



Metagenomic insights into the diet and past Kolyma Lowland habitat from the intestinal content of a Late Pleistocene steppe bison

Chenyu Jin ^{a,b,c,*}, Adrian Forsythe ^d, Miklós Bálint ^{e,f,g}, Frank Kienast ^h, Pavel A. Nikolskiy ⁱ, Rafał Kowalczyk ^j, Katerina Guschanski ^{k,l}, Tom van der Valk ^{a,b,m,**}

^a Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

^b Centre for Palaeogenetics, Stockholm, Sweden

^c Department of Zoology, Stockholm University, Stockholm, Sweden

^d Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

^e Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325, Frankfurt am Main, Germany

^f Department of Insect Biotechnology, Justus-Liebig University, Heinrich-Buff-Ring 26-32, 35392, Gießen, Germany

^g LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325, Frankfurt am Main, Germany

^h Senckenberg Research Institute and Natural History Museum, Research Station for Quaternary Palaeontology Weimar, Germany

ⁱ Geological Institute, Russian Academy of Sciences, Moscow, Russia

^j Mammal Research Institute, Polish Academy of Sciences, Białowieża, Poland

^k Institute of Ecology and Evolution, School of Biological Sciences, The University of Edinburgh, Edinburgh, United Kingdom

^l Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

^m Scilifelab, Stockholm, Sweden

ARTICLE INFO

Handling Editor: Dr Yan Zhao

Keywords:

Ancient DNA

Palaeogenetics

Bison priscus

Siberia

Flora

Mammoth steppe

ABSTRACT

During the last cold stage (115–11.7 cal kya BP), the non-glaciated part of Arctic Eurasia exhibited a distinctive plant composition consisting of steppe and xerophilous arctic plants. This vegetation formed the Pleistocene mammoth steppe, which provided a habitat for large herbivores like mammoths (*Mammuthus primigenius*) and steppe bison (*Bison priscus*). In this study, we analyzed the DNA sequences of well-preserved plant material from the intestinal tract of the frozen mummy of an extinct steppe bison discovered in the Kolyma Lowland, Northeastern Russia in 2009. Radiocarbon dating placed the specimen at >48 kyr. We leveraged genome-wide and skinned genome accessions to maximize the retrieval of sequenced reads for identification. In line with macrofossil evidence, we identified alkali grass (*Puccinellia*), indicating high salinity, and Larch (*Larix*) in the DNA data of the steppe bison sample. The occurrence of larch north of the current tree line indicates warm temperatures during the growing season. Additionally, we confirmed the presence of plant genera typically associated with steppes and productive meadows under open and arid soil conditions such as *Artemisia*, *Tanacetum*, *Koeleria*, *Festuca*, *Puccinellia*, *Taraxacum*, *Alopecurus*, and *Poa*, while *Phippisia* suggests high soil water content, at least locally. We also attempted to infer the likely most closely related species for each genus. The most abundant authenticated genera, *Artemisia* and *Puccinellia*, align with their known prevalence in the mammoth steppe, reflecting broader environmental signals beyond the steppe bison's diet. Hence, the use of metagenomic approaches in combination with comprehensive reference databases offered insights into steppe bison diet, allowed us to obtain a more detailed view of the mammoth steppe vegetation, and provided extended interpretations of soil conditions.

1. Introduction

In 2009, a remarkably well-preserved, frozen carcass of a steppe bison (*Bison priscus*) was discovered at the Alazeya river bank of the

Kolyma Lowland in Northeastern Russia (Fig. 1), also known as Alazeya bison (Kienast and Nikolskiy, 2023). Based on the stratigraphic context of the carcass, this bison individual lived during Marine Isotope Stage 3 (MIS 3, 57–29 cal kyr BP) (Lisiecki and Raymo, 2005; Kienast and

* Corresponding author. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

** Corresponding author. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

E-mail address: chenyu.jin@nrm.se (C. Jin).

Nikolskiy, 2023). Radiocarbon dating of its bone and skin tissues yielded age estimates of 48 kyr, which is out of the reliable detection range of radiocarbon dating and the specimen is thus estimated to be 57–48 cal kyr BP. The exceptional preservation kept the intestinal content of the individual intact, which offers the potential to recover ancient dietary DNA from the plant species consumed in the final meals of this steppe bison. By sequencing and analyzing this DNA, we can not only gain valuable insights into the diet of the steppe bison but also improved understanding of the plant communities existing in northeastern Russia during MIS 3. During the last ice age, encompassing MIS 3, the now lost mammoth steppe biome was characterized by the coexistence of steppe plants and xerophilous (aridity-adapted) arctic plants, resembling vegetation complexes described as glacial relicts in central and northeast Yakutia (Yurtsev, 1982, 2001). This combination of plant communities was also revealed in plant macrofossil studies on permafrost deposits at sites in the non-glaciated sector of the Arctic spanning from the New Siberian Archipelago to the Taymyr Peninsula in Beringia (Kienast et al., 2005; Kienast, 2013; Ashastina et al., 2018).

Permafrost and lake sediment samples from diverse regions in the Arctic, covering the last 50 thousand years, indicate a relatively uniform steppe tundra ecosystem (Wang et al., 2021). This study also showed that metagenomics can uncover plant and animal diversity, including the detection of extinct species (Wang et al., 2021). In contrast to sediment samples reflecting preservation or sampling variability within the same lake (Wang et al., 2023), the intestinal content of a steppe bison, a mobile herbivore, could provide a snapshot of plant communities not bound to a particular catchment. As one of the most characteristic megafauna species inhabiting the northern hemisphere during

the mid- and late Pleistocene, steppe bison (*Bison priscus*) survived till the early Holocene (10.7–8.2 cal kyr BP) (MacPhee et al., 2002; Boeskorov et al., 2016). While the dietary preferences of steppe bison limit the sampling scope, their mobility allowed them to interact with a wide range of vegetation types, making them valuable indicators of plant community composition in the mammoth steppe.

Previous metabarcoding ancient DNA dietary studies presented intestinal/stomach and coprolite specimens of woolly mammoths and bison, dated to about 21.8 and over 55 kya respectively, showing that forbs (herbaceous flowering plants) and graminoids (herbaceous plants with a grass-like morphology) were the dominant growth forms during the last cold stage (Willerslev et al., 2014). Steppe bison coprolites from Siberia and Alaska revealed a diverse vegetation type in the diet, ranging from meadow, tundra, snow patches to woods (Willerslev et al., 2014; Polling et al., 2021). Although the metabarcoding methods used on these studies obtained genus-level resolution, a comprehensive metagenomic analysis could potentially offer a more detailed understanding by utilizing all the preserved DNA sequence reads. A metagenomic approach also allows for the incorporation of DNA damage patterns and sequence divergence estimates, which are not captured by barcode regions used in metabarcoding, enabling downstream authentication of the ancient DNA. In this study, we performed metagenomic analyses on ancient DNA sequences obtained from the intestinal remains of the Alazeya bison to achieve two main objectives: (1) to assess the detection resolution of plant taxa and interpretability in their niche using a database of Arctic and general plant genomes, and (2) to obtain insight into the plant composition of the steppe bison diet, complementing prior macrofossil inferences.



Fig. 1. (A) Sampling location of the steppe bison in the Kolyma Lowland in northeast Yakutia, Russia. (B) The microbiological and DNA sampling was conducted in a clean chamber in Moscow in 2013. (C) The specimen following sampling. (D) The profile of sampled gut content analyzed in this study and other bison tissues. All the images of the Alazeya steppe bison above are contributed by Pavel A. Nikolskiy, along with the clarification that the bison was traced to Kolyma Lowland rather than Anyuy River as previously reported (Nikolskiy and Shidlovsky, 2014).

2. Methods

2.1. Sample information and location

The carcass of the Alazeya steppe bison was discovered in permafrost sediments at a steep frozen bank (precise coordinates unknown) of the Alazeya River in the Kolyma Lowland, Yakutia, Russia in 2009 (Kienast and Nikolskiy, 2023). A sample of partly digested material was then removed from the digestive tract of the animal, comprising the intestine and possibly part of the abomasum (Fig. 1C and D). Alazeya steppe bison was dated to Marine Isotope stage 3 (MIS 3, ~60–27 cal kyr BP), a period characterized by stadial–interstadial transitions and fluctuating temperature (Van Meerbeeck et al., 2011). Therefore, the potential time frame of the last meal of the steppe bison (57–48 cal kyr BP) could be during either stadial or interstadial as it spans a few thousands years.

For quality control in authentication, we additionally used the stomach content of six sequenced contemporary European bison (*Bison bonasus*) individuals from the Białowieża Forest, Poland, the closest living relative of the steppe bison. Samples were obtained by planned culling for population control ($n = 3$), freshly deceased individuals with health complications ($n = 2$, lung infection and gastrointestinal complications), and an individual that had died due to an accident (roadkill) in December 2020 to March 2021 (Table S1). It is noteworthy that this bison population was provided with supplementary feed during the winter period (although not all individuals are using the feeders), with individuals also dispersing to agricultural areas, mainly to rape fields, for feeding (Kowalczyk et al., 2011). With a single exception, four samples were collected from each individual along the digestive tract: rumen, reticulum, omasum and abomasum. The abomasum sample was missing in one individual (ID1428). Samples were collected from fresh individuals (as far as possible in case of road kills), frozen at -80°C and shipped to Uppsala University for DNA extraction and library preparation.

2.2. Laboratory methods

Samples of the steppe bison were processed in a dedicated ancient DNA laboratory at Uppsala University. After removing the surface layer to reduce external contamination, we sampled four replicates (BB1–4) from the same frozen bulk material (Table S2). Two subsamples (BB1 and BB2) were extracted following the Dabney protocol frequently used for ancient DNA material (Dabney et al., 2013) with slight modifications as described (Brealey et al., 2021), except that no surface decontamination was performed. The subsamples BB3 and BB4 were extracted using the Qiagen Powersoil kit, following the manufacturer's protocol with some modifications: After the addition of Solution C1 and proteinase K, incubation was performed at 37°C overnight in a rotating incubator, followed by an incubation at 65°C for 1 h and vortexing for 10 min, we used 250 μl of Solution C3, the eluates were further processed through the Zymo-Spin III-HRC Columns to remove potential inhibitors. Samples of the European bison were extracted in a laboratory dedicated to faecal sample processing using the Qiagen Powersoil kit following the manufacturer's protocol with an additional inhibitor removal step with the Zymo-Spin III-HRC Columns, as described above. Following DNA extraction, European bison samples were fragmented with a Covaris E220 sonicator, aiming for fragments of 350 base pairs in length. We included extraction negative controls for both steppe and European bisons. For all DNA extracts, including the negative controls, we built double-stranded double-barcoded and double-indexed sequencing libraries following the protocols (Meyer and Kircher, 2010; van der Valk et al., 2020), also included separate library preparation negative controls. Samples were pooled with other sequencing libraries unrelated to the project and sequenced on the Illumina NovaSeq 6000 platform, 150 PE, with ancient bison samples being re-sequenced across multiple lanes (Table S2).

2.3. Bioinformatic analysis

2.3.1. Kraken 2 classification

We used fastp v0.23.4 (Chen et al., 2018) to merge all the paired-end sequence reads and trim adapter sequences, only keeping merged fragments of at least 30 base pairs. Next, to taxonomically classify sequencing reads, we constructed genome-wide databases for microbes and plants using Kraken 2 (Wood et al., 2019). First, we used all 402,709 available assemblies from the Genome Taxonomy Database (GTDB) v214.0 (Parks et al., 2022), which comprises 80,789 bacterial and 4416 archaeal species clusters to construct a microbial Kraken 2 database using Kraken 2-build on default settings. A second microbial Kraken 2 database was built using all the archaea, viruses, bacteria, protozoa, UniVec and fungi genomes released on March 13th, 2023 from NCBI Refseq v.216 (O'Leary et al., 2016). Last, we prepared a plant database. To do so and enable the classification of sequence data on 1 TB compute nodes, we downsampled the 1427 *Viridiplantae* (green plant) assemblies available in NCBI as of April 2023. To this end, we estimated the pairwise genetic distances between each genome using Mash v2.3 (Ondov et al., 2016). We then downsampled the genomes to maximize total genetic diversity across the broadest possible taxonomic distribution represented in the database. For species with multiple assemblies, we retained only the assembly with the highest contig N50. Next, we constructed a Kraken 2 plant database by including the 1257 NCBI genome accessions from downsampling NCBI assemblies (Kitts et al., 2016), all RefSeq mitochondrial genomes available in April 2023 (O'Leary et al., 2016) and all assemblies from the PhyloNorway database, which contains skimmed genomes of 1323 species (retrieved on April 13, 2023) covering the majority of plants found in Iceland, Fennoscandia, and the Arctic (Alsos et al., 2020). In total, the database comprises 2376 plant species from 960 genera (Table S3).

For sequence data processing, we used a sequential approach. First, we classified the sequence reads against the GTDB microbial database using Kraken 2 and removed reads that could be assigned microbial taxonomy. Sequence reads that did not match the GTDB database, were classified against the microbial Refseq database, and again we only retained the unclassified reads. This was done to only retain the sequencing data for non-microbial reads that are informative about plant communities. The remaining unclassified reads were then classified against the database of plant genomes.

2.3.2. Competitive mapping

Taxonomic classification results from Kraken 2 report the number of unique minimizers (subset of kmers), which are sequence matches between the query sequence data and the reference genomes in the database. In the classification step, we used the Kraken 2 default settings with a minimizer length of 31 base pairs. As plant species with large genomes tend to have a high number of assigned unique minimizers, we controlled for the effect of reference genome by calculating a ratio between the number of unique minimizers and reference genome size after log₁₀ transformation of both values. We then selected 60 species with the highest ratio, retaining a single species per genus belonging to *Viridiplantae* (green plants). We further added the *Larix gmelinii* reference to this set, as this species was previously identified from the macrofossil analysis of this steppe bison sample (Kienast and Nikolskiy, 2023).

This set of 61 plant genomes was used to build a single bowtie2 v2.5.1 index. The merged sequence data of the steppe and all European bison samples, prior to filtering out microbial reads, were then mapped to this index using bowtie2 with the –very-sensitive setting (Langmead and Salzberg, 2012). Unmapped reads were removed by filtering out reads with a mapping quality of 0. Duplicate reads were then removed using samtools markdup v1.17 (Li et al., 2009). We extracted the read names of reads classified to GTDB and microbial Refseq Kraken 2 database and filtered out these reads in the reads mapped by competitive mapping.

One potential concern is that, due to the broad origin of the sequence

data, reads may align to the reference in the competitive mapping index by chance, even when the reference species (or closely related species) are not actually present. To test this, we constructed a second Bowtie index comprising the 25 genomes with the highest number of classified sequences in the steppe-bison data, along with five model species genomes: *Drosophila melanogaster*, *Gallus gallus*, *Oryza sativa*, *Zea mays*, and *Arabidopsis thaliana* (Fig. S1). After aligning all steppe bison reads to this index, we found that virtually no reads mapped to any of the model species, whereas many of the other genomes showed substantial read alignment (Fig. S1). This suggests that random misalignment from highly diverged taxa does not significantly affect downstream genetic patterns.

2.3.3. Authentication based on ancient DNA damage pattern

The sequence reads of European bison are processed in the same manner as those of steppe bison, which serve as the negative control for damage pattern in selecting the tools and strategy for authentication. Using the bowtie-mapped reads, for each species we calculated the post-mortem DNA damage scores (PMD scores) using PMDtools (Skoglund et al., 2014). Simultaneously, the deamination patterns were visualized using mapDamage v2.0.8 (Jónsson et al., 2013). By using visual inspection of the DNA damage patterns, we identified the species that exhibited increased C > T transitions at the 5' end and G > A transitions at the 3' end that exceeded the frequency of all the other substitutions. Additionally, we used the likelihood ratio test based on C > T transitions in Pydamage v0.7 to discriminate truly ancient reads from modern reads, which may originate from sample contamination (Borry et al., 2021). We tested different window lengths (5, 8, 10 bp) apart from the default 13 base pair used in Pydamage to maximize the detection overlapping with mapDamage results but also provide a standardized statistics-based threshold. Species with a likelihood ratio test p-value <0.05 were considered as putatively ancient. To increase the robustness of ancient species identification, we only considered species as authenticated if DNA damage patterns were identified with both tools, mapDamage and PyDamage.

2.3.4. Estimating sequence divergence between reads and references

To infer the most closely related species, we estimated sequence divergence between the reference genomes and the steppe bison reads. We first extracted reads that were competitively mapped to the 61 target species in our mapping plant database. These extracted reads in bam format were then converted to fFASTQ files format using samtools bam2fq, and then mapped to all available NCBI and PhyloNorway species assemblies of other species under the same genus using bowtie2. A consensus fasta sequence of the remapped reads for each species was then generated using angsd –dofasta v0.940, sampling a random read at each mapped genomic position (Korneliussen et al., 2014). To minimize biased estimates of sequence divergence due to ancient DNA damage, this script removes all C-to-T transitions on the forward strand and G-to-A transitions on the reverse strand. Next, we calculated the total number of bases in the consensus sequence as well as the total number of base pair differences between the reference and the consensus sequence using a custom Python3 script, where the sequence divergence was then estimated as the number of mutations divided by the total number of sites covered by mapped reads.

3. Results

3.1. Taxonomic classification results

We extracted DNA from the steppe bison intestine and part of the abomasum and additionally used the intestinal content of six contemporary European bison (*Bison bonasus*) individuals from Poland, the closest living relative of the steppe bison, as control samples. After sequencing, we merged the reads and removed duplicates resulting in 300M (million) sequence reads for the steppe bison and between 153

and 230M reads for the European bison samples (Table S4). To taxonomically classify the sequence data, we built three large Kraken 2 databases, a database for bacteria and archaea using the Genome Taxonomy Database (GTDB) v214.0 (Parks et al., 2022), a database containing all archaea, viruses, bacteria, protozoa, UniVec and fungi genomes of Refseq v.216 (O'Leary et al., 2016), and a database with 2376 plant species from NCBI and Phylonorway (Kitts et al., 2016; Alsos et al., 2020). Next, we filtered out all sequence reads classified to one of the microbial databases and then classified the remaining 7.7M reads (2.55 % of merged reads) for the steppe bison sample and between 4.0 and 7.9M (2.15 %–4.45 % of merged reads) for the European bison samples to taxa in the custom-built green plant (*Viridiplantae*) database (Table S3, Fig. S2). Out of all classified reads of the steppe bison sample, there are 2.8M reads classified to the genus *Artemisia* (0.95 % of all merged reads) and 1.1M reads assigned to *Puccinellia* (0.38 % of all merged reads) (Table S4). These two genera had the highest abundance in the steppe bison sample.

3.2. Competitive mapping results

Based on the classification result of the steppe bison sample, we calculated for each genome in the database the number of assigned unique minimizers normalized by their genome size (see methods). For the 60 most abundant genera in the sample, we then selected one species genome for each genus, from here on referred to as the representative genome and built a mapping index from the concatenated genomes of the selected 60 species. We also added the plant *Larix gmelini* to this index, as this species was previously identified in the steppe bison gut through macrofossil analysis (Kienast and Nikolskiy, 2023). We then competitively mapped all the raw reads of bison samples against this index. For the steppe bison sample, we aligned a total of 2.1M reads (0.69 % of all merged reads) uniquely to one of these 61 reference assemblies. However, a subset of the reads was previously classified as microbial against the GTDB Kraken 2 database, which we filtered out, resulting in 1.7M reads (0.57 % of all merged reads) (Table S6). This suggests that some of the plant reference genomes, most of which were made by skinned genome sequencing of voucher specimens, contain significant microbial sequence contamination in the reference. We also mapped the European bison sequence data to the reference index and aligned between 55 and 300 thousand reads (corresponding to 0.03 %–0.17 % of merged reads) (Table S6). Confirming the classification results, the genera with the most mapped reads in the steppe bison were *Artemisia* and *Puccinellia* accounting for 55.4 % and 32.7 % of all mapped reads, respectively (Fig. 2A). However, for the majority of the 61 species used in the competitive mapping, many more reads were classified using the Kraken2 kmer classification (Fig. 2B), suggesting that the alignments are more conservative. We did not detect any mapped reads to the 61 plant species in either the extraction or library preparation blanks that were sequenced alongside the steppe bison, demonstrating that little to no DNA contamination was introduced during the lab processing of the sample.

3.3. Authentication results for DNA damage and sequence divergence

To validate the species detection, we first estimated ancient DNA damage patterns in the mapped reads using both mapDamage and Pydamage (Jónsson et al., 2013; Borry et al., 2021). We also tested PMDtools (Skoglund et al., 2014) to assess post-mortem damage scores (PMD scores), however, found that the modern European bison samples also showed signs of DNA-damage and we therefore did not use the PMD scores for species authentication (Fig. 2A, Fig. S3). In total, we found 18 representative genomes showing clear signs of ancient DNA damage (Fig. 2C–Table S7). The genus with the lowest number of mapped reads (N = 567) that showed evident DNA damage was *Scalesia* (*S. atractyloides*) (Fig. 2A).

Since all the reference genomes of modern plants have diverged from

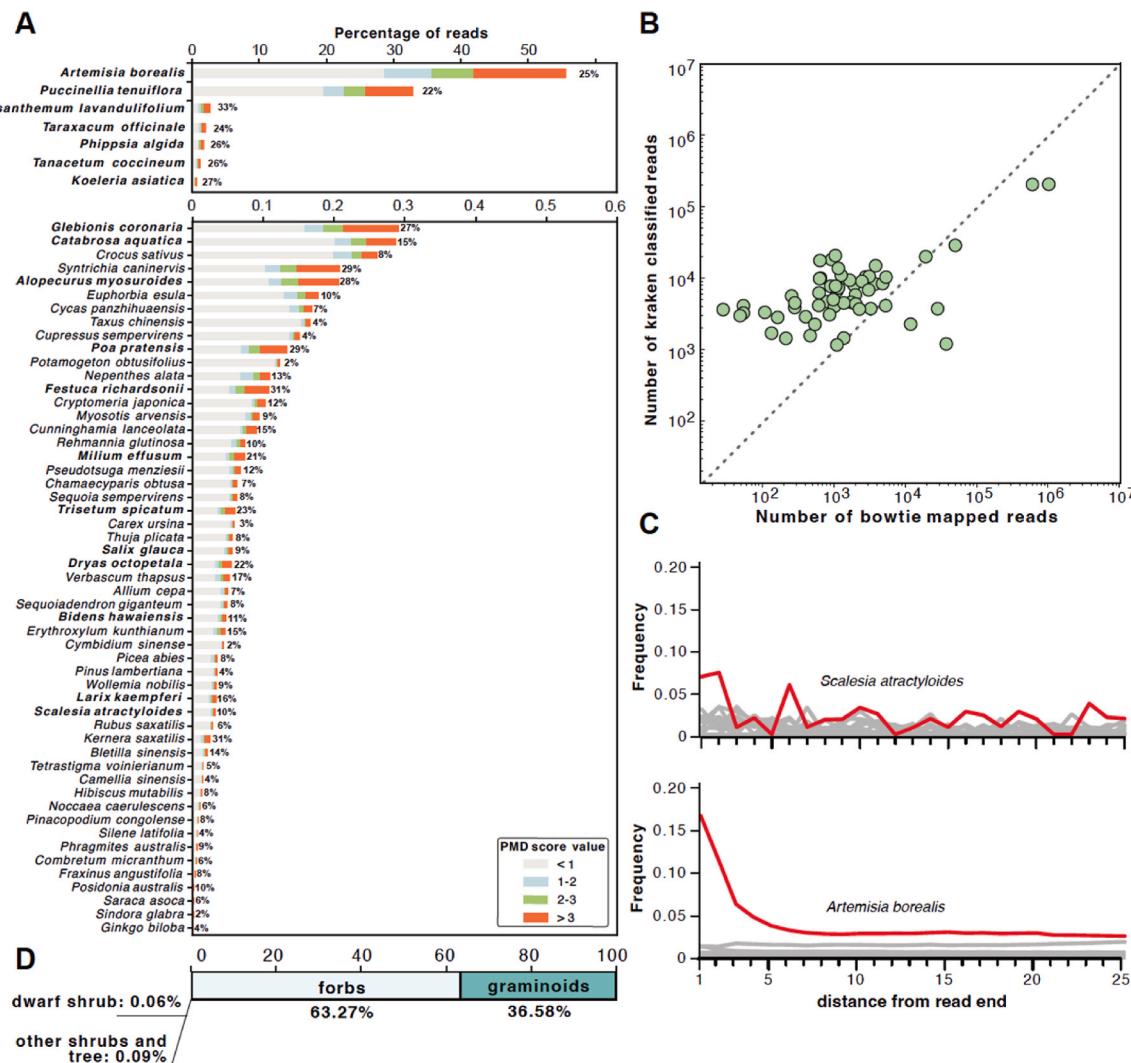


Fig. 2. (A) For the steppe bison sample, the percentages of reads mapped to the 61 species are presented in descending order across two separate panels, each displaying different percentage ranges. The species displaying a clear DNA damage pattern are highlighted in bold. The corresponding PMD score distributions are illustrated using colours, with the percentage of reads with PMD score more than 3 out of the total number of reads mapped to this species shown correspondingly. (B) The number of reads mapped to the 61 species in competitive mapping was compared with those classified to the same species in Kraken 2. (C) The mapDamage plot at the 5' end for two authenticated ancient species from the steppe bison gut with the highest and lowest number of mapped reads, *Artemisia* and *Scalesia* (Table S11), with C-to-T damage pattern denoted in red and the rest of the substitutions in grey in forward strands. (D) The percentage of reads out of total reads mapped to different species types, such as forbs, graminoids, dwarf shrubs, erect shrubs and trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the taxa present in the steppe bison sample by around fifty thousand years, the reads may map to related taxa with varying degrees of divergence. To avoid being constrained by only having one representative species for each genus in competitive mapping, we extracted all the mapped reads per genus, i.e. the reads mapped to the representative genome, and then remapped these reads to all available reference genomes within the same genus. We then calculated the total number of sequence mismatches between the reads and each reference genome, normalized by the total number of sites covered to obtain sequence divergences (Fig. 3, Table S8). Based on these calculations, we infer that the reference with the lowest sequence divergence to the steppe bison data is most likely the closest modern analogue in the database.

4. Discussion

4.1. Vegetation composition in the mammoth steppe of Beringia and steppe bison diet

Plant macrofossil spectra previously detected in Pleistocene permafrost deposits in northern Yakutia have shown that aquatic, littoral pioneer, meadow, and steppe associations coexisted together with high arctic plant communities (Kienast, 2013). Together they formed a patchy, grassland-dominated vegetation mosaic, which constituted Beringia's mammoth steppe (Kienast, 2013; Ashastina et al., 2018). This now-vanished biome with close modern analogues in Yakutia's relict steppe vegetation (Yurtsev, 1982, 2001) was associated with the existence of large herbivores such as woolly mammoth, woolly rhinoceros, musk ox, horse, and steppe bison (Guthrie, 1982). Even though it formed a locally diverse mosaic of plant communities in response to local environmental conditions such as topography, exposition, and soil moisture, it was structurally more or less homogeneous over large areas.

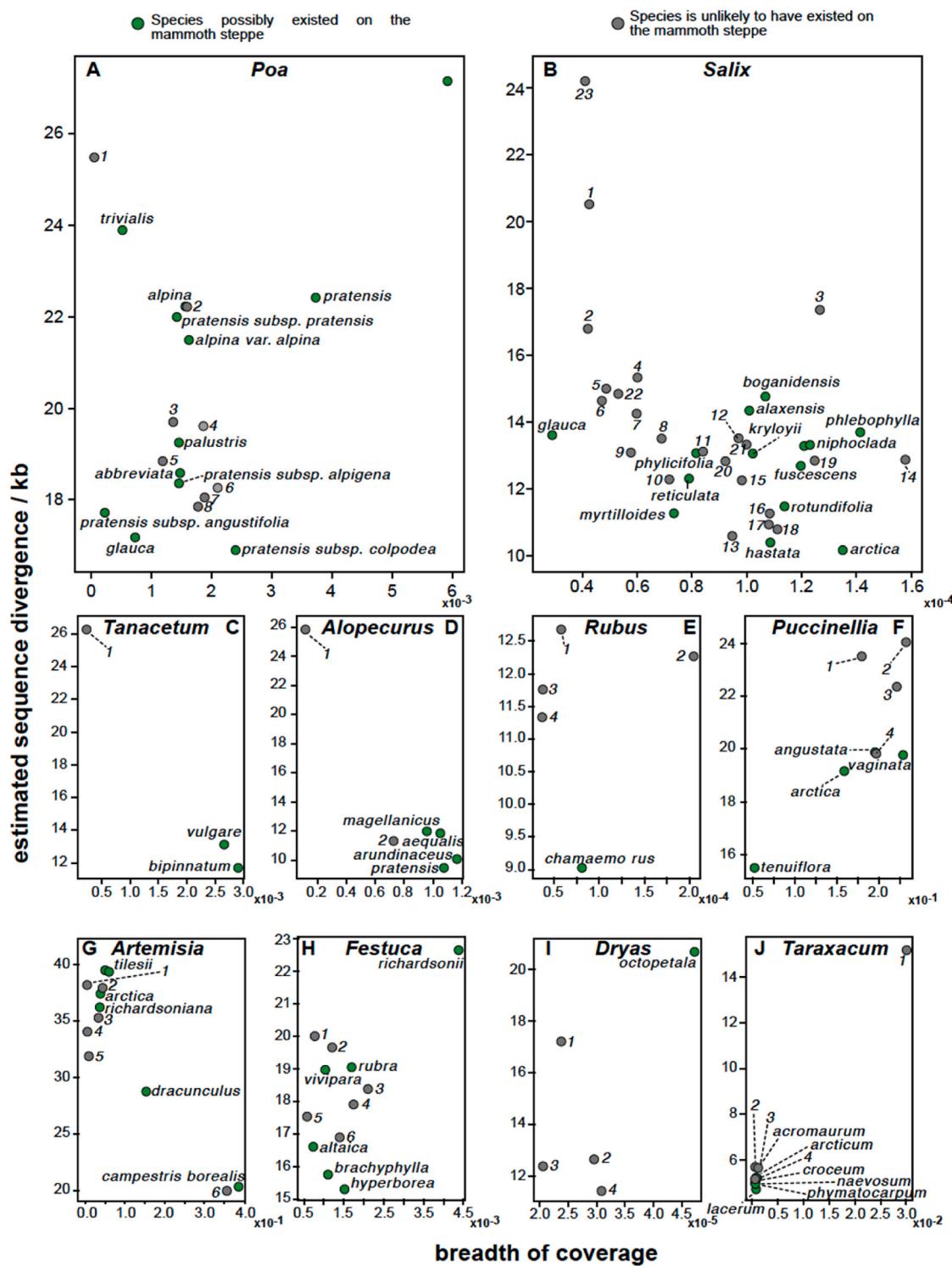


Fig. 3. The sequence divergence (in base pair difference to the reference per thousand base pairs) for the identified species, plotted against their corresponding breadth of coverage. In green are the species that are likely to occur in the mammoth steppe, based on the ecology of their modern analogue species and macro-/microfossil evidence, whereas in grey are species unlikely to have ever existed in this region. Each panel depicts the competitive mapping of a separate genus. (A) 1: annua, 2:bulbosa, 3: chaixii, 4: alpina subsp. vivipara, 5: laxa subsp. flexuosa, 6: pratensis subsp. irrigate, 7: leptocoma, 8: hartzii. (B) 1: arbutifolia, 2: exigua, 3: serissima, 4: udensis, 5: purpurea, 6: suchowensis, 7: viminalis, 8: bangongensis, 9: brachista, 10: discolor, 11: lanata, 12: caprea, 13: arctophila, 14: myrsinifolia, 15: fuscescens, 16: borealis, 17: myrsinites, 18: uva-ursi, 19: caprea, 20: prolixa, 21: pulchra, 22: koriyanagi, 23: dunnii. (C) 1: coccineum. (D) 1: myosuroides, 2: geniculatus. (E) 1: saxatilis, 2: grabowskii, 3: parviflorus, 4: idaeus. (F) 1: nutkaensis, 2: maritima, 3: capillaris, 4: svalbardensis. (G) 1: tridentata, 2: vulgaris, 3: norvegica, 4: argyi, 5: annua, 6: campestris camp. (H) 1: altissima, 2: trachyphylla, 3: prolifera, 4: baffinensis, 5: pratensis, 6: edlundiae. (I) 1: drummondii, 2: integrigolia, 3: ajanensis, 4: alaskensis. (J) 1: officinale, 2: brachyceras, 3: tenebriancis, 4: recedens. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

This landscape homogeneity is not only true for cold stage vegetation but also for interglacials, when the same plant communities continued to exist in Northeast Siberia and were complemented by boreal woodlands with larch, tree or shrub birch both north and south of the modern treeline (Kienast et al., 2008, 2011; Ashastina et al., 2018). Thus, a sharp latitudinal vegetation zonation like tundra, taiga, and steppe, as it is characteristic of the current interglacial (Holocene), presumptively did not exist during earlier warm stages (Sher, 1997).

Based on the ice-rich permafrost sequences in western Lena Delta, Arctic Siberia, pollen and macrofossil records suggest a tundra steppe during MIS 4 and 3 (Schirrmeister et al., 2011). In the layer dated to be around 48 kyr BP (during MIS 3) of the Mamontovy Khayata permafrost sequence from the Bykovsky Peninsula, few taxa were detected in the macrofossil assemblage, which represents steppe and meadow plants, floodplain meadows, littoral zone vegetation, aquatic, and mire communities (Kienast et al., 2005). Our sample, also dated to approximately >48 kyr BP, shows overlap in the detection of Dryas, Puccinellia, and *Salix* genera with those found in the Mamontovy Khayata sequence. The identified species in our sample represents steppe, meadow, littoral and aquatic environments as well. The detection of genera in our study also aligns well with the steppe-like and tundra-like composition and an open landscape with sparse tree or shrub inferred for MIS3 periods later than our sample (57–48 cal kyr BP). For the Mamontovy Khayata sequence layer dated to ~45.3 + 1.2/-1.1 kyr BP, the macrofossils show more species of the plant communities as around 48 kya (Kienast et al., 2005). The macrofossil of the peat sequence dated ~42.4–43.8 thousands years BP at Duvanny Yar, Kolyma Lowland, northeast Siberia starts with poorly-drained soil conditions with dwarf willow-sedge-horsetail vegetation, followed by the domination of Sphagnum mosses, sedges and *Comarum palustre* (Zanina et al., 2011). The larch trunk and pollen records both suggested an open larch forest with willow, alder and dwarf birch, while a polygonal ground and hence plant communities in mosaic is inferred from the association of the peat with the ice wedges (Zanina et al., 2011).

Macrofossil analysis of the Alazeya steppe bison intestinal tract revealed that its diet was composed of alkali grass (*Puccinellia* species) indicating soil salinity under arid climate, as well as *Chenopodium prostratum* and *Chenopodium album*, plants also associated with brackish soil (Kienast and Nikolskiy, 2023). Additionally, a macrofossil identified as *Larix gmelinii* implies the survival of tree taxa during cold stages and hence, confirming that a rapid recolonization of trees occurred from formerly postulated cryptic refugia in Northern Siberia during Pleistocene cold stages (Kienast and Nikolskiy, 2023). However, macrofossil identification is limited to preserved remains with identifiable morphology. DNA-based insights can not only validate but also complement macrofossil findings, as demonstrated in this study.

The diet of another Yakutian steppe bison sample living approximately 10.5 cal kyr BP, as revealed by DNA barcoding, showed 57 % of the reads identified as non-graminoid herbs, including forbs, less than 4 % as graminoids, and 38.6 % as woody taxa (Van Geel et al., 2014). A later metabarcoding analysis of the same bison rumen sample revealed a diet composition of 48 % forbs, 25 % graminoids and 25 % shrubs (Polling et al., 2021). Despite significant differences in the proportion of graminoids detected using different barcode regions, the two studies suggest that the steppe bison was a mixed feeder, rather than an exclusive grazer (Hofman-Kamińska et al., 2024). The European bison has a diet consisting of 59.4 % woody species and 33.6 % forbs, based on chloroplast *trnL* analysis (Kowalczyk et al., 2019). Both European and American bison (*Bison bison*), the next closest relative to steppe bison, exhibit diets influenced by seasonal biomass availability (Bergmann et al., 2015; Kowalczyk et al., 2019). Based on the dietary patterns of its closest relatives, the steppe bison's diet likely reflected both its preferences and the seasonal availability of plants in the Mammoth Steppe.

Out of the mapped reads to genera authenticated with DNA damage pattern and empirical possibility based on the distribution of fossils or modern sample of the reference species, we found that 63.27 % of reads

mapped to forbs, 36.58 % to graminoid, 0.06 % to dwarf shrub, and 0.09 % to other shrubs and tree (Fig. 2D). The large proportion of reads mapping to forbs and graminoids matches the previous study based on metabarcoding from the Yakutian steppe bison (Van Geel et al., 2014). The low number of reads mapping to woody species in our sample could be a result of taphonomy effects on DNA preservation, the dietary preferences of the steppe bison, or the limitations of our database, which included only 60 genera (61 species as with 2 *Larix* species) for competitive mapping.

4.2. Bioinformatic cautions and limitations

Our metagenomic DNA analysis of the intestinal content revealed additional insights in mammoth steppe plant communities and the diet of the resident steppe bison. However, the limited number of species in the reference database and variations introduced by the low number of sites covered by mapped reads make it challenging to identify the exact taxa (Fig. 3, Table S8). Comparing k-mer classification to our mapping approach, we find a higher number of positive hits in the k-mer approach due to high sensitivity of matching short (31bp) minimizers with the reference sequence rather than matching whole reads as is performed in the mapping approach (Wood et al., 2019). The higher number of k-mer classified reads than mapped reads indicates that downstream mapping of sequence reads is important to further validate Kraken 2 classification results (Fig. 2B).

We observed that ancient DNA damage signals, such as those identified using *PMDtools*, can result in false positives. For instance, sequence data from modern European bison stomach samples showed a proportion of reads with PMD scores above three—a threshold typically associated with ancient DNA (Fig. S4 and Fig. S5) (Skoglund et al., 2014). This underscores the need to use multiple tools for reliable DNA damage estimation. An additional, unexplored factor may be the accumulation of DNA damage unique to gastrointestinal samples, as all ancient and modern bison gastrointestinal samples in this study were derived from such sources. Further analysis using more gastrointestinal metagenomic data could shed light on the impact of these environmental conditions on DNA damage patterns.

Another concern is reference bias introduced by using a single representative species per genus, which could artificially inflate sequence similarity to the representative reference species in competitive mapping for read retrieval, rather than reflecting the true divergence within the genus. However, we found that among the 18 genera that were further analyzed, the genus-wise extracted reads of 16 genera showed the lowest sequence divergence to a different species within the corresponding genus than the one that was used as genus representative in competitive mapping (Fig. 3). It suggests that reference bias has a limited impact on our divergence estimates in identifying the most likely closely related species within the genus (Table S8). To further investigate this, we compared mapping results using two species from the genus *Koeleria*. Initially 10,995 reads were uniquely assigned to *Koeleria asiatica*, while 1021 were uniquely assigned to *Koeleria spicata* in competitive mapping. For the 1021 reads mapped to *K. spicata*, the sequence divergence was 15.4 bp of mismatches per kb of mapped region (bp per kb). When we remapped these reads to *K. asiatica*, the divergence estimate decreased to 12.1 bp per kb. Similarly, the 10,995 reads mapped to *K. asiatica* showed a divergence of 18.5 bp per kb, which increased to 21.9 bp per kb when remapped to *K. spicata*. The consistently lower divergence estimate to *K. asiatica* suggests that reference bias does not significantly affect sequence divergence estimates. Since the reference genomes of the two species are incomplete skimmed-genome assembly, the 1021 reads only mapped to *K. spicata* are likely mapped to genomic regions that are not assembled in the *K. asiatica* reference.

Using *Koeleria* as a case study, we also examined the impact of removing potential ancient DNA damage (C-to-T on the forward strand and G-to-A on the reverse strand) for sequence divergence estimates.

Regardless of whether the potential damages were removed, *K. asiatica* consistently exhibited the lowest sequence divergence (Table S9), showing that using the lowest sequence divergence as proxy for the correctly assigned taxa is likely robust against mismatches caused by post-mortem DNA damage. Second, since bases associated with indels are excluded from sequence divergence calculations, this can result in an underestimation of the true divergence. We therefore also assessed the impact of indels on the sequence divergence calculations. To do so, we remapped the reads that aligned to *K. asiatica* in the competitive mapping to *K. spicata*, and found 13.7 times as many indels that are supposed to lower the sequence divergence estimate to *K. spicata* (Table S10). Regardless, we obtained the lowest sequence divergence to *K. asiatica* consistently. These findings support the conclusion that the identification of *K. asiatica* as the most closely related species is robust and not an artifact of damage removal or indel-related underestimation.

Based on the competitive mapping result, similar to what has been observed for other metagenomic samples (Mann et al., 2023), a relatively low number of reads, in this study sometimes to as low as a few hundred reads, can impede the accurate detection of plant and animal taxa. Additionally, although we initially prioritized studying up to 60 different genera by using k-mer classification, using the competitive mapping data resulted in the absence of ancient DNA damage in 43 of these taxa. This shows the challenge of authenticating ancient DNA reads from a low number of mapped reads and consequently ultra-low breadth of coverage (median 0.000018X across all plant species) for the metagenomic data as analyzed here (Table S11).

As reported here for many genera, despite the thousands of assemblies available from NCBI and skim genomes from PhyloNorway, there are still many species with a (former) distribution in Yakutia for which no reference genomes are available and could thus not be directly identified. Additionally, species that were present in Yakutia during the late Pleistocene but are locally extirpated today are also missing from the reference databases. Therefore, in most cases we could only infer the most closely related species among the range of available genomes, to the actual taxa present in the steppe bison sample.

Additional factors such as the completeness, size and quality of reference genomes and the low number of total covered sites made it challenging to authenticate all the species we identified here (Fig. S6, Fig. S7). A high breadth of coverage reduces the variation for the estimated sequence divergences of the mapped reads, and simulations showed that a minimum breadth of coverage of 0.001X is required to reduce the variance of sequence divergence to below 1 bp per kb and a minimum breadth of coverage of 0.1X for 0.1 bp per kb (Fig. 4, Fig. S8).

In our data, only species in the genera *Artemisia* and *Puccinellia* had an abundance of reads that cover the reference genome at a breadth of coverage above 0.01X. For the other genera where there was only a small difference in sequence divergence between the taxa, we did not infer the most closely related species to the taxa in the sample, taking the impact of breadth of coverage on the sequence divergence estimates into account.

Finally, due to computational limitations when analysing metagenomic data, we focused on the 60 genera which were deemed as most abundant in the sample. In the future, to expand the range of genera investigated, one could build independent mapping indexes, each containing all the reference genomes per taxa and input these mapping results to tools such as ngsLCA that allow to obtain summary statistics across all genera despite mapping the reads against separate databases (Wang et al., 2022).

4.3. Ecological interpretations of the taxa detected based on DNA findings

Among the 18 genera identified and authenticated out of the 60 genera in competitive mapping, 14 are coherent with a presence in Yakutia based on the distribution of modern analogue species and macro-/microfossil evidence. For the most abundant genera, *Artemisia*, *Alopecurus*, *Dryas*, *Phippia*, *Koeleria*, *Larix*, and *Tanacetum*, we tried to infer the most closely related species in our database to the taxa in the sample and make ecological inferences at both the species and genus levels. For taxonomically complex groups such as *Festuca*, *Poa*, *Salix*, *Taraxacum*, and *Puccinellia*, we limited our ecological interpretations to the genus level as there is limited species-level support from modern distributions. For two genera, *Bidens* and *Chrysanthemum*, there were no reference genomes available in our database for the likely modern analogue species *Bidens radiata*, *B. tripartita*, and *Chrysanthemum zawadzki*. Consequently, the two genera were not included in further discussion.

We also identified genera in the metagenomic data that are unlikely to have been present in Yakutia, Russia. For example, the genus *Glebionis* (4972 number of mapped reads) is not known to have ever occurred in Siberia. Additional genera with modern distributions outside Yakutia include *Scalesia*, *Catabrosa*, and *Milium*. These genera are likely all false positive identifications, highlighting that cross-validation using both genetic inference and empirical knowledge of distribution and niche is necessary.

In the subsequent sections we discuss the reads mapped to each genus using the competitive mapping approach, unless otherwise

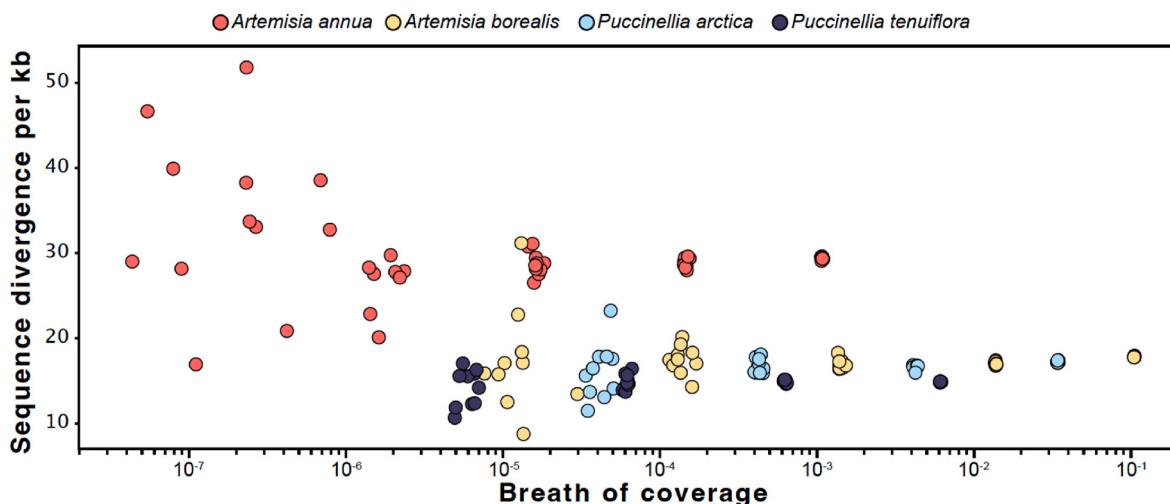


Fig. 4. The relationship between breadth of coverage and sequence divergences estimated from subsampling the competitively mapped or remapped reads. For each species we randomly subsampled reads to lower breadths of coverages and re-estimated the sequence divergence. Lower breadth of coverage results in a higher variance of the sequence divergence estimates.

specified. Then, for each genus, following remapping to all sister species with a genus, we infer the most closely related species to the reads in the sample based on estimated sequence divergences and discuss the ecological niche interpretation of the identified species or genera.

4.3.1. Shrubs and trees suggest warm temperature during the growing season

Among the woody plant genera, we detected reads mapped to larch (*Larix* sp.) and willow (*Salix* sp.). The presence of larch indicates a warm temperature during the growing season and aerated soil conditions. For *Larix*, competitive mapping after filtering out potential microbial reads resulted in 573 reads uniquely mapping to *Larix kaempferi* (Japanese larch) with a skinned reference genome size of 13,492 Mb, while no reads mapped to *Larix gmelinii* (Dahurian larch), which has a reference size of only 1.13 Mb. However, when we extracted the 573 reads mapped to the Japanese larch and independently mapped these to all available *Larix* reference genomes, the sequence divergence to the Dahurian larch (*Larix gmelinii*) was the lowest among all genomes at 3.5 bp per kb. Thus, despite the low number of covered sites in the Dahurian larch (N = 1140), related to its incomplete small reference genome, the Dahurian larch is the most closely related species to the reads in our sample. This also matches with the reported past distribution of the Dahurian larch in eastern Siberia before the last glacial maximum (Schulte et al., 2022; Haupt et al., 2024) and its identification in the macrofossil analysis of the steppe bison gastrointestinal sample (Kienast and Nikolskiy, 2023). Today, Dahurian larch is the primary tree species in the northern larch woodlands that constitutes the vegetation at and south of the modern tree line in Yakutia (Kuznetsova et al., 2010). It is tolerant to extremely cold winter temperatures and competitive against other tree species only under the extreme cold-continental climate of Yakutia.

Another woody taxon with a high number of authenticated mapped reads was the low shrub willow (*Salix glauca*: 970 reads), which usually occurs in river floodplains and is widespread in moist lowland tundra (CAVM team, 2003; Kienast et al., 2005). The genus *Salix* is a taxonomically complex group with some species having a distribution in the arctic tundra today. After remapping the 970 reads to all other *Salix* species, between 5385 bp and 50 kb were covered by mapped reads among a range of *Salix* species including *Salix arctica* (sequence divergence 10.2 bp per kb), *S. hastata* (10.4 bp per kb) and *S. rotundifolia* (11.5 bp per kb) (Fig. 3B–Table S8), which are all present in Yakutia today, as well as *S. myrtilloides* (11.3 bp per kb) which is present in West Siberia. Given the low breadth of coverage (0.0001X) and the minor difference in sequence divergence between *S. arctica* (10.2 bp per kb) and *S. hastata* (10.4 bp per kb), and since they are both species that are present in Yakutia, the exact species identification remains unclear. *S. arctica* tends to inhabit tundra, while *S. hastata* occurs in floodplain shrubberies along riverbanks and lakeshores in boreal and arctic regions (Dierßen, 1996).

Although we also found other coniferous taxa with mapped reads such as spruce (*Picea abies*: 602 reads) and pine (*Pinus lambertiana*: 590 reads), both genera did not show typical signs of ancient DNA damage. In addition to the lack of supporting evidence for the presence of these species in Yakutia during the last cold stage, our data also does not support the detection of spruce and pine.

4.3.2. Aridity and continental climate

We identified the genera: *Artemisia* (945,484 reads), *Tanacetum* (17,879 reads), *Koeleria* (10,995 reads) as well as *Trisetum* (1021 reads), *Festuca* (1852 reads), and *Dryas* (945 reads) in the steppe bison sample.

Artemisia was among the most abundant genera in the steppe bison data, accounting for 55.4 % of all the competitively mapped reads. As *Artemisia* has also been reported as highly abundant in Pleistocene pollen spectra obtained from permafrost sections and lake cores in the same study region (Andreev et al., 2011) as well as Late Pleistocene macrofossil assemblages of the study region (Kienast et al., 2005;

Schirrmeister et al., 2008; Ashastina et al., 2018), we hereby obtain additional support for the abundant presence of this genus on the mammoth steppe. These results also suggest that *Artemisia* was a significant component of the steppe bison diet. The subspecies *Artemisia campestris* ssp. *campestris* and *A. campestris* ssp. *borealis* had the breadth of coverage of 0.35 and 0.38 respectively, thus providing highly accurate reference divergence estimates. Among all *Artemisia* genomes in the database, we found the lowest sequence divergence to *A. campestris* ssp. *campestris* (20.0 bp per kb), followed by *A. campestris* ssp. *borealis* (20.4 bp per kb) (Fig. 3G), suggesting that the steppe bison gut contained plants closely related to the species complex of *Artemisia campestris* (Fig. 3G–Table S8). *Artemisia campestris* ssp. *borealis* is still occurring in Yakutia today, typically in well-drained or dry sites on coarsely clastic substrates in south-exposed slopes and tundra (Yurtsev, 1982; Krasnoborov, 2007). Several other closely related species for which we did not have a reference genome available, such as *A. camp.* ssp. *bargusinensis*, *A. commutata*, or *A. pubescens*, might also have been present in the sample as these species are known to grow in Central Yakutia today, and *A. frigida*, which is still occurring in relict steppe patches in Northeast Siberia (Yurtsev, 2001). All *Artemisia* species occurring in the Siberian Arctic (*A. camp.* *borealis*, *A. subarctica*, *A. arctisibirica*, *A. tilesii*) grow at dry or well drained sites like rock screes and pebble beds.

The species *Tanacetum bipinnatum* (sequence divergence 11.8 bp per kb) and *Tanacetum vulgare* (sequence divergence 13.2 bp per kb) (Fig. 3C–Table S8), exhibit the lowest sequence divergence among the *Tanacetum* taxa. However, based on modern distribution ranges, the closely related *Tanacetum vulgare* ssp. *boreale* is the most likely true species, for which currently no reference genome is available and could thus not be directly detected in our analysis. *T. vulgare*, including *T. vulgare* ssp. *boreale*, is typically found in meadows at the edges of open woodlands and forests or other nutrient-rich, disturbed sites (Dierßen, 1996), whereas *T. bipinnatum* is an Arctic species that thrives in dry habitats such as open sandy xeric grasslands on pingos or dunes where it co-occurs with *Koeleria asiatica* (Hulten, 1968; Yurtsev, 2001).

Our analysis consistently supported *Koeleria asiatica* as the species with the lowest sequence divergence (Table S8) within the genus *Koeleria*. The frost- and drought-adapted (cryoxerophilous) species *Koeleria asiatica* is a tufted grass that is widespread in the present-day steppe environments in Beringia and northeastern Yakutia (Yurtsev, 2001) and occurs in tundra steppes at dry, exposed sites (Reinecke and Troeva, 2017).

We found a large number of reads mapped to *Festuca*, with 23 kb–167 kb among the *Festuca* species genomes covered, providing a good statistical basis for the estimation of sequence divergence. The lowest sequence divergence was to *Festuca hyperborea* (15.3 bp per kb), but the species *Festuca brachyphylla* also showed a similar sequence divergence (15.8 bp per kb) (Fig. 3H–Table S8), both species are present in Yakutia nowadays (GBIF Secretariat, 2023) and adapted to dry climate (xerophilous) and when the climate is not dry, for example in the modern tundra, occur at well-drained (hence locally dry) sites (Aleksandrova, 1980; Malyshev and Peschkova, 2001). Generally speaking, *Festuca* is a taxonomically complex group forming key constituents of Eurasian steppe grasslands, particularly forb-dominated meadow steppes, which represent the north easternmost steppe outposts in Yakutia (Reinecke and Troeva, 2017).

Among the genus *Dryas*, the reads mapped to *Dryas alaskensis* (11.5 bp per kb) and *D. ajanensis* (12.5 bp per kb) showed the lowest sequence divergence (Fig. 3I–Table S8). The most likely species present in the steppe bison sample, *Dryas alaskensis*, exists nowadays from N. Russian Far East to Subarctic America, majorly in subalpine or subarctic biome (IPNI, 2025; Govaerts, 2024). This species tends to inhabit a wide range of niches: solifluction soil, sand-gravel beaches, sandy lakeshores, old beach ridges, wet meadows, stream banks, dry rocky knolls, rocky heath and ericaceous tundra (Flora of North America Editorial Committee, 1993). All Siberian *Dryas* species occur on well drained or dry, coarsely clastic ground, habitats exposed to wind with little snow cover in arctic

tundra (Aleksandrova, 1980), and tundra-steppes (Kucherov and Daniëls, 2005).

4.3.3. Locally high soil moisture and water bodies

In the competitive mappings we identified mapped reads to the genus *Phippsia* (26,168 reads). The most closely related species in this genus was *Phippsia algida* ssp. *algida* (sequence divergence 18.0 bp per kb) (Table S8), which is a high-arctic species with circumpolar distribution (GBIF Secretariat, 2023). However, the sequence divergence is relatively high compared to the inferences from other genera, suggesting that the reads come from a different, but closely related (sub)species to *Phippsia algida*. *Phippsia algida*, as one of the northernmost vascular species, occurs at poorly vegetated, barren, and high-arctic wet sites such as snow beds, coastal saline sites, or sometimes in polar deserts (Aleksandrova, 1980; Rebristaya et al., 1995). They exhibit tolerance to, but are not dependent on, brackish conditions (Dierßen, 1996).

4.3.4. Highly productive saline meadows

We also found and authenticated the presence of ancient DNA from the grassland genera: *Puccinellia* (558,270 reads), *Taraxacum* (34,564 reads), *Alopecurus* (3528 reads), and *Poa* (2284 reads).

We mapped a high number of reads to *Puccinellia tenuiflora*, the genome we used as representative for the alkali grasses of the genus *Puccinellia*. After competitively mapping reads to all available *Puccinellia* sp. reference genomes, we obtained a range of covered sites between 9.2 kb and 55.3 kb, which allow for reliable sequence divergence estimates. We observed the lowest sequence divergence to *Puccinellia tenuiflora* (15.5 bp per kb) (Fig. 3F–Table S8), which was also identified in the macrofossil analysis of the same sample (Kienast and Nikolskiy, 2023). The species *Puccinellia tenuiflora* is a forage grass high in nutritional value and highly resilient for multiple stress conditions including chilling, disease and drought, mainly distributed in Asian grasslands (Guo et al., 2020). *Puccinellia* is a taxonomically complex group. All species within the genus are halophytes (salt-tolerant plants) commonly found along the coast or in arid inland areas of Asia. They are characteristic of brackish floodplain meadows, pastures, and other halophytic communities. They can also tolerate seasonally fluctuating water levels. Therefore, *Puccinellia* species are indicators of high evaporation and, thus, an arid climate when they occur far from the coast. The latter was the prevailing condition when the steppe bison consumed its last meal in what was then the centre of the Beringian landmass, now the coastal lowlands.

We also identified reads belonging to the genus *Taraxacum*, with the lowest sequence divergence to *Taraxacum lacerum* (4.8 bp per kb, also known as *Taraxacum ceratophorum*) (Fig. 3J–Table S8), a species nowadays native to the region spanning from North European Russia to Belgium (IPNI, 2025; Govaerts, 2024) and also occurs in Yakutia (GBIF Secretariat, 2023). Other species that showed low sequence divergence and are found in Yakutia today are *Taraxacum phymatocarpum* (5.0 bp per kb), *T. croceum* (5.2 bp per kb), and *T. arcticum* (5.3 bp per kb) (Fig. 3J). Because the breadth of coverage, at 10^{-4} , is much lower than the required 10^{-1} to maintain a standard deviation of 0.1 bp per kb from sampling effects, we restrain from suggesting that *Taraxacum phymatocarpum* is the most likely species present in the steppe bison sample, despite supported by the modern distribution range and relatively low sequence divergence. Yakutian *Taraxacum* species are grassland plants occurring in floodplain meadows along riverbanks, saline meadows, steppes, tundra, or stony and pebbly slopes in the Arctic.

In genus *Alopecurus*, the species with the lowest sequence divergence was *Alopecurus pratensis* (9.6 bp per kb) (Fig. 3D–Table S8). However, *A. arundinaceus* (10.2 bp per kb) also showed a similarly low sequence divergence and both species are present in Central Yakutia nowadays (Malyshev and Peschkova, 2001). Therefore, our inference remains at genus level. *A. pratensis* is a productive meadow plant, while *A. arundinaceus* occurs in Central Yakutia in meadows under brackish conditions, often co-occurring together with *Puccinellia*. In general,

these species reflect well-drained (hence locally dry) soil.

Based on the current geographic distributions of the genus *Poa* (GBIF Secretariat, 2023) and sequence divergences, *Poa pratensis* ssp. *colpodea* (16.9 bp per kb), *Poa glauca* (17.2 bp per kb), and *Poa pratensis* ssp. *angustifolia* (17.7 bp per kb) could each be species present in the steppe bison sample (Fig. 3A–Table S8). Due to limited resolution to distinguish between these samples and the taxonomical complexity, we refrain from further inferring the exact present species within this genus. Generally, *Poa* species are common across various grassland types in the Northern Hemisphere, from ruderal vegetation and steppes to tundra, meadows, and wetlands.

Altogether, the listed grassland taxa might have been components of a productive saline meadow, which was grazed by the steppe bison shortly before it died.

4.3.5. Genera with the most DNA sequences were prevalent in the mammoth steppe environment

There is substantial palaeobotanical evidence on the composition of the ancient mammoth steppe ecosystem. An overview is given by (Kienast, 2013). To understand the vegetation composition of the Pleistocene grasslands in Beringia, it is necessary to consider the soil conditions, which are strongly affected by the underlying ice-rich permafrost. Especially in the lowlands, this so-called ice complex is penetrated by thick polygonal ice wedge systems, which form a characteristic surface structure with low polygonal walls that impede the surface water runoff. In the resulting depressions, precipitation and melt water together with solutes accumulate and, due to the underlying permafrost, they can not percolate deeper than into the active layer - the seasonally thawing top 0.5–1.5 m of the ground. Under the continental and arid Pleistocene climate in Beringia, the moisture in such polygonal shallow depressions greatly fluctuated seasonally from wet during the snow melt to dry in the course of the summer. In such closed polygonal depressions, the high evaporation and resulting capillary rise of solutes led to salt accumulation in the upper soil. Brackish conditions, accordingly, developed due to the combined effect of aridity, lacking drainage and the permafrost table preventing percolation of solutes. *Puccinellia* is perfectly adapted to salt-influenced plains and depressions with seasonally fluctuating moisture and it is one of the most abundant taxa detected not only in this study but in most macrofossil assemblages of Pleistocene cold (and also warm) stages from the Arctic lowlands of Yakutia (Andreev et al., 2009; Kienast et al., 2005, 2008, 2011).

Steppes and tundra steppes in the plant sociological sense formed within the mammoth steppe in constantly dry habitats such as the polygonal walls and other elevations in the terrain. The abundance of *Artemisia* in the intestinal contents does not necessarily reflect the dietary preferences of the bison. *Artemisia* was a widespread plant in the mammoth steppe biome, present in all grassland communities like steppes, tundra-steppes, saline meadows and tundra as this genus is still occurring in such habitats with dozens of species in Siberia today. The high proportion of *Artemisia* in pollen records (and possibly in genomic data too) might be due to the fact that *Artemisia* is a prolific pollen producer and, consequently, its pollen, as it is wind-dispersed, is ubiquitous in such open landscapes.

Several *Artemisia* species are tolerant to brackish conditions. Following East Siberian *Artemisia* species can occur in saline habitats (on solonet or solonchak soils): *A. leucophylla*, *A. mongolica*, *A. adamsii*, *A. laciniata*, *A. rupestris*, *A. obtusilobia*, *A. jacutica*, *A. anethifolia*, and *A. schrenkiana* (Krasnoborov, 2007). It is therefore very probable that *Artemisia* was actually widespread on the last grazing ground of the bison rather than merely specially selected by the dietary preference of the steppe bison.

4.4. Concluding remarks

In this study, we not only confirmed but also expanded upon prior macrofossil discoveries using metagenomic data and large classification

reference databases. Based on the following mapping database containing the 60 selected genera, we found that the reads mapping to authenticated and empirically possible genera are mostly derived from forbs (63.27 %) and graminoids (36.58 %). Apart from the overlap between macrofossil findings in detecting *Larix* and *Puccinellia*, we also detected 12 other genera that were further validated through the presence of ancient DNA damage. For those genera that, based on ecological interpretations, are plausible to have occurred in Yakutia and that are not part of taxonomically complex groups, we identified the most closely related species based on estimated sequence divergence. In future work, for samples with high coverage and abundant reads for a given taxon, one could explore even higher resolution detection using marker genes, enabling precise phylogenetic placement.

We integrated whole genomes and skimmed genomes of plant species into our classification and the subsequent mapping approaches, maximizing the retrieval of reads from different genera and the detection of taxa. We found that *Artemisia* and *Puccinellia* were the most abundant genera and, together with *Alopecurus*, thus likely important components of the steppe bison diet. *Puccinellia* is commonly found in environments with brackish soils and the presence of *Larix* suggests warm temperatures during the growing season. *Tanacetum*, *Koeleria*, *Festuca* suggest habitats characterized by open, dry ground, while *Phippia* is associated with soils that are at least locally moist. The results show many grassland taxa, both graminoids like *Puccinellia*, *Alopecurus*, *Festuca*, *Koeleria*, and *Poa* and forbs like *Artemisia*, *Tanacetum* and *Taraxacum* potentially co-occurring in the meadow visited by the bison. The obtained species spectrum allows for the reconstruction of the steppe bison's habitat as a grassland composed of saline meadows, steppes and tundra steppes existing under arid conditions and a cold continental climate with relatively warm summers due to a large seasonal temperature gradient. The varieties of different niches we identified in this single sample support previous findings about the presence of diverse plants and niches in the mammoth steppe (Kienast, 2013; Willersley et al., 2014; Kienast and Nikolskiy, 2023).

Data accessibility statement

All the raw sequence data is available in ENA under PRJEB83219, All the raw sequence data is available in ENA under PRJEB83219. The competitive mapping workflow is available at https://github.com/Amend-1634/Competitive_mapping_bowtie2.git. The sequence divergence time workflow is available at https://github.com/Amend-1634/divergence_to_alternative_species.git.

Credit authorship

Chenyu Jin: data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing-original draft, writing-review, editing. Adrian Forsythe: data curation, formal analysis, methodology, writing-review, editing. Miklós Bálint: writing-review, editing. Frank Kienast: conceptualization, formal analysis, investigation, writing-original draft, writing-review, editing. Pavel A. Nikolskiy: data curation, writing-review, editing. Rafał Kowalczyk: writing-review, editing. Katerina Guschanski: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, writing-review, editing. Tom van der Valk: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing-original draft, writing-review, editing.

Declaration of generative AI in scientific writing

During the preparation of this work the authors used ChatGPT v.4.0 and ChatGPT v3.1 to assist in structuring the text. The authors made use of GitHub copilot v2024 for improved clarity and efficiency of the written Python and bash scripts.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Open access funding is provided by Stockholm University. TvdV acknowledges the support from the SciLifeLab and Wallenberg Data Driven Life Science Program (KAW 2020.0239). KG was supported by the Swedish Research Council VR (2020–03398). MB was supported by the LOEWE Programme of the Hessian Ministry of Higher Education, Research and the Arts (LOEWE/1/10/519/03/03.001(0014)/52). The authors also acknowledge support from Science for Life Laboratory (SciLifeLab), the National Genomics Infrastructure (NGI) funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for access to the UPPMAX computational infrastructure. The database constructions were performed on the Luxembourg national supercomputer MeluXina. The authors gratefully acknowledge the LuxProvide teams for their expert support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.quascirev.2025.109604>.

Data availability

A link to the data and/or code is provided as part of this submission.

References

- Aleksandrova, V.D., 1980. The Arctic and Antarctic, Komarov Lectures. Cambridge University Press, Cambridge, England.
- Alsos, I.G., Lavergne, S., Merkel, M.K.F., Boleda, M., Lammers, Y., Alberti, A., Pouchon, C., Denoeud, F., Pitelkova, I., Puçsaş, M., Roquet, C., Hurdu, B.-I., Thuiller, W., Zimmermann, N.E., Hollingsworth, P.M., Coissac, E., 2020. The treasure vault can be opened: large-Scale genome skimming works well using Herbarium and Silica Gel dried material. Plants 9. <https://doi.org/10.3390/plants9040432>.
- Andreev, A.A., Grosse, G., Schirmeister, L., Kuznetsova, T.V., Kuzmina, S.A., Bobrov, A.A., Tarasov, P.E., Novenko, E.Y., Meyer, H., Derevyagin, A.Y., Kienast, F., Bryantseva, A., Kunitsky, V.V., 2009. Weichselian and Holocene palaeoenvironmental history of the Bol'shoy Lyakhovsky Island, New Siberian Archipelago, Arctic Siberia. Boreas 38, 72–110.
- Andreev, A.A., Schirmeister, L., Tarasov, P.E., Ganopolski, A., Brovkin, V., Siegert, C., Wetterich, S., Hubberten, H.-W., 2011. Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. Quat. Sci. Rev. 30, 2182–2199.
- Ashastina, K., Kuzmina, S., Rudaya, N., Troeva, E., Schoch, W.H., Römermann, C., Reinecke, J., Otte, V., Savvinov, G., Wesche, K., Kienast, F., 2018. Woodlands and steppes: pleistocene vegetation in Yakutia's most continental part recorded in the Batagay permafrost sequence. Quat. Sci. Rev. 196, 38–61.
- Bergmann, G.T., Craine, J.M., Robeson, M.S., Fierer, N., 2015. Seasonal shifts in diet and gut Microbiota of the American bison (bison bison). PLoS One 10, e0142409, 2nd.
- Boeskorov, G.G., Potapova, O.R., Protopopov, A.V., Plotnikov, V.V., Agenbroad, L.D., Kirikov, K.S., Pavlov, I.S., Shchelchkova, M.V., Belolyubskii, I.N., Tomshin, M.D., Kowalczyk, R., Davydov, S.P., Kolesov, S.D., Tikhonov, A.N., van der Plicht, J., 2016. The Yukagir Bison: the exterior morphology of a complete frozen mummy of the extinct steppe bison, *Bison priscus* from the early Holocene of northern Yakutia, Russia. Quat. Int. 406, 94–110.
- Borry, M., Hübner, A., Rohrlach, A.B., Warinner, C., 2021. PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. PeerJ 9, e11845.
- Brealey, J.C., Leitão, H.G., Hofstede, T., Kalthoff, D.C., Guschanski, K., 2021. The oral microbiota of wild bears in Sweden reflects the history of antibiotic use by humans. Curr. Biol. 31, 4650–4658.e6.
- CAVM team, 2003. Circumpolar Arctic Vegetation Map, Scale 1:7 500 000. Conservation of Arctic Flora and Fauna (CAFF) Map No. 1. U.S. Fish and Wildlife Service. Anchorage, Alaska.
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete

- mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA.* 110, 15758–15763.
- Dierßen, K., 1996. Vegetation Nordeuropas: 112 Tabellen. Ulmer.
- Flora of North America Editorial Committee, 1993. In: Flora of North America North of Mexico. 19+ Vols. New York and Oxford.
- GBIF Secretariat, 2023. GBIF backbone taxonomy. <https://doi.org/10.15468/39OMEI>
- Govaerts, R., 2024. WCVP: World Checklist of Vascular Plants. The Royal Botanic Gardens, Kew. <http://sftp.kew.org/pub/data-repositories/WCVP/>, n.d.
- Guo, R., Zhao, L., Zhang, K., Gao, D., Yang, C., 2020. Genome of extreme halophyte *Puccinellia tenuiflora*. *BMC Genom.* 21, 311.
- Guthrie, R.D., 1982. 19 - mammals of the mammoth steppe as paleoenvironmental indicators. In: Hopkins, D.M., Matthews, J.V., Schweger, C.E., Young, S.B. (Eds.), *Paleoecology of Beringia*. Academic Press, pp. 307–326.
- Haupt, S., Bernhardt, N., Killing, S., Meucci, S., Herzschuh, U., Zakharov, E.S., Harpke, D., Pestryakova, L.A., Kruse, S., 2024. Biogeography of larches in eastern Siberia – using single nucleotide polymorphisms derived by genotyping by sequencing. *Ecography*. <https://doi.org/10.1111/ecog.07092>.
- Hofman-Kamińska, E., Merceron, G., Bocherens, H., Boeskorov, G.G., Krotova, O.O., Protopopov, A.V., Shpansky, A.V., Kowalczyk, R., 2024. Was the steppe bison a grazing beast in Pleistocene landscapes? *R. Soc. Open Sci.* 11, 240317.
- Hulten, E., 1968. *Flora of Alaska and Neighboring Territories*. Stanford University Press, Palo Alto, CA.
- IPNI, 2025. International Plant Names Index. The Royal Botanic Gardens, Kew. Harvard University Herbaria & Libraries and Australian National Herbarium. <https://www.ipni.org>.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684.
- Kienast, F., 2013. Plant macrofossil records - arctic Eurasia. In: *Encyclopedia of Quaternary Science*. Elsevier, Elias, Scott A., Mock, Cary, pp. 733–745.
- Kienast, F., Nikolskiy, P., 2023. Plant Material from the Intestinal Tract of a Frozen Steppe Bison (Bison priscus) - a Window into ice-age Environments.
- Kienast, F., Schirrmeister, L., Siegert, C., Tarasov, P., 2005. Palaeobotanical evidence for warm summers in the East Siberian Arctic during the last cold stage. *Quat. Res.* 63, 283–300.
- Kienast, F., Tarasov, P., Schirrmeister, L., Grossé, G., Andreev, A.A., 2008. Continental climate in the East Siberian Arctic during the last interglacial: implications from palaeobotanical records. *Global Planet. Change* 60, 535–562.
- Kienast, F., Wetterich, S., Kuzmina, S., Schirrmeister, L., Andreev, A.A., Tarasov, P., Nazarova, L., Kossler, A., Frolova, L., Kunitsky, V.V., 2011. Paleontological records indicate the occurrence of open woodlands in a dry inland climate at the present-day Arctic coast in western Beringia during the Last Interglacial. *Quat. Sci. Rev.* 30, 2134–2159.
- Kitts, P.A., Church, D.M., Thibaud-Nissen, F., Choi, J., Hem, V., Sapojnikov, V., Smith, R.G., Tatusova, T., Xiang, C., Zherikov, A., DiCuccio, M., Murphy, T.D., Pruitt, K.D., Kimchi, A., 2016. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* 44, D73–D80.
- Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15, 356.
- Kowalczyk, R., Taberlet, P., Coissac, E., Valentini, A., Miquel, C., Kamiński, T., Wójcik, J.M., 2011. Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *For. Ecol. Manage.* 261, 821–828.
- Kowalczyk, R., Wójcik, J.M., Taberlet, P., Kamiński, T., Miquel, C., Valentini, A., Craine, J.M., Coissac, E., 2019. Foraging plasticity allows a large herbivore to persist in a sheltering forest habitat: DNA metabarcoding diet analysis of the European bison. *For. Ecol. Manage.* 449, 117474.
- Krasnoborov, I.M., 2007. Asteraceae. *Asteraceae. Flora of Siberia* 13.
- Kucherov, I.B., Daniëls, F.J.A., 2005. Vegetation of the classes carici-kobresietea and cleistogenetea squarrosoe in central chukotka. *Phytocoenologia* 35, 1019–1066.
- Kuznetsova, L.V., Zakharov, V.I., Sosina, N.K., Nikolin, E.G., Ivanova, E.I., Sofronova, E.V., Poryadina, L.N., Mikhalyova, L.G., Vasiliyeva, I.I., Remigailo, P.A., Gabyshev, V.A., Ivanova, A.P., Kopyrina, L.I., 2010. Flora of yakutia: composition and ecological structure. In: Troeva, E.I., Isaev, A.P., Cherosov, M.M., Karpov, N.S. (Eds.), *The far North: Plant Biodiversity and Ecology of Yakutia*. Springer, Netherlands, Dordrecht, pp. 24–140.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lisiecki, L.E., Raymo, M.E., 2005. A pliocene-pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* 20. <https://doi.org/10.1029/2004pa001071>.
- MacPhee, R.D.E., Tikhonov, A.N., Mol, D., de Marlave, C., van der Plicht, H., Greenwood, A.D., Flemming, C., Agenbroad, L., 2002. Radiocarbon chronologies and extinction dynamics of the late Quaternary mammalian megafauna of the taimyr peninsula, Russian Federation. *J. Archaeol. Sci.* 29, 1017–1042.
- Malyschev, L.I., Peschkova, G.A., 2001. Poaceae. *Poaceae. Flora of Siberia* 2.
- Mann, A.E., Fellows Yates, J.A., Fagernäs, Z., Austin, R.M., Nelson, E.A., Hofman, C.A., 2023. Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus. *Quat. Int.* 653–654, 33–46.
- Meyer, M., Kircher, M., 2010. Illumina sequencing Library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, prot5448 db.
- Nikolskiy, P., Shidlovsky, F., 2014. Preliminary data from the study of the intact 50 000 BP frozen mummy of the Anyuy steppe bison (Anyuy River, Arctic Far East). In: Abstract Book of the Vith International Conference on Mammoths and Their Relatives, SASG, Special, p. 141.
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetverin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Magliotti, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745.
- Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Philippy, A.M., 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132.
- Parks, D.H., Chuochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P., 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794.
- Polling, M., ter Schure, A.T.M., van Geel, B., van Bokhoven, T., Boessenkool, S., MacKay, G., Langeveld, B.W., Ariza, M., van der Plicht, H., Protopopov, A.V., Tikhonov, A., de Boer, H., Gravendeel, B., 2021. Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna. *Quat. Sci. Rev.* 267, 107084.
- Rebristaya, O.V., Skvortsov, A.K., Tolmachev, A.I., Tsvelev, N.N., Yurtsev, B.A., 1995. Polypodiaceae-Graminae. *Flora of the Russian Arctic* 1.
- Reinecke, J., Troeva, E., 2017. Extrazonal steppes and other temperate grasslands of northern Siberia – Phytosociological classification and ecological characterization. *Phytocoenologia* 47, 167–196.
- Schirrmeister, L., Grossé, G., Kunitsky, V., Magens, D., Meyer, H., Dereviagin, A., Kuznetsova, T., Andreev, A., Babiy, O., Kienast, F., Grigoriev, M., Overduin, P.P., Preusser, F., 2008. Periglacial landscape evolution and environmental changes of Arctic lowland areas for the last 60 000 years (western Laptev Sea coast, cape mamontov Klyk). *Polar Res.* 27, 249–272.
- Schirrmeister, L., Grossé, G., Schnelle, M., Fuchs, M., Krbetschek, M., Ulrich, M., Kunitsky, V., Grigoriev, M., Andreev, A., Kienast, F., Meyer, H., Babiy, O., Klimova, I., Bobrov, A., Wetterich, S., Schwamborn, G., 2011. Late Quaternary paleoenvironmental records from the Western Lena Delta, Arctic Siberia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 299, 175–196.
- Schulte, L., Meucci, S., Stoof-Leichsenring, K.R., Heitkam, T., Schmidt, N., von Hippel, B., Andreev, A.A., Diekmann, B., Biskaborn, B.K., Wagner, B., Melles, M., Pestryakova, L.A., Alsos, I.G., Clarke, C., Krutovsky, K.V., Herzschuh, U., 2022. Larix species range dynamics in Siberia since the last Glacial captured from sedimentary ancient DNA. *Commun. Biol.* 5, 570.
- Sher, A.V., 1997. A brief overview of the late-cenozoic history of the Western beringian lowlands. In: Edwards, M.E., Sher, A.V., Guthrie, R.D. (Eds.), *Terrestrial Paleoenvironmental Studies in Beringia*, Proceedings of a Joint Russian-American Workshop, June 1991. The Alaska Quaternary Center, Fairbanks, Alaska, pp. 3–6, 3–6.
- Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., Jakobsson, M., 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. USA.* 111, 2229–2234.
- van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L., Guschanski, K., 2020. Index hopping on the Illumina HiSeqX platform and its consequences for ancient DNA studies. *Mol. Ecol. Resour.* 20, 1171–1181.
- Van Geel, B., Protopopov, A., Bull, I., Duijim, E., Gill, F., Lammers, Y., Nieman, A., Rudaya, N., Trofimova, S., Tikhonov, A.N., Vos, R., Zhilich, S., Gravendeel, B., 2014. Multiproxy diet analysis of the last meal of an early Holocene Yakutian bison. *J. Quat. Sci.* 29, 261–268.
- Van Meerbeek, C.J., Renssen, H., Roche, D.M., Wohlfarth, B., Bohncke, S.J.P., Bos, J.A. A., Engels, S., Helmens, K.F., Sánchez-Goñi, M.F., Svensson, A., Vandenberghe, J., 2011. The nature of MIS 3 stadial-interstadial transitions in Europe: new insights from model-data comparisons. *Quat. Sci. Rev.* 30, 3618–3637.
- Wang, Y., Korneliussen, T.S., Holman, L.E., Manica, A., Pedersen, M.W., 2022. *ngsLCA*—A toolkit for fast and flexible lowest common ancestor inference and taxonomic profiling of metagenomic data. *Methods Ecol. Evol.* 13, 2699–2708.
- Wang, Y., Pedersen, M.W., Alsos, I.G., De Sanctis, B., Racimo, F., Prohaska, A., Coissac, E., Owens, H.L., Merkl, M.K.F., Fernandez-Guerra, A., Rouillard, A., Lammers, Y., Alberti, A., Denoeud, F., Money, D., Ruter, A.H., McColl, H., Larsen, N. K., Cherezova, A.A., Edwards, M.E., Fedorov, G.B., Haile, J., Orlando, L., Vinner, L., Korneliussen, T.S., Beilmann, D.W., Björk, A.A., Cao, J., Dockter, C., Esdale, J., Gusuarova, G., Kjeldsen, K.K., Mangerud, J., Rasic, J.T., Skadhauge, B., Svendsen, J.I., Tikhonov, A., Wincker, P., Xing, Y., Zhang, Y., Froese, D.G., Rahbek, C., Bravo, D.N., Holden, P.B., Edwards, N.R., Durbin, R., Meltzer, D.J., Kjær, K.H., Möller, P., Willerslev, E., 2021. Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nature* 600, 86–92.
- Wang, Y., Wessels, M., Pedersen, M.W., Epp, L.S., 2023. Spatial distribution of sedimentary DNA is taxon-specific and linked to local occurrence at intra-lake scale. *Commun. Earth Environ.* 4, 1–13.
- Willerslev, E., Davison, J., Moora, M., Coissac, E., Edwards, M.E., Lorenzen, E. D., Vestergård, M., Gusuarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Bräthen, K.A., Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I.G., Bellemain, E., Brysting, A.

- K., Elven, R., Sønstebo, J.H., Murton, J., Sher, A., Rasmussen, M., Rønn, R., Mourier, T., Cooper, A., Austin, J., Möller, P., Froese, D., Zazula, G., Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R.G., MacPhee, R.D.E., Gilbert, M.T.P., Kjær, K.H., Orlando, L., Brochmann, C., Taberlet, P., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506, 47–51.
- Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257.
- Yurtsev, B., 2001. The Pleistocene “Tundra-Steppe” and the productivity paradox: the landscape approach. *Quat. Sci. Rev.* 20, 165–174.
- Yurtsev, B.A., 1982. 9 - RELICS of the xerophyte vegetation of beringia in Northeastern ASIA. In: Hopkins, D.M., Matthews, J.V., Schweger, C.E., Young, S.B. (Eds.), *Paleoecology of Beringia*. Academic Press, pp. 157–177.
- Zanina, O.G., Gubin, S.V., Kuzmina, S.A., Maximovich, S.V., Lopatina, D.A., 2011. 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.