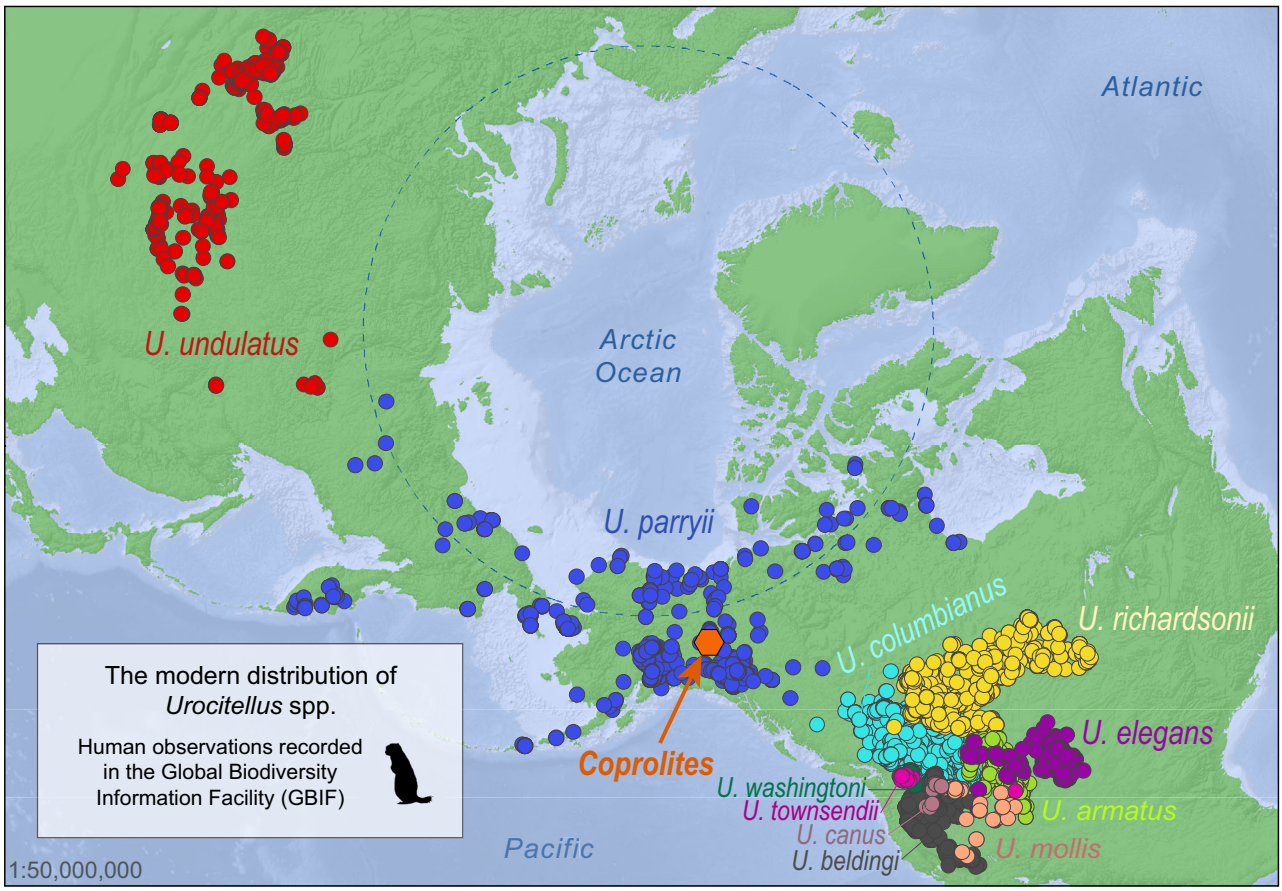


Fig. 2 | Extant distribution of Urocitellus spp. See Table S5 for DOI search data. Source data are provided as a Source Data file.

Geological setting and chronology

The central Yukon Klondike region of eastern Beringia (Yukon and Alaska) remained unglaciated throughout the Pleistocene, with intact ground ice preserving ecological records since at least the Middle Pleistocene[63]. The oldest sample in our set (SC-14) was recovered from the Gold Run site, with the associated tephra at that site dating to 688 ± 44 kya[64,65]. This age falls within either the glacial transition between MIS 16c and 17a (Fig. 1), or could also potentially originate from MIS 18a, because the ecosystem preceding the Gold Run tephra appear consistent with glacial conditions (ruling out MIS 17, as this was an interglacial period). Plant and insect macrofossils from just above and below the Gold Run tephra identify a dry, cooler-than today shrub tundra environment with dwarf birch (*Betula nana-glandulosa* type), willow (*Salix* sp.), and sedges (*Carex* spp.), along with grasses (Poaceae), cinquefoil (*Potentilla* sp.), mosses, and infrequent spruce (*Picea* sp.)[19]. This is further supported by remains of the rove beetle (*Holoboreaphilus nordenskioldi*), and the weevil *Lepidophorus thulius* (synonym *Vitavitus thulius*) that are restricted to dry tundra/steppe environments. Faunal remains temporally associated with the Gold Run tephra in central Yukon include pika (*Ochotona* sp.), chipmunks (*Neotamias* sp.), shrews (*Sorex* sp.), lemmings (*Lemmus* sp.), voles (Arvicolinae), ground squirrels (*Urocitellus* spp.), weasels (*Mustela* sp.), wolves (*Canis*), mammoths (Elephantidae), muskox (Caprinae), and horses (*Equus*) [19,66].

The other ground squirrel coprolites variously date between MIS 1 and 4, which includes the Glacial-to-Interglacial transition (13–10 kya) and the Last Glacial Maximum (LGM, ca. 26.5–19 kya), which are both subsumed within the Last Glacial Period (ca. 115–11.7 kya)[67] (Fig. 1). Ecosystems in central Yukon during the Last Glacial Period were primarily a graminoid (grass) and forb (non-woody flowering herb) dominated dry steppe-tundra with willow (*Salix* spp.) and birch (*Betula* spp.) shrub patches and a diverse abundance of megafaunal animals including: woolly mammoths (*Mammuthus primigenius*), saiga antelopes (*Saiga tatarica*), American lion (*Panthera atrox*), American cheetah (*Miracinonyx trumani*), caballine and stilt-legged horses (*Equus caballus/ferus/lambei* and *Haringtonhippus francisci*), steppe bison (*Bison priscus*), helmeted muskox (*Bootherium bombifrons*), short-faced bears (*Arctodus simus*), and sabre-toothed cats (*Homotherium serum*) among many others[68]. This ecosystem, which has no modern analogue, is widely referred to as the 'mammoth-steppe'[69]. Regional variants of the mammoth-steppe covered much of the northern hemisphere during the Last Glacial Period, stretching from the Iberian Peninsula across Eurasia and into North America. The overall catalogue of flora and fauna in eastern Beringia from the LGM to the Pleistocene-Holocene transition has been relatively well characterized from many independent efforts using a diverse range of palaeoecological proxies [58,59,70–72].

A total of seven radiocarbon dates were obtained for samples expected to fall within the age-limits of radiocarbon dating (Tables S3-S4). Two samples produced non-finite ages (SC-6 and SC-8) and sample SC-12 is the only Holocene-aged material presented with a calibrated age of ~680 cal yr BP. The remaining radiocarbon dated samples (SC-3, SC-10, SC-11, and SC-13) range between ~17,200 and 36,000 cal yr BP.

The remaining six samples were found in stratigraphic contexts associated with known tephra deposits from the Klondike. Sample SC-14 was collected below the Gold Run tephra which has a ^{206}U–^{238}Pb Bayesian model age of 688 ± 44 kya[65]. Samples SC-5, SC-15, SC-16, and SC-17 are associated, or were collected directly within the Dawson tephra, which has an age of 25,300 ^{14}C years BP (29,055–29,470 calibrated years BP)[73–76]. The remaining samples (SC-6, SC-8, and SC-9) are associated with the Sheep Creek Klondike tephra that has been aged to ~80 kya using single-grain optically stimulated luminescence dating [63,77].

Macrofossils

A total of 45 plant and animal macrofossil types were sorted from 11 samples and comprised of 39 unique types, 17 unique plant families, and two animal taxa. All samples contained indeterminate vegetative material, such as fragmentary stems, which is excluded from types here. Of the 45 types, there were 25 (21 unique) plant and animal taxa identified to genus or species level (55% of total types), 11 (eight unique) taxa identified to family or higher taxonomic level (24% of total types), and nine indeterminate (20% of total types). Only nine types were found to occur in multiple samples. Therefore, the samples were represented mainly by unique assemblages of macrofossils, with each sample represented by three to nine types (Figs. S18–26, see Supplementary Note 3 for further details and Supplementary Data 1).

Metagenomics

in silico simulations. To validate the performance of our primary metagenomic classifications (map-filtering to a regional organelle database, *BLASTn* aligning those mapped reads against a 2024 local copy of NCBI-NT, and taxonomically binning with strict *MEGAN7* parameters)[78–81] we performed in silico testing using simulated aDNA data[82] of species detected in our classifications to assess rates of false positives/negatives, and true positives/negatives. Our simulation data supports the empirical taxonomic identifications of our coprolites, confirming that the *MEGAN* classifications have the lowest rates of false positive classifications among tested classifiers with simulated data (Fig. S27–31, see Supplementary Note 5 for further details).

Metagenomics: coprolite aeDNA

A diverse range of taxa were identified metagenomically from megafauna (e.g. woolly mammoths, steppe bison, and horses), to cervids, canids, felids, rodents, birds, insects, ~200 plant clades, fungi, and bacteria/archaea. The diversity found within each kingdom is detailed individually here.

Animals. The most prominent aeDNA eukaryote signal in the enriched Pleistocene coprolites is the tribe Marmotini (ground squirrel) followed by the genus *Urocitellus* (Fig. 3). Within *Urocitellus*, there are hits to two primary species, *U. richardsonii* (Richardson's ground squirrel) and *U. parryii* (Arctic ground squirrel). Hits were also observed to multiple other ground squirrel clades including *Callospermophilus lateralis* (golden-mantled ground squirrel), *Cynomys* (prairie dogs), *Ictidomys tridecemlineatus* (thirteen-lined ground squirrel), Marmota (alpine marmot), *Spermophilus* (susliks), and *Tamias* (chipmunks). Within Arvicolinae, signals of *Dicrostonyx groenlandicus* (northern collared lemming), *Lemmus trimucronatus* (Canadian lemming), and *Microtus miurus* (singing vole) are also present.

A diverse range of large animal aeDNA signals were identified from all 13 of the putative ground squirrel Pleistocene coprolites, which included: *Mammuthus primigenius* (woolly mammoth), *Ochotona* (pika), *Bison priscus* (steppe bison), *Ovis* (likely *Ovis dalli*, thinhorn sheep), *Rangifer tarandus* (caribou), *Canis* (wolf/coyote), Mustelidae (perhaps American marten or ferret), *Puma* (cougar or *Miracinonyx*), Microchiroptera (microbats), *Equus* (cabelline horse), *Lagopus* (likely willow ptarmigan), and Corvoidea (superfamily of songbirds). The Holocene-aged coprolite has abundant hits to *Lepus americanus* (snowshoe hare), with smaller signals from Mustelidae and *Lagopus*, and contains none of the ground squirrel or Pleistocene faunal signals. *Homo* is observed in multiple samples, including the extraction negative controls and Holocene positive control, as well as the 700,000-year-old coprolite (SC-14). We can therefore classify this human signal as stemming from modern contamination.

The Holocene-aged coprolite (SC-12) was confidently identified as originating from *L. americanus* with over 100,000 unique mitochondrial DNA sequences classified. To non-specialists, identifying a defacator from faeces alone can be difficult (especially with ancient faeces), but here we highlight the utility of aDNA for taxonomic coprolite identification from an ambigious source. See Supplementary Note 7 for a phylogenetic assessment of this *L. americanus* coprolite (Fig. S36).

Plants. Over 200 plant LCA clades were identified in the ground squirrel and snowshoe hare coprolites (Fig. 4). The Pleistocene-aged ground squirrel coprolites were dominated by expected mammoth-steppe flora, namely graminoids (grasses) and forbs (non-woody flowering herbs). The most abundant plant signal to a genus rank LCA node was *Plantago* (*P. canescens* [Arctic plantain]). Other abundant taxa include Asteraceae (*Artemisia* [sagebrush] and *Achillea* [yarrow]), Boraginaceae (*Lappula*, likely *L. occidentalis* [western stickseed] based on distribution), Caryophyllaceae (*Silene* [catchfly] likely *Silene* cf. *involucrate* subsp. *tenella*), Polygonoideae (*Bitum*), Hedysareae (*Hedysarum alpinum* [alpine sweetvetch]), Salicaceae (*Salix* [willow]), Brassicaceae (Alysseae, *Draba* [whitlow-grass], *Braya*), Poaceae (*Poa*, *Bromus* [grasses]), *Carex* (sedges), Papaveroideae, and Ranunculoideae (*Anemone* and *Ranunculus* [buttercups]). By contrast, the Holocene-aged snowshoe hare coprolite had predominantly woody plant species such as *Picea* sp. (spruce), *Betula* sp. (birch), *Alnus* sp. (alder), *Rhododendron* sp. (heath), with a much more significant proportion of *Salix*. None of these plant taxa appear in any of the negative controls.

Off-target organisms. A total of 404 bacteria and archaea genera were identified in the MegaBlast data (Fig. 5a, b). The coprolites have distinct signals from the negative controls in all cluster analyses, with *Pseudomonas*, *Acidovorax*, and *Staphylococcus* being the most prominent microbes identified that form the microbial background signal of our wet-lab reagents. The Holocene-aged snowshoe hare coprolite has strong signals from *Clostridium* compared to the ground squirrel coprolites, whereas the squirrel coprolites have abundant signals from taxa such as: *Pedobacter*, *Clavibacter*, *Flavobacterium*, *Arthrobacter*, *Nitrosospira*, *Mycobacterium*, *Listeria*, *Bacillus*, *Sporosarcina*, *Carnobacterium*, *Polaromonas*, *Comamonas*, *Streptococcus*, *Nocardioides*, *Psychrobacillus*, *Caulobacteraceae*, *Mycolicibacterium*, *Cryobacterium*, and *Streptomyces*. When compared with microbial aeDNA in sediment samples from Murchie et al[36,56–58]. that are spatial-temporary related, it becomes clear that the microbes in the coprolites are distinct from those found in the surrounding sediments (Fig. 5d, e, S32). Sample SC-15 was a mixture of coprolite and sediment (no intact pellets), compared with the other SC samples that had intact and discrete faecal pellets. This is clear in the microbial data where it sits in an intermediary position between the ground squirrel coprolites and Pleistocene-aged sediment (Fig. 5d, S32). Likewise, SC-12 (the Holocene snowshoe hare coprolite) is microbially distinct, as are the negative controls. This data suggests that despite the presumed 'porous' nature of coprolites, these entities do an extremely good job of retaining molecules endogenous to the material, while minimizing leaching from the external environment, suggesting little movement of DNA from above or below the deposits.

Several arthropod clades were identified, including those associated with spiders, ants, moths, beetles, and grasshoppers (albeit all at high taxonomic ranks as these taxa were not specifically targeted with the capture enrichment [see Table S6 for bait panel overview] and a curated insect genetic reference database needs to be built). In addition, the parasitic nematode genus *Parastrongyloides* sp. was identified. Of the 13 samples presented, four were associated with cache-bearing middens that were analysed for both plant and invertebrate remains. Morphological identifications of the subfossil invertebrate remains show that several taxa were present in both the aeDNA data and in their associated middens. Of particular interest are the remains of grasshoppers in sample SC-2. Until recently, grasshoppers had never been recovered from Quaternary deposits in Beringia, but a recent study from Cocker et al[83]. recovered three short-horned grasshopper legs (Acrididae: cf. *Pseudochorthippus curtipennis*) from seed caches

https://doi.org/10.1038/s41467-026-72977-6

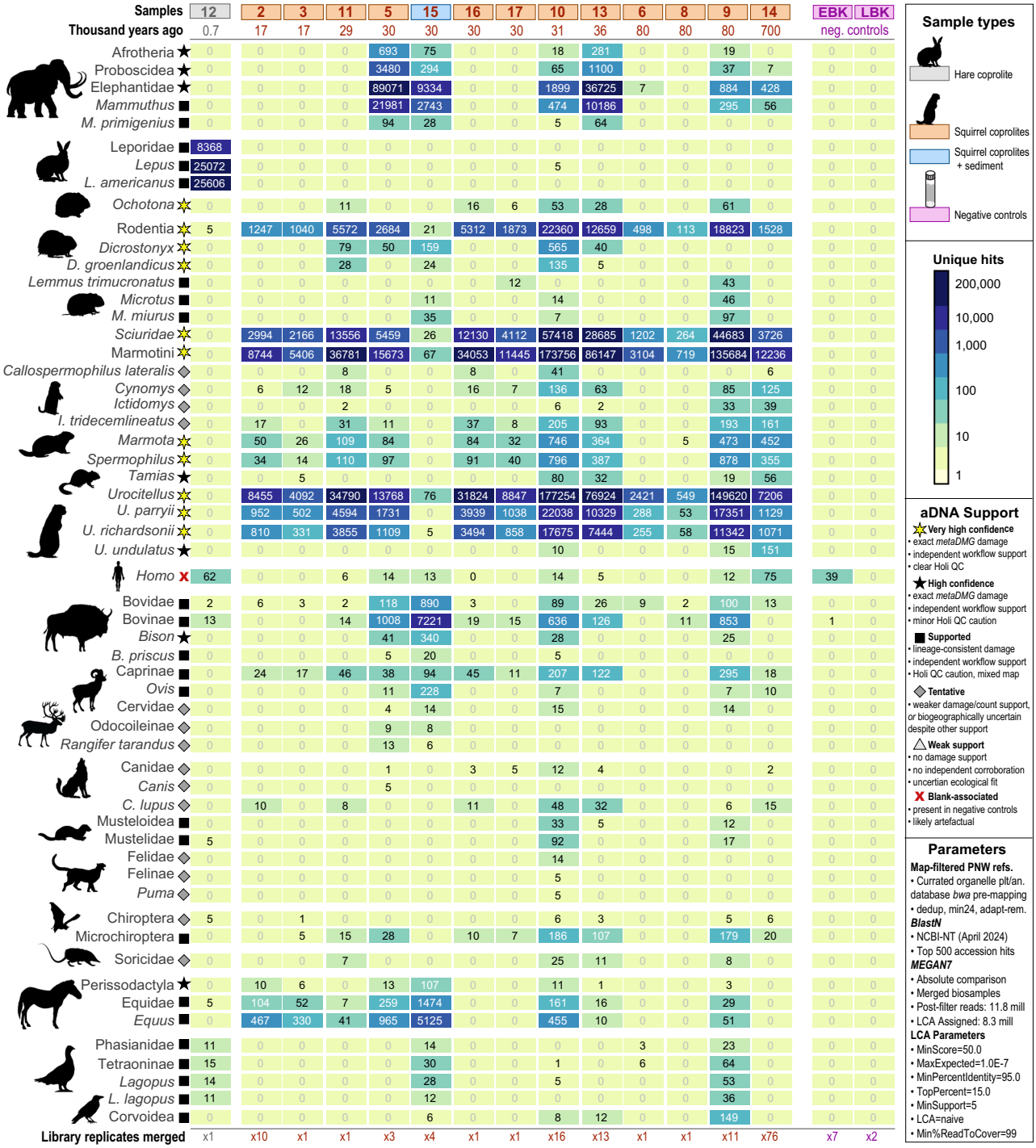


Fig. 3 | Metagenomic classifications of animal aeDNA. Symbol shapes indicate relative aDNA support based on Holi/metaDMG damage evidence, independent taxonomic classification workflow support, ecological/biogeographic plausibility, and negative-control screening. Taxa classified as tentative, weakly supported, or blank-associated are retained for completeness and further testing, not ecological inference. Source data are provided as a Source Data file.

associated with samples SC-2 and SC-5. SC-2 has a strong aeDNA signal of grasshopper (Fig. 5c). Beetles are the most numerous macrofossils recovered from the midden samples and are commonly represented by grounzd beetles (Carabidae: e.g., *Bembidion* spp. and *Harpalus vittatus* cf. *alaskensis*) weevils (Curculionidae: e.g., *Connatichela artemisiae* and *Lepidophorus lineaticollis*), and dung beetles (Scarabaeidae: *Aphodius* sp.), amongst lesser represented families like the carrion beetles (Silphidae: *Thanatophilus* cf. *coloradensis*). Although ants and spiders were not present as macrofossils in these middens, both have been recovered from other studies in the region with only a single spider (Thomisidae: *Xysticus* sp.) recovered from Pleistocene middens to date[83]. See Supplementary Note 6 for a description of fungal taxa.

Genome assembly and phylogenetics

We were able to reassemble 18 mitochondrial genomes from these coprolites, including *Urocitellus* ($n = 12$), *Lepus* ($n = 1$), *Bison* ($n = 2$), and *Equus* ($n = 3$) (Table S7). Our assemblies of six additional *Mammuthus* mitogenomes and their phylogenetic analysis are detailed in a separate forthcoming report.

https://doi.org/10.1038/s41467-026-72977-6

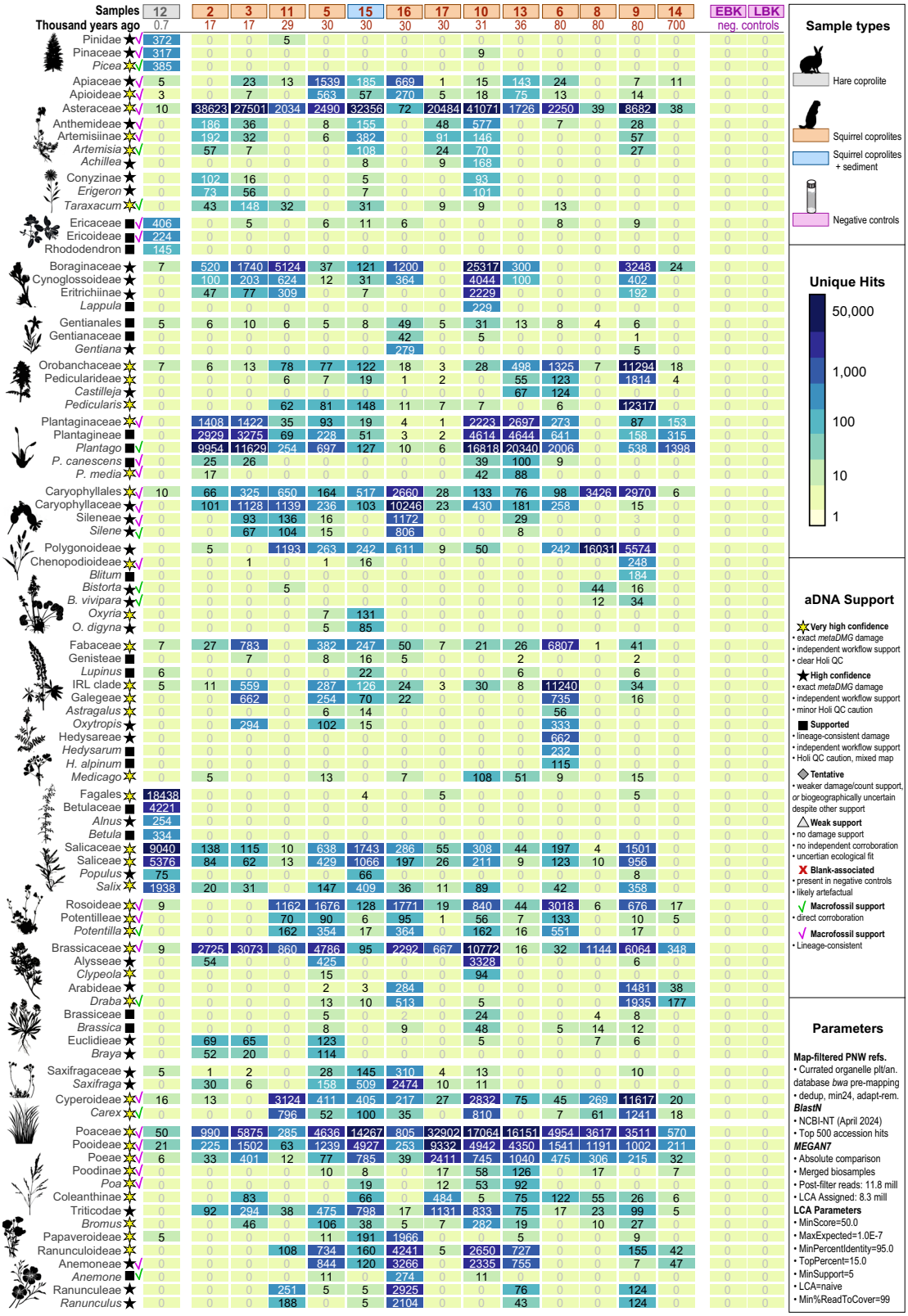


Fig. 4 | Metagenomic classifications of plant aeDNA. Symbol shapes indicate relative aDNA support based on Holi/metaDMG damage evidence, independent taxonomic classification workflow support, ecological/biogeographic plausibility, macrofossil corroboration, and negative-control screening. Taxa classified as tentative, weakly supported, or blank-associated are retained for completeness and further testing, not ecological inference. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-026-72977-6

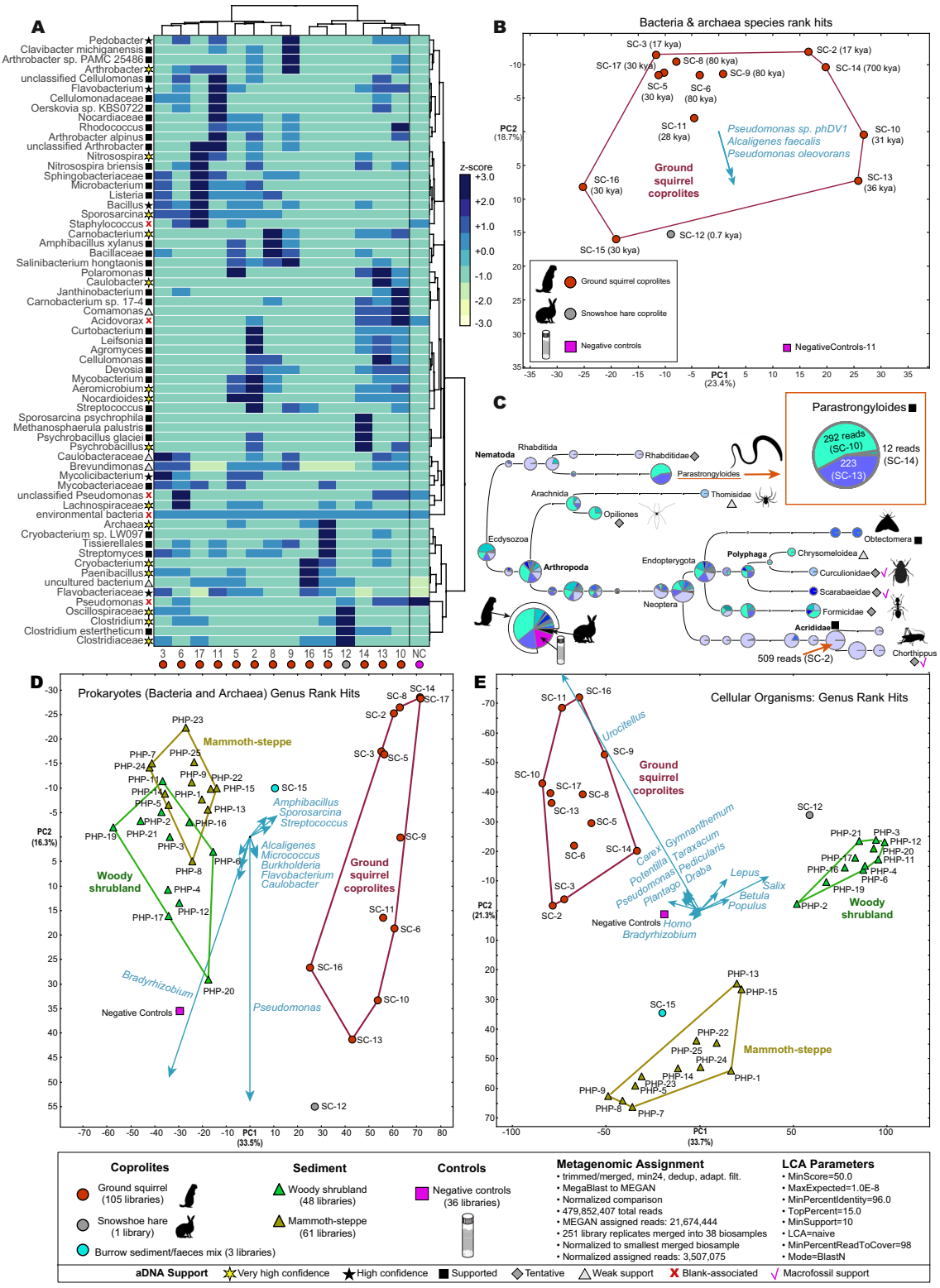


Fig. 5 | Metagenomic classifications of off-target microbes and invertebrates. **A** Hierarchically clustered heat-map of top taxonomic hits within bacteria and archaea for coprolites. **B** Principal Coordinate Analysis (PCoA) of prokaryote assignments with a χ^2 ecological index, normalized to the smallest library. **C** Nematode and arthropod assignments. **D** PCoA (normalized, χ^2) of prokaryote genera including local permafrost sediment libraries[36,56–59,90] merged by core ID; **E** PCoA (normalized, χ^2) of all taxonomically binned cellular organisms with permafrost sediment libraries. Symbol shapes indicate relative aDNA support based on Holi/metaDMG damage evidence, ecological/biogeographic plausibility, macrofossil corroboration, and negative-control screening. Taxa classified as tentative, weakly supported, or blank-associated are retained for completeness and further testing, not ecological inference. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-026-72977-6

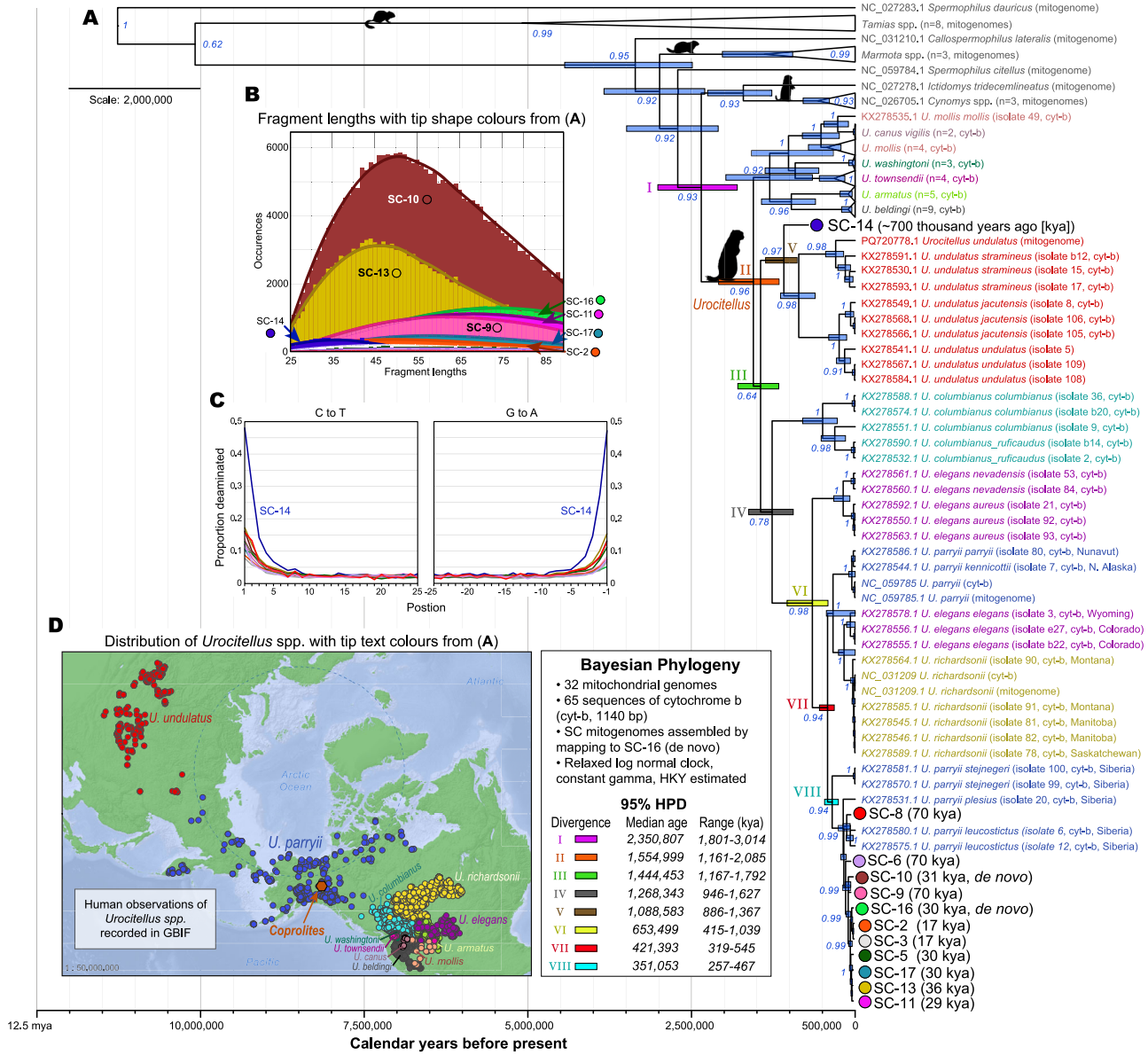


Fig. 6 | Evolutionary history of Holarctic ground squirrels. A Bayesian maximum clade credibility tree of mitochondrial genomes and cytochrome-b sequences from all publicly available Marmotini taxa. Divergence ages and scale are estimated as median calendar/calibrated years before present using a relaxed log normal clock rate, constant tree prior, and 100 million iterations. Divergence ages are shown as at nodes with full range of HPDs blue bars with a 95% highest posterior distribution (HPD). Node labels indicate posterior support. Ancient samples noted with black text and coloured circles that correspond with the colours in panels (**B** and **C**). **B** Fragment length distribution of assembled mapped reads by midden showing expectedly short fragment lengths. **C** mapDamage deamination plot of terminal base modifications with expected smile pattern. **D** Distribution of Urocitellus spp. recorded in GBIF represented as coloured dots that correspond with the tip colours of species in panel (**A**). Source data are provided as a Source Data file.

Urocitellus. These coprolites had exceptional DNA preservation, allowing for mitogenomic assembly from all ground squirrel coprolites (Fig. 6a). The only sample where a squirrel mitogenome could not be reassembled was midden SC-15, a sample that was predominantly sediment, no longer discrete faecal pellets. SC-16 and SC-10 had sufficient endogenous content to allow for de novo assembly. The remaining libraries were reference mapped to the SC-16 de novo genome and placed phylogenetically (Fig. 6), in addition to reference mapping to *U. parryii* and *U. richardsonii* to monitor for reference mapping bias. All three references produced the same phylogenetic placement, indicating that reference bias had not impacted their position in the trees (Fig. 6, S34-35). Unexpectedly, the coprolite assembled mitogenomes do not cluster with those from extant ground squirrel sequences in the Yukon today (*U. parryii*) (Figs. S34-35). When adding data from the cytochrome-b (cyt-b) gene to our Bayesian phylogeny, where many more species sequences are publicly available, our data splits *U. parryii* into 2 primary clades, supported with a 0.94 posterior (Bayesian tree, Fig. 6a) and 100% bootstrap support (maximum likelihood tree, 1000 bootstrap, Figs. S34–35). One clade is composed of the *U. parryii* NCBI reference sequence along with the Arctic subspecies *U. parryii parryii* and *U. p. kennicottii*, and is a sister clade with *U. elegans* and *U. richardsonii*, which is consistent with McLean et al[84]. However, we identify a second clade of Alaskan and Siberian subspecies (*U. p. stejnegeri/plesius/leucostictus*) that also contain our Yukon coprolite assemblies. The estimated divergence between these two clades occurred ~421 kya (319–545 kya 95% highest posterior density [HPD]), indicating that the current taxonomy of *U. parryii* is not monophyletic and that our coprolites and several current

U. parryii subspecies are likely a distinct population (or species) from the RefSeq *U. parryii* individual. Our 700 kya sample (SC-14) sits nearly basal to Arctic/Subarctic *Urocitellus* with both mitogenome and cyt-b phylogenies (Fig. 6, S34–S35), posterior 0.97, bootstrap 100% and is a sister clade to *U. undulatus*—a species currently extant only in Eurasia (China, Kazakhstan, Mongolia, and Siberia). This group diverged ~1.4 mya (posterior 0.65, bootstrap 100%) with the SC-14 clade having diverged from *U. undulatus* ~1.1 mya (posterior 0.97, bootstrap 100%, 886–1367 kya HPD), approximately 300 ky before the lifetime of squirrels from the SC-14 midden, suggesting that this sample is also likely a previously undescribed population (or species) of *Urocitellus*.

Equus. Three *Equus* mitogenomes were assembled from the ground squirrel coprolites dating to 17.2 and 30 kya (Fig. 7). The squirrel coprolite middens (SC-2, SC-5) appear to have better preserved short equid fragments than the predominantly sediment midden sample (SC-15). They are phylogenetically placed within the North American clade of *Equus caballus* alongside other Yukon horses (referred to in some cases as *Equus lambei*) and the sedaDNA *Equus* mitogenomes from Murchie et al[57]. The 700 kya horse from Thistle Creek[85] sits basal to this branch, but within the same clade as other Yukon horses. Only five Yukon horse mitogenomes assembled from bone are publicly available. With the coprolite aeDNA mitogenomes we present here, and the sediment aeDNA from Murchie et al[57]. environmental samples have now more than doubled the mitogenomic references for Yukon horse. These assemblies all have expected terminal deamination and fragment length distributions expected from ancient remains.

Bison. Two steppe bison mitochondrial genomes could be reassembled from midden samples dating to 30 kya and fall within the expected North American clade of *Bison priscus* mitogenomes reassembled from bones and sediment (Fig. 7).

Thermal age

The rates of deamination and depurination are significantly correlated with age and GISP2 isotopes[86] as a palaeo-temperature proxy (see Supplementary Note 8 for further details).

Authentication

There are several components of our research design that support the authenticity of our aeDNA data. The first of these are our 13 sequenced negative controls processed alongside each batch of samples that contain none of the metagenomic signals observed in the coprolite samples. Additional negative controls were used during qPCR assays that were not sequenced as they resulted in no detectible signals. In total, this includes 8 subsampling/extraction controls (with a microcentrifuge tube left open during subsampling and processed alongside the sample batch thereafter) and 5 library negative controls across all batches. These controls were consistently negative on qPCR assays throughout aDNA processing, and despite pooling the entirety of the negative control indexed libraries (which were also run through additional indexing cycles in attempt to increase sequence signal in the negatives), these controls only received a small fraction of the sequencing reads compared to actual samples (Fig. S1, Table S1). The only detectable eukaryote signal in the negative controls is a small portion of human DNA (Fig. 3), which suggests that the sporadic human signal in the other metagenomic samples is most likely background contamination and not ancient human DNA. The lack of any other DNA hits in the 13 negative controls indicates that the signals we see in the samples originate from those coprolites themselves and are not the result of modern or cross contamination.

In addition, the inclusion of a positive control in the form of a single Holocene-aged snowshoe hare coprolite processed alongside the ground squirrel coprolites helps confirm that the ecological signals we observe are consistent with temporally expected flora and fauna. The Holocene coprolite contains Holocene typical woody plant species and a strong signal of snowshoe hare, whereas the Pleistocene coprolites contain flora and fauna diagnostic of a mammoth-steppe ecosystem and are lacking in several Holocene specific taxa (including *Lepus americanus*, *Picea* sp., *Alnus* sp., *Betula* sp., and *Rhododendron* sp.). This sample essentially functions here as a control for false positive signals stemming from the movement of molecules up and down the permafrost margins during their depositional history (in addition to the potential for wet-lab cross contamination) and again indicates that the metagenomic spectra observed in the coprolites originated from the samples themselves. This sample indicates that there has been no temporal mixing of signatures between Pleistocene- and Holocene-aged deposits through any of our processing, and suggests that bioinformatic false-positives are not a substantive problem with this dataset.

All taxa with sufficiently high mapping coverage for mitogenome assembly show clear patterns of deamination (C to T and G to A terminal miscoding lesions)[87] and depurination (aDNA expected short fragment lengths with mean distributions <50 bp) (Figs. 7,8). These patterns have a rough temporal signal where younger middens have less deamination and depurination than older middens. The Holocene-aged hare faecal sample has low rates of deamination (~5%, Fig. S36) compared to Pleistocene-aged samples that range from ~10 to 50%, again demonstrating that the coprolite aeDNA signals are ancient. These *mapDamage* results are also supported by *metaDMG* plots of the total metagenomic classification by midden (Figs. S37-S97, see Fig. S102 for low count taxa *mapDamage* plots), and the ensemble support framework from our two classification workflows, *Holi*/*metaDMG* damage signals, macrofossil corroborations, and ecological plausibility add variable levels of support confidence to each metagenomic classification (Figs. 3–5).

Simulations for assessing true/false positive rates among sample communities with varied taxonomic representations, damage profiles, and taxonomic assignment software add further support for our bioinformatic pipeline approach (see Supplementary Note 5). Furthermore, while it is difficult to ground truth any aeDNA sample, abundant macrofossils from these middens help support the metagenomic classifications by finding many of the same species and overall ecological compositions (Figs. S18–26, see Supplementary Data 1). These complementary identifications help support the authenticity of our diverse aeDNA classifications.

Finally, the authenticity of our results is strongly supported by the mitogenome assemblies. Sufficient coverage was available for the assembly of >18 mitochondrial genomes—including two de novo assemblies—that make phylogeographic sense. These results are consistent with other sedimentary and bone aDNA genomic data, lending support again to the claim that these signals originate from the coprolites themselves.

aDNA from bones, sediments, and coprolites

We compared the on-target (organelle) proportion of aeDNA from these coprolites with bone[88,89] and sedimentary aDNA[36,56–58,90] from permafrost sites in Yukon and Alaska that had been enriched using the same bait-set (Fig. 8a, b). The sediment and coprolite libraries were mapped against the Pacific Northwest Quaternary database (Table S6), and the Alaskan/Yukon bone libraries were mapped against the RefSeq woolly mammoth mitogenome (also included in the PNW database). When simply calculating based on on-target read proportion, unsurprisingly, permafrost bone samples have the most endogenous host aDNA, followed by coprolites and sediment extracts at the far end. When normalized for raw input however, coprolite aeDNA of on-target organisms (plants/animals) far exceeds the preservation of both bone and sediment from eastern Beringia. Additionally, when total diversity is considered, bones perform poorly (as expected) generally only containing a single (host) organism, whereas coprolites and sediments

https://doi.org/10.1038/s41467-026-72977-6

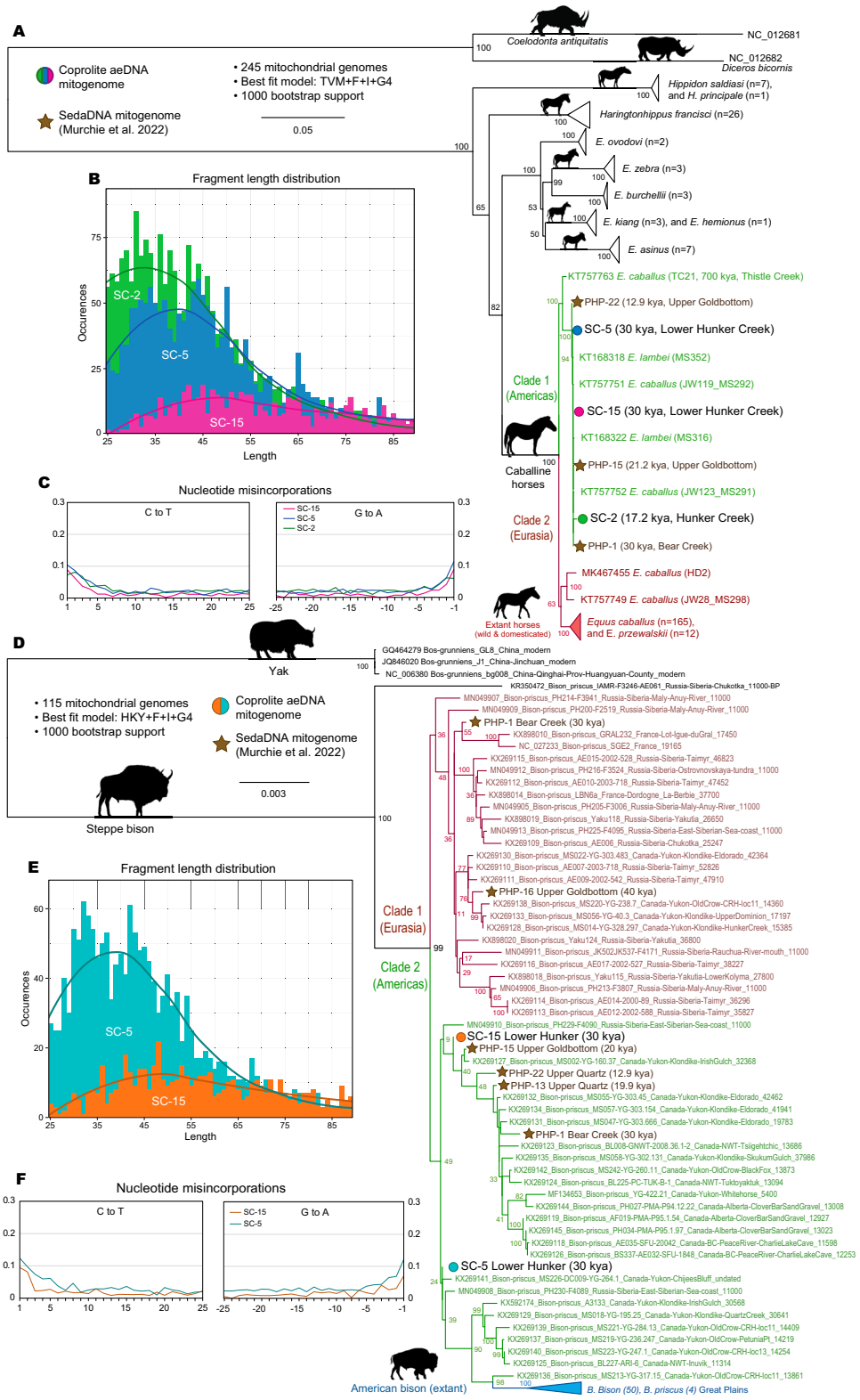


Fig. 7 | Maximum likelihood trees of Equus and Bison evolutionary history. **A** Mitochondrial phylogeny with all available Equidae taxa. **B** Fragment length distributions. **C** mapDamage deamination rates. **D** Mitochondrial phylogeny with all available Bison spp. individuals. **E** Fragment length distributions. **F** mapDamage deamination rates. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-026-72977-6

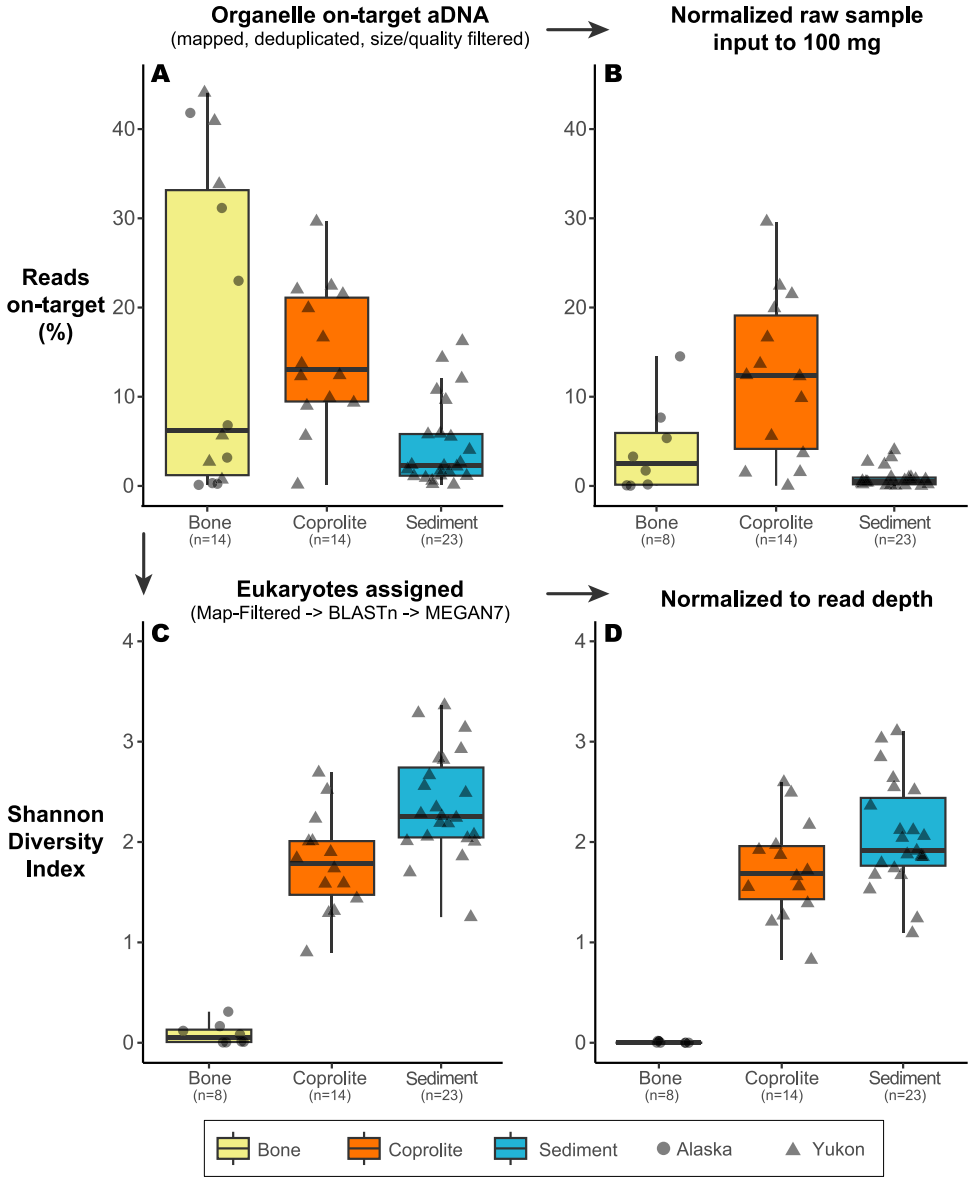


Fig. 8 | On-target aDNA reads (post-enrichment) for permafrost preserved bones, coprolites, and sediments from the Yukon (Klondike) and Alaska (Tanana Valley). Centre lines indicate medians, box bounds indicate the 25th and 75th percentiles, and whiskers extend to the most extreme values within 1.5× the interquartile range. Points represent individual samples, jittered horizontally for visibility. **A** Reads mapped to Pacific Northwest Quaternary Panel references. Bone data from Enk et al[89]. and Rowe et al[88]. sediment data from Murchie et al[36,57,58]. coprolite data from this paper. **B** Proportions scaled to 100 mg of starting input. Only the Alaskan bone on-target proportions used because sample specific raw input weights were not reported in Enk et al[89]. **C** Shannon diversity index of taxonomically assigned reads in MEGAN7. **D** Read depth normalized comparison from (**C**). Source data are provided as a Source Data file.

preserve a wide diversity of aeDNA (Fig. 8c, d). For context, all of these samples (bone and sediment) produced genomic scale data with metagenomic and phylogenetic applications. This underscores the exceptional preservation and information value of coprolites, even when compared with other permafrost preserved materials that themselves may be considered exceptional.

Summary of discussion

Permafrost-preserved ground squirrel coprolites provide a robust, multi-taxon aeDNA archive that captures biomolecular signals of plants, fungi, arthropods, microbes, birds, and mammals across multiple glacial–interglacial intervals, and can preserve genome-scale signals from both the host and co-occurring vertebrates. By combining a conservatively validated metagenomic classification with multiple authentication controls, we show that coprolites can complement (and in some cases, outperform) coeval sediments and skeletal material for reconstructing Beringian palaeoecology and for recovering phylogenetically informative mitochondrial genomes, including a Middle Pleistocene ground squirrel lineage. What follows is a discussion of the evolutionary implications for ground squirrels, a disentangling of dietary DNA vs exogenous inputs, the palaeoecological implications of this dataset, a summary of authenticity checks and caveats, and future directions now possible by leveraging the growing power of aeDNA with frozen archives of ground squirrel coprolites.

The evolutionary history of Beringian ground squirrels

One common assumption is that fossil remains of Pleistocene Arctic ground squirrels recovered from central Yukon likely belong to the same species currently present in northern and southern Yukon, *Urocitellus parryii*[14,18,19,41] (Fig. 1). The metagenomic classifications (Fig. 3) and phylogenetic placement (Fig. 6) of Pleistocene ground squirrel mitogenomes presented here question this species classification.

Subspecies of *U. parryii* become paraphyletic with the addition of our coprolite aeDNA data, and the coprolite mitogenomes do not cluster with mitogenomes of *U. parryii* or *U. richardsonii*. When cyt-b data are added to the mitogenome phylogeny, one *U. parryii* clade distributed in the Arctic is more closely related to *U. elegans* and *U. richardsonii* from the Great Plains and Rocky Mountains. This evolutionary relationship is consistent with Herron et al[38]. and their taxonomic revision of *Spermophilus*. However, when we add the 11 new ancient sequences assembled here to the combined mitogenome and cyt-b data, Bayesian analyses split *U. parryii* into another more distantly related clade that is distributed today primarily in Subarctic Siberia and Alaska (Fig. 6). There are only three publicly available mitochondrial genomes available within *Urocitellus* for comparison, however this same split is clear and strongly supported in the mitogenome phylogeny where all ten MIS 4–1 individuals form a distinct clade with strong support (0.94 posterior, 100% ML bootstrap) from *U. parryii* and *U. richardsonii* (Figs. S34–35). With the cyt-b data, we can observe that *U. parryii* itself is paraphyletic with an estimated divergence date of ~421 kya (319–545 kya HPD) between the two major clades (0.94 posterior, Fig. 6; and 100% bootstrap support). A large portion of reads were classified as originating from *U. richardsonii* over *U. parryii* in the metagenomics (Fig. 3), and this undescribed population was hinted at in previous sedimentary aeDNA work that found Pleistocene ground squirrel DNA of *U. richardsonii*, not *U. parryii*, in central Yukon[58]. This incongruity stems from both an incomplete taxonomy for *Urocitellus*, and incomplete reference data of their past and present genetic diversity. A similar inconsistency exists for *U. elegans* where that species again appears paraphyletic with two distinct populations that may be best classified as separate species (Fig. 6). Previous barcoding work sequencing a 125 bp region of cyt-b from Yukon ground squirrel middens found that their sequences were identical to *U. parryii*[16]. However, this short region lacks sufficient variation to capture the larger evolutionary patterns evident when using full genes and genomic data (Fig. S100).

There are three main potential explanations for the *Urocitellus* paraphyly observed here. (1) There existed a previously undescribed population (perhaps species) of *Urocitellus* in the Yukon during the Pleistocene that our new aeDNA data has uncovered—meaning that their descendants have been misclassified as a subspecies of *U. parryii*. (2) There is insufficient data available for comparison to capture the genetic diversity of *Urocitellus* today and in the past, and perhaps this paraphyletic relationship will resolve with additional sampling and genomic resolution. (3) The previously documented, widespread hybridization behaviour within *Urocitellus* challenges robust taxonomic classifications, especially with ancient individuals restricted to mitochondrial genomic data[40,91–100]. At minimum, these findings call for further nuclear genomic research on extant and extinct *Urocitellus* populations in Beringia to resolve this taxonomic incongruity within *U. parryii*.

Ground squirrel burrows during the Pleistocene tend to concentrate during MIS 4 (~80 kya) and MIS 2 (~30 kya) and are almost entirely absent from MIS 3 (with SC-13 at 36 kya being a relative exception). Cocker et al[41]. suggest that this is likely due to relatively warmer interstadial conditions during MIS 3 that reduced permafrost active layer depth—a feature necessary for robust *Urocitellus* populations, which are currently absent from central Yukon. Cocker et al[41]. identified that the repopulation of *Urocitellus* at 80 kya and 30 kya presents a continuity question as to whether populations persisted in localized refugium after 80 kya and redistributed during colder periods. Alternatively, there may have been a complete extirpation of *Urocitellus* during MIS 3 followed by an establishment from external populations. Given the three unknown factors regarding taxonomy discussed above (a potentially undescribed species, insufficient comparative data, and widespread hybridization), it is difficult to resolve this question entirely. However, our phylogenies suggest continuity between the 80 kya and 30 kya mitogenomic and cyt-b phylogenies (Fig. 6). The 80 kya samples (SC-6/8/9) cluster together, however they remain within the same clade as the MIS 2 middens with relatively little separation. This suggests continuity and the persistence of refugial MIS 3 populations during the Pleistocene. Notably, the same may not be said for Holocene populations where there may have been an extirpation of Late Pleistocene populations.

The very old (700 kya) MIS 16c/18a mitogenome (SC-14) sits basal to the clade of northern *Urocitellus* species and predates radiations from much of the genus (Fig. 6). This specimen is chronologically tied for the third oldest mitogenome assembled to date from discrete materials and the oldest from ancient feaces[85,101,102]. Notably, the top five oldest mitogenomes (from woolly mammoth, horse, and now ground squirrel) have all been recovered from Beringia— highlighting the exceptional preservation of this region for archiving aDNA of the Quaternary. A similar incongruity exists with the 700 kya coprolite mitogenome as those from MIS 4–2 where we again see evidence of a previously undescribed population whose closest extant relatives (*U. undulatus*) exist currently in Siberia, China, Kazakhstan, and Mongolia. This Middle Pleistocene mitogenome helps us calibrate the *Urocitellus* mitochondrial molecular clock to estimate the divergences among *Urocitellus* spp. Using this well calibrated tip date, we estimate that northern and southern species of Urocitellus diverged ~1.5 mya (0.96 posterior, 1,161–2,085 kya HPD), that our 700 kya mitogenome diverged from *U. undulatus* ~1.1 mya (0.97 posterior, 886–1,367 kya HPD), and that our MIS 4–2 coprolite specimens diverged from the *U. parryii*, *U. elegans*, *U. richarsonii* species complex ~421 kya (0.94 posterior, 319–545 kya HPD), during the transition from MIS 12–11. This old mitogenome helps us refine the taxonomy of *Urocitellus* and resolve their deep evolutionary history relative to glacial-interglacial transitions given the robust tephrochronological age of deposits associated with the Gold Run Tephra.

Diet versus exogenous midden leaching

The breadth of taxa identified via aeDNA in these ground squirrel coprolites necessitates questioning to what degree these signals are the result of diet (i.e. are endogenous to the coprolites themselves), or are the result of exogenous aeDNA leaching from the surrounding burrow sediments or midden cache (which have been found to contain macrofossils from a diverse range of plant and animal tissues). Packrat-like behaviour (primarily food/nest caching) is well known within *Urocitellus* middens[14,18,41], and there is evidence of ground squirrels having a broad omnivorous diet[41,45,48,49,52]. Macrofossil remains of *Dicrostonyx groenlandicus*, *Microtus* spp., and *Lemmus* sp., have all been found in Klondike middens, which are species identified in the coprolites processed here (Fig. 3), as have remains of grasshoppers, spiders, ants, and beetles (Fig. 5C). This macrofossil evidence helps support the aeDNA classifications of these taxa that are also in the aeDNA dataset. We also observe microbiome dietary differences between snowshoe hare and ground squirrel coprolites, with *Clostridium* being significantly more abundant in the snowshoe hare coprolite (Fig. 5a, S32). *Clostridium* functions as a means of cellulose breakdown, with Clostridales having been observed in 16S rRNA sequencing to constitute 33.1% of modern snowshoe hare gut microbiomes [103].

The quantity of megafaunal aeDNA (i.e. animals with a mass ≥45 kg; *Mammuthus*, *Bison*, *Equus, Ovis, Rangifer*, and *Canis*) is surprising, but not entirely unexpected. SC-15, the degraded coprolite and sediment mixture, contains the highest proportions of bovid and equid DNA, indicating that the sediments in which the middens are constructed contain large quantities of megafaunal aeDNA. However, this does not rule out these signals from being dietary as the microbial communities (and overall aeDNA signal) observed in the coprolites are consistently distinct from those of contemporary Klondike sediments (Fig. 5D). If there had been significant leaching of sedimentary aeDNA

into the coprolites, we would expect a comparable leaching of microbial DNA that would conflate the coprolite and sedimentary communities. This is not what we observe. Rather, all discrete coprolite pellets cluster together at the exclusion of Pleistocene and Holocene microbial sediment communities from several nearby Klondike sites[56]. SC-15, the coprolite/sediment mix, sits in an intermediary position between the two main clusters as we would expect, suggesting that this sample does contains a more significant signal of sedimentary microbial input than do any of the other coprolite pellets. We cannot entirely rule out midden tissue aeDNA leaching into the coprolites however. Contemporary squirrel metagenomic analyses of faecal pellets would help characterize similarities in microbial communities to help quantify the degree to which these coprolites may contain exogenous sedaDNA. Follow-up work on Pleistocene pellet aeDNA would benefit from doing single pellet extractions if feasible, subsampling only the interior coprolite portion (excluding the exteriors). In addition, sediment samples from outside the middens and hibernacula would allow for direct metagenomic comparison with the pellets to determine the degree to which the diverse aeDNA signals observed here are primarily dietary, exogenous midden aeDNA, or both. In retrospect, we would have done this had we known how well these coprolites would perform. Regardless of the functional origin of aeDNA in the coprolites, the fauna spectrum observed here reconstructs a diverse timeseries of Pleistocene ecosystems that is highly consistent with other palaeontological and palaeogenomic research.

Palaeoecology through the lens of squirrel faeces

The coprolite aeDNA data presented here reconstructs an expected spectrum of mammoth-steppe organisms that corresponds well with analogous sedimentary aeDNA datasets from the region[36,56–58,90] with a distinct transition towards woody Holocene ecosystems first occurring ~13 kya with the Bølling–Allerød Interstadial (Fig. 9). There is an overall continuity between MIS 4 and MIS 2 coprolite sites with relatively minor variation in faunal/floral presence and absence across the Mid-to-Late Pleistocene (relative to the admittedly small sample sizes investigated here). The most notable coprolite difference comes from the Holocene-aged positive control where there is a distinct signal of woody species diagnostic of the boreal forest today and a comparative absence of graminoids, forbs, and megafauna. The most common plant taxa identified with macrofossils by Zazula et al[18]. were *Poa, Carex, Bistorta, Ranunculus, Draba, Erysimum*, and *Potentilla* with additional taxa like *Plantago* and *Phlox* abundant in middens identified in Cocker et al[17]. With the exception of *Phlox*, all of these taxa are identified in the metagenomic aeDNA data here, as well as the barcoding aeDNA data of Langeveld et al[16]. highlighting a continuity in palaeoecological proxies.

Several high rank arthropod clades were identified that were not specifically targeted with our bait-set. These include parasitic helminths, arachnids, beetles, moths, ants, and grasshoppers. *Parastrongyloides* was identified in three samples (SC-10, SC-13, and SC-14), but this taxon has not been previously identified as an endoparasite species infecting ground squirrels. Known endoparasitic fauna associated with ground squirrels include the roundworms *Strongyloides robustus*[104] and *Baylisascaris laevis*[105], the cestodes *Paranoplocephala wigginsi* and *Diandrya* spp.[106,107], and five species of apicomplexans in the genus *Eimeria*[108]. To date, only individuals of the flea, *O. alaskensis*, have been recorded from Pleistocene-aged middens[83]. The genus *Parastrongyloides* are facultative parasites of mammals, capable of consecutive free-living life cycles[109]. *Parastrongyloides winchesi* was described from European moles and shrews[110] and has since been identified as far afield as Japan[111] and in one report, eastern North America[112]. The hits to *Parastrongyloides* here may suggest an undocumented parasitic load within Pleistocene ground squirrels, the presence of free-living *Parastrongyloides* in the local environment, or a misidentification as a result of database incompleteness within helminth palaeo-diversity. Regardless of the taxonomic specificity, this signal indicates that Pleistocene helminths are also amenable to aeDNA study via coprolites and that ground squirrels have had a gastro-intestinal burden of parasitic worms for millennia. This data parallels other work demonstrating the value of aeDNA for studying ancient parasites [113,114].

It is difficult to determine at this time the degree to which the diverse signal of Marmotini taxa are true positives (either through burrow reoccupation, predation, or scavenging) or whether those hits represent some degree of false positive classifications of ancient *Urocitellus* populations due to database incompleteness amongst Quaternary rodent biodiversity and well documented ground squirrel hybridization[40,91–100]. The simulation data (Figs. S17-18) demonstrates that despite aDNA damage patterns, our metagenomic pipeline will accurately identify *Urocitellus* and other ancient faunal/floral sequences if they have database representation. However, such databases do not represent the total biodiversity present in ancient populations, and as such, are susceptible to false positive/negative classifications from related species if the true population has no genetic representation. All of the Marmotini taxa identified here may have reasonably lived in central Yukon during the Late Pleistocene. Although, the breadth of their signal here seems excessive, and it is reasonable to postulate that some portion of their taxonomic binning here is the result of database incompleteness and interspecies hybridization driving hits to closely related taxa. We therefore encourage caution in interpreting the presence of specific ground squirrel taxa beyond *Urocitellus* given these known limitations.

Identified carnivore aeDNA includes mustelids (martens/weasels/ferrets), grey wolf (*Canis lupus*), and cougar (*Puma*). Canid/felid DNA must be interpreted with caution as companion animals are common laboratory contaminants and read counts are low (see Fig. S102 for *mapDamage* plots for low read count taxa). These signals are absent from the negative controls however, and we observe hits specifically to both *Puma concolor* and two hits to *Miracinonyx trumani* (American cheetah) in sample SC-10 dating to 31 kya, that are subsumed within the summed 24 hits to Felidae that could not be called specifically due to the LCA parameters. At this time, there is no publicly available *Miracinonyx* mitogenome on Genbank (only 6 small gene fragments ranging from 143-393 bp), which challenges taxonomic classification as *Miracinonyx* is a sister clade to *Puma*, and is more distantly related to the African cheetah (*Acinonyx jubatus)*[115]. This lack of publicly available genomes for ancient felids also challenges classifications of *Homotherium serum* (saber-toothed cat) with a single 457 bp sequence of cyt-b that severely limits aeDNA detection. Genetic evidence has found that North American *Puma* today are descendants of a small founder population from eastern South America that recolonized North America during early the Holocene following the extinction and extirpation of large Holarctic felids[66,116]. *Puma* are very recent migrants to the Yukon (first confirmed sighting in 2000), which has historically been considered outside of their range. Our *Puma* aeDNA classifications here dating to 31 kya may be explained by several hypotheses. If the felid signals are accurate, their aeDNA may originate from: (1) Pleistocene lineages of *Puma* that were extirpated during the Pleistocene-Holocene extinctions; (2) *M. trumani*—a close relative of *Puma* with insufficient genetic representation to taxonomically identify; or (3) another Pleistocene felid entirely that cannot be correctly identified due to reference database incompleteness. As databases continue to improve with time, aeDNA datasets such as those presented here may become ideal for reanalysis to resolve these ambiguities. All three carnivore signals are most prominent in sample SC-10. This suggests that burrow SC-10 either has authentic aeDNA from scavenging by squirrels, or perhaps more likely, predation by carnivores attempting to enter the hibernaculum.

Top plant family rank hits include: Asteraceae, Boraginaceae, Plantaginaceae, Caryophyllaceae, Polygonaceae, Brassicaceae, Poaceae, Ranunculoideae, and Ranunculacea (Fig. 4), with genus and

https://doi.org/10.1038/s41467-026-72977-6

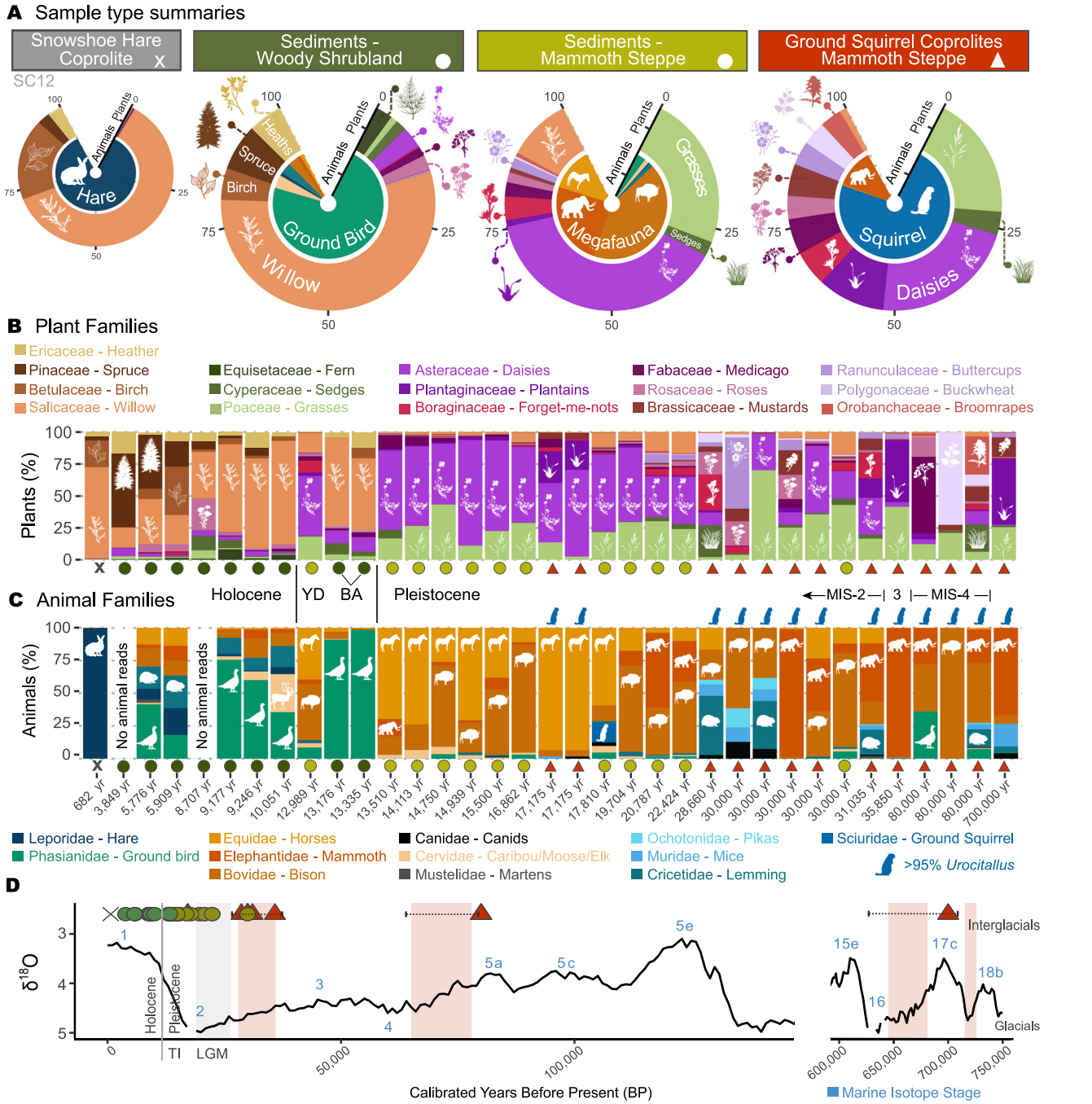


Fig. 9 | Ecological change through the Quaternary archived in sedimentary and coprolite aeDNA. A Summary of four main sample types in circularized stacked bar plots. **B** Major plant and **C** animal families in proportional stacked bar plots. See Fig. S101 for read counts. Ground squirrel signal removed from samples with >95% Urocitellus aeDNA. **D** Benthic $\delta^{18}O$ record[196] with MIS substages[197] rescaled from Fig. 1. BA Bølling–Allerød interstadial, YD Younger Dryas stadial. Source data are provided as a Source Data file.

species-specific hits therein at lower read depths. This distribution of specificity is to be expected as the PalaeoChip Arctic RNA bait enrichments targeted the well-studied chloroplast genes *trnL*, *matK*, and *rbcL*, which contain conserved regions shared across many large clades. In the same manner that hits to Afrotheria, Proboscidea, and Elephantidae represent DNA from variable conserved regions of the *Mammuthus primigenius* mitogenome, hits likewise to Asteraceae likely originate from known high abundance forbs like *Artemisia* that simply have few genetic regions that are genus or species diagnostic.

An increasingly notable result amongst aDNA and aeDNA research on Beringian equids is the phylogenetic placement of caballine horses between Eurasian and North American clades. Historically, palaeontologists have classified the caballine Yukon horse as *Equus lambei* (Yukon wild horse)[117–119], implying a distinction from *E. przewalskii* (Mongolian wild horse) and *E. caballus/ferus* (Eurasian domestic horses)[118–121]. There is a clear distinction between Eurasian and American caballine horses that was thoroughly documented by Vershinina et al[117]. and can be observed here (Fig. 7). However, this same pattern of genetically isolated Eurasian and American populations is also observed among *B. priscus* (Fig. 7) and M. *primigenius*[57]—which, at this time, are not designated as being separate species on either side of the Bering Land Bridge. If species designations among Beringian

megafauna are to be applied consistently, it should be considered whether *E. lambei* is truly deserving of a species designation. If so, Eurasian mammoths and steppe bison may necessitate the same treatment. If not, *E. lambei* may be best collapsed within *E. caballus/ferus*.

Authenticity and caveats

There are several caveats to be mindful of regarding these data, results, and interpretations. First, most samples represent the homogenization of several faecal pellets (~2–8). This was initially intended to facilitate a methodological comparison for improving the cold spin extraction, where we expected mostly *Urocitellus* and microbial DNA. These pellets were not subsampled from their interiors and were not externally UV-C decontaminated. It is therefore difficult to estimate the diet versus midden leaching of aeDNA to explain the taxonomic diversity of our metagenomic classifications. The microbial signal suggests this leaching is minimal, but we lack a direct on-site 1:1 comparison of sediments and coprolites. Furthermore, these sites have been perennially frozen in permafrost conditions for tens of millennia and show no signs of reworking. These factors lend support to the interpretation that aeDNA mobility from vertical leaching and reworking have not meaningfully impacted the results here[55]. Second, the reconstructed mitogenomes likely do not originate from single individuals. Like other aeDNA genomic reconstructions, it is likely that DNA from several individuals is represented in the assemblies, particularly given the first caveat. Third, there are several sporadic hits to nonsensical taxa that cannot be easily removed from the metagenomic classifications by increasing LCA stringencies without also losing a large proportion of the sensical aeDNA taxa. These appear to be due to short fragment sizes (~30 bp), conservation, homoplasy, database bias/errors and incompleteness, deamination causing terminal base errors, algorithmic limitations, and potential modern labware/reagent contamination. Some of these nonsensical taxa include Whippomorpha (suborder of whales/dolphins/porpoises), Suidae (family of pigs/hogs), Colobinae (old world monkeys), and Elephas (extant elephants). Read counts for hits to these false positive taxa are all low ($\lesssim$10). A full list of classified organisms is included in the supplementary materials.

Despite these caveats, there is strong support for the authenticity of our data for all the reasons discussed previously, including: consistently clean negative controls across all batches, a clean positive control producing data with an entirely different ecological period, high read coverage that facilitated mitogenomic assembly and biogeographically reasonable phylogenetic placement, clear deamination and fragmentation characteristics indicative of ancient DNA, microbial community support for a lack of leaching, sensical palaeoenvironmental reconstructions based on previous research, complementary proxy support of taxonomic presence from macro-fossils and palynology, and robust tephrochronological control.

Coprolites as aeDNA archives of Quaternary ecosystems

The diversity and abundance of aeDNA recovered from the permafrost preserved, ground squirrel coprolites presented here underscores the immense value of Arctic rodent middens as repositories of Quaternary ecosystems. These data contain a wealth of ecological and evolutionary information amenable to many deeper investigations outside of the scope of this report. Furthermore, it is reasonable to expect that with further improvements in DNA extraction recovery and isolation, the use of single-stranded library preparation[122–124], the expanded targeting scope of RNA probes for Quaternary organisms, reference database improvements, and improved bioinformatic tools, that the information potential of these middens will continue to expand. Initial extractions yielded limited recoverable DNA, underscoring that iterative wet-lab optimizations can be essential for converting challenging coprolites into informative metagenomic archives. The ecological and evolutionary power of coprolites would appear to exceed that of both bone and sediment. With normalized inputs, coprolites exceed both bones and sediment for their aeDNA preservation, and are equal to sediment for their breadth of aeDNA diversity. Coprolites not only safeguard aDNA from the host and their gut microbiome, but also a wide swath of their ecological contemporaries. Hundreds of such middens have been identified in the Yukon[14,18,41]. An aeDNA analysis of such a collection would robustly resolve Mid-to-Late Pleistocene ecological dynamics, while molecular clocks—calibrated using tephrochronologically constrained middens—could anchor stratigraphic contexts beyond radiocarbon limits for the phylogenomics of numerous Quaternary lineages yielding unprecedented insights into evolution, adaptation, and extinction.

Methods

Our sampling and research comply with all relevant Yukon Government regulations. All samples were collected under licences granted by the Yukon Government Science and Explorers Act (Duane Froese, Britta Jensen, Scott Cocker, and team) with Yukon Government scientists (Grant Zazula, Elizabeth Hall, and Susan Hewitson) and with permission from the Tr'ondëk Hwëchin First Nation whose Traditional Territory includes our study region. Ethical approval was not required for this research as no archaeological remains were processed or analyzed and no human DNA was targeted or analyzed.

Field recovery

Samples were collected during field seasons spanning 2007–2021 in the Klondike goldfields of central Yukon Territory, Canada. Samples associated with Arctic ground squirrels include middens (frozen and thawed) and isolated latrines. All samples, irrelevant of whether they were frozen or thawed, were transported to freezers in Dawson City, then to storage facilities at the Yukon Palaeontology Programme in Whitehorse and subsequently to the Permafrost Archives Science Laboratory at the University of Alberta in Edmonton.

Chronology

A total of 14 coprolite samples were analysed with their chronology established using either radiocarbon dating and/or stratigraphic association with volcanic tephra. Seven samples were processed for radiocarbon dating (Table S3-S4). An eighth sample, SC-2 (DF09-HC-30), was collected in association with SC-3 (DF09-HC-29) and so only the latter was radiocarbon dated. Pre-treatment was completed at the University of Alberta following an acid-base-acid methodology. Samples were then freeze-dried, sealed in sterilised vials, and shipped to either the W. M. Keck Carbon Cycle Accelerator Mass Spectrometer Facility at the University of California Irvine (UCIAMS) or to the André E. Lalonde National Facility in Accelerator Mass Spectrometry at the University of Ottawa (UOC). CO_2 production, graphitisation, and measurement of radiocarbon abundances was completed at both facilities. Radiocarbon dates (^{14}C) were calibrated using OxCal version 4.4[125] and the IntCal20 radiocarbon calibration curve[126]. Calibrated ranges are presented as both median values and their 2σ uncertainty range and are rounded to the nearest five years. The remaining six samples were dated based on their association with tephra. This was particularly important when samples are older than the limit of radiocarbon dating (> 50,000 years BP). Samples were prepared and geochemically analyzed at the University of Alberta on a JEOL 8900 R electron microprobe by wavelength dispersive spectrometry using standard methodologies[127]. We confirmed the correlations to the Dawson, Sheep Creek-Klondike and Gold Run tephras using concurrent analysis of known reference samples housed in the John Westgate Tephra Collection at the University of Alberta.

Ancient DNA: general processing and facilities

Initial aDNA wet-lab work (batches 1–3) was conducted in dedicated aDNA clean rooms at the McMaster Ancient DNA Centre (McMaster

University), which is subdivided into facilities for sample preparation (with physical separation between labs for discrete tissues and environmental samples), mastermix reaction setup, and DNA extraction through library preparation. The post-indexing clean room (used for capture enrichment) is in a physically isolated workspace, and the subsequent high-copy PCR workspace is in an adjacent 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 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 min of UV irradiation at >100 mJ/cm^2.

Additional wet-lab work (batch 4) was conducted both at McMaster and in the newly built Palaeogenomics Facility at the Hakai Institute (Quadra Island, BC). This facility is likewise dedicated to aDNA research and is physically separated from lab spaces with modern DNA and post-PCR amplicons. The lab has a positive-pressure HEPA filtration system where the air is filtered every 60 s, there are five zones of air pressure gradients between internal rooms, multi-point ceiling UV-C irradiation for full room sterilizations, dedicated rooms for scrub and clean-suit donning and decontamination, reagent preparation, subsampling, DNA extraction, and library preparation, and full facility live-track monitoring with Zabbix for CO_2, air pressure gradients, and temperature (by room and for all refrigerated equipment). Equipment, tabletops, and work spaces are cleaned before and after use with either in-house preparations of 1.2% sodium hypochlorite (subsampling and decontamination rooms) or *DNA Away*™ (Thermo Fisher Scientific)[128,129] and *ELIMINase*® (Decon Labs, Inc., reagent and extraction rooms). Bleach alternatives are used in rooms with routine guanidine salt use to minimize the risk of accidental chemical reactions that form toxic gases (i.e. chloramine, chlorine, and hydrogen cyanide). Workstations and rooms are UV irradiated after each session and repeat use plasticware (tube racks) are rinsed with water, submerged in 1.2% sodium hypochlorite for >2 min, rinsed in Barnstead smart2-pure water, and UV irradiated for 30 min. The Hakai PalaeoGenomics facility has a unidirectional workflow progressing from low-copy to high-copy workspaces. The post-indexing and enrichment workstation, along with the modern environmental DNA cleanrooms and sequencing facility are in a physically separated building from the PalaeoGenomics facility.

aDNA extraction and qPCR inhibition

It has been observed in multiple studies that aDNA is heterogeneously distributed and preserved in discrete tissues and environmental deposits[130–138]. For this reason, we decided to homogenize several available faecal pellets recovered within the same midden context to maximize aeDNA recovery in the case of single pellet poor preservation. Pellet subsampling was carried out in the sedimentary DNA lab at the McMaster Ancient DNA Centre. The day prior to subsampling, metal tools were bleached (6% NaClO) thoroughly and cleaned with ultrapure water, then baked at ~200 °C for 6 h overnight and left to cool in the oven. The next day, work surfaces were wiped with bleach followed by ultrapure water, and sample tubes/bags were removed from the freezer to thaw.

For each sample, a sterile tin foil sheet was placed on the table, which was wiped with bleach and rinsed with ultrapure water. Weigh boats and plastic/metal scoopulas were placed in the work area, then a mobile table top UV was placed above the worksurface. The work area and tools were UV-C irradiated for ~20 min, flipping the tools and moving the benchtop UV after ~10 min to maximize coverage. Fresh bleach sterilized nitrile gloves were used for each sample. Faecal pellets were poured into a sterile weigh boat and the contextually associated pellets were homogenized manually by tearing/grinding with the metal tools. Subsamples of 0.2–0.1 g were transferred to PowerBead Pro tubes (Qiagen), and the remaining homogenized pellet material was poured into 50 mL tubes for long term cold storage (−20 °C). The work area was thoroughly cleaned with bleach and rinsed with ultrapure water, and the same setup procedure was repeated for each sample again, including ~20 min of UV-C irradiation between samples to minimize cross-sample airborne contamination. Air blanks (empty PowerBead Pro tubes) were left open during all sediment handling phases. Lids of these negative controls were closed during UV irradiation so that any airborne contamination captured in the blanks would be retained throughout processing. These air blanks ($n = 6$) became extraction negative controls to monitor for background contamination with each batch, which were carried through all subsequent processing steps to sequencing.

Initial homogenized extracts (batch 1) (Fig. 1) used 0.2 grams of subsampled coprolite, which was processed using a disintegration phase with an overnight proteinase K digestion, followed by silica spin column purifications with a high-volume guanidinium binding buffer[36,139]. During subsequent tests (batch 2–4), we found that an input of 0.1 grams performed better due to overwhelming inhibitor co-elution at higher inputs. During disintegration/lysis, subsamples were loaded into PowerBead Pro tubes already containing 1 mL of a custom digestion buffer composed of: Tris-Cl 0.02 M, SDS 0.5%, CaCl2 0.01 M, PowerBead Solution 50%, DTT 100 mM, PTB 5 mM, PVP 2.5%[5,36]. Samples were physically disrupted using a TissueLyser I for 10 min set to 20/S. They were then centrifuged at 10,000 rcf for 1 min to pellet the lysate/beads/coprolite and proteinase K (0.25 mg/mL) was individually added to each tube (12.5 µL). Tubes were manually agitated by hand until the pellets were dislodged and both the beads and subsample could flow freely in the tube when inverted. Tubes were arranged in a vertically rotating incubator set to 37 °C and ran overnight for ~20 h. The next day, tubes were centrifuged at 10,000 rcf for 5 min. The supernatant was removed and transferred to fresh 1.5 mL tubes for immediate purification or frozen for later processing (depending on sub-batch size given centrifuge capacity).

For purification, 13 volumes (per volume of lysis supernatant) of high-volume binding buffer was created following Dabney et al[139]. containing: guanidinium HCL 5 g/mol, isopropanol 40%, sodium acetate 0.1 M, Tween-20 0.05%. This binding buffer was transferred to individual 15/50 mL tubes and the $\lessapprox$ 1 mL of lysis supernatant was pipetted into the 13 mL of buffer and mixed by repeated inversion. These were then placed in a refrigerated centrifuge, with all buckets carefully balanced to within ~0.01 grams (by slowly squirting water into the bucket sleeve on a high-precision scale tared to the heaviest bucket), and centrifuged overnight (~18 h) at 1500 rcf, and set to 4 °C[36]. The next day during setup, the centrifuge speed was increased to 3500 rcf for ~1–2 h as the work area was re-sterilized and tubes were labelled. Once removed, the 14 mL of supernatant was carefully poured into high-capacity silica column binding tubes from the Roche High Pure Viral Nucleic Acid Kit, taking care to not disrupt the inhibitor pellet that had formed at the base of the 15/50 mL tubes from the overnight cold spin. Purification proceeded as per Dabney et al[139]. with 2 washes (710 mL of PE Buffer, Qiagen) to remove residual salts, a one minute dry spin (rotated 180° after 30 s to maximize ethanol removal), and eluted in 25 µL of EBT twice for a total extract volume of 50 µL.

Inhibition content was tested using the Inhibition Index qPCR assay[36]. This qPCR used 8 µL of mastermix (10x PCR Buffer II 1X, MgCl2 2.5 mM, dNTP mix 250 µM, BSA 1 mg/mL, forward/reverse primers 0.25 µM each, EvaGreen 0.5X, AmpliTaq Gold 0.5 U/µL) with 1 µL of extract template spiked with an additional 1 µL of synthetic DNA at 1000 copies/µL to test for delays to PCR amplification. The inhibition observed in duplicate qPCRs—through a delay in Cq values and change to the slope of the amplification curve—was quantified using a modified version of the Inhibition Index[36]. This test compared the average duplicate Cq and max RFU of the standard amplifications (1 µL 1000

copies/μL standard, 1 μL water, 8 μL mastermix) with the average duplicate standard+sample Cq and max RFU amplifications (1 μL standard, 1 μL water, 8 μL mastermix). Unlike in Murchie et al[36]. hill-slope was not used in the calculation in order to streamline the analysis, and in our case, we subtracted the values from 1 so that here, a value of 1 indicates an extract PCR reaction is entirely inhibited (no detectable enzymatic function), and a score of 0 indicates that there was no noticeable decline in amplification relative to the standard (no inhibition). Samples with Inhibition Indexes $\lesssim$0.4 tend to successfully undergo library preparation, although ideally the less inhibition the better so long as aDNA loss is minimized (a challenge with DNA purification kits designed for modern DNA [i.e. high quantity/quality]).

The supplementary materials detail our experimental attempts to further optimize inhibition removal while maximizing aDNA retention. Therein, we tested 8 variant extraction methodologies. Several of the non-cold spin variants were processed through to sequencing for batch 4 to maximize the chances of genomic reconstructions from our 700,000-year-old sample. Our attempt to use ultra-sonication paired with the cold spin to improve inhibitor removal was unsuccessful in the final experimental batch (Fig. S16-S17) despite that approach repeatedly being the top performing variant in preceding experiments (Figs. S7-S15). The reason for this performance switch is unclear at this time and will necessitate further testing. However, the cold spin inhibition removal method paired with high-volume binding buffer consistently out-performed alternative approaches in terms of total library adapted DNA[36,139], and so the primary libraries utilized for sequencing here originate from this extraction batch (method c).

Library adaptation

Double-stranded library preparation (dsLP), Batch 2/3/4. Most extracts were library prepared using dsLP[140,141]. A subset from batch 4 (processed at the Hakai Institute) were prepared with the same protocol, but with the addition of Uracil-DNA glycosylase (UDG) treatment to repair uracil miscoding lesions caused by hydrolytic deamination, which are typical of ancient DNA[87]. Library negative controls were introduced for each batch during this stage.

Step 1, blunt end repair: non-UDG mastermix: 1X NE Buffer 2.1, 1 mM DTT, 100 μM dNTP mix, 1 mM ATP, 0.5 U/μL T4 polynucleotide kinase, and 0.1 U/μL T4 DNA polymerase with ultrapure water. This mastermix (30 μL) was distributed into high-profile strip tubes (0.2 mL), into which 10 μL of aDNA extract was loaded. These strip tubes were placed on a thermocycler at 25 °C for 15 min, followed by 12 °C for 15 min.

Blunt end repair (step 1) UDG mastermix: 1X NE Buffer 2.1, 1 mM DTT, 100 μM dNTP mix, 1 mM ATP, 0.5 U/μL T4 polynucleotide kinase, 0.1 U/μL Uracil-DNA glycosylase, and 0.5 U/μL Endonuclease VIII with ultrapure water. This mastermix (30 μL) was distributed into high-profile 0.2 mL strip tubes, followed by 10 μL of sedaDNA extract. These strip tubes were placed on a thermocycler at 37 °C for 3 h. Afterwards, tubes were centrifuged and 0.2 U/μL T4 DNA polymerase (2.7 μL) was individually pipetted into each well (new tips between wells). The strips were mixed, spun down, and placed on a thermocycler at 25 °C for 15 min, followed by 12 °C for 15 min.

Step 1 clean-up. Both the non-UDG and UDG blunt end repair mixes were purified using a QIAquick PCR Purification Kit (Qiagen, following manufacturers specifications, eluting in 20 μL of EB) to maximally retain small fragments (the kit claims to retain fragments down to at least 40 bp). All subsequent steps were identical for both variants.

Step 2, adapter ligation mastermix (20 μL): 1X T4 DNA Ligase Buffer, 5% PEG-4000, 0.25 μM library adapter mix (P5/P7), and 0.125 U/μL T4 DNA Ligase. Samples were mixed and placed on a thermocycler at 16 °C for >15 h. The next day, samples were purified using a MinElute PCR Purification Kit (Qiagen) following manufacturer's specifications and eluting in 20 μL of EB.

Step 3, adapter-fill in mastermix (20 μL): 1X ThermoPol Rxn Buffer, 250 μM dNTP mix, 0.4 U/μL BST Polymerase. Libraries were run on a thermocycler at 37 °C for 30 min followed by 20 min at 80 °C to heat-inactivate the polymerase without needing to subject libraries to another purification (as silica column purifications have been observed anecdotally to result in a ~50% loss of DNA [data not shown]).

Quantitative PCR (qPCR) on a BioRad CFX96 was used to assess library conversion success using a short amplification assay with priming sites on the ligated adapters (mastermix: 1X KAPA SYBR FAST qPCR Master Mix, 0.2 μM Meyer IS7 primer [P5F], 0.2 μM Meyer IS8 primer [P7R])[140,141] run in duplicates against custom GBlock (IDT) standards containing the same adapter priming sites outside of the synthetic insert region run as a duplicate dilution series (E^6–E^0 copies/μL). This assay was run in post-PCR workspaces.

Indexing. Once confirmed that the libraries converted successfully, 12.5 μL of library template per sample was indexed with dual unique 8-bp index sequences using fluorescence guided qPCR (mastermix [27.5 μL]: 1X KAPA SYBR FAST qPCR Master Mix, 750 nM Illumina P5 unique forward index primer, 750 nM Illumina P7 unique reverse index primer, and ultrapure water). Fluorescence guided qPCR was run on a BioRad CFX96 for 6–12 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). This technique was used so that each library received only as many PCR cycles as necessary. Amplifying into plateau (running out of PCR reagents) results in heteroduplexes—molecules with single- and double-stranded regions—which can impact size selection via gel excision following electrophoresis[123]. Indexing qPCR was carried out in post-PCR workspaces using the following cycling conditions: denaturation at 95 °C for 5 min, then 30 s; extension at 60 °C for 45 s, repeated for 6–12 cycles, then final extension at 60 °C for 5 min. All subsequent steps were performed in aDNA dedicated dead air hoods in those spaces with the same UV irradiation and decontamination procedures, but in areas physically separated from pre-PCR ancient DNA labs. Indexed libraries were purified using a MinElute PCR Purification Kit (Qiagen) following manufacturer specifications, eluting in 15 μL of EBT.

Single-stranded library preparation (ssLP). A subset of libraries were also prepared with the ssLP method ssDNA3.0[122,123] to improve library conversion for capture enrichment. This method is summarized here. Prior to library preparation, extracts were treated with 0.025 U/μL USER enzyme for 15 min at 37 °C to reduce the amount of damage within the molecules while leaving the ends relatively unaffected.

Step 1, dephosphorylation mastermix: T4 RNA Ligase Buffer (1X), Tween-20 (0.05% vol/vol), FastAP (0.0125 U/μL), and ultrapure water dispersed and mixed with 5.25 μL of extract, mixed and run on a thermocycler at 37 °C for 10 min, and 95 °C for 2 min.

Step 2, ligation of first adapter mastermix: PEG-8000 (20% wt/vol), ATP (0.5 mM), ssAdapter (0.125 μM, CL78/TL110), and T4 DNA Ligase (0.375 U/μL) mixed by repeated inversion on a rotor for 15 min total (5 min before adding enzyme, and 10 min after), and 34.4 μL of mix added directly to each library from the step 1 mix. This new mix was then run on a thermocycler at 37 °C for 1 h, and 95 °C for 2 min and then placed in a freezer overnight.

Step 3, bead preparation and immobilization of ligation products: the next day an aliquot Dynabeads MyOne Streptavidin C1 (20 μL per library) was magnet pelleted and the storage supernatant was removed, followed by two washes with the B&W Buffer (NaCl 0.1 M, Tris-HCl pH 8.0 0.01 M, EDTA pH 8.0 0.001 M, Tween-20 0.05%, SDS 0.5%, ultrapure water) and resuspended in B&W Buffer, from which 100 μL was added to the ligation reaction from the previous day and rotated at room temperature rotation for 20 min. Afterwards, beads were pelleted with a magnetic rack and the supernatant was discarded. The beads were then washed with 200 μL of B&W Buffer, again followed by pelleting of the magnetic beads and removal of the

supernatant. Next, the beads were washed with 100 µL of Stringency Wash buffer (SDS 0.1 % wt/vol, SSC 0.1 X, ultrapure water) and incubated for 3 min at 45 °C. The reaction was pelleted and the supernatant discarded, followed by resuspension in 200 µL of Wash Buffer II (NaCl 0.1 M, Tris-HCl 0.01, EDTA 0.001 M, Tween-20 0.05%, ultrapure water), followed by magnet pelleting and supernatant removal.

Step 4, primer annealing/extension: mastermix of Klenow reaction buffer (1X), dNTPs (0.2 mM), Tween-20 (0.05% vol/vol), CL130 µM, ultrapure water, and Klenow DNA polymerase (U/µL) added (50 µL per reaction) to step 3 washed beads, mixed, and transferred to thermomixer and incubated at 35 °C for 20 min with shaking at 800 rpm. After the incubation, the reaction was washed following the same bead pelleting procedure as step 3.

Step 5, second adapter ligation and elution: a mastermix of T4 DNA Ligase buffer (1 X), PEG-4000 (5% wt/vol), Tween-20 (0.05 % vol/vol), double-stranded adapter (CL53/73 2 µM), T4 DNA Ligase (0.1 Weiss U/µL), and ultrapure water was added to bead reaction from step 4 (100 µL of mix per reaction), mixed, then transferred to a thermomixer at 22 °C for 1 h shaking at 800 rpm. We used a modified version of CL53/73 as described in ref.[142], which does not require a custom sequencing primer. After the incubation, the reaction was washed following the same bead pelleting wash procedure as in step 3/4. Libraries were then eluted in 50 µL EBT, incubated at 95 °C for 1 min, followed by cooling to 25 °C, then transferred to low-retention 0.6 mL tubes and stored frozen. Libraries were treated identically to those generated with dsLP thereafter, including the same qPCR assay and indexing protocols.

Capture enrichment

Probe design. In-solution RNA hybridization enrichments were carried out using 3 bait-sets. First in batch 2 and 3 using the PalaeoChip Arctic v1.0 combination bait-set (myBaits Arbor Biosciences)[36]. This 2-part bait set targets whole mitochondrial genomes from approximately 180 extinct and extant Holarctic fauna, and the chloroplast genes *trnL*, *rbcL*, and *matK* from approximately 2100 species of northern plants. Both the plant (18,672 probes) and animal (57,529 probes) bait-sets were used together in the same reaction with proportions of 25% plant and 75% animal. Second, Batch 5 used a bait-set to capture mitochondrial genomes from ground squirrels, with emphasis on *Urocitellus*, *Spermophilus*, and *Marmota*, including baits designed to capture ancestral members of Xerinae. The ancestral sequences were inferred with FastML[143] with the approach described in Vogel et al[144]. We inferred ancestral sequences for the two species in *Urocitellus* as well as connecting ancestral members between the three genera, resulting in a set of five mt reference genomes with NCBI accession numbers: NC_059785.1, NC_031209.1, NC_048490.1, NC_042243.1, NC_080739.1, and four ancestral sequences. PalaeoChip and the Xerinae bait-sets were designed and synthesized in collaboration with myBaits Arbor Biosciences. Third, we also enriched with Proboscidean baits. Their design and results are detailed in a parallel forthcoming report.

Enrichment, wet-lab. In solution enrichments were carried out using two rounds of the myBaits v5.03 high-sensitivity protocol (Daicel Arbor Biosciences). For each library, 7.5 µL of indexed library template was combined with 4.5 µL of the library block master mix to block enrichment of repetitive and synthetic adapter DNA (using 8 bp xGen blocking oligos synthesized by IDT [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 [50X Denhardt's Sol], 0.25% Hyb S [10% SDS], 1.56 X Hyb R [RNAsecure], 200 ng/rxn of RNA baits [variable by set]) was pre-warmed to 60 °C before being combined with the library-block mixture. The combined mix was then incubated for ~20 h overnight at 58 °C for bait-library hybridization. See[36,58] and the myBaits v5.03 manual for additional protocol details.

Following overnight hybridization, streptavidin beads (Dynabeads MyOne Streptavidin C1) were dispensed (20 µL per reaction), washed with 200 µL of binding buffer (1 M NaCl, 0.5 mM Tris-HCl, 0.1 mM EDTA, ultrapure water) per reaction, then resuspended in 20 µL binding buffer per reaction and aliquoted into PCR strips. Baits were captured with 20 µL of the bead binding buffer suspension per library, incubated at 58 °C for 2.5 min, quickly manually agitated and spun down, then incubated again for another 2.5 min. Beads were pelleted and the supernatant (the non-captured library fraction) was removed and stored at −20 °C as per Klunk et al[145]. The beads were resuspended in 180 µL of 60 °C Wash Buffer X (0.08% HYB S, 0.2X Wash Buffer [0.1X SSC, 0.1% SDS, 1 mM EDTA]) per tube and washed four times following the myBaits v5.03 protocol. Beads were resuspended in 18.8 µL EBT, PCR reamplified for 14 cycles, then purified with MinElute columns following manufacturer's protocols and eluted in 9.05 µL EB. The entire enrichment process was repeated—starting with the hybridization and library mastermixes—reamplifying during the second round for 8 cycles and eluting post-purification in 15 µL EBT.

Pooling, size-selection, and sequencing

Enriched indexed libraries were quantified using the long-amplification total library qPCR assay on a BioRad CFX96 with priming sites located outside the dual-unique index adapters (mastermix: 1X KAPA SYBR FAST qPCR Master Mix, 0.2 µM Meyer IS5 primer [P5F], 0.2 µM Meyer IS6 primer [P7R])[140,141]. We used Illumina PhiX Sequencing Control V3 as the qPCR standard in a 10-fold dilution series from 100–0.001 pM. In the case of several batch 4 enriched libraries, because the samples were more age-degraded (700 kya) and the target was species specific (unlike the PalaeoChip enrichments where thousands of species are targeted simultaneously), the post-enriched quantifications were too low to pool for a NextSeq given anticipated losses from a post-pooling MinElute purification and a gel excision size-selection purification. This necessitated reamplifying a subset of enriched batch 4 libraries again for 8 cycles prior to pooling.

Enriched and shotgun libraries were pooled by batch to equimolar concentrations. All pools were size-selected with gel excision following electrophoresis for library adapted molecules ranging between 150–500 bp. Gel plugs were purified using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's protocol. Batch 1–3 and part of 4 went through library quality control with an Agilent Tapestation using a High Sensitivity D1000 ScreenTape and were sequenced on an Illumina NextSeq 2000 with 2×50 paired-end sequencing chemistry at the Farncombe Metagenomics Facility (McMaster University). The other part of batch 4 went through library quality control with a QIAxcel Connect System with a DNA High Sensitivity kit (Qiagen) and was sequenced on a NextSeq 1000 with XLEAP-SBS 2×50 paired-end sequencing chemistry at the Marna Genome Lab at the Hakai Institute.

Bioinformatics

Sequenced reads were demultiplexed in all batches with Illumina DRAGEN secondary analysis software (BCL-Convert v4.3.6). Three primary analysis pathways were utilized with this dataset: (1) metagenomic classification of eukaryote and prokaryote taxa with *BLASTn*/*MEGAN7* and *holi*; (2) single-target mitogenome assembly and phylogenetics; and (3) aDNA authentication through both metagenomic and single-target damage analyses.

Metagenomic classification

The metagenomic taxonomic classifications reported in the main text were selected after testing several alternative pipelines as reported in the supplementary materials (Figs. S18-S19). Our goal was to select the approach that minimized false positive taxonomic binning due to database or algorithmic bias. This preference against false positives likely resulted in an abundance of false negatives (see Fig. S18) from

true taxa in the final data-set not being called due to strict lowest common ancestor (LCA) parameters. At this time, there does not appear to be a perfect pipeline for balancing these limitations. As such, we have selected the most reasonably conservative approach in our estimation with strict LCA parameters intended to remove the majority of false positives (largely due to database bias and incompleteness).

MEGAN. The main-text eukaryote metagenomics followed a 3-stage approach (summarized here first with a detailed description thereafter). First, mapping all filtered *FASTA*s to a comprehensive set of organelle reference genomes from regionally sensical taxa to remove prokaryotes, nuclear DNA, and any other taxa with poor reference coverage from the filtered set of reads. Then, after deduplicating the map-filtered *FASTA*s, running those reads through *BLASTn* to retrieve the top 500 alignments against a locally stored 2024 copy of the National Centre for Biotechnology Information GenBank Nucleotide Database (NCBI-NT)[146,147]. A large 'top alignments' value was used to mitigate the problems of database bias towards organisms of commercial/scientific interest, and the problems of *BLASTn* top sequence recovery limitations[148]. This approach has the challenge of generating very large *BLAST* files (often each being >100 gb), but these can be compressed as *gz* files that *MEGAN7* will accept. Finally, we processed the map-filtered *FASTA* and *BLASTn* files through *MEGAN7* (v7.0.9-beta)[78,79] to taxonomically classify the reads with a strict set of LCA parameters. We have found in general that database bias is a significant driver of false positives, and even a 'comprehensive' local database lacks the palaeo-breadth to avoid false positives driven by undescribed ancient diversity. By filtering down with a mapped set from locally curated references and then aligning that against a redundant and massively extensive database (NCBI-NT), we are able to better minimize the associated limitations of both curated and non-curated databases.

Stage 1, demultiplexed *FASTQ* files were trimmed and merged with *leeHom*[149] using ancient DNA parameters (--ancientdna) then mapped (map-filtered) to a curated Pacific Northwest Quaternary Reference panel (Table S6) with *network-aware-BWA* (maximum edit distance of 0.01 [-n 0.01], allowing for a maximum of two gap openings [-o 2], and with seeding effectively disabled [-l 16500]). All reads that mapped (merged or unmerged but properly paired) were then extracted with *libbam* (https://github.com/grenaud/libbam), deduplicated based on 5' and 3' mapping positions with *biohazard* (https://bitbucket.org/ustenzel/biohazard), and converted to *FASTA* files and size restricted to a minimum length of 24 bp. These map-filtered *FASTA* read files were filtered to remove any lingering sequence similarity to the Illumina adapter sequences (sequences with an edit distance of 1 to the common adapter sequence AGATCGGAA and its reverse complement), and were then string deduplicated using the *NGSXRemoveDuplicates* module of *NGSeXplore* (https://github.com/ktmeaton/NGSeXplore).

Stage 2, map and quality filtered *FASTA*s were used as the input for *BLASTn*[150], which were aligned against an April 9, 2024 local copy of NCBI-NT set to return the top 500 alignments (unique accession hits) per read with e-values less than 1.0E-5 (flags: -num_alignments 500 -max_hsps 1 -evalue 0.00001).

Stage 3, *BLASTn* output files were then passed to *MEGAN7* (v7.0.9-beta) 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-7, Minimum percent identity = 95%, Top percent consideration of taxonomic hits based on bit-score = 20%, Minimum read support = 5, naive LCA at 100% coverage, Minimum percent read to cover = 99%. Extraction replicate libraries were merged in *MEGAN* using the absolute comparison tool. All libraries (merged replicates and single libraries processed without replicates) were compared using the absolute comparison tool and nodes with assigned reads were plotted with heat maps by kingdom.

Non-target. A secondary set of metagenomic classifications were also carried out for non-target taxa (e.g. prokaryotes, fungi, arthropods)[56]. In this set, *FASTQ* files were again trimmed and merged with *leeHom*, *BAM* files were converted to *FASTA* format, then subset to a minimum length of 24 bp, filtered for any lingering similarity to sequencing adapters, and string deduplicated with the *NGSXRemoveDuplicates* module of *NGSeXplore* (https://github.com/ktmeaton/NGSeXplore). These filtered *FASTA* files were queried with *MegaBlast*[150] using a July 2020 local copy of NCBI-NT set to return the top 50 alignments (unique accession hits) per read with e-values less than 1.0E-8 (flags: -num_alignments 50 -max_hsps 1 -evalue 0.00001). *MegaBlast* was used in this case over *BlastN* because the speed of the *BlastN* algorithm is too slow for a dataset of this size that has not been filtered to a set of curated genomic targets. The outputs were passed to *MEGAN7* where the *BLAST* results were filtered through a lowest common ancestor (LCA) algorithm using the following parameters: MinScore = 50.0; MaxExpected = 1.0E-8; MinPercentIdentity = 96.0; TopPercent = 15.0; MinSupport = 10; LCA = naive; MinPercentReadToCover = 98; mode = BlastN.

holi. The *holi* pipeline was run with in-house modifications to the Kap København protocol[151,152]. Individual raw *FASTQ* files were first converted to *FASTA* format with *seqtk* v1.3-r114-dirty (https://github.com/lh3/seqtk), then mapped to the RefSeq bacteria database with *BlastN* v2.10.0 + [150] (flags: -task megablast, -evalue 1e-5, -max_target_seqs 5, -perc_identity 98, -qcov_hsp_perc 90). The resulting list of bacterial mapping reads were compiled for forward and reverse reads and subsequently removed from the *FASTQ* files using *bbmap* v39.06 (https://sourceforge.net/projects/bbmap/). This removal significantly decreased the number of putative false positive assignments in the LCA results, likely due to bacterial contamination present in many reference genomes. Poly-A and poly-T tails and remaining adapter sequences were removed with *fastq-tools* v0.8.3 (https://github.com/dcjones/fastq-tools/tree/master). Low complexity and short reads were removed and the remaining reads were indexed and deduplicated using *sga* v0.10.15[153] with a dust threshold of 1 and a minimum read length of 30. All filtered *FASTQ* files belonging to the same sample were merged, then indexed and deduplicated using *sga*.

Pre-processed reads from each sample were then mapped to 11 databases using *bowtie2* v2.5.0[154]. Databases used include Arctic Fauna[151], PhyloNorway[72], NCBI NT[155], and the following RefSeq databases: bacteria, invertebrates, chloroplast, mitochondria, vertebrate mammals, other vertebrates, other eukaryotes, and others[156]. Resulting *sam* files were sorted using *sambamba* v0.8.2[157] and merged for each sample using *samtools* v1.18[158]. The merged *sam* file for each sample was run through *metaDMG-core* v0.38.0[159] to generate LCA and damage profile results.

Metagenomic classification damage support. Taxonomic classifications identified from the *BLASTn*/*MEGAN* and *MegaBlast* workflows were evaluated along three parallel evidence axes: (i) damage authentication based in *Holi*/*metaDMG*, (ii) independent macrofossil corroboration, and (iii) ecological plausibility. These evaluations and the full taxonomic classifications are included as a Supplementary Data 3.

For each taxon, read-distribution support was summarized from the *MEGAN* count matrices as (i) the maximum number of reads observed in any ancient library, (ii) the number of ancient libraries in which the taxon occurred, and (iii) its occurrence in the merged negative controls. A strict read-distribution rule was defined as at least 50 reads in at least one ancient library with the *MEGAN* workflow, presence in at least two ancient libraries, and absence from the merged negative controls. Library-specific *Holi*/*metaDMG* outputs were then used to evaluate damage-based support for the same taxonomic classifications. Exact *Holi*/*metaDMG* support was recorded only when

the same taxon was present in the corresponding *Holi/metaDMG* library output, in a library where that taxon was also present in the corresponding *MEGAN* count matrix with damage > 0.01, significance > 2, and MAP_valid = TRUE. A separate strong single-library rule was used to recognise temporally or ecologically restricted taxa that reached at least 250 reads in one ancient library while remaining absent from the merged negative controls.

Because *Holi/metaDMG* can accumulate evidence at taxonomic nodes that attract multiply aligned reads from related taxa in aeDNA samples, exact damage/significance support was further filtered using a simple multi-mapping/fit QC overlay derived from the *Holi/metaDMG* output itself. For the best exact-support library for each taxon, QC was summarized from alignments per read (N_alignments / N_reads), rho_Ac, and MAP_valid. Alignments per read were treated as a proxy for elevated multi-mapping or reference overlap, whereas rho_Ac was treated as a simple fit-caution indicator. QC was classified as clean when alignments per read were <= 10 and rho_Ac <= 0.3, caution when either metric exceeded those values but remained <= 25 and <= 0.5 respectively, and strong caution when alignments per read were > 25, rho_Ac > 0.5, or MAP_valid was FALSE. Exact damage/significance support that failed these thresholds was retained as evidence, but downgraded in the final classification. These cautions are indicative of reads from multiple species mapping in the *Holi/metaDMG* workflow, making it difficult to measure taxon-specific rates of damage in those instances without extensive competitive mapping validations.

Lineage-consistent support was evaluated separately from exact authentication. Lineage support was restricted to direct ancestor/descendant relationships at family rank or lower (family, subfamily, tribe, subtribe, genus, subgenus, species, and subspecies) inferred from the *Holi/metaDMG* tax_path hierarchy. Broad clades and higher-rank nodes such as order, class, and no-rank groupings were not treated as lineage support. Lineage-consistent support can help classify a taxon as "Supported" but not by itself yield an authenticated (high or very high confidence) class.

Blank association was evaluated from both the merged negative-control counts and the *Holi/metaDMG* blanks output. A taxon was classified as showing strong blank association when exact *Holi/metaDMG* damage/significance support was also present in blanks, when the merged negative controls contained at least 10 reads, or when control reads reached at least 10% of the maximum ancient-library count and at least 2 reads. Lower-level control signal was recorded as blank caution, and all remaining taxa were classed as blank clean.

The final aDNA support ladder was defined as follows. Very high confidence supported required exact *Holi/metaDMG* support, independent support from the conservative *MEGAN* count matrix (strict read-distribution support or strong single-library support), no blank association, and at least one exact-support library with *Holi/metaDMG* N_reads >= 100 and clean *Holi/metaDMG* QC. This category applied to 17.1% ($n = 121$) of taxa in the targeted *BLASTn* approach and 11.4% ($n = 87$) of taxa in the *MegaBlast* off-target approach. High confidence supported required exact *Holi/metaDMG* support, independent *MEGAN* count-matrix support, and no blank association, but without the additional very high-confidence thresholds. This category applied to 7.5% ($n = 53$) of taxa in the targeted *BLASTn* approach and 1.6% ($n = 12$) of taxa in the *MegaBlast* off-target approach. Supported was applied to taxa with exact *Holi/metaDMG* damage/significance support but insufficient independent count support or downgraded QC (due to mixed mapping from related taxa), taxa with strong *MEGAN* count-matrix support but no exact *Holi/metaDMG* support, and taxa with low-rank lineage support in libraries where the focal taxon was present in the count matrix. Within this class, exact damage-based support was flagged in the support field as damage-supported (exact). This category of "Supported" includes organisms with very strong genomic assembly and phylogenetic support, along with very high read counts (such as bison, horse and mammoth), but are only classified as "Supported" because of the *Holi* QC caution from mixed-mapping which requires competitive mapping to remove. Since mix-mapping is expected with aeDNA, this category is the largest of the dataset, and was applied to 61% ($n = 431$) of taxa in the targeted *BLASTn* approach and 47.4% ($n = 362$) of taxa in the *MegaBlast* off-target approach. Tentative was reserved for taxa lacking the evidence required for the higher aDNA support classes but remaining plausible in light of ecological context, and/or independent corroboration, provided that they were not strongly blank-associated. This category also includes ecologically suspect organisms (far outside of known range), but that have damage and count support that would otherwise be sufficient for higher categories if not for the ecological uncertainty. This category applied to 2.4% ($n = 17$) of taxa in the targeted *BLASTn* approach and 17.7% ($n = 135$) of taxa in the *MegaBlast* off-target approach. Weak support and ecologically uncertain was used as a precautionary category for low-support taxa that were not strongly blank-associated but lacked convincing count/damage, ecological, or corroborative support. This category applied to 12.0% ($n = 85$) of taxa in the targeted *BLASTn* approach and 12.5% ($n = 95$) of taxa in the *MegaBlast* off-target approach. Blank-associated or likely artefactual was reserved for taxa with strong negative control overlap and without sufficient exact *Holi/metaDMG* support or independent count support to overcome that blank signal (applicable primarily to microbes). This category applied to 0.0% ($n = 0$) of taxa in the targeted *BLASTn* approach and 9.4% ($n = 72$) of taxa in the *MegaBlast* off-target approach.

Independent corroboration was reported separately from DNA authentication in Fig. 4, 5. Corroboration was assigned from the macrofossil dataset (Supplementary Data 1). Species- and genus-level overlaps were treated as direct macrofossil corroboration, whereas higher-rank overlaps were treated as lineage-consistent macrofossil corroboration. Ecological plausibility was evaluated separately from both DNA authentication and independent corroboration. Pleistocene ground-squirrel coprolites were interpreted against a Beringian mammoth-steppe setting, whereas the Holocene SC-12 snowshoe hare coprolite was treated as a distinct ecological control with boreal taxa.

Single-target genome assembly and phylogenetics

Urocitellus mitochondrial genomes were assembled using both de novo and reference-based assembly. For the de novo assembly we used *SPAdes* (v3.13.0)[160], a specified short read aligner. The reference-based assembly was performed using *endoCaller* from *Schmutzi* (v1.5.6)[161] with mapping performed using *BWA aln* (v0.7.17) (parameters were set as described previously) and *bam2prof* (https://github.com/grenaud/bam2prof) providing the damage estimation profiles used by *endoCaller*. Assembled mitogenomes from our samples were aligned with all representatives of the subfamily Xerinae, including the two available mitogenomes from the genus *Urocitellus*—*Urocitellus richardsonii* (NCBI accession number: NC_031209.1) and *Urocitellus parryii* (NCBI accession number: NC_059785.1). The multiple sequence alignments were done using *MAFFT* (v7.526)[162] on automatic mode and the phylogenetic trees were constructed using *RAxML* (v8.2.12)[163] with a GTR substitution model and the Daurian ground squirrel (*Spermophilus dauricus*, NCBI accession number: NC_027283.1) as an outgroup as well as 1000 bootstraps for the tree verification.

First, we aimed to assemble all samples de novo using the trimmed unmerged reads from stage 1 as input. Only two samples resulted in fully assembled sequences: SC-10 and SC-16 with a k-mer size of 33. SC-10 has an average coverage of 37.76X and was fully assembled with a length of 16,479 bp but with a warning referring to low coverage. SC-16 was assembled with a coverage 588.74X and a total sequence length of 16,332 bp. To verify our de novo assemblies, we additionally performed reference-based assembly for both samples and reached a sequence identity between the assemblies of 99.96% for SC-16 and 99.98% for SC-10.

For all samples where de novo assembly only resulted in individual contigs, we chose reference-based assembly. First, we extracted all reads for each sample that were assigned to the subfamily Xerinae using *euka* (v2.0.2)[164]. These extracted reads were run through *soibean* to identify the species. For all samples except SC-14, *soibean's*[144] estimate landed on the reference genome for *Urocitellus richardsonii* close to the root connecting it with the reference for *U. parryii*. The close proximity made us use both references for independent assemblies with *endoCaller*. In both trees we see our assembled genomes fall in a separated cluster to both *Urocitellus* reference genomes, which led us to construct a third reference-based assembly, where we used our SC-16 de novo assembled mitogenome as the reference guidance (Fig. 6, S22). We preferred the SC-16 de novo assembly due to its higher coverage. An overview of reference-based assembly stats can be found in Table S7.

For the 700 kya sample (SC-14), the original shotgun-sequences sample only contained 249 unique reads mapping to the family of Xerinae. After extracting and remapping these reads using *soibean*, we could observe the most likely position of the data in the Xerinae tree to be on the ancestral branch N8, representing the ancestral reconstructed internal tree node between both *Urocitellus* sequences. After target enrichment we aimed for de novo assembly of SC-14, however contigs were insufficient in length and coverage. The reference-based assembly for SC-14 resulted in an average coverage of 366.7X and 94% of the reference genome covered. As indicated from the shotgun data analysis with *soibean*, the assembled mitogenome for SC-14 lands basal to the genus *Urocitellus*.

The lack of available reference genomes for the genus *Urocitellus* limits the range of our analysis because the genus of *Urocitellus* includes several unrepresented species. Publicly available references for additional *Urocitellus* species only include the Cytochrome B site in the mitogenome. We wanted to place our reconstructed sequences within the wider phylogenetic context of these sequences. We aligned all available Cytochrome B site sequences from the genus *Urocitellus* with our assembled sequences using *MAFFT*. The resulting multiple sequence alignment (MSA) was cut at the starting and end position of the Cytochrome B sites. We observe that SC-14 again sits basal to all sequences.

For the *Urocitellus* Bayesian phylogenetics, all available complete mitochondrial genomes for Marmotini were downloaded from GenBank as well as all cytochrome b sequences for *Urocitellus*. Sequences were aligned using *MAFFT* v7.490 with the --auto option[162]. Phylogenetic analysis was performed using *BEAST* 1.10.4[165] on the Cipres Science Gateway[166]. As tip dates for our newly sequenced samples, we either used direct radiocarbon dates or dates of the tephra layers the samples were associated with (Tables S3–S4). We used the HKY + G + I substitution model[167] and chose an uncorrelated relaxed clock[168] with a lognormal prior distribution with a mean of 5E-8 and a standard deviation of 2E-8. As a tree prior we used a coalescent constant size prior with a gamma distribution shape of 10 and a scale of 800,000. The Markov chain Monte Carlo was run for 100 Mio iterations, sampling every 10,000th iteration. Effective sample size was checked using Tracer v. 1.7.2[169], with all scores being over 200. A maximum clade credibility tree was built using *TreeAnnotator*, removing the first 10% of trees as burn-in. The final tree was visualized in *FigTree* v. 1.4.4 (https://github.com/rambaut/figtree/releases).

Mitochondrial genomes for *Equus*, *Bison*, and *Lepus* were assembled with the following reference mapping approach. Demultiplexed *FASTQ* files were converted to *bam* files with *fastq2bam* (https://github.com/grenaud/BCL2BAM2FASTQ), then trimmed and merged with *leeHom*[149] using ancient DNA specific parameters (--ancientdna). Reads were mapped to a regionally sensical species reference with network-aware-BWA (https://github.com/mpieva/network-aware-bwa)[170] with a maximum edit distance of 0.01 (-n 0.01), allowing for a maximum two gap opening (-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), filtered to only retain reads with a > 30 map quality and minimum fragment length of 24 bp using *samtools* (https://github.com/samtools/samtools). These mapped bam files were then imported into Geneious Prime 2024.0.7 (https://www.geneious.com/) for visualization where the contigs were manually curated using pre-assembled alignments and NCBI references as guides to reduce the influence of stacked, non-specific reads (common in aeDNA sample types) from impacting consensus calls of false positive single nucleotide polymorphisms and indels. 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 and adjacent to these high coverage regions were N-masked. Consensus sequences were called with an 85% nucleotide identity threshold and degenerate bases called in ambiguous positions.

Mitogenome alignments for Equus and Bison were obtained from[57], which were curated for sequences with ≥80% completeness and >3X coverage. New consensus sequences were added and aligned with *muscle* (v3.8.425)[171]. The alignment was inspected, and end gaps were replaced with Ns. Model testing and maximum likelihood analyses were conducted using *IQ-Tree* (v1.6.1283) with outgroup rooting and 1000 bootstrap replicates.

Authentication

Our assessment of aeDNA authenticity is based on (1) wet-lab negative controls, quantifications assays, and rigorous cleanroom operating procedures; (2) DNA damage analyses of deamination (nucleotide misincorporations) and depurination (strand fragmentation); (3) in silico testing of alternative metagenomic classifiers with simulated and real data; (4) analyses of off-target taxonomic assignments; (5) comparison with complementary methods; and (6) genome assembly and phylogenetic placement.

Negative controls and SOPs. Negative controls were used at every stage of processing, starting with open tubes during subsampling to monitor for airborne contamination which then were processed alongside during extraction, new extraction negatives added during lysing, new library negatives added during each batch of dsLP and ssLP, indexing negatives, and separate replicate non-template controls on each qPCR assay. More than 15 negative controls were processed alongside our aeDNA samples. Across all batches, 107 extracts (including replicates) were processed, and a final total of 144 sample libraries were sequenced. With the 15+ negatives, we had a ratio of 1 negative for each ~8 coprolite samples. Of the 15+ negative controls, 13 negatives had sufficient adapter and background microbial DNA to be PCR detectable and sequenced and those results are displayed alongside all coprolite samples. During indexing, the negatives were run for additional cycles to increase their concentration to allow for representation during pooling and sequencing. Despite this, all negatives had to be pooled in their entirety during equimolar pooling because their concentrations remained too low. All forms of negatives were closely monitored during qPCR assays (completed after each stage of processing), in addition to monitoring for successful amplification of samples. In addition, rigorous aDNA ultra-cleanroom operating procedures (as described above) were utilized throughout to minimize contamination risk. This includes full body suits, hairnets, masks, and face shields in a DNA decontaminated cleanroom, frequent glove changing and glove sterilization, using pre-sterilized plastic wear and UV-C irradiating surfaces, equipment, and other non-sterile plasticware, decanting/aliquoting stock reagents, decontaminating surfaces with bleach/DNA-Away/ELIMINase, and strict 1-way travel of personnel, equipment, reagents, and samples from low-to-high copy workspaces.

DNA damage. Nucleotide misincorporations and short fragment lengths are useful characteristics for confirming that sequence data originates from old DNA and not contemporary contaminants[87,172–175]. We used *mapDamage* (v2.3.0)[176,177] and *metaDMG* (v0.37)[159]. Plots associated with these damage analyses are included with their associated phylogenetic trees metaDMG plots for all samples are included in the supplementary materials (Figs. S36-S97, S102). For the *metaDMG* significance/damage plots, these highlight taxa with a sufficient amount of C-T and G-A misincorporations to exceed thresholds that these taxa have strong support of being ancient. Taxa that fall below these thresholds are not necessarily contaminants, but rather simply lack the abundance of data to demonstrate that they have the statistical significance to be robustly classified as being ancient. See the preceding section on metagenomic classifications for how the *Holi/metaDMG* data was used to support the primary (PNW map-filtered, *BLASTn/MEGAN* and *MegaBlast/Megan*) taxonomic classifications

Thermal age. The deamination and depurination analyses followed previously published methods[56,58,178]. Read lengths were filtered out from the linear regression if it accounted for less than 10 reads to prevent any outliers or insufficient data from affecting the results. Estimates of the effect of climatological changes on the damage process were estimated using δO^{18} values from the Greenland ice sheet[179]. Linear interpolation of the δO^{18} was required as the estimated dates for our samples did not match the sampling date from the Greenland ice sheet. The SC-14 sample was excluded from the damage analysis as the sampling date was older than the data available from the Greenland ice sheet.

In silico metagenomic simulations. We wanted to assess rates of false/true positives and false/true negatives that may be observed in our (and alternative) metagenomic classification strategies using simulated organelle DNA with RefSeq taxa identified our ancient samples. This test is not a 1:1 test of our coprolite aeDNA data as each simulated set of genomic aDNA directly exists in our reference databases, whereas with ancient samples the vast majority of diversity in our samples is unrepresented in even the most comprehensive of reference databases. The purpose of this analysis was to confirm that if we give the classifiers known data, we can ensure that it classifies those reads at minimum somewhere along the correct taxonomic rank hierarchy as we expect it to, if not to the exact species. This ensures that in an ideal 1:1 scenario, it functions as intended. To do this, we generated three main datasets with *gargammel*[82]: *Urocitellus parryii* (mitogenome) alone, *U. parryii* with bacteria, and *U. parryii* with bacteria and 11 animal mitogenomes and 10 plant chloroplast genomes (Table S8).

To model bacterial taxa likely to occur in our empirical data, bacterial reads identified in the preprocessing step of the *holi* pipeline (see *holi* methods) from a single sample (SC-10c_TCMP) were run through the eager pipeline v2.4.7[180] using the kraken2 standard database. In R v4.3.3[181] all non-Homo sapiens species-level reads in the kraken2 output were selected and randomly sampled 100 times to determine a representative list of taxa and the frequency of each species in the example empirical sample for the bacterial only contamination file. Reference genomes for each species were downloaded from NCBI using Entrez Direct.

To model other eukaryotic samples, taxa were chosen that are common in North American Pleistocene aeDNA datasets (Table S8). To determine relative proportions of plants and animals to model, we ran the raw SC-10c_TCMP *FASTQ* files through the *eager* metagenomic pipeline, classifying with *kraken2* to the *RefSeq* mitochondrial database (see Supplementary Data 2) and observed relative eukaryotic proportions of approximately 25% plants and 75% plants. The proportion of each species within each kingdom was equal (i.e., there were 11 animal species, the eukaryotic proportion of each was modelled at 1/11 * 0.25). These proportions along with the bacterial contaminant species proportions were then scaled 80% bacteria and 20% eukaryotes for the combined contamination file. Reference mitochondrial genomes for each species were downloaded from NCBI using *edirect* v16.8 [182].

Six different simulations were run in *gargammel* v1.1.2[82] with 1.15 M 51 bp reads per read direction (2.3 M total), loosely mirroring the example empirical file SC-10c_TCMP. As *gargammel* does not model the exact Illumina technology used in our empirical files (NextSeq 1000 or 2000), we used the default settings of HiSeq 2500 reads and adapters. The simulations were then run with either 100% *U. parryii* reads; 80% bacterial contamination and 20% *U. parryii* reads; or 80% bacterial and eukaryotic contamination reads and 20% *U. parryii* reads. These three variants were run with and without deamination simulations using the suggested Briggs Model parameters[82,173] with the same rates of damage for all taxa (-damagee 0.03,0.4,0.01,0.3).

1) *Eager* pipeline with *kraken2*: The *gargammel* simulated forward and reverse reads were run through the eager v2.4.7 metagenomic pipeline[183]. Briefly, this included removing adapters and merging paired reads with *AdapterRemoval* v2.3.2, mapping to the *Homo sapiens* GRch38 reference genome to remove human contamination with *bwa* v0.7.17-r1188, and classifying non-Homo sapiens reads with the RefSeq mitochondrial database using *kraken2* (see Supplementary Data 2).

2) *BLASTn/MEGAN7*: See main metagenomic analysis summary.

3) *MALT* + local database pipeline: A *Malt* database was built using mitochondrial and chloroplast reference genomes of plant and animal taxa local to the Pacific Northwest, including all simulated eukaryotic taxa references. Duplicates were removed from the non-Homo sapiens *FASTA* output files from the eager pipeline with *seqkit* v0.12.1[184]. Low-complexity reads were removed from the files using prinseq-lite v0.20.4[185] with a *DUST* entropy threshold of 7. Processed reads were then mapped to the Pacific Northwest taxa database with *MALT* v0.6.1[186] using settings designed to match the *MEGAN7* settings detailed above (--mode BlastN --minBitScore 50 --maxExpected 0.00000001 --topPercent 15 --minSupport 5 --minPercentIdentityLCA 96 --numThreads 8 --alignmentType SemiGlobal -iu --verbose).

4–5) *aMeta* pipeline: The forward and reverse *gargammel* simulation reads were merged with *fastp* v0.23.4 then run through the *aMeta* pipeline[187], mapping to the NCBI NT database with *krakenuniq*[188] and *bowtie2*[189]. We altered the default *aMeta MALT* settings to more closely match the settings we used in our malt with local database run of the simulated data (-sup 5 -top 15 -mpi 96.0 -b 50 -e 0.00000001 -v).

For the *eager+kraken2*, *BLASTn/MEGAN7*, *MALT*+local database, and *aMeta* pipelines, the taxonomy assigned to each read was compared to the taxonomy of the NCBI accession the read was modelled from in order to determine how close the pipeline got to the correct assignment. In R[181], reads were categorized as true positive (called taxon equal to simulated taxon), partial true positive called taxon was not called to the simulated species, but to a higher rank with the simulated species' taxonomy (i.e., Mammalia or Marmotini or *Urocitellus* instead of *Urocitellus parryii*), false positive called taxon was not within the simulated species' taxonomy (i.e. *Urocitellus richardsonii* instead of *Urocitellus parryii*), or false negative (the pipeline was not able to assign taxon to the read). The sum of reads corresponding to each category and proportion of all reads in each category were calculated for each pipeline. For partial true positives, the ratio of correct ranks:total ranks of simulated reads was recorded to determine how close the partial true positives were to the ground truth. Similarly, for false positives a correctness ratio for each read was calculated as (1):

$$correctness = \left(\frac{n\,correctly\,called\,ranks}{n\,ranks\,in\,simulated\,sample}\right) - \left(\frac{n\,incorrect\,ranks}{n\,ranks\,in\,simulated\,sample}\right)$$

In some cases, the number of incorrect ranks was higher than the number of simulated ranks; for example, if a read was called as *U. parryii* with 30 ranks but the read came from a bacterial taxon with only 9 total ranks. In these cases the proportion of incorrect to total ranks was substituted with 1.

Macrofossils

Macrofossil analysis was conducted on the remaining 11 faecal samples after aDNA extraction (SC-2, SC-5, SC-6, SC-8, SC-9, SC-10, SC-11, SC-12, SC-14, SC-16, SC-17). All samples processed for macrofossils were exhausted. To disaggregate samples, samples SC-5 and SC-8 were stored in 10% trisodium phosphate (Na_3PO_4) and all other samples, except SC-14 which was fully disaggregated, were treated in a 2-h cold bath of 10% sodium hydroxide (NaOH). Samples were then wet screened for macrofossils using a mesh size of 125 μm and washed with deionized water. Reserved wash water was spun down and the residue was processed for pollen and spores. Screened macrofossils were scanned under magnification (Lecia S9i stereo microscope, 20 X – 120 X), along evenly spaced transects, and all identifiable and unique macrofossil types were pulled for identification.

The Seed Reference Collection (SRC, Quaternary Environments, Royal Alberta Museum) and published keys and papers[18,190–192] were referenced for taxonomic identification. Imaging of macrofossils was completed using a stereo microscope with built in camera (Lecia M 165 C, IC 90 E) and processed using Zerene Stacker and Adobe Photoshop.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw sequence data have been uploaded to the NCBI SRA (BioProject: PRJNA1180709) and made publicly available. The metagenomic classifications, damage support, and all other sample data are included as Supplementary Data files. All other intermediary files derived from the raw sequence data (e.g., *bam, fasta, blast, rma6*) that are not necessary for the recreation of these results are stored on servers at the Hakai Institute and McMaster University and will be available upon request to Tyler Murchie (tyler.murchie@hakai.org) if any intermediary files are of interest. Sedimentary raw sequence data used here for analysis from Murchie et al. (2021/2022/2023) are also available on the NCBI SRA under BioProjects: PRJNA1004851, PRJNA752360 and PRJNA722670. The mammoth bone raw sequence data from Rowe et al. (2024) are available under BioProject PRJNA1405728. The PNW Quaternary References are available on Zenodo https://doi.org/10.5281/zenodo.16621723 along with the PalaeoChip Arctic v1.0 bait sequences used for capture enrichment https://doi.org/10.5281/zenodo.5643845. Source data are provided as a Source Data file. Source data are provided with this paper.

References

1. Yang, L., Zhang, X., Zhao, X. & Xiang, H. The technological advance and application of coprolite analysis. *Front. Ecol. Evol.* **9**, 797370 (2022).
2. Green, E. J. & Speller, C. F. Novel substrates as sources of ancient DNA: Prospects and hurdles. *Genes 2017* **8**, 180 (2017).
3. Sutton, M. Q., Malik, M. & Ogram, A. Experiments on the determination of gender from coprolites by DNA analysis. *J. Archaeol. Sci.* **23**, 263–267 (1996).
4. Poinar, H. N. et al. Molecular coproscopy: Dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Sci. (1979).* **281**, 402–406 (1998).
5. Karpinski, E., Mead, J. I. & Poinar, H. N. Molecular identification of paleofeces from Bechan Cave, southeastern Utah, USA. *Quat. Int.* **443**, 140–146 (2016).
6. Hagan, R. W. et al. Comparison of extraction methods for recovering ancient microbial DNA from paleofeces. *Am. J. Phys. Anthropol.* **171**, 275–284 (2020).
7. Bon, C. et al. Coprolites as a source of information on the genome and diet of the cave hyena. *Proc. R. Soc. B: Biol. Sci.* **279**, 2825–2830 (2012).
8. Hofreiter, M. et al. A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol.* **9**, 1975–1984 (2000).
9. Akeret, Ö & Rentzel, P. Micromorphology and plant macrofossil analysis of cattle dung from the Neolithic lake shore settlement of Arbon Bleiche 3. *Geoarchaeology* **16**, 687–700 (2001).
10. Speller, C. F. et al. Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. *Proc. Natl. Acad. Sci. Usa.* **107**, 2807–2812 (2010).
11. Palacio, P. et al. Genome data on the extinct Bison schoetensacki establish it as a sister species of the extant European bison (Bison bonasus). *BMC Evol. Biol.* **17**, 1–11 (2017).
12. Kuch, M. & Poinar, H. Extraction of DNA from Paleofeces. in *Methods in Molecular Biology* vol. 840 37–42 (Humana Press, 2012).
13. Wood, J. R. et al. Mid-Holocene coprolites from southern New Zealand provide new insights into the diet and ecology of the extinct little bush moa (Anomalopteryx didiformis). *Quat. Sci. Rev.* **263**, 106992 (2021).
14. Zazula, G. D., Froese, D. G., Elias, S. A., Kuzmina, S. & Mathewes, R. W. Early Wisconsinan (MIS 4) Arctic ground squirrel middens and a squirrel-eye-view of the mammoth-steppe. *Quat. Sci. Rev.* **30**, 2220–2237 (2011).
15. Zanina, O. G., Gubin, S. V., Kuzmina, S. A., Maximovich, S. V. & Lopatina, D. A. Late-Pleistocene (MIS 3-2) palaeoenvironments as recorded by sediments, palaeosols, and ground-squirrel nests at Duvanny Yar, Kolyma lowland, northeast Siberia. *Quat. Sci. Rev.* **30**, 2107–2123 (2011).
16. Langeveld, B. W. et al. A multidisciplinary study of a Late Pleistocene arctic ground squirrel (Urocitellus parryii) midden from Yukon, Canada. *Quat. Res.* **89**, 333–351 (2018).
17. Cocker, S. L. et al. Latest Pleistocene (17,500–13,500 cal yr BP) Arctic ground squirrel (Sciuridae: Urocitellus parryii) middens record late persistence of steppe-tundra in central Yukon Territory. *Quat. Res.* 1–19. https://doi.org/10.1017/QUA.2025.10032 (2025).
18. Zazula, G. D., Froese, D. G., Elias, S. A., Kuzmina, S. & Mathewes, R. W. Arctic ground squirrels of the mammoth-steppe: Paleoecology of Late Pleistocene middens (~24 000–29 450 14 C yr BP), Yukon Territory. *Can. Quat. Sci. Rev.* **26**, 979–1003 (2007).
19. Westgate, J. A. et al. Gold run tephra: A middle pleistocene stratigraphic and paleoenvironmental marker across west-central Yukon Territory, Canada. *Can. J. Earth Sci.* **46**, 465–478 (2009).
20. Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. Molecular caving. *Curr. Biol.* **13**, 693–695 (2003).
21. Poinar, H., Kuch, M., McDonald, G., Martin, P. & Pääbo, S. Nuclear gene sequences from a late pleistocene sloth coprolite. *Curr. Biol.* **13**, 1150–1152 (2003).
22. Rawlence, N. J. et al. Using palaeoenvironmental DNA to reconstruct past environments: Progress and prospects. *J. Quat. Sci.* **29**, 610–626 (2014).
23. Shillito, L. M., Blong, J. C., Green, E. J. & van Asperen, E. N. The what, how and why of archaeological coprolite analysis. *Earth. Sci. Rev.* **207**, 103196 (2020).
24. Blong, J. C. & Shillito, L. M. Coprolite research: Archaeological and paleoenvironmental potentials. *Archaeol. Anthropol. Sci.* **13**, 1–7 (2021).

25. Hunt, A. P. & Lucas, S. G. Coprolites. *Encycl. Geol.: Vol. 1-6, Second Ed.* **3**, 532–544 (2020).
26. Poinar, H. N. et al. A molecular analysis of dietary diversity for three archaic native americans. *Proc. Natl. Acad. Sci. USA* **98**, 4317–4322 (2001).
27. Iñiguez, A. M., Reinhard, K. J., Araújo, A., Ferreira, L. F. & Vicente, A. C. P. Enterobius vermicularis: ancient DNA from north and south American human coprolites. *Mem. Inst. Oswaldo Cruz* **98**, 67–69 (2003).
28. Iñiguez, A. M., Araújo, A., Ferreira, L. F. & Vicente, A. C. P. Analysis of ancient DNA from coprolites: a perspective with random amplified polymorphic DNA-polymerase chain reaction approach. *Mem. Inst. Oswaldo Cruz* **98**, 63–65 (2003).
29. Loreille, O., Roumat, E., Verneau, O., Bouchet, F. & Hänni, C. Ancient DNA from Ascaris: extraction amplification and sequences from eggs collected in coprolites. *Int. J. Parasitol.* **31**, 1101–1106 (2001).
30. King, C. E., Debruyne, R., Kuch, M., Schwarz, C. & Poinar, H. N. A Quantitative Approach to Detect and Overcome PCR Inhibition in Ancient DNA Extracts. *Biotechniques* **47**, 941–949 (2009).
31. Kreader, C. A. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl. Environ. Microbiol.* **62**, 1102–1106 (1996).
32. McKee, A. M., Spear, S. F. & Pierson, T. W. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* **183**, 70–76 (2015).
33. Opel, K. L., Chung, D. & McCord, B. R. A Study of PCR Inhibition Mechanisms Using Real Time PCR*. *J. Forensic Sci.* **55**, 25–33 (2010).
34. Haile, J. Ancient DNA extraction from soils and sediments. in *Methods in Molecular Biology* (eds. Shapiro, B. & Hofreiter, M.) vol. 840 57–63 (Humana Press, 2012).
35. Epp, L. S.,Zimmermann, H. H. & Stoof-Leichsenring, K. R. Sampling and Extraction of Ancient DNA from Sediments. in Ancient DNA: Methods and Protocols (eds. Shapiro, B. et al.) 31–44 https://doi.org/10.1007/978-1-4939-9176-1_5. (Humana Press, New York, 2019).
36. Murchie, T. J. et al. 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 (2021).
37. Heintzman, P. D. et al. The Sedimentary Ancient DNA Workflow. in *Tracking Environmental Change Using Lake Sediments. Developments in Paleoenvironmental Research, vol 21*. (eds Capo, E., Barouillet, C. & Smol, J. P.) 53–84 (Springer, Cham, 2023).
38. Herron, M. D., Castoe, T. A. & Parkinson, C. L. Sciurid phylogeny and the paraphyly of Holarctic ground squirrels (Spermophilus). *Mol. Phylogenet. Evol.* **31**, 1015–1030 (2004).
39. Helgen, K. M., Cole, F. R., Helgen, L. E. & Wilson, D. E. Generic revision in the holarctic ground squirrel genus spermophilus. *J. Mammal.* **90**, 270–305 (2009).
40. Eddingsaas, A. A., Jacobsen, B. K., Lessa, E. P. & Cook, J. A. Evolutionary History of the Arctic Ground Squirrel (Spermophilus parryii) in Nearctic Beringia. *J. Mammal.* **85**, 601–610 (2004).
41. Cocker, S. L. et al. Predation, reoccupation, cannibalism, and scavenging? Records of small mammals in arctic ground squirrel middens from east Beringia. *Arct. Antarct. Alp. Res.* **56**, (2024).
42. Mclean, B. S. & Org, W. M. Urocitellus parryii (Rodentia: Sciuridae). *Mamm. Species* **50**, 84–99 (2018).
43. Buck, C. L. & Barnes, B. M. Temperatures of Hibernacula and Changes in Body Composition of Arctic Ground Squirrels over Winter. *J. Mammal.* **80**, 1264–1276 (1999).
44. Hobbie, E. A. et al. Stable isotopes and radiocarbon assess variable importance of plants and fungi in diets of arctic ground squirrels. *Arct. Antarct. Alp. Res.* **49**, 487–500 (2017).
45. Boonstra, R., Krebs, C. J. & Kanter, M. Arctic ground squirrel predation on collared lemmings. *Can. J. Zool.* **68**, 757–760 (1990).
46. Schuette, P., Ebbert, S., Droghini, A. & Nawrocki, T. Small mammal diet indicates plant diversity, vegetation structure, and ecological integrity in a remote ecosystem. *Biodivers. Conserv.* **31**, 909–924 (2022).
47. O'Donoghue, M. Early Survival of Juvenile Snowshoe Hares. *Ecology* **75**, 1582–1592 (1994).
48. Geist, O. W. Habits of the Ground Squirrel Citellus Lyratus on St. Lawrence Island, Alaska. *J. Mammal.* **14**, 306–308 (1933).
49. Cade, T. Carnivorous Ground Squirrels on St. Lawrence Island, Alaska. *J. Mammal.* **32**, 358–360 (1951).
50. McLean, I. G. Paternal behaviour and killing of young in Arctic ground squirrels. *Anim. Behav.* **31**, 32–44 (1983).
51. McLean, I. G. Seasonal patterns and sexual differences in the feeding ecology of arctic ground squirrels (Spermophilus parryii plesius). *Can. J. Zool.* **63**, 1298–1301 (1985).
52. Ebbert, S. E. & Byrd, G. V. Eradications of invasive species to restore natural biological diversity on Alaska Maritime National Wildlife Refuge. in *Turning the Tide: The Eradication of Invasive Species. Proceedings of the International Conference On Eradication of Island Invasives* (eds Veitch, C. R. & Clout, M. N.) 102–109 (IUCN, Gland, Switzerland and Cambridge, UK, 2002).
53. Smith, J. E. et al. Vole hunting: Novel predatory and carnivorous behavior by California ground squirrels. *J. Ethol.* **43**, 3–12 (2024).
54. Barnes, B. M. Freeze Avoidance in a Mammal: Body Temperatures Below 0 °C in an Arctic Hibernator. *Sci. (1979).* **244**, 1593–1595 (1989).
55. Murchie, T. J., Giguet-Covex, C., Heintzman, P. D., Slon, V. & Wang, Y. Terrestrial Fauna and Hominin DNA from Sedimentary Archives. in *Tracking Environmental Change Using Lake Sediments. Developments in Paleoenvironmental Research, vol 21*. (eds. Capo, E., Barouillet, C. & Smol, J.) 299–378 https://doi.org/10.1007/978-3-031-43799-1_11. (Springer, Cham, 2023).
56. Murchie, T. J., Long, G. S., Lanoil, B. D., Froese, D. & Poinar, H. N. Permafrost microbial communities follow shifts in vegetation, soils, and megafauna extinctions in Late Pleistocene NW North America. *Environ. DNA* **5**, 1759–1779 (2023).
57. Murchie, T. J. et al. Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost sediments. *Curr. Biol.* **32**, 851–860.e7 (2022).
58. Murchie, T. J. et al. Collapse of the mammoth-steppe in central Yukon as revealed by ancient environmental DNA. *Nat. Commun. 2021 12:1* **12**, 1–18 (2021).
59. Monteath, A. J. et al. Relict permafrost preserves megafauna, insects, pollen, soils and pore-ice isotopes of the mammoth steppe and its collapse in central Yukon. *Quat. Sci. Rev.* **299**, 107878 (2023).
60. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Sci. (1979).* **356**, 605–608 (2017).
61. Mamanova, L. et al. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **7**, 111–118 (2010).
62. Carpenter, M. L. et al. Pulling out the 1%: Whole-Genome capture for the targeted enrichment of ancient dna sequencing libraries. *Am. J. Hum. Genet.* **93**, 852–864 (2013).
63. Froese, D. G., Westgate, J. A., Reyes, A. V., Enkin, R. J. & Preece, S. J. Ancient permafrost and a future, warmer arctic. *Sci. (1979).* **321**, 1648 (2008).
64. Westgate, J. A., Preece, S. J. & Jackson, L. E. Revision of the tephrostratigraphy of the lower Sixtymile River area, Yukon Territory, Canada. *Can. J. Earth Sci.* **48**, 695–701 (2011).

65. Buryak, S. D. et al. Laser-ablation ICP-MS zircon U-Pb ages for key Pliocene-Pleistocene tephra beds in unglaciated Yukon and Alaska. *Quat. Geochronol.* **73**, 101398 (2022).
66. Harington, C. R. Pleistocene vertebrates of the Yukon Territory. *Quat. Sci. Rev.* **30**, 2341–2354 (2011).
67. Clark, P. U. The Last Glacial Maximum. *Sci. (1979).* **325**, 710–714 (2009).
68. Stuart, A. J. Late Quaternary megafaunal extinctions on the continents: a short review. *Geol. J.* **50**, 414–433 (2015).
69. Guthrie, D. R. Mammals of the mammoth steppe as paleoenvironmental indicators. in *Paleoecology of Beringia* (eds. Hopkins, D. M., Matthews, J. V. Jr., Schweger, C. E. & Young, S. B.) 307–326 (Academic Press, New York, 1982).
70. Guthrie, R. D. *Frozen Fauna of the Mammoth Steppe: The Story of Blue Babe*. (University of Chicago Press, Chicago, 1990).
71. Willerslev, E. et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).
72. Wang, Y. et al. Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nat. 2021 600:7887* **600**, 86–92 (2021).
73. Davies, L. J., Jensen, B. J. L., Froese, D. G. & Wallace, K. L. 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 (2016).
74. Demuro, M. et al. Optically stimulated luminescence dating of single and multiple grains of quartz from perennially frozen loess in western Yukon Territory, Canada: Comparison with radiocarbon chronologies for the late Pleistocene Dawson tephra. *Quat. Geochronol.* **3**, 346–364 (2008).
75. Froese, D. G., Zazula, G. D. & Reyes, A. V. Seasonality of the late Pleistocene Dawson tephra and exceptional preservation of a buried riparian surface in central Yukon Territory, Canada. *Quat. Sci. Rev.* **25**, 1542–1551 (2006).
76. Froese, D., Westgate, J., Preece, S. & Storer, J. Age and significance of the Late Pleistocene Dawson tephra in eastern Beringia. *Quat. Sci. Rev.* **21**, 2137–2142 (2002).
77. Demuro, M., Arnold, L. J., Froese, D. G. & Roberts, R. G. OSL dating of loess deposits bracketing Sheep Creek tephra beds, northwest Canada: Dim and problematic single-grain OSL characteristics and their effect on multi-grain age estimates. *Quat. Geochronol.* **15**, 67–87 (2013).
78. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
79. Huson, D. H. et al. MEGAN community edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).
80. Camacho, C. et al. BLAST + : Architecture and applications. *BMC Bioinforma.* **10**, 1–9 (2009).
81. NCBI Resource Coordinators Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, 8–13 (2018).
82. Renaud, G., Hanghøj, K., Willerslev, E. & Orlando, L. gargammel: a sequence simulator for ancient DNA. *Bioinformatics* **33**, 577–579 (2017).
83. Cocker, S. L. et al. Pleistocene grasshoppers, fleas, thrips, and mites: rare and new records from Arctic ground squirrel middens in east Beringia in Yukon Territory, Canada. *Can. Entomol.* **157**, e21 (2025).
84. McLean, B. S., Jackson, D. J. & Cook, J. A. Rapid divergence and gene flow at high latitudes shape the history of Holarctic ground squirrels (Urocitellus). *Mol. Phylogenet. Evol.* **102**, 174–188 (2016).
85. Orlando, L. et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nat. 2013 499:7456* **499**, 74–78 (2013).
86. Zielinsk, G. A. & Mershon, G. R. Paleoenvironmental implications of the insoluble microparticle record in the GISP2 (Greenland) ice core during the rapidly changing climate of the Pleistocene-Holocene transition. *Bull. Geol. Soc. Am.* **109**, 547–559 (1997).
87. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
88. Rowe, A. G. et al. A female woolly mammoth's lifetime movements end in an ancient Alaskan hunter-gatherer camp. *Sci. Adv.* **10**, eadk0818 (2024).
89. Enk, J. et al. Mammuthus population dynamics in Late Pleistocene North America: Divergence, Phylogeogrpaphy and Introgression. *Front. Ecol. Evol.* **4**, 1–13 (2016).
90. Murchie, T. J. Ancient environmental DNA as a means of understanding ecological restructuring during the Pleistocene-Holocene transition in Yukon, Canada. (McMaster University, 2021).
91. Tukhbatullin, A. et al. Surrounded by Kindred: Spermophilus major Hybridization with Other Spermophilus Species in Space and Time. *Biol. (Basel)* **12**, 880 (2023).
92. Zimmerman, E. G. & Cothran, E. G. Hybridization in the mexican and 13-lined ground squirrels, Spermophilus mexicanus and Spermophilus tridecemlineatus. *Experientia* **32**, 704–706 (1976).
93. Cothran, E. G. & Honeycutt, R. L. Chromosomal differentiation of hybridizing ground squirrels (Spermophilus mexicanus and S. tridecemlineatus). *J. Mammal.* **65**, 118–122 (1984).
94. Spiridonova, L. N. et al. RAPD-PCR Analysis of Ground Squirrels from the Tobol-Ishim Interfluve: Evidence for Interspecific Hybridization between Ground Squirrel Species Spermophilus major and S. erythrogenys. *Russ. J. Genet.* **41**, 991–1001 (2005).
95. Cothran, E. G. Morphologic Relationships of the Hybridizing Ground Squirrels Spermophilus mexicanus and S. tridecemlineatus. *J. Mammal.* **64**, 591–602 (1983).
96. Thompson, C. W., Stangl, F. B. & Bradley, R. D. Ancient Hybridization And Subsequent Mitochondrial capture in Ground Squirrels (Genus Ictidomys). *Occasional Papers, Museum of Texas Tech University* (2015).
97. Giboulet, O., Chevret, P., Ramousse, R. & Catzeflis, F. DNA–DNA hybridization evidence for the recent origin of marmots and ground squirrels (Rodentia: Sciuridae). *J. Mamm. Evol.* **4**, 271–284 (1997).
98. Tsvirka, M. V., Chelomina, G. N. & Korablev, V. P. Genetic evidence of hybridization between paletailed Spermophilus pallidicauda Satunin, 1903 and alashanic S. alaschanicus Büchner, 1888 ground squirrels in Mongolia. *Russ. J. Genet.* **42**, 421–428 (2006).
99. Staff, T. P. O. Correction: Implications of Hybridization, NUMTs, and Overlooked Diversity for DNA Barcoding of Eurasian Ground Squirrels. *PLoS One* **10**, e0120631 (2015).
100. Ermakov, O. A. et al. A Molecular Genetic Study of Hybridization in Four Species of Ground Squirrels (Spermophilus: Rodentia, Sciuridae). *Russ. J. Genet.* **38**, 796–809 (2002).
101. Dalén, L., Heintzman, P. D., Kapp, J. D. & Shapiro, B. Deep-time paleogenomics and the limits of DNA survival. *Science* **382**, 48–53 (2023).
102. van der Valk, T. et al. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).
103. Finlayson-Trick, E. C. L. et al. Taxonomic differences of gut microbiomes drive cellulolytic enzymatic potential within hindgut fermenting mammals. *PLoS One* **12**, e0189404 (2017).
104. McGee, S. G. Helminth parasites of squirrels (Sciuridae) in Saskatchewan. *Can. J. Zool.* **58**, 2040–2050 (1980).
105. Sapp, S. G. H. et al. Beyond the raccoon roundworm: The natural history of non-raccoon Baylisascaris species in the New World. *Int. J. Parasitol. Parasites Wildl.* **6**, 85–99 (2017).
106. Rausch, R. L. Rediscription of Diandrya composita Darrah, 1930 (Cestoda: Anoplocephalidae) from nearctic marmots (Rodentia:

Sciuridae) and the relationships of the genus Diandrya emend. *Proc. Helminthol. Soc. Wash.* **47**, 157–164 (1980).
107. Barrera, M. A., Janes, J. K. & Gorrell, J. C. Molecular phylogenetics and systematics of two enteric helminth parasites (Baylisascaris laevis and Diandrya vancouverensis) in the Vancouver Island marmot (Marmota vancouverensis). *Int. J. Parasitol. Parasites Wildl.* **19**, 301–310 (2022).
108. Seville, R. S., Oliver, C. E., Lynch, A. J., Bryant, M. C. & Duszynski, D. W. Eimeria species (Apicomplexa: Eimeriidae) from arctic ground squirrels (Spermophilus parryii) and red squirrels (Tamiasciurus hudsonicus) in Alaska and in Siberia, Russia. *J. Parasitol.* **91**, 857–862 (2005).
109. Grant, W. N. et al. Parastrongyloides trichosuri, a nematode parasite of mammals that is uniquely suited to genetic analysis. *Int. J. Parasitol.* **36**, 453–466 (2006).
110. Morgan, D. O. Parastrongyloides winchesi gen. et sp. nov. A Remarkable New Nematode Parasite of the Mole and the Shrew. *J. Helminthol.* **6**, 79–86 (1928).
111. Yokohata, Y., Abe, H., Jiang, V. P. & Kamiya, M. Gastrointestinal helminth fauna of Japanese moles Mogera spp. *Jpn. J. Vet. Res.* **37**, 1–13 (1989).
112. Vaucher, C. & Durette-Desset, M. C. Nouvelles données sur les Helminthes parasites de la Musaraigne Blarina brevicauda (Say). *Rev. suisse de. zoologie* **85**, 361–378 (1978).
113. Wood, J. R. et al. Ancient parasite DNA from late Quaternary Atacama Desert rodent middens. *Quat. Sci. Rev.* **226**, 106031 (2019).
114. Ledger, M. L. et al. Sedimentary ancient DNA as part of a multimethod paleoparasitology approach reveals temporal trends in human parasitic burden in the Roman period. *PLoS Negl. Trop. Dis.* **19**, e0013135 (2025).
115. Barnett, R. et al. Evolution of the extinct Sabretooths and the American cheetah-like cat [2]. *Curr. Biol.* **15**, R589–R590 (2005).
116. Culver, M., Johnson, W. E., Pecon-Slattery, J. & O'Brien, S. J. Genomic ancestry of the American puma (Puma concolor). *J. Heredity* **91**, 186–197 (2000).
117. Vershinina, A. O. et al. Ancient horse genomes reveal the timing and extent of dispersals across the Bering Land Bridge. *Mol. Ecol.* **30**, 6144–6161 (2021).
118. Kefena, E. et al. Discordances between morphological systematics and molecular taxonomy in the stem line of equids: A review of the case of taxonomy of genus Equus. *Livest. Sci.* **143**, 105–115 (2012).
119. Barrón-Ortiz, C. I. et al. What is Equus? Reconciling taxonomy and phylogenetic analyses. *Front. Ecol. Evol.* **7**, 458879 (2019).
120. Barrón-Ortiz, C. I. et al. Cheek tooth morphology and ancient mitochondrial DNA of late Pleistocene horses from the western interior of North America: Implications for the taxonomy of North American Late Pleistocene Equus. *PLoS One* **12**, e0183045 (2017).
121. Flack, N. et al. The genome of Przewalski's horse (Equus ferus przewalskii). *G3 Genes|Genomes|Genetics* **14**, (2024).
122. Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, gkx033 (2017).
123. Gansauge, M. T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
124. Kapp, J. D., Green, R. E. & Shapiro, B. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *J. Heredity* **112**, 241–249 (2021).
125. Bronk Ramsey, C. Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* **51**, 337–360 (2009).
126. Reimer, P. J. et al. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0-55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
127. Jensen, B. J. L., Beaudoin, A. B., Clynne, M. A., Harvey, J. & Vallance, J. W. A re-examination of the three most prominent Holocene tephra deposits in western Canada: Bridge River, Mount St. Helens Yn and Mazama. *Quat. Int.* **500**, 83–95 (2019).
128. Fischer, M. et al. Efficacy assessment of nucleic acid decontamination reagents used in molecular diagnostic laboratories. *PLoS One* **11**, e0159274 (2016).
129. Champlot, S. et al. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* **5**, e13042 (2010).
130. Massilani, D. et al. Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. *Proc. Natl. Acad. Sci. USA.* **119**, e2113666118 (2022).
131. Emery, M. V. et al. Targeted enrichment of whole-genome SNPs from highly burned skeletal remains. *J. Forensic Sci.* **69**, 1558–1577 (2024).
132. Emmons, A. L., Davoren, J., DeBruyn, J. M. & Mundorff, A. Z. Inter and intra-individual variation in skeletal DNA preservation in buried remains. *Forensic Sci. Int. Genet.* **44**, 102193 (2020).
133. Haarkötter, C. et al. A comparison between petrous bone and tooth, femur and tibia DNA analysis from degraded skeletal remains. *Electrophoresis* **44**, 1559–1568 (2023).
134. Hansen, H. B. et al. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS One* **12**, e0170940 (2017).
135. Sarhan, M. S. et al. Ancient DNA diffuses from human bones to cave stones. *iScience* **24**, 103397 (2021).
136. Campos, P. F. et al. DNA in ancient bone – Where is it located and how should we extract it? *Ann. Anat. - Anatomischer Anz.* **194**, 7–16 (2012).
137. Pinhasi, R. et al. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* **10**, e0129102 (2015).
138. Wang, Y., Wessels, M., Pedersen, M. W. & Epp, L. S. Spatial distribution of sedimentary DNA is taxon-specific and linked to local occurrence at intra-lake scale. *Commun. Earth Environ. 2023 4:1* **4**, 1–13 (2023).
139. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.* **110**, 15758–15763 (2013).
140. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **5**, pdb.prot5448 (2010).
141. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, 1–8 (2012).
142. Psonis, N., Vassou, D. & Kafetzopoulos, D. Testing a series of modifications on genomic library preparation methods for ancient or degraded DNA. *Anal. Biochem.* **623**, 114193 (2021).
143. Ashkenazy, H. et al. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584 (2012).
144. Vogel, N. A. et al. soibean: High-Resolution Taxonomic Identification of Ancient Environmental DNA Using Mitochondrial Pangenome Graphs. *Mol. Biol. Evol.* **41**, msae203 (2024).
145. Klunk, J. et al. 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 (2019).
146. Agarwala, R. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
147. Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **41**, (2013).
148. Shah, N., Nute, M. G., Warnow, T. & Pop, M. Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics* **35**, 1613–1614 (2019).
149. Renaud, G., Stenzel, U. & Kelso, J. LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).

150. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
151. Kjær, K. H. et al. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nat. 2022 612:7939* **612**, 283–291 (2022).
152. McLaren, D. et al. Late Pleistocene Faunal Assemblages from Karst Cave Settings on Northern Vancouver Island, Canada. *PaleoAmerica* **9**, 216–236 (2023).
153. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* **22**, 549–556 (2012).
154. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
155. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
156. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
157. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
158. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
159. Michelsen, C. et al. metaDMG – A Fast and Accurate Ancient DNA Damage Toolkit for Metagenomic Data. *bioRxiv* https://doi.org/10.1101/2022.12.06.519264 (2022).
160. Bankevich, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
161. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
162. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
163. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
164. Vogel, N. A. et al. euka: Robust tetrapodic and arthropodic taxa detection from modern and ancient environmental DNA using pangenomic reference graphs. *Methods Ecol. Evol.* **14**, 2717–2727 (2023).
165. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
166. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *Gateway Computing Environments Workshop, (GCE)* 1–8 https://doi.org/10.1109/GCE.2010.5676129. (New Orleans, L.A., 2010).
167. Hasegawa, M., Kishino, H. & Yano, T. aki. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
168. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
169. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
170. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
171. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
172. Mitchell, D., Willerslev, E. & Hansen, A. Damage and repair of ancient DNA. *Mutat. Res.* **571**, 265–276 (2005).
173. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104**, 14616–14621 (2007).
174. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered from highly degraded samples - A case study on DNA in faeces. *Front. Zool.* **3**, 11 (2006).
175. Corinaldesi, C., Beolchini, F. & Dell'Anno a. Damage and degradation rates of extracellular DNA in marine sediments: Implications for the preservation of gene sequences. *Mol. Ecol.* **17**, 3939–3951 (2008).
176. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
177. Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153–2155 (2011).
178. Kistler, L., Ware, R., Smith, O., Collins, M. & Allaby, R. G. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* **45**, 6310–6320 (2017).
179. Grootes, P. M. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10-3- to 105-year time resolution. *J. Geophys. Res. Oceans* **102**, 26455–26470 (1997).
180. Fellows Yates, J. A. et al. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* **9**, e10947 (2021).
181. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* https://www.R-project.org/ (2024).
182. Kans, J. Entrez Direct: E-utilities on the Unix Command Line. *Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for Biotechnology Information* https://www.ncbi.nlm.nih.gov/books/NBK179288/ (2013).
183. Peltzer, A. et al. EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
184. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962 (2016).
185. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
186. Herbig, A. et al. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* https://doi.org/10.1101/050559 (2016).
187. Pochon, Z. et al. aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* **24**, 1–30 (2023).
188. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 1–10 (2018).
189. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
190. Cody, W. J. *Flora of the Yukon Territory*. (National Research Press, Ottawa, Ontario, 2000).
191. Flora of North America Editorial Committee. Flora of North America North of Mexico [Online]. 25+ vols. http://beta.floranorthamerica.org (2025).
192. Kersahw, L. & Allen, L. *Vascular Flora of Alberta: An Illustrated Guide*. (Amazon Digital Services LLC, 2020).
193. Dyke, A. S. An outline of the deglaciation of North America with emphasis on central and northern Canada. *Dev. Quat. Sci.* **2**, 373–424 (2004).
194. Clark, P. U. & Mix, A. C. Ice sheets and sea level of the Last Glacial Maximum. *Quat. Sci. Rev.* **21**, 1–7 (2002).

195. Bond, J. D. *Paleodrainage Map of Beringia, Yukon Geological Survey, Open File 2019-2*. https://data.geology.gov.yk.ca/publication/81642 (2019).
196. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18 O records. *Paleoceanography* **20**, 1–17 (2005).
197. Railsback, L. B. et al. Marine isotope stages and substages. *Some Fundamentals of Mineralogy and Geochemistry* http://railsback.org/Fundamentals/SFMGSubstages01.pdf (2015).

Acknowledgements

Thank you to the Tr'ondëk Hwëch'in First Nation and the placer gold mining community of the Klondike for their continued support of our research and access to study sites in central Yukon. Thanks to Brian Golding for providing access to his computing cluster at McMaster University, and thanks to the University of Victoria and University of British Columbia for access to their computing resources. Thanks to the CANA foundation for funding TJM during his postdoctoral work at McMaster University. Thanks to all members and affiliates of the Hakai Institute and Tula Foundation for their financial, facility, and computational support—with special thanks to Eric Peterson and Christina Munck. Thanks to Simon Goring at Neotoma for his help with pulling species lists for the Pacific Northwest reference panel. Thanks to the staff at the Farncombe Metagenomics Facility for their assistance with sequencing at McMaster University, and Rute Clemente-Carvalho along with Biodiversity Genomics support staff for sequencing at the Hakai Institute. Thanks to Richard Caners (Curator of Botany, Royal Alberta Museum) for assistance with bryophyte identifications, as well as Svetlana Kuzmina for insect fragment identification assistance, and Kelly Rozanitis (Grad Research Asst Fellowship, Earth & Atmospheric Sciences, University of Alberta) for assistance with HF processing for the pollen samples. This work was funded by the Tula Foundation, CANA Foundation, Belmont Forum, NSERC Alliance (to HNP and TJM), BiodivERsA grants (to DGF and HNP for the Future ArcTic Ecosystems [FATE] research consortium), NSERC Discovery grants (to DGF and HNP), and University of Alberta Northern Research Awards (to SLC).

Author contributions

Conceptualization: T.J.M., S.L.C., D.G.F., H.N.P. Methodology: T.J.M., S.L.C., S.B., N.A.V., L.N., E.K., D.T., M.A.B., D.M.G., E.M., G.S.L., G.D.Z., B.J.J., D.G.F. Software: T.J.M., L.N., S.B., N.A.V., E.M. Validation: T.J.M., S.L.C., L.N., S.B., N.A.V. Formal analysis: T.J.M., S.L.C., L.N., S.B., N.A.V., D.T. Investigation: T.J.M., S.L.C., L.N., M.A.B., D.M.G., S.B., N.A.V. Resources: L.Y.R., D.G.F., H.N.P. Data curation: T.J.M., L.N., S.B., N.A.V.. Writing: T.J.M., all. Writing—Review: All. Visualization: T.J.M., S.L.C., L.N., D.M.G. Supervision: L.Y.R., G.D.Z., D.G.F., H.N.P. Project administration: D.G.F., H.N.P., T.J.M. Funding: D.G.F., H.N.P., L.Y.R., S.L.C., T.J.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-72977-6.

Correspondence and requests for materials should be addressed to Tyler J. Murchie, Duane G. Froese or Hendrik N. Poinar.

Peer review information : *Nature Communications* thanks Mikkel Pedersen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.