DISSERTATION | DOCTORAL THESIS

Titel | Title

Mineralogical Preservation of the Human Biome from the Depth of Time

verfasst von | submitted by

Victoria Oberreiter-Moritz BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Biologie
Betreut von | Supervisor:	Univ.-Prof. Ron Pinhasi PhD Univ.-Prof. Mag. Dr. Thomas Rattei

ACKNOWLEDGEMENTS

To begin with**,** I would like to express my gratitude to my supervisor, Dr. Ron Pinhasi, for giving me the opportunity to work and learn in his group even before starting my PhD. Thank you for believing in my abilities and providing me with the chance to grow both as a researcher and as an individual throughout this journey.

I also want to extend my heartfelt gratitude to Dr. Susanna Sawyer, who guided me through the challenges of academia and helped me navigate its tough waters.

I am deeply grateful to the Pinhasi Lab group members and the Research Network MINERVA with its members who made this research possible through funding and their guidance. I also wish to thank my collaborators for contributing their expertise to my projects and for warmly welcoming me to their countries and excavations. A special thank you goes to Dr. Ivor Karavanić, who always integrated me as part of his team whenever I came to take samples and taught me so much about archaeology and the history of the sites. It was a wonderful experience.

I am grateful to have been a part of our Research Network HEAS, which trusted me with funding and the incredible opportunity to produce a podcast – Thank you, Maeve and Tom!

My deepest appreciation goes to my PhD mates, Brina Zagorc, Petra Šimková, Constanze Schattke, Arne Bielke, and Thomas Beard, for their endless support, motivation, and for being wonderful listeners and friends throughout this process.

I am beyond thankful to my loving and supportive family, especially my parents, for making this journey possible and for their eternal love and moral uplift. And last, but never least, my love and husband Mišo – you are my happiness, my rock, and my greatest supporter during moments of self-doubt. With you by my side, I feel capable of achieving anything.

LIST OF PUBLICATIONS

This thesis is based on the following two published articles, and one manuscript in revision.

Publication I

Pere Gelabert, Susanna Sawyer, Anders Bergström, Ashot Margaryan, Thomas C. Collin, Tengiz Meshveliani, Anna Belfer-Cohen, David Lordkipanidze, Nino Jakeli, Zinovi Matskevich, Guy Bar-Oz, Daniel M. Fernandes, Olivia Cheronet, Kadir T. Özdogan, Victoria Oberreiter, Robin N.M. Feeney, Mareike C. Stahlschmidt, Pontus Skoglund, and Ron Pinhasi. Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Current biology*, *31*(16), pp.3564-3574. https://doi.org/10.1016/j.cub.2021.06.023

Publication II

Victoria Oberreiter, Pere Gelabert, Florian Brück, Stefan Franz, Evelyn Zelger, Sophie Szedlacsek, Olivia Cheronet, Fernanda Tenorio Cano, Florian Exler, Brina Zagorc, Ivor Karavanić, Marko Banda, Boris Gasparyan, Lawrence Guy Straus, Manuel R. Gonzalez Morales, John Kappelman, Mareike Stahlschmidt, Thomas Rattei, Stephan M. Kraemer, Susanna Sawyer & Ron Pinhasi. Maximizing Efficiency in sedimentary ancient DNA Analysis: A Novel Extract Pooling Approach. *Scientific reports*, *14*(1), 19388. https://doi.org/10.1038/s41598-024-69741-5

Publication III

Pere Gelabert & Victoria Oberreiter, Lawrence Guy Straus, Manuel Ramón González Morales, Susanna Sawyer, Ana B. Marín-Arroyo, Jeanne Marie Geiling, Florian Exler, Florian Brueck, Stefan Franz, Fernanda Tenorio Cano, Sophie Szedlacsek, Evelyn Zelger, Michelle Hämmerle, Brina Zagorc, Alejandro Llanos-Lizcano, Olivia Cheronet, José-Miguel Tejero, Thomas Rattei, Stephan M. Kraemer, Ron Pinhasi. A sedimentary ancient DNA perspective on human and carnivore persistence through the Last Glacial Maximum in El Mirón Cave, Spain. (Awaiting decision on second revision with Nature Communications)

Table of Contents

Abstract

Sedimentary ancient DNA (sedaDNA) has become an invaluable tool for reconstructing past environments and understanding the dynamics of ancient human and animal populations in the absence of their physical remains. This thesis presents novel methodologies and significant findings from three pivotal studies conducted at Paleolithic cave sites across Europe, Africa, and Asia. This research sheds light on ancient ecosystems and faunal dynamics and provides insights into the human past by employing high-throughput sequencing, targeted hybridization capture, and a new extract pooling technique.

The first study applies shotgun sequencing to explore 25 ka old sediments from Satsurblia Cave, Georgia, revealing human, wolf, and bison DNA. We uncovered a previously unknown human lineage with basal Eurasian ancestry, an extinct wolf lineage, and significant changes in bison populations since the Last Glacial Maximum (LGM).

The second study focuses on optimizing sedaDNA analysis through an efficient extract pooling approach. Pooling multiple sediment extracts significantly reduces the cost and time associated with large-scale screening while maintaining high sensitivity for detecting ancient DNA. Applied across multiple Paleolithic sites in Europe, Asia, and Africa, it demonstrated that aDNA signals could be reliably detected even when pooled with samples devoid of DNA.

The third study, conducted at El Mirón Cave in Spain, uses high-throughput sequencing and targeted hybridization capture to identify a wide range of faunal DNA from sediment layers dating from the Solutrean to the Mousterian period. This research reveals a remarkable diversity of 28 animal taxa, including 15 species previously undetected in the zooarcheological record, and demonstrates genetic continuity in human mitochondrial DNA at a time when Iberia served as a crucial refugium during the LGM.

This research demonstrates that sedaDNA can greatly enhance archeological studies by complementing traditional methods, offering a comprehensive understanding of past ecosystems and human history.

Zusammenfassung

Sedimentäre alte DNA (sedaDNA) hat sich als ein wertvolles Instrument zur Rekonstruktion vergangener Umwelten erwiesen und liefert wichtige neue Einblicke in die Menschheitsgeschichte sowie die Tierwelt der Vergangenheit. Diese Dissertation präsentiert eine neue, effiziente Labormethode sowie bedeutende aDNA-Ergebnisse aus drei zentralen Studien, die an bedeutenden archäologischen Fundorten in Europa, Afrika und Asien durchgeführt wurden. Durch den Einsatz von Shotgun-Sequenzierung und gezielten, innovativen Labormethoden wird das komplexe Zusammenspiel von Arten in verschiedenen Ökosystemen beleuchtet und ein tieferes Verständnis für Überlebensstrategien während kritischer klimatischer Veränderungen vermittelt.

In der ersten Studie werden die Ergebnisse der Shotgun-Sequenzierung von 25.000 Jahre altem Sediment aus der Satsurblia-Höhle in Georgien vorgestellt. Dabei konnten DNA-Spuren von Menschen, Wölfen und Bisons extrahiert und analysiert werden. Die Studie enthüllt eine bisher unbekannte menschliche Abstammungslinie mit eurasischer Herkunft, eine ausgestorbene Wolfslinie sowie signifikante Veränderungen in den Bisonpopulationen seit dem letzten glazialen Maximum (LGM).

Die zweite Studie stellt eine innovative Methode des Extrakt-Poolings vor, welche die Effizienz der sedaDNA-Analyse signifikant steigert. Durch die parallele Laboranalyse von Sedimentproben können die Kosten und der Zeitaufwand für großangelegte Screenings erheblich reduziert werden, ohne dabei die Sensitivität für den Nachweis alter DNA zu verlieren. Diese Technik wurde an mehreren paläolithischen Fundstätten in Europa, Asien und Afrika angewandt und zeigte, dass aDNA-Signale auch dann zuverlässig nachweisbar sind, wenn sie in einem „Pool“ mit bis zu vier Proben ohne DNA gemeinsam analysiert werden.

Die dritte Studie, durchgeführt an Sedimentproben der El Mirón-Höhle in Spanien, nutzt gezielte Hybridisierungstechniken und High-Throughput-Sequenzierung, um eine breite Palette tierischer DNA aus Sedimentschichten vom Solutréen bis zum Moustérien zu identifizieren. Die Untersuchung zeigt eine bemerkenswerte Diversität von 28 Tierarten, darunter 15 Arten, die in vorherigen morphologischen Analysen unentdeckt geblieben waren. Die nachgewiesene menschliche mitochondriale DNA weist auf eine genetische Kontinuität während des letzten glazialen Maximums hin.

Diese Forschung zeigt, dass sedaDNA archäologische Studien erheblich bereichern kann, indem sie traditionelle Methoden ergänzt und ein umfassendes Verständnis vergangener Ökosysteme und der Menschheitsgeschichte ermöglicht.

1. Chapter 1: Introduction

1.1. Ancient DNA (aDNA)

Ancient DNA (aDNA) refers to all the genomic material (nuclear and mitochondrial) that withstood the ravages of time and can be retrieved from archeological or environmental material (Hofreiter, Serre, Poinar, Kuch, & Pääbo, 2001). After the death of an organism, DNA gradually accumulates chemical and physical damage. A common type is the deamination of cytosine to thymine bases, referred to as C to T transitions (Briggs et al., 2007), frequently occurring at the terminal bases of fragments. Additionally, DNA strand breaks happen due to exposure to environmental factors via hydrolysis and oxidative damage, resulting in very short fragments, commonly below 100 base pairs in length (Dabney, Meyer, & Pääbo, 2013; Willerslev & Cooper, 2005). These characteristics can be used as authentic indicators for the identification of aDNA in modern paleogenomic studies (Ginolhac, Rasmussen, Gilbert, Willerslev, & Orlando, 2011; Green et al., 2009; Sawyer, Krause, Guschanski, Savolainen, & Pääbo, 2012).

1.1.1. aDNA Studies on Paleolithic Hominins

Since the earliest attempts to extract DNA from ancient remains in the 1980s (Higuchi, Bowman, Freiberger, Ryder, & Wilson, 1984; Pääbo, 1985) and the complete sequencing of the human genome in the early 2000s through the collaborative Human Genome project (International Human Genome Sequencing Consortium, 2004), the field of ancient DNA (aDNA) has seen a surge of advancements in sequencing technologies and wet lab applications. Over the past years, aDNA has made great strides in researching human evolution. A ground-breaking achievement was the sequencing of the first Neanderthal draft genome in 2010 (Green et al., 2010). This study was succeeded by the genomic identification of the Denisovans, a previously unknown hominin group that, until then, had not been recognized morphologically from the archaeological record (Reich et al., 2010). Further advancements included sequencing the oldest hominin DNA retrieved thus far (Meyer et al., 2016). These hominins from the site of Sima de los Huesos in Spain were dated to 430 ka and genomically assigned to early Neanderthals. Another notable contribution was the sequencing of a first-generation Neanderthal-Denisovan hybrid, providing the first genomic evidence of interbreeding between these two groups (Viviane Slon et al., 2018).

1.1.2. aDNA Studies on Modern Humans and Ancient Fauna

The application of these novel aDNA methodologies enabled the investigation of human migration patterns, ancestries, and population dynamics (e.g., Fu et al., 2014, 2015; Gamba et al., 2014; Hajdinjak et al., 2021; Jones et al., 2015; Lazaridis et al., 2014; Raghavan et al., 2014; Seguin-Orlando et al., 2014; Whittle, Pollard, & Greaney, 2023). For instance, one of the fundamental aDNA studies on the Upper Paleolithic involved 51 ancient Eurasian genomes dating back to ca. 45-7 ka (Fu et al., 2016). The results revealed that all Europeans between 37-14 ka descended from a single founder population, shaping the genetic makeup of present-day Europeans. Additionally, they observed a significant reduction in Neanderthal ancestry over time due to natural selection, and gene flow from Near Eastern populations became prominent in Europe after 14 ka.

(Posth et al., 2023) analyzed 356 ancient hunter-gatherer genomes between 35 - 5 ka and revealed an ancestry profile that survived the Last Glacial Maximum (LGM; ca. 20 - 26.5 ka) (Clark et al., 2009) and persisted in southwestern Europe. This ancestry later contributed to the genetic makeup of populations associated with the Magdalenian culture. In addition to Paleolithic studies, subsequent aDNA research has extensively explored later periods, such as the Neolithic, the advent of farming, and the Bronze Age, revealing crucial insights into human migration, the spread of agriculture, and population dynamics (e.g., Allentoft et al., 2015; Haak et al., 2015; Lazaridis et al., 2014, 2016; Mathieson et al., 2018).

Besides the investigation of hominin evolution and population genomics, aDNA was also employed to shed light on the genetic history of paleofauna, for instance, the extinct woolly mammoth (Miller et al., 2008), arctic fox (*Alopex lagopus*) (Dalén et al., 2007), a 700 ka old horse (*Equus ferus caballus*) (Orlando et al., 2013), grey wolf (*Canis lupus*) (Bergström et al., 2022) and bison (*Bison bonasus*) (Massilani et al., 2016).

Although aDNA provides crucial insights into evolutionary processes and the genetic relationships among ancient populations, its utility is constrained by the scarcity of the skeletal record. Rare fossil finds limit the information available to specific stratigraphic layers and, hence, the time periods these remains are assigned to. Thus, these data only represent isolated time points.

Another constraint in aDNA research is DNA preservation, as discussed in point 1.2.3. A combination of environmental factors, including humidity, pH, exposure to sunlight (i.e., UV radiation), and microbial activity influences the long-term survival of DNA, with temperature being one of the primary contributors to its persistence (Smith, Chamberlain, Riley, Stringer, & Collins, 2003). Low temperatures slow down DNA degradation by reducing the rates of enzymatic activity, hydrolysis, and oxidation, as well as decreased microbial activity (Hofreiter et al., 2001; Lindahl, 1993; Smith et al., 2001; Willerslev, Hansen, & Poinar,

2004). Cold and arid environments, such as permafrost and deep caves, are optimal for DNA preservation (e.g., Loreille et al., 2001; Lorenzen et al., 2011; Poinar et al., 2006; Reich et al., 2010; Shapiro et al., 2004; Viviane Slon et al., 2018). Hence, aDNA data is not only limited by fossil finds but also restricted to temperate regions (Viviane Slon et al., 2022). Many archeological sites lack morphological evidence or sufficient DNA preservation in bones. This limited geographical coverage makes it difficult to reconstruct the genomic history of our past.

1.2. Sedimentary Ancient DNA (sedaDNA)

To address these limitations, aDNA research has increasingly adapted methods from the field of environmental DNA (eDNA), which enable the detection and analysis of contemporary genetic material found in environmental samples, such as soil (Allen et al., 2023; Ariza et al., 2023), water (Antich et al., 2021; Blattner, Ebner, Zopfi, & von Fumetti, 2021) or air (Johnson, Fokar, Cox, & Barnes, 2021; Lynggaard et al., 2022). These sources allow for the detection of organisms even in the absence of skeletal remains. Archeological sediments have proven to be a valuable source to help fill knowledge gaps in the field of paleogenomics and complement the data retrieved from classical aDNA sources, such as bones (Hagelberg, Sykes, & Hedges, 1989; Pinhasi et al., 2015), teeth (Harney et al., 2021; Höss, Dilling, Currant, & Pääbo, 1996), hair (Gilbert et al., 2007), and permafrost conserved specimens (Lorenzen et al., 2011; Orlando et al., 2013; Poinar et al., 2006; Shapiro et al., 2004). The immediate source of DNA in sediments has yet to be determined. Different sources have been suggested, including feces and urine, skin, hair, and microscopical organic inclusion (i.e., bones and coprolites) (Andersen et al., 2012; Hofreiter, Betancourt, Sbriller, Markgraf, & McDonald, 2003; Lydolph et al., 2005; Massilani et al., 2022; Willerslev et al., 2003). Whether extracellular DNA is directly bound to minerals or is being released from organic inclusion contained in the sediment during the extraction process remains a topic of debate (Massilani et al., 2022; Ogram, Sayler, Gustin, & Lewis, 1988).

Mineralogical repositories, with a special focus on archeological sediments, have become increasingly important as a source for aDNA in paleogenomic studies over the past decade, leading to the emergence of a field known as “sedimentary ancient DNA” (sedaDNA). Stratigraphic profiles revealed through archeological excavations offer a continuous record spanning extensive periods of time. Extracting aDNA from sediment layers allows researchers to study temporal changes in ancient ecosystems, track population turnovers, and identify taxa not represented in the morphological record. For example, sedaDNA has been employed to trace shifts in vegetation and megafauna as a response to climatic changes (Clarke et al., 2024; Laura Parducci et al., 2017; Willerslev et al., 2003)

Initially, sedaDNA studies focused on the PCR-based detection of the presence or absence of species within a given environment, providing a basic understanding of ancient biodiversity (Hofreiter, Mead, Martin, & Poinar, 2003; Willerslev et al., 2003). The rise of Next Generation Sequencing (NGS) technologies and advanced analytical methods have allowed for more detailed investigations, including the reconstruction of population dynamics and phylogenies (Gelabert et al., 2021; Graham et al., 2016; Mikkel W. Pedersen et al., 2016; Vernot et al., 2021; Willerslev et al., 2014; Zavala et al., 2021). These developments have expanded the potential of sedaDNA research, enabling a more detailed exploration of the broad spectrum of genomic signatures within sediments.

1.2.1 Paleoenvironmental Reconstruction and Ancient Faunal Dynamics

The pioneering study reporting on ancient environmental DNA was published in 2003, showing that DNA from extinct fauna and flora could be recovered from a non-skeletal repository, in this case from permafrost soil material, as well as from temperate cave sediments (Willerslev et al., 2003). Another PCR-based study on ancient cave sediments produced mtDNA sequences revealing the presence of four animal taxa at Rampart Cave, Arizona (Hofreiter, Mead, et al., 2003). Not long after this breakthrough, ancient insect and conifer DNA was discovered in ice cores through high throughput sequencing (Willerslev et al., 2007), enabling the reconstruction of a forested paleoenvironment in southern Greenland. Ancient plant DNA was successfully extracted from a range of different repositories including pollen (L. Parducci, Suyama, Lascoux, & Bennett, 2005; Suyama et al., 1996), sediments (Willerslev et al., 2014), speleothems (Stahlschmidt et al., 2019), and lake sediments (Clarke et al., 2020; S. Liu et al., 2021; Laura Parducci et al., 2017, 2019; Mikkel Winther Pedersen et al., 2013). Comparative studies on permafrost sediments (Jørgensen et al., 2012) and lake sediments (Clarke et al., 2020; Laura Parducci et al., 2017; Mikkel Winther Pedersen et al., 2013) even reported that ancient plant communities could be predicted more accurately through the use of sedaDNA compared to the analysis of pollen and macrofossils. Pre-NGS studies were constrained by early sequencing methods (Sanger, Nicklen, & Coulson, 1977), struggling to recover sufficient DNA from complex environmental samples. The advent of NGS in the late 2000s revolutionized genomic studies, including sedaDNA research, by increasing speed and reducing sequencing costs, facilitating the simultaneous decoding of millions of DNA fragments (Shendure & Ji, 2008). In addition to genome-wide approaches, DNA metabarcoding is still a widely applied and efficient method that uses specific genetic markers to confirm the presence of predetermined taxa in diverse genome pools produced by environmental samples. For instance, a recent study on 15,000-year-old sediments employing metabarcoding revealed a detailed vegetation transition in Alaska. This data provided a more

comprehensive and accurate record of environmental changes compared to traditional pollen analyses (Clarke et al., 2024).

Further applications of sedaDNA involve refining the known extinction dates of species and faunal turnovers, especially since rare macrofossils provide limited information on the spatial and temporal patterns of extinction (for a review of sedaDNA applications, see Özdoğan et al., 2024). For instance, (Haile et al., 2009) provided evidence that North American mammoths and horses went extinct more recently than previously estimated based on the zooarcheological record, shifting the extinction dates forward by several thousand years. The extinction time of the St. Paul Island mammoths in Alaska was precisely dated to 5.6 ± 100 ka based on several parameters, including the presence of mammoth sedaDNA in layers up until this time (Graham et al., 2016). Environmental sources can also provide information on human migrations and colonization, as illustrated by (Mikkel W. Pedersen et al., 2016). This ground-breaking study retrieved metagenomic DNA from lake sediment cores taken from a Pleistocene ice-free corridor, which was long hypothesized to be the main migration route during the colonization of the Americas (Heintzman et al., 2016; Ives, Froese, Supernant, & Yanicki, 2014; Meiri et al., 2014). The sedaDNA-based evidence indicated that the corridor became ecologically viable for human migration by approximately 12.6 cal. ka, through the emergence of fauna and flora. Consequently, human groups before 12.6 ka likely took an alternative route, such as a coastal pathway, to North America. The corridor between the ice sheets of Beringia and North America, however, may have been used for later north-south migrations, such as by populations associated with the Clovis culture around 13.4 ka (Sanchez et al., 2014). The study by Pedersen et al. effectively combined sedaDNA evidence with radiocarbon dating, pollen analysis, and macrofossil data, providing a comprehensive reconstruction of postglacial processes and human migration in the absence of hominin remains.

1.2.2. Hominin sedaDNA and Methodological Advancements

For the first time in 2017, mitochondrial DNA from Neanderthals and Denisovans was extracted directly from clay-rich Pleistocene cave sediments, demonstrating the significant potential of sedaDNA for studying human evolution (V. Slon, Hopfe, Weiß, & Mafessoni, 2017). This framework introduced the ability to identify hominin and mammalian species and reconstruct phylogenetic trees directly from archaeological stratigraphy without targeting macroscopic fossils. Slon et al. analyzed samples from seven sites with evidence of hominin occupation. They performed shotgun sequencing alongside human and mammalian mitochondrial capture with a success rate of 90% for Late Pleistocene and 42% for Middle Pleistocene samples, defined by the presence of mtDNA fragments assigned to at least one family,

displaying authentic aDNA damage. The success rate for retrieving hominin mtDNA sufficient for further analyses was roughly 17% (V. Slon et al., 2017).

Mitochondrial sedaDNA results from Baishiya Karst Cave on the Tibetan Plateau indicate the presence of Denisonans at 100 ka, 60 ka, and potentially even 45 ka (Zhang et al., 2020). This evidence, together with a 160 ka old Denisovan mandible (Zhang et al., 2020), supports the long-term occupation of high altitudes by this homin group and suggests potential genetic adaptations to such environments.

Three studies on cave sediments followed in 2021, including the first paper in this thesis (Chapter 2), in which our lab used shotgun sequencing to recover nuclear DNA of wolf, bison, and a human genome from a 25 ka sediment sample from Satsurblia Cave in Georgia (Gelabert et al., 2021). The retrieved nuclear human DNA was sufficient to reveal a previously unknown Caucasian pre-LGM ancestry that contributed to later Eurasian populations. However, the data resolution was insufficient to investigate phylogenetic relationships with individuals it clusters with in more detail.

Vernot et al. developed a new and efficient method for enriching and analyzing ancient hominin nuclear DNA from cave sediments, thereby reducing faunal misalignments (Vernot et al., 2021). This approach facilitated detailed genomic analyses of Neanderthals and Denisovans. It revealed significant events such as a Neanderthal population turnover, along with the loss of mtDNA diversity at Galería de las Estatuas in northern Spain around 100,000 years ago (Vernot et al., 2021). They further produced phylogenetic results for Chagyrskaya and Denisova Cave in Siberia, consistent with previous results achieved through skeletal material (Mafessoni et al., 2020; Meyer et al., 2012; Prüfer et al., 2014).

Similarly, Zavala et al. identified distinct shifts in hominin and faunal DNA at Denisova Cave, uncovering multiple occupation phases and migrations by both Denisovans and Neanderthals during the Pleistocene (Zavala et al., 2021). Their success rate of retrieving hominin mtDNA from loose sediment was approximately 24%, and 94% for retrieving mammalian mtDNA. They also found that the mtDNA retrieved from cave sediments after 80 ka closely matched the mtDNA from the skeletal remains of Denisova 3 and 4 (Krause et al., 2010; Meyer et al., 2012; Sawyer et al., 2015) and Denisovan mtDNA from Baishiya Karst Cave dated to 70–45 ka (Zhang et al., 2020).

Zavala et al. and Vernot et al. developed and used capture-based approaches. These techniques involve synthetic DNA or RNA fragments designed to match specific sequences in the target genomes, such as mammalian or hominin DNA. These baits are introduced into the DNA pool, where they bind to complementary DNA fragments. Once the target DNA is bound, it can be separated from non-target DNA through repeated washing steps. This process significantly enriches the sample for desired DNA before sequencing, reducing the need for deep sequencing and increasing the recovered amount of target DNA.

The innovation in both of these studies (Vernot et al., 2021; Zavala et al., 2021) lies in the new techniques that enabled a high success rate in retrieving sufficient genomic data from sediments, coupled with a dense, grid-like sampling strategy designed to capture a complete time-series of a site's genetic history (**Figure 1**).

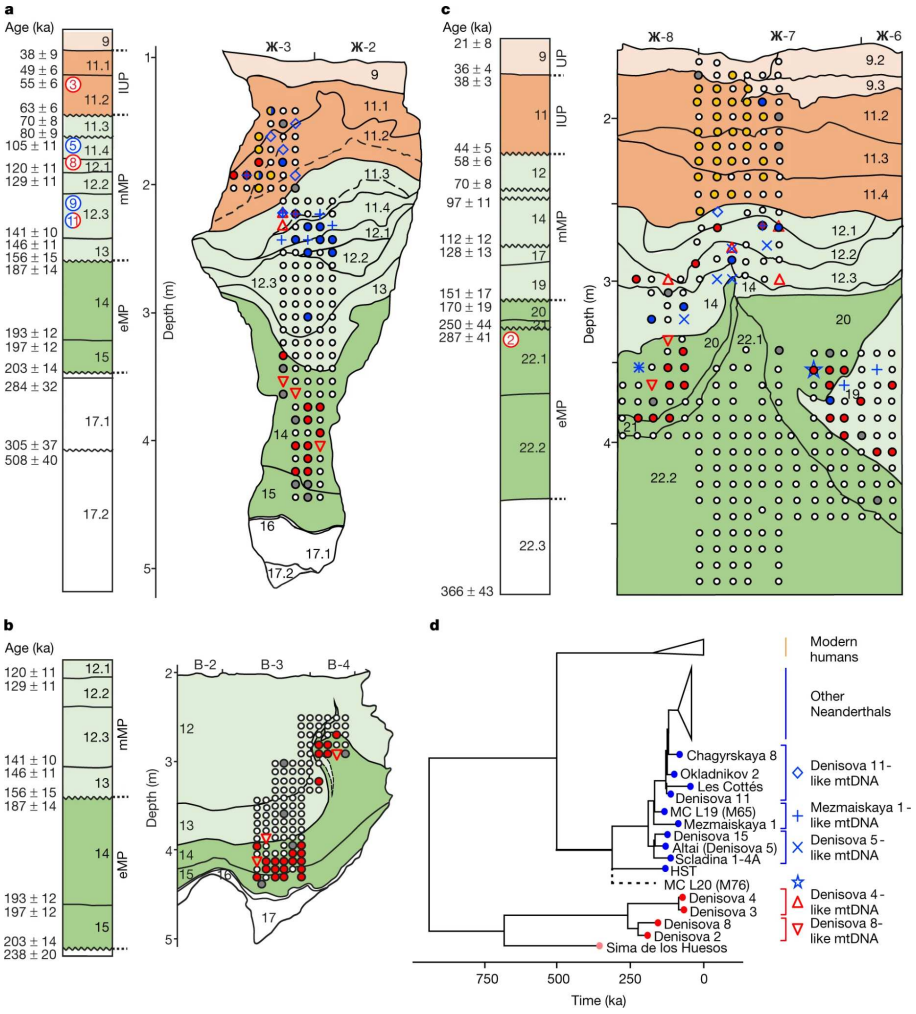


Figure 1: a) East Chamber, southeast profile. **b)** East Chamber, northwest profile. **c)** Main Chamber, southeast profile. **d)** Phylogenetic tree including mtDNA genomes from Zavala et al. 2021 and previously published ancient genomes. **a-c:** Circles indicate sampling locations, colors correspond to type of hominin mtDNA (red to Denisovan, blue to Neanderthal, yellow to ancient modern human, gray to unidentified hominin, and white indicates no hominin mtDNA). The figure is taken and the figure description is adapted from Zavala et al. (2021).

Besides taking loose sediment samples directly from the stratigraphy in a grid-like manner, an alternative method that preserves the sediment and its inclusions in situ was introduced by (Massilani et al., 2022).

This study used resin-impregnated sediment blocks and performed micromorphological analyses and DNA extraction on various components ("microfeatures") such as coprolites, bones, and matrix, as well as inorganic features identified within the blocks. They compared DNA recovery of loose, dehydrated, and impregnated sediment blocks from several Pleistocene caves with known DNA preservation and found that, overall, resin-impregnated blocks performed best at DNA yield and library efficiency. Their results show that resin impregnation, which involves high-temperature steps lasting up to several days, does not negatively impact aDNA retrieval. They further found that the extraction from targeted microsampling (drilling holes of ca. 1 mm with up to ca. 8 mg of sample weight) was significantly more efficient in regards to library efficiency and the amount of DNA retrieved compared to regular samples (drilling to retrieve ca. 30 mg of sample weight), likely due to a reduction in co-extracted inhibitors. From the targeted microfeatures, bone and coprolites yielded the highest quantities of mammalian mtDNA, respectively, while producing fewer mammalian taxa than the matrix. The recovery of aDNA from the matrix showed that it is not sterile, and DNA was preserved within it. At the same time, it does not answer the question of whether the DNA the matrix yielded stems from organic components (e.g., finely ground bones and coprolites, which were identified as part of the matrix) or if it is adsorbed to mineral particles (Massilani et al., 2022). However, the observed heterogeneity in the taxonomic distribution of mammalian species across the blocks may suggest that organic inclusions are the primary locations for preservation.

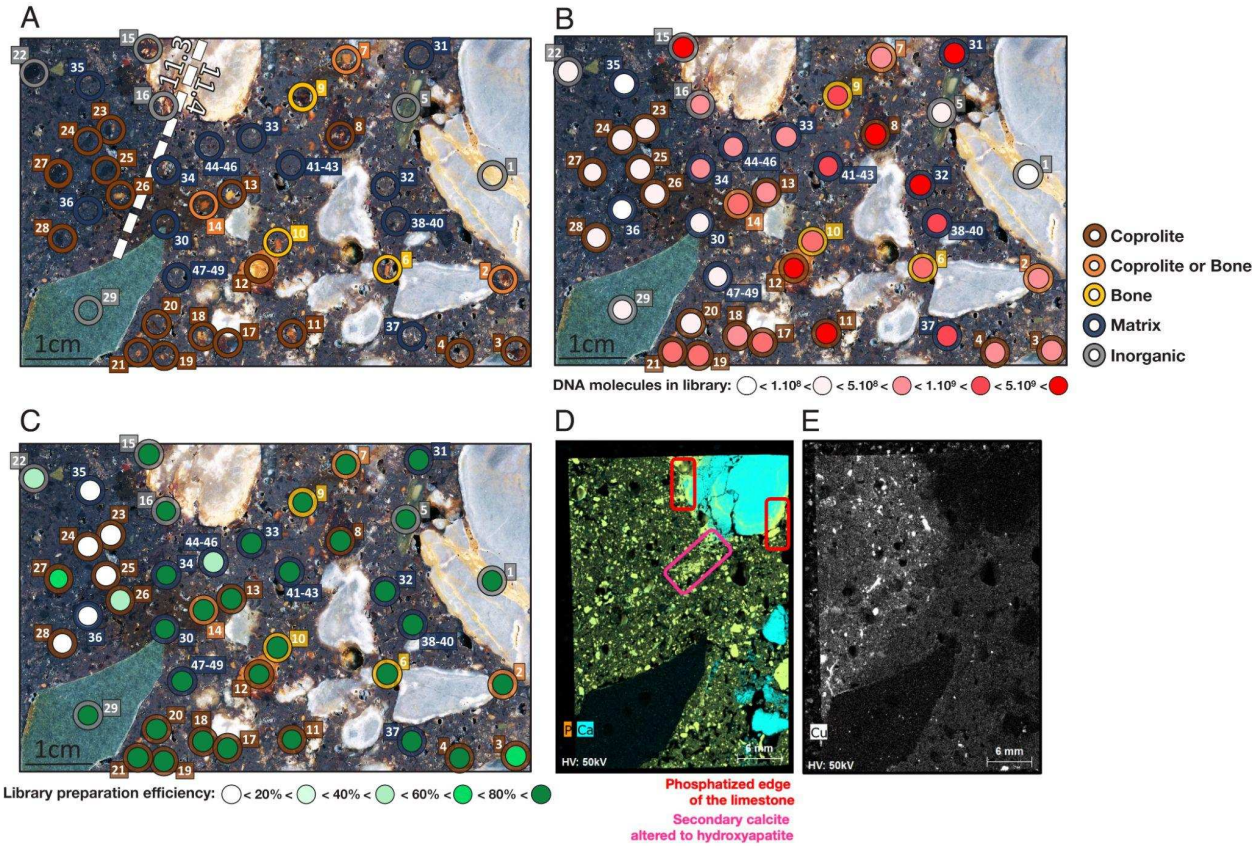


Figure 2: Targeted microsampling from resin-impregnated sediment block (DCE5C). A) Sampling spots indicated by circles, color of circles corresponds to the type of microfeature (brown to coprolite, light-brown to either coprolite or bone, yellow to the bone, blue to the matrix, and gray to inorganic components). B) Color fill indicates the number of DNA molecules retrieved per sample. C) Color fill indicates the library efficiency per sample. D) µXRF surface scan. Turquoise corresponds to calcium, orange to phosphate, yellow to calcium phosphate-rich areas. Red and magenta frames correspond to phosphatized limestone and secondary calcite, respectively. E) µXRF surface scan of the same area as E). The color white shows copper-rich areas. Figure taken and figure description adapted from Massilani et al., 2022.

Sampling from solidified sediments via polymerization comes with the advantage of minimizing the potential of translocation of sediment and offers the possibility to visually identify post-depositional disturbance such as bioturbation (Araujo, 2013; Grave & Kealhofer, 1999). This advantage is also provided by other, naturally sealed, solid systems, such as cave speleothems. A metagenomic study revealed floral and mammalian DNA from multiple layers and an enclosed bear bone in two stalagmites from Solkota

Cave, Georgia (Stahlschmidt et al., 2019). The DNA-producing layers were U-series dated to approximately 84 - 56 ka.

A study conducted on Bronze Age human skeletal remains demonstrated the diffusion of DNA from human bones into surrounding calcite stones (Sarhan et al., 2021). Interestingly, the cave stones exhibited better DNA preservation in terms of fragmentation than the bones themselves, which the authors attribute to the possibility that leached DNA might be less exposed to microbial decay than DNA within the bones. Unlike the results produced by (Massilani et al., 2022), the data produced by Sahran et al. support the idea that minerals can act as preserving agents for nucleic acids, although Massilani et al. do not entirely exclude this possibility.

Subsequent research has opened up the possibility of identifying individual genomes directly from personal ornaments. Essel et al. demonstrated this by successfully extracting the wearer's or producer's DNA from a 25 - 19 ka old deer tooth pendant from Denisova Cave (Essel et al., 2023). They managed to extract non-endogenous human DNA from the skeletal remains of another species by introducing a novel, non-invasive extraction protocol. This protocol involves a series of increasingly hot sodium phosphate washes up to 90 °C, effectively releasing foreign human DNA that had adsorbed to deeper layers of the tooth pendant. This study is particularly significant because, for the first time in the history of aDNA research, it was possible to precisely assign an artifact to a specific human individual without physical association. As bone is a porous material, it allows DNA to seep through and become protected by its mineral component hydroxyapatite, which was previously shown to be a potential protective agent for DNA (Brundin, Figdor, Sundqvist, & Sjögren, 2013). This research illustrated that minerals, specifically hydroxyapatite, can effectively bind and protect extracellular DNA.

Environmental samples can provide valuable information about human life history, subsistence, and disease. For instance, lake sediments revealed anthropogenic impacts on land use and fauna (Thomson-Laing et al., 2024). Also, dietary insights into megafauna and humans, along with evidence of ancient pathogens, have been uncovered through sedaDNA studies (Søe et al., 2018; Willerslev et al., 2014).

1.2.3. DNA Preservation in Minerals

Recently, a pivotal study reported on the oldest DNA ever retrieved, extracted from a two million-year-old environment from Greenland (Kjær et al., 2022). This research pushes the known limits of DNA preservation back by nearly a million years (van der Valk et al., 2021) but also raises crucial questions about the mechanisms that allow DNA to survive for such extended periods. The decay of DNA typically starts post-mortem due to the lack of intrinsic cell repair pathways of living organisms and accumulates

over time (Lindahl, 1993). Extracellular DNA, released from lysed cells, can potentially survive millions of years when it becomes adsorbed to and subsequently protected by mineral surfaces. Generally, DNA can bind to various mineralogical components due to its inherent negative charge (Freeman, Dieudonné, Collins, & Sand, 2020). This adsorption causes the modification of the DNA's chemical conformation. Hence, enzymes fail to recognize its altered structure (Cai, Huang, & Zhang, 2006; Melzak, Sherwood, Turner, & Haynes, 1996; Nguyen & Elimelech, 2007). Various studies have shown that under laboratory conditions, DNA can strongly bind to mineral surfaces, especially clays (Cai et al., 2006; Greaves & Wilson, 1969; Khanna, Yoder, Calamai, & Stotzky, 1998; Ogram et al., 1988), sand (Lorenz & Wackernagel, 1987), humic acids (Crecchio & Stotzky, 1998), and iron oxides (B. Liu & Liu, 2014; Sodnikar, Parker, Stump, ThomasArrigo, & Sander, 2021). Several factors, including the environment's pH level, the structural properties of the DNA molecules, ionic strength, and temperature, can influence adsorption. For instance, (Melzak et al., 1996) demonstrated that DNA can adsorb to silica in the presence of certain salts. These salts neutralize the negative charge of the DNA, which enhances adsorption. They further found that adsorption is enhanced at higher temperatures and decreased pH values and that linearized DNA, which can form more extensive contact with the silica surface, adsorbs more strongly than supercoiled DNA (Melzak et al., 1996). Another study on clay minerals confirmed that the adsorption of DNA was influenced by factors such as ionic strength, pH level of the medium, and type and amount of clay present (Cai et al., 2006). This research further showed that fine clays are more effective at adsorbing DNA, while coarse clays display a higher binding affinity. Hydroxyapatite, a calcium phosphate, is the main inorganic component of bone and is also known to bind and protect DNA from decay as it reduces the effects of hydrolysis and nuclease activity (Brundin et al., 2013; Lindahl, 1993; Weiner & Wagner, 1998).

1.2.4. The Effects of Mitigation, Leaching, and post-depositional Movement

Leaching and mitigation of DNA refers to the movement of nucleic acids between sediment layers, which must be considered when screening a site for aDNA preservation. In frozen sediments, vertical DNA leaching seems insignificant, as the macrofossil record is mainly congruent with sedaDNA profiles (Lydolph et al., 2005; Willerslev et al., 2003). A study on two New Zealand caves tested the downward movement of DNA in sediment by looking at the genomic presence of non-indigenous, mammalian species such as sheep (*Ovis aries*) in layers pre-dating their introduction to the island. They further screened for moa DNA in post-extinction strata, which would indicate upward movement of DNA (Haile et al., 2007). The results revealed no significant amount of moa DNA in modern layers. However, vertical leaching of sheep DNA did indeed play a role in the unsaturated, temperate sediments. The authors speculate that sheep

produced larger amounts of urine than birds, which, in turn, carried DNA vertically downwards. This scenario seems more likely than the movement of sediment per se.

DNA leaching is expected to vary by the amount of bioturbation and by the type of sediment. It occurs less frequently in dense, clay-rich sediments due to the ability of clay to bind and stabilize DNA, compared to sandy, more permeable sediments that allow more water flow (Cai et al., 2006; Willerslev et al., 2003). Post-depositional movements such as bioturbation can be assessed with the help of micromorphological analyses (Araujo, 2013; Grave & Kealhofer, 1999). However, the movement of DNA alone (i.e., leaching or migration) is rather challenging to detect. A study conducted in zoological environments showed that soil structure, chemistry, and properties influence the extent of DNA leaching. The movement of DNA was less pronounced in soils with pH levels above 6 (favoring adsorption), soils with fine textures, and clay-rich horizons, as clay tends to act as a barrier (Andersen et al., 2012). Andersen et al. detected leaching in areas with increased deposition of organic material (e.g., below latrines or areas with high volumes of urine deposition) to a maximum depth of 70 cm (below an elephant enclosure). Notably, there was no evidence of DNA leaching under a tiger trail or an ostrich farm, indicating that certain environmental and behavioral factors, such as urine concentration in small areas, can exacerbate DNA movement through soil layers (Andersen et al., 2012).

Slon et al. found minimal evidence of DNA leaching at Sefunim Cave, Israel. Stable environmental conditions and limited post-depositional movement between layers, detected by archeological analyses, as well as partial cementation of sediment, played pivotal roles in preserving the stratigraphy in situ (Viviane Slon et al., 2022).

DNA leaching is reduced to a minimum in stalagmites due to their closed structure, which restricts fluid movement and protects embedded DNA. (Stahlschmidt et al., 2019) demonstrated that DNA enclosed and preserved in stalagmites is less susceptible to contamination and leaching compared to loose sediment. DNA is effectively sealed and protected from environmental degradation by calcite layers in stalagmites, providing a reliable and datable source for genomic material. Similarly, resin-impregnated microstratigraphic blocks preserve sediment in situ, limiting the mixing of layers (Massilani et al., 2022).

Overall, the post-depositional movement of sediment needs to be individually assessed for each archeological site through geological and micromorphological methods to ensure that DNA leaching does not compromise the accuracy of paleobiological reconstructions.

In this thesis, I present the sedaDNA results of two Paleolithic cave sites in Eurasia, employing a metagenomic approach and targeted capture, and one methodological paper focussing on the wet lab pooling of sediment extracts.

2. Chapter 2: Ancient Sedimentary DNA from the Caucasus

2.1. Publication I

Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment

Status: Published

Citation: Pere Gelabert, Susanna Sawyer, Anders Bergström, Ashot Margaryan, Thomas C. Collin, Tengiz Meshveliani, Anna Belfer-Cohen, David Lordkipanidze, Nino Jakeli, Zinovi Matskevich, Guy Bar-Oz, Daniel M. Fernandes, Olivia Cheronet, Kadir T. Özdogan, Victoria Oberreiter, Robin N.M. Feeney, Mareike C. Stahlschmidt, Pontus Skoglund, and Ron Pinhasi. Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. Current biology, 31(16), pp.3564-3574. https://doi.org/10.1016/j.cub.2021.06.023

Article

Current Biology

Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment

Highlights

- A single shotgun-sequenced Pleistocene sediment yielded genomic data of three mammals
- Sediment genome sequencing can produce data comparable to that from skeletal remains
- A pre-LGM human lineage from the Caucasus was an ancestral component of West Eurasia
- ~0.01X wolf and bison environmental genomes, suggesting reshaping of populations

Authors

Pere Gelabert, Susanna Sawyer, Anders Bergström, ..., Mareike C. Stahlschmidt, Pontus Skoglund, Ron Pinhasi

Correspondence

pere.gelabert@univie.ac.at (P.G.), anders.bergstrom@crick.ac.uk (A.B.), pontus.skoglund@crick.ac.uk (P.S.), ron.pinhasi@univie.ac.at (R.P.)

In brief

Cave sediments preserve ancient DNA for thousands of years. Gelabert et al. retrieve human and mammalian nuclear and mitochondrial environmental "shotgun" genomes from a single 25,000-year-old sediment sample from Satsurblia cave. Their results provide insights into the population genetic histories of three mammalian species.

Gelabert et al., 2021, Current Biology *31*, 3564–3574
August 23, 2021
https://doi.org/10.1016/j.cub.2021.06.023

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment

Pere Gelabert,[1,11,13,*] Susanna Sawyer,[1,11] Anders Bergström,[2,11,*] Ashot Margaryan,[3] Thomas C. Collin,[4] Tengiz Meshveliani,[5] Anna Belfer-Cohen,[6] David Lordkipanidze,[5] Nino Jakeli,[5] Zinovi Matskevich,[7] Guy Bar-Oz,[8] Daniel M. Fernandes,[1,9] Olivia Cheronet,[1] Kadir T. Özdoğan,[1] Victoria Oberreiter,[1] Robin N.M. Feeney,[4] Mareike C. Stahlschmidt,[10] Pontus Skoglund,[2,12,*] and Ron Pinhasi[1,12,*]

[1]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria
[2]Ancient Genomics Laboratory, Francis Crick Institute, London, UK
[3]Center for Evolutionary Hologenomics, University of Copenhagen, Copenhagen, Denmark
[4]School of Medicine, University College Dublin, Dublin, Ireland
[5]Georgian National Museum, Institute of Paleoanthropology and Paleobiology, Tbilisi, Georgia
[6]Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem, Israel
[7]Israel Antiquities Authority, Jerusalem, Israel
[8]Zinman Institute of Archaeology, University of Haifa, Haifa, Israel
[9]CIAS, Department of Life Sciences, University of Coimbra, Coimbra, Portugal
[10]Department of Human Evolution, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany
[11]These authors contributed equally
[12]These authors contributed equally
[13]Lead contact
*Correspondence: pere.gelabert@univie.ac.at (P.G.), anders.bergstrom@crick.ac.uk (A.B.), pontus.skoglund@crick.ac.uk (P.S.), ron.pinhasi@univie.ac.at (R.P.)
https://doi.org/10.1016/j.cub.2021.06.023

SUMMARY

Cave sediments have been shown to preserve ancient DNA but so far have not yielded the genome-scale information of skeletal remains. We retrieved and analyzed human and mammalian nuclear and mitochondrial environmental "shotgun" genomes from a single 25,000-year-old Upper Paleolithic sediment sample from Satsurblia cave, western Georgia:first, a human environmental genome with substantial basal Eurasian ancestry, which was an ancestral component of the majority of post-Ice Age people in the Near East, North Africa, and parts of Europe; second, a wolf environmental genome that is basal to extant Eurasian wolves and dogs and represents a previously unknown, likely extinct, Caucasian lineage; and third, a European bison environmental genome that is basal to present-day populations, suggesting that population structure has been substantially reshaped since the Last Glacial Maximum. Our results provide new insights into the Late Pleistocene genetic histories of these three species and demonstrate that direct shotgun sequencing of sediment DNA, without target enrichment methods, can yield genome-wide data informative of ancestry and phylogenetic relationships.

INTRODUCTION

Ancient DNA fragments sequenced from bone,[1] teeth,[2] and hair[3] have revolutionized our understanding of natural history and the human past.[4,5] When skeletal material is not available, ancient environmental DNA has been used to determine the presence or absence of different species. Several studies based on PCR methods demonstrated the presence of ancient DNA in sediments,[6] including in caves,[7] and more recently, high throughput sequencing techniques have been applied.[8–11] Cave sediment ancient DNA has been used to track the presence or absence of species across a range of environments and time periods, primarily through targeted amplification or capture of single genetic regions.[12] A ground-breaking study showed DNA preservation in clay-rich sediments since ~240 ky[13] and used targeted enrichment to recover sufficient numbers of fragments to reconstruct mtDNA phylogenies of Neanderthals and Denisovans. A similar study recovered Denisovan mitochondrial DNA from sediments deposited ~100 kya and ~60 kya from Baishya Karst Cave on the Tibetan Plateau.[14] A recent study used targeted enrichment of 1.6 million loci to recover Neanderthal and Denisovan nuclear DNA from three Paleolithic sites. This yielded enough DNA to allow for some analyses of genome-wide ancestry, including the finding of a Neanderthal population replacement at one of the sites, thereby demonstrating the possibility of large-scale nuclear DNA recovery from sediments.[15]

Here, we report results from shotgun sequencing and genomic analysis of a sediment sample from the Upper Paleolithic site of

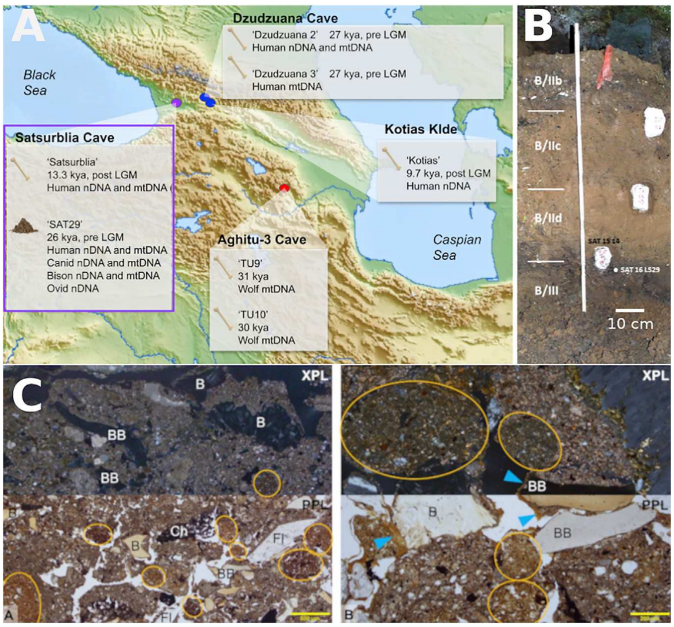


Figure 1. Site description: an overview of the location of Satsurblia cave with contextual information

(A) Map of the Caucasus region showing relevant sites that have yielded ancient DNA from humans (blue dots), animals (red dot), or both (purple dot). Only sites with remains from the Mesolithic or older are shown.

(B) Layer B of Satsurblia cave with the SAT29 sampling location shown.

(C) Two microphotographs (A and B) from block sample SAT 15 14. The microphotographs were taken adjacent to SAT 16 LS29, in cross- (XPL) and plane-polarized light (PPL), showing dominant components and processes. Typical anthropogenic components are bone (B), burnt bone (BB), charcoal (Ch), and flint (Fl). Note also the occurrence of rounded soil aggregates (yellow circles) that were transported into the cave from soils forming outside and that exhibit cross-striations of the clay component resulting from repeated wetting and drying cycles during their formation. Similarly, clay coatings in voids (blue arrow) result from water percolating through the sediment. See also Data S1.

Satsurblia Cave, southern Caucasus, dating to the Last Glacial Maximum (LGM, 25,000 years ago [kya]). In most of the Caucasus, and particularly in western Georgia, karst systems hold low and stable year-round temperatures and low acidity (no guano deposits in most systems). The sediment sample yielded up to several million sequence reads from human, wolf (*Canis lupus*), and bison (*Bison bonasus*), corresponding to genome-wide data comparable to low-coverage sequencing obtained from skeletal remains.

RESULTS

We analyzed six sediment samples from different layers of areas A (pre-LGM) and B (pre and post-LGM) of Satsurblia cave (Figure 1A)[16] and performed shotgun sequencing to screen them for mammalian DNA (Data S1A). One of the samples, SAT16 LS29 (SAT29) from layer BIII (Figure 1B), which is radiocarbon dated to 25.4–24.5 ka cal BP,[16] contained substantial amounts of DNA from humans as well as from other mammals and was therefore sequenced to greater depth.

We sequenced 561,263,536 reads from the SAT29 sample, and after filtering, we retained 226,880,778 reads. Metagenomic screening with centrifuge[17] indicated that 1.3% of the reads were of eukaryotic origin, and four main mammalian genera were identified: *Ovis* (28%), *Homo* (9%), *Canis* (5.5%), and *Bos* (2.1%) (Figure 2A). After competitive mapping to the sheep, human, dog, and domestic cattle reference genomes (Method Details), we assigned a total of 4,956,676 reads as follows: *Canis* (2,378,237 reads, 48.0% of assigned reads), *Bos* (1,811,555 reads, 36.5% of assigned reads), *Homo* (661,765 reads, 13.5% of assigned reads), and *Ovis* (105,119 reads, 2.1% of assigned reads). We then used BLAST+ and MEGAN to verify the accuracy of the mapping process and show that it is unlikely that any other animal species are represented in substantial amounts in the sample (Figures 2B, 2C, S1, S2, S3, S4, S5, and S6; Data S1B; Method details).

Mitochondrial capture and evidence for multiple individuals

Through targeted capture,[20] we recovered 3,447 human mitochondrial DNA (mtDNA) reads, which when aligned to rCRS resulted in a 10-fold coverage of the human mtDNA. We similarly recovered 5,809 *Canis* mtDNA reads, yielding a 13-fold coverage of the *C. lupus* mitochondrial genome. Capture of cattle mtDNA also proved successful, but the recovered sequences showed high similarly to bison mitochondria, suggesting that the DNA identified as deriving from *Bos* in the metagenomic screening in fact derived from *Bison sp.* (that was not included in the metagenomic reference set). We recovered 2,448 reads providing an 8-fold coverage of the *B. bonasus* mitochondrial genome. All the recovered environmental genomes show elevated deamination values and short fragment length distributions, consistent with an ancient origin (Figures S1A and S1B; Data S1C).

To investigate whether the recovered human, wolf, and bison DNA derive from single or several individuals from each of these species, we focused on the mitochondrial sequences. Due to their haploid state in single individuals, sequence polymorphisms among mitochondrial reads provide evidence for the presence of multiple individuals (Figure S2; Table 1).[13,21–23] SAT29 displays more than one allele at many known human, wolf, and bison mitochondrial variants, raising the possibility of

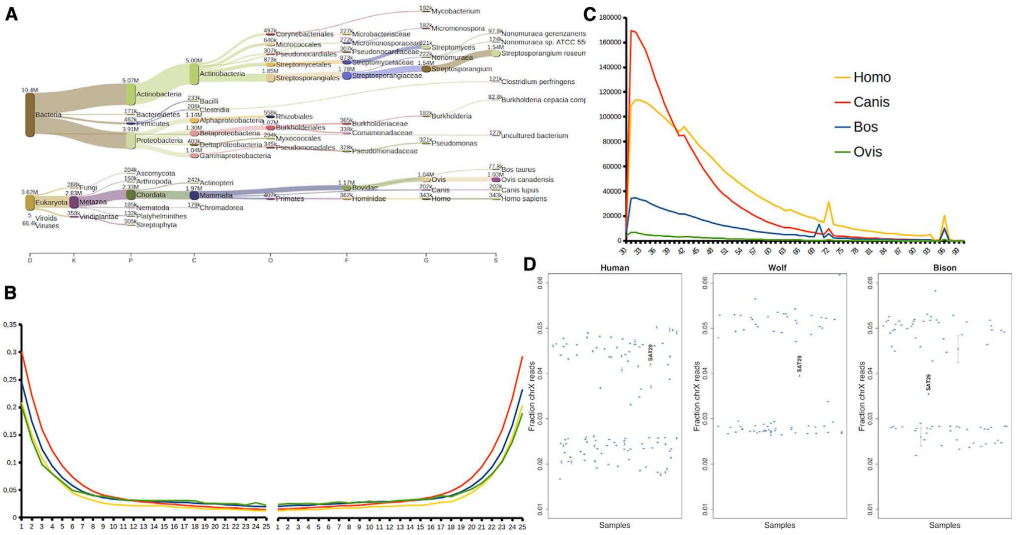


Figure 2. Sequencing data properties

(A) Metagenomic prediction with centrifuge.

(B) Deamination patterns for the filtered reads mapped against the four reference genomes: *Ovis aries*, *Homo sapiens*, *Bos taurus*, and *Canis lupus*.

(C) Fragment length distributions for the four aligned mammalian species.

(D) X chromosome read proportions. For each of the three species, the fraction of nuclear reads mapping to the X chromosome is displayed for the SAT29 sample as well as a number of previously published genomes for comparison. For wolf and bison, the comparative genomes are the modern and ancient genomes used for the ancestry analyses, while for human they are 101 ancient genomes from a study of the Eurasian steppe.[18] Bars denote 95% binomial confidence intervals.[19] See also Figures S1 and S6 and Data S1.

diversity in the sample (Figure S1C). Applying CALICO,[23] we estimated that 12% of the human mitochondrial reads and 25% of the wolf reads come from minority sources. Schmutzi estimated a lower fraction (1%, 0%–5%) of human minor sources. Finally, contamMix estimated a 4% minority fraction for the human reads and a 31% fraction for the wolf reads. ContamMix estimated no diversity in the bison data, but this could be limited by the low coverage of the mtDNA genome and the low number of potential contaminant mtDNA sequences. Based on the above, we can confirm the presence of DNA from more than a single individual for the wolf data, and likely also the human data, whereas the bison data are inconclusive. However, because of the complexity of the data and the limited coverage, the presented estimates of minority fractions should not be taken as more than rough indications.

Next, we assessed whether modern contamination could explain the detection of mitochondrial polymorphism in the sample (Table 1). Modern sequences are expected to be less fragmented, but we find no difference in the length distribution between deaminated and other reads (method details). Manual inspection of phylogenetically diagnostic positions suggests the presence of minority sources in the human data, including in the deaminated reads (Table S1). However, these variants are present at very low frequencies, in line with the estimated low proportion of minority sources (Tables 1 and S1). All human reads support haplogroup N, suggesting any minority sources would not be from haplogroup M and derivatives. Similar detailed inspection of the wolf data shows that the evidence of substantial polymorphism persists when restricting to deaminated reads (Figure S2B). The bison data are too limited to allow for a similar robust assessment. The diversity signal in the SAT29 sample is thus consistent with DNA deriving from multiple human and wolf individuals, rather than reflecting modern contamination. The contDeam.pl tool from Schmutzi estimates 0% contamination for all three species; in the three files, we have obtained values of 0, which indicates that there is no evidence of contamination based on deamination patterns.[21]

Finally, we examined the sex of the human, canid, and bison genomic data. When comparing the amount of reads mapping to the X chromosome relative to the autosomes, we find that the human data are consistent with deriving from a female individual, or multiple female individuals. In contrast, the wolf and bison X chromosome read fractions are intermediate between those expected for male and female karyotypes, suggesting it may derive from individuals of both sexes, again indicating multiple source individuals (Figure 2D).

Phylogenetic dating of mitochondrial DNA

With the consensus sequence of the human mtDNA reads, we performed a multiple sequence alignment and generated a Bayesian phylogenetic tree with BEAST2 (Figures 3C and S3). The SAT29 sequence is positioned within the diversity of

Table 1. Point estimates and confidence intervals (in parentheses) of minority mitochondrial sequence proportions obtained for the captured mitochondrial data for each of the three analyzed taxa using Schmutzi, ContamMix, and Calico

	Calico	ContamMix	Schmutzi
Homo sapiens	0.12 (0.03–0.21)	0.04 (0.01–0.1)	0.01 (0.0–0.05)
Canis lupus	0.246 (0.21–0.27)	0.31 (0.25–0.39)	–
Bison bonasus	–	0.01	–

See also Figure S2 and Table S1.

haplogroup N, close to the Dzudzuana-3 (25.5 kya) genome from Dzudzuana cave and basal to the modern samples enclosed in haplogroups N, X, and W. Haplogroup N originated outside of Africa from haplogroup L3 between 50 and 60 kya, and is common among present day Near Eastern populations but rare among present-day European populations.[24–26] We then estimated tip dates of the SAT29 human, wolf, and bison consensus mtDNA sequences. The human SAT29 mtDNA consensus has a mean age of 28,543 BP (95% HPD interval, 15,928–41,867 BP), whereas the wolf mtDNA consensus mean age is 28,257 BP (95% HPD interval, 18,083–38,265 BP). The bison mtDNA consensus has an age estimate of 21,928 BP (HPD interval of 14,954–27,987 BP). The phylogenetic positions of the SAT29 consensus sequences were also confirmed in maximum likelihood trees performed with the three datasets (Figure S3). The radiocarbon date of layer BIII of Satsurblia cave is 25.5.–24.2 ka cal. BP, and thus falls within the confidence intervals, and within a few thousand years of the mean estimates, of the mtDNA tip dates for all the three species. Although the mean estimate for the bison is somewhat younger than those for the human and wolves, the 95% confidence intervals overlap. This thus provides strong support for the Pleistocene origin of the recovered DNA.

Ancestry of the SAT29 human nuclear DNA

Using the 661,765 human nuclear reads, we genotyped 11,116 pseudo-haploid positions from the 1240K dataset.[27] To explore the human ancestry of SAT29 within the context of pre- and post-LGM diversity, we performed a principal components analysis (PCA) on 2,335 modern Eurasian genomes and projected 82 ancient individuals onto the resulting components (Figures 3A and S4; Table S2). Previous studies have revealed two different ancient human lineages from the Caucasus that were distinct from the rest of Pleistocene and early Holocene diversity. A late Upper Paleolithic (13.3 kya) genome from Satsurblia cave and a Mesolithic (9.7 kya) genome from the nearby cave of Kotias Klde revealed "Caucasus Hunter Gatherer" (CHG) ancestry, a distinct ancient lineage that split from western hunter-gatherers ~45 ka BP, shortly after the expansion of modern humans into Western Eurasia.[28] A second, older, pre-LGM lineage is represented by genome-wide data from two individuals dated to ~26 ka BP from Dzudzuana cave, southern Caucasus, and likely contributed at least half of the ancestry of later populations in Europe, the Near East, and North Africa.[29] We find that the SAT29 sample clusters with Dzudzuana2 in the PCA and not with the late Upper Paleolithic and Mesolithic genomes from the Caucasus or with any other published pre-LGM Eurasian genomes. Unsupervised ADMIXTURE clustering[30] further supports the similarity between SAT29 and Dzudzuana2 (Figures 3B and S4; Tables S2 and S3).

We used outgroup f_3-statistics to quantify the amount of shared genetic drift between SAT29 and other ancient genomes.[31] SAT29 shares more drift with Villabruna (Italy, 12,140 ± 70 bp)[32] and Dzudzuana2 than with other ancient individuals (Figure S4B), including the post-LGM individuals from the Caucasus (Satsurblia and Kotias). Among present-day Eurasian populations, SAT29 shows higher genetic affinity to Northern and Western Europeans rather than Central and South Asians (Figure S4C). Our results for the SAT29 human autosomal data are thus consistent with the results reported by Lazaridis et al.,[29] revealing a previously undocumented pre-LGM human ancestry from the Caucasus that contributed to various later Eurasian populations. The low coverage of the SAT29 environmental genome, however, did not allow us to further analyze possible differences in the ancestry between Dzuzuana2 and SAT29.

Next, we investigated if the amount of Neanderthal ancestry in the SAT29 human environmental genome could be estimated. Using an f_4-ratio,[31] we estimated 1% Neanderthal ancestry, with confidence intervals of 0%–6.6%. The point estimate is similar to that of Dzudzuana2 and likely lower than that of Paleolithic European and present-day West Eurasian populations due to dilution from large amounts of Basal Eurasian ancestry.[29] However, the large uncertainty of the estimate precludes any strong conclusions. Our results thus suggest that, unless substantially larger amounts of autosomal DNA can be recovered than what is analyzed here, sediment DNA is unlikely to allow for confident estimates of archaic ancestry proportions.

Ancestry of the SAT29 wolf nuclear DNA

We analyzed the 2,378,237 *Canis* reads using a set of variants among 722 modern wolves, dogs, and other canid species.[33] We randomly sampled one SAT29 read at each position, resulting in a genotype call at 439,426 transversion variants. In ADMIXTURE analyses (Figures 4A and S5) the SAT29 sample clusters with Eurasian wolves, and using f_4-statistics, we found that it clearly shares genetic drift with wolves and dogs, to the exclusion of coyotes, golden jackals, and other canids (Z >30 for all species, Method details). It does not, however, display stronger affinity to wolves over dogs or vice versa (Figure S5).

We next used admixture graphs to further investigate the relationship of the SAT29 environmental genome to present-day wolves and dogs, as well as two Pleistocene wolf genomes from Siberia (35–33 ka), which have ancestries that are basal to modern wolves and dogs.[34,35] We tested all possible topologies without admixture relating a coyote, SAT29, a modern wolf, a modern dog, and the two Pleistocene Siberian wolves, while explicitly accounting for reference bias in the ancient genomes (Method details). Only three of the 100 graphs provide good fits and feature the Siberian Pleistocene wolves on a branch basal to modern populations. The graphs fit equally well and differ only in that SAT29 is placed either basal to the Siberian Pleistocene branch, on this branch, or downstream of this branch (Figure 4B). Previous studies have found that present-day wolf population structure has mostly formed after the LGM.[23,34,36–39] Our results are consistent with this scenario because the SAT29 environmental genome harbored an

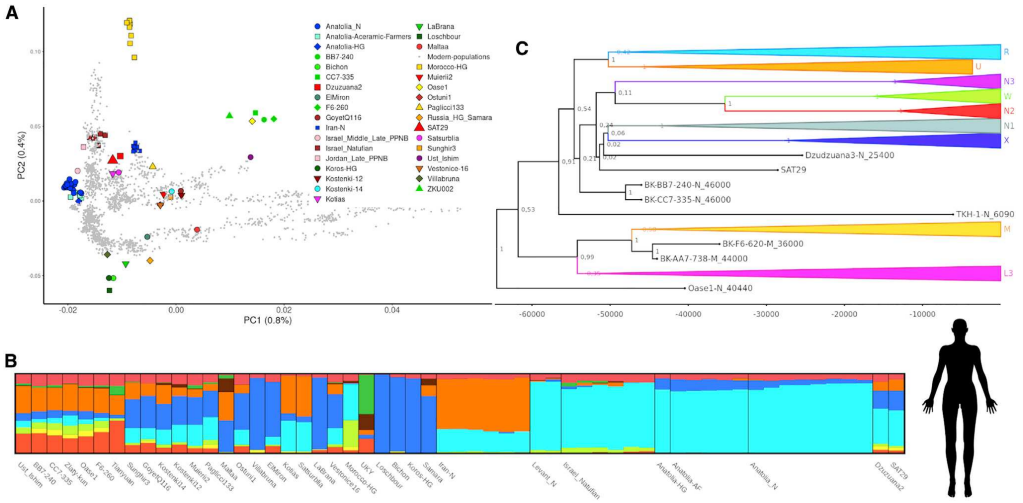


Figure 3. SAT29 human genomics

(A) A principal component analysis built with 2,335 modern Eurasian samples into which 82 ancient samples were projected. SAT29 appears closest to Dzudzuana2.

(B) In ADMIXTURE clustering with 82 ancient genomes, SAT29 displays a profile similar to that of Dzudzuana2.

(C) Bayesian tree built with 69 ancient genomes and 167 modern mitochondrial genomes. SAT29 falls close to the Dzudzuana2 genome and within the N haplogroup diversity. The values in the tree nodes represent the posterior probabilities.

See also Figures S3 and S4 and Tables S2 and S3.

ancestry that diverged from the ancestors of modern wolves and dogs before these diversified. Although Late Pleistocene wolves in the Caucasus were not closely related to those in Siberia, they thus had a similarly basal ancestry that has either gone extinct or been transformed by later population processes.

These autosomal ancestry results are consistent with the mitochondrial capture results, in which the SAT29 wolf consensus sequence falls on a branch together with two ancient wolves from the Aghitu-3 cave in Armenia (Figures 4C and S3; Table S4), dated to 31–30 ka, on a seemingly extinct West Eurasian branch of the wolf mtDNA phylogeny.[37] To further characterize the phylogenetic identity of the SAT29 consensus sequence, we recovered mtDNA from two wolf bones from Satsurblia cave through capture: a 4.9-fold genome from sample Y7d from layer B IVa (25.5–24.4 ka cal BP) and a 5.2-fold genome from sample T20a from layer A IIb (17.9–16.2 ka cal BP). These sequences fall close to SAT29 in the phylogeny, supporting the endogenous origin of the SAT29 environmental genome and demonstrating mitochondrial genetic continuity of wolf populations in the Caucasus for at least 10,000 years through the end of the LGM. When adding these sequences to the phylogenetic tip dating, the SAT29 age estimate comes out slightly younger (mean of 19,937). However, it is possible that the low quality of these two sequences hurts the accuracy of the inference in some way.

Ancestry of the SAT29 bison nuclear DNA

We compiled a number of Bovine genomes,[40–42] identified 1.4 million heterozygous transversion sites in a gaur genome, and assigned genotypes to all individuals at these sites by randomly sampling one read per position. This resulted in a genotype call at 27,724 transversion variants for the SAT29 bovid sample. Using ADMIXTURE clustering (Figure 5A) and f_4-statistics, we find that the SAT29 environmental genome shares genetic drift with bison (*Bison sp.*), to the exclusion of aurochs, domestic cattle, and Asian bovid species (Z >20 for all species, Method details). This is thus consistent with the mitochondrial sequence being of *Bison sp.* origin. It also provides important authentication of the soil metagenomic approach, because the identified environmental genome is a different species from the cattle (*Bos taurus*) used as a reference genome.

We next used f_4-statistics and admixture graphs to further investigate the relationship of the SAT29 environmental bison genome to present-day populations as well as early 20th century European bison (*Bison bonasus;* wisents) from Poland and the Caucasus. SAT29 is closer to these historical European individuals, as well as to modern European bison, than to a North American bison ($|Z| > 6$), implying that the divergence between European and American populations predates the age of SAT29. The Caucasian and the Polish populations have been classified as separate subspecies of European bison, but the SAT29 sample is not detectably closer to one over the other. Furthermore, these two recent populations share genetic drift to the exclusion of the SAT29 sample ($|Z| = 3.4$ and 4.5 for the two relevant f_4-statistics).

We tested all possible topologies without admixture to summarize the relationships between these bison genomes. The

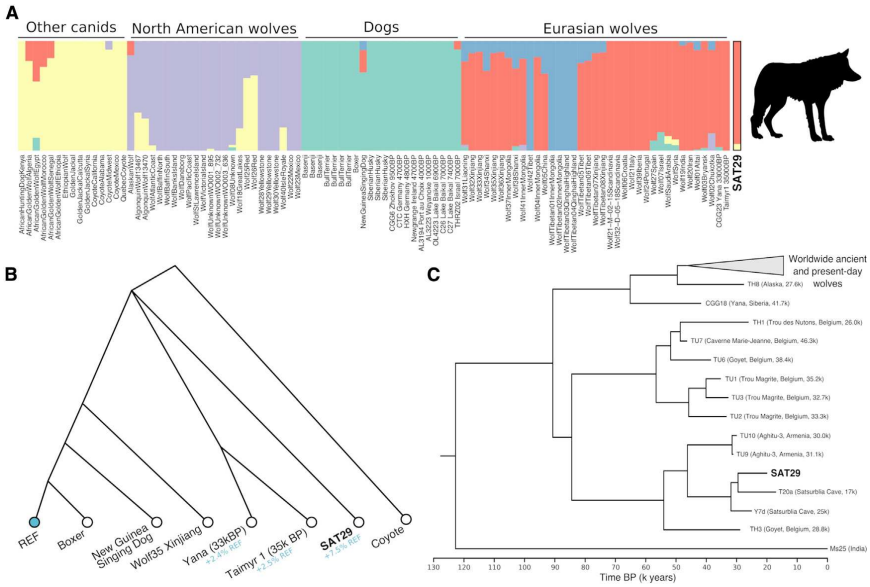


Figure 4. SAT29 wolf genomics

(A) In ADMIXTURE clustering with wolves, dogs, and other canid species, the SAT29 sample clusters with Eurasian wolves.

(B) Population history model relating the SAT29 sample to modern wolves, dogs, and Pleistocene Siberian wolves. Inferred admixture proportions from the dog reference genome (REF) to account for reference bias, are shown in blue. The trifurcation point indicates that it cannot confidently be determined on which side of this point the SAT29 sample falls.

(C) Bayesian tree of 68 modern and 39 ancient wolf mitochondrial genomes.

See also Figures S3 and S5 and Table S4.

best-fitting topology has SAT29 as basal to the historical genomes from Poland and the Caucasus, and the American bison as basal to all of these (Figure 5B). This model explains all f_4-statistics except a signal of some excess affinity between the American and Polish genomes. We do not attempt to discriminate between more complex models involving possible admixture events. Overall, the results tentatively suggest that the history of bison in western Eurasia might share some features with wolf history, in that Late Pleistocene ancestries appear basal to present-day populations, suggesting that population structure has been substantially reshaped since the LGM. This autosomal ancestry is also consistent with the mitochondrial phylogeny, in which the SAT29 bison consensus sequence falls on a branch with other ancient west Eurasian bison within a clade that has been called "Bb2,"[43] closest to the sequence of a Holocene Armenian individual (Figures 5C and S3; Table S5).[42]

We additionally recovered 72,100 reads aligning to *Ovis aries* and explored these reads with a dataset of 103 *Ovis* and *Capra* genomes from 10 species. Tests of the form f_4(Oreamnos,-SAT29;C,D) indicated that the SAT29 sample is more similar to *Ovis* than to *Capra*, but it does not cluster with any of present-day *Ovis* species (Figure S6). We were thus unable to elaborate on the ancestry of this *Ovis* DNA.

DISCUSSION

Our results demonstrate that unbiased shotgun sequencing of sediment ancient DNA can yield genome-wide data that is informative about the ancestry of several taxa. The DNA retrieved here is lower in quantity, and hence resolution, compared to what is often obtained from various well-preserved bones and teeth. Nonetheless, it provided information largely comparable to low-coverage ancient genome sequences and allowed us to apply complementary analyses of multiple mammalian species to reconstruct some aspects of their population histories. Our results thus point to new possibilities in the study of sediment ancient DNA, demonstrating that it can serve as an additional or alternative source of genome-wide information to skeletal remains.

The damage characteristics, the mitochondrial tip dates, and the fact that all three environmental genomes represent ancestries that no longer exist among the given species strongly support a Pleistocene origin of the recovered DNA. Moreover, the deamination patterns and fragment length distributions are similar for the human, wolf, and bison DNA. Additionally, the results of the population genetics analyses are in accordance with those published by other studies on skeletal genomes from the same region and time period.[29] These observations thus suggest

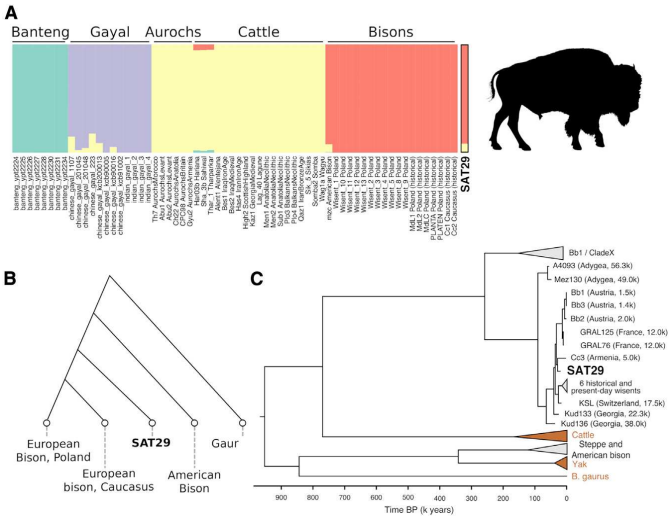


Figure 5. SAT29 bison genomics

(A) In ADMIXTURE clustering with cattle, aurochs, bison, gayals, and bantengs, the SAT29 sample clusters with bison.

(B) A population history model relating the SAT29 sample to present-day American and historical (early 20th century) European bison from Poland and the Caucasus.

(C) Bayesian tree built with SAT29 and other 70 bovid mitochondrial genomes.

See also Figure S3 and Table S5.

that the DNA has not been significantly affected by modern contamination or leaching through archaeological layers.

Although promising, the study of whole genomes from archeological sediments also has several limitations. We report evidence for the presence of sequences deriving from multiple individuals in our environmental genomes. The specific nature of environmental genomes, which in most cases likely compose a mosaic of several individuals from a given taxon, complicates certain analyses, in particular the reconstruction of mitochondrial consensus sequences. On the other hand, allele frequency-based analyses of genome-wide relationships are well-suited to sequences deriving from multiple individuals, as long as these are of largely the same ancestry. Despite these limitations, genome-wide ancient sediment DNA might open new directions for the study of whole ecosystems, including interactions between different species and aspects of human practices linked to the use of animals or plants.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archeological context
- METHOD DETAILS
  - DNA extraction
  - PCR, quality control and next generation sequencing
  - Mitochondrial capture
  - Human DNA screening
  - Bioinformatic processing of sample SAT29
  - Damage pattern analysis
  - BLASTN analysis
  - Human population genetics
  - Sex determination of SAT29 human reads
  - Neanderthal ancestry in SAT29
  - Human mitochondrial analysis
  - mtDNA mixture estimate
  - Presence of multiple individuals identification
  - mtDNA consensus calling
  - Human mitochondrial tip dating
  - Wolf genome: comparative dataset
  - Wolf genome: ancestry analyses
  - Canid bone testing
  - Canid mitochondrial tip dating
  - Presence of multiple individuals
  - Bison genome: comparative dataset
  - Bison genome: ancestry analyses
  - Bison mitochondria capture and analysis
  - Ovis genomic analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.06.023.

ACKNOWLEDGMENTS

This study makes use of data generated by the NextGen Consortium. We acknowledge Gabriel Renaud for advice on Schmutzi and on the analysis of the mitochondrial genome. We acknowledge Spencer Sawyer, Manuela Alscher, and Odin for modern DNA. We acknowledge David Reich and Iosif Lazaridis for sharing the data of the Dzudzuana2 genome and helping us with its analysis. This work has been supported by the "Mineralogical Preservation of the Human Biome from the Depth of Time" (MINERVA) research platform, code AGB326800, from the University of Vienna. P.S. was supported by the European Research Council (852558), a Wellcome Trust Investigator

award (217223/Z/19/Z), and the Vallee Foundation. P.S. and A.B. were supported by Francis Crick Institute core funding (FC001595) from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust. This research was funded in whole, or in part, by the Wellcome Trust (FC001595). For the purpose of open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission T.C.C. was supported by the Medical Trainee PhD Scholarship, UCD.

AUTHOR CONTRIBUTIONS

R.P., S.S., and P.G. conceived the study. A.B.-C., D.L., T.M., N.J., Z.M., and G.B.-O. provided samples. M.C.S. collected the samples. S.S., O.C., D.F., T.C.C., V.O., K.T.Ö., and R.N.M.F. performed the experimental work. P.G., S.S., A.B., A.M., and T.C.C. analyzed the data. P.G., S.S., A.B., R.P., and P.S. wrote the manuscript with inputs from all coauthors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 6, 2021
Revised: April 23, 2021
Accepted: June 9, 2021
Published: July 12, 2021

REFERENCES

1. Hagelberg, E., Sykes, B., and Hedges, R. (1989). Ancient bone DNA amplified. Nature *342*, 485.
2. Höss, M., Dilling, A., Currant, A., and Pääbo, S. (1996). Molecular phylogeny of the extinct ground sloth Mylodon darwinii. Proc. Natl. Acad. Sci. USA *93*, 181–185.
3. Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., et al. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science *317*, 1927–1930.
4. Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H., and Lalueza-Fox, C. (2020). Beyond broad strokes: sociocultural insights from the study of ancient genomes. Nat. Rev. Genet. *21*, 355–366.
5. Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. Nat. Rev. Genet. *2*, 353–359.
6. Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., and Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science *300*, 791–795.
7. Hofreiter, M., Mead, J.I., Martin, P., and Poinar, H.N. (2003). Molecular caving. Curr. Biol. *13*, R693–R695.
8. Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L.Z., Beaudoin, A.B., Zutter, C., Larsen, N.K., et al. (2016). Postglacial viability and colonization in North America's ice-free corridor. Nature *537*, 45–49.
9. Søe, M.J., Nejsum, P., Seersholm, F.V., Fredensborg, B.L., Habraken, R., Haase, K., Hald, M.M., Simonsen, R., Højlund, F., Blanke, L., et al. (2018). Ancient DNA from latrines in Northern Europe and the Middle East (500 BC-1700 AD) reveals past parasites and diet. PLoS ONE *13*, e0195481.
10. Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature *506*, 47–51.
11. Graham, R.W., Belmecheri, S., Choy, K., Culleton, B.J., Davies, L.J., Froese, D., Heintzman, P.D., Hritz, C., Kapp, J.D., Newsom, L.A., et al. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. Proc. Natl. Acad. Sci. USA *113*, 9310–9314.
12. Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Sarkissian, C.D., Haile, J., Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E., et al. (2015). Ancient and modern environmental DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130383.
13. Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., et al. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. Science *356*, 605–608.
14. Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., Jacobs, Z., Dai, Q., Massilani, D., et al. (2020). Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science *370*, 584–587.
15. Vernot, B., Zavala, E.I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., et al. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science *372*, eabf1667.
16. Pinhasi, R., Meshveliani, T., Matskevich, Z., Bar-Oz, G., Weissbrod, L., Miller, C.E., Wilkinson, K., Lordkipanidze, D., Jakeli, N., Kvavadze, E., et al. (2014). Satsurblia: new insights of human response and survival across the Last Glacial Maximum in the southern Caucasus. PLoS ONE *9*, e111271.
17. Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.
18. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.
19. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.
20. Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.
21. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.
22. Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., and Pääbo, S. (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature *505*, 403–406.
23. Bergström, A., Frantz, L., Schmidt, R., Ersmark, E., Lebrasseur, O., Girdland-Flink, L., Lin, A.T., Storå, J., Sjögren, K.-G., Anthony, D., et al. (2020). Origins and genetic legacy of prehistoric dogs. Science *370*, 557–564.
24. Soares, P., Alshamali, F., Pereira, J.B., Fernandes, V., Silva, N.M., Afonso, C., Costa, M.D., Musilová, E., Macaulay, V., Richards, M.B., et al. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. Mol. Biol. Evol. *29*, 915–927.
25. Fernandes, V., Alshamali, F., Alves, M., Costa, M.D., Pereira, J.B., Silva, N.M., Cherni, L., Harich, N., Cerny, V., Soares, P., et al. (2012). The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. Am. J. Hum. Genet. *90*, 347–355.
26. Drake, N.A., Blench, R.M., Armitage, S.J., Bristow, C.S., and White, K.H. (2011). Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. Proc. Natl. Acad. Sci. USA *108*, 458–462.
27. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.
28. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.
29. Lazaridis, I., Belfer-Cohen, A., Mallick, S., Patterson, N., Cheronet, O., Rohland, N., Bar-Oz, G., Bar-Yosef, O., Jakeli, N., Kvavadze, E., et al.

(2018). Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry. bioRxiv, 423079.

30. Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics *12*, 246.
31. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
32. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
33. Plassais, J., Kim, J., Davis, B.W., Karyadi, D.M., Hogan, A.N., Harris, A.C., Decker, B., Parker, H.G., and Ostrander, E.A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nat. Commun. *10*, 1489.
34. Skoglund, P., Ersmark, E., Palkopoulou, E., and Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr. Biol. *25*, 1515–1519.
35. Sinding, M.S., Gopalakrishnan, S., Ramos-Madrigal, J., de Manuel, M., Pitulko, V.V., Kuderna, L., Feuerborn, T.R., Frantz, L.A.F., Vieira, F.G., Niemann, J., et al. (2020). Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. Science *368*, 1495–1499.
36. Freedman, A.H., Gronau, I., Schweizer, R.M., Ortega-Del Vecchyo, D., Han, E., Silva, P.M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., et al. (2014). Genome sequencing highlights the dynamic early history of dogs. PLoS Genet. *10*, e1004016.
37. Loog, L., Thalmann, O., Sinding, M.S., Schuenemann, V.J., Perri, A., Germonpré, M., Bocherens, H., Witt, K.E., Samaniego Castruita, J.A., Velasco, M.S., et al. (2020). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. Mol. Ecol. *29*, 1596–1610.
38. Ramos-Madrigal, J., Sinding, M.-H.S., Carøe, C., Mak, S.S.T., Niemann, J., Samaniego Castruita, J.A., Fedorov, S., Kandyba, A., Germonpré, M., Bocherens, H., et al. (2021). Genomes of Pleistocene Siberian Wolves Uncover Multiple Extinct Wolf Lineages. Curr. Biol. *31*, 198–206.e8.
39. Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A.S., Schweizer, R.M., Ramirez, O., Pollinger, J., Galaverni, M., Ortega Del-Vecchyo, D., et al. (2016). Worldwide patterns of genomic variation and admixture in gray wolves. Genome Res. *26*, 163–173.
40. Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J.M., Zhang, Y., Tokarska, M., Li, Y., Wang, M.-S., Faruque, O., Nielsen, R., et al. (2018). Pervasive introgression facilitated domestication and adaptation in the Bos species complex. Nat. Ecol. Evol. *2*, 1139–1145.
41. Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., et al. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science *365*, 173–176.
42. Węcek, K., Hartmann, S., Paijmans, J.L.A., Taron, U., Xenikoudakis, G., Cahill, J.A., Heintzman, P.D., Shapiro, B., Baryshnikov, G., Bunevich, A.N., et al. (2017). Complex Admixture Preceded and Followed the Extinction of Wisent in the Wild. Mol. Biol. Evol. *34*, 598–612.
43. Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E.A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., et al. (2016). Past climate changes, population dynamics and the origin of Bison in Europe. BMC Biol. *14*, 93.
44. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.
45. Hannon, G.J. (2010). FASTX-Toolkit.
46. Simpson, J.T., and Durbin, R. (2010). Efficient construction of an assembly string graph using the FM-index. Bioinformatics *26*, i367–i373.
47. Breitwieser, F.P., and Salzberg, S.L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics *36*, 1303–1304.
48. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
49. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
50. Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics *32*, 292–294.
51. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
52. Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., et al. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nat. Commun. *7*, 10408.
53. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
54. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
55. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
56. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823.
57. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. *17*, 377–386.
58. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
59. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.
60. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44* (W1), W58-63.
61. Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.
62. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
63. Kalandadze, A.N., and Kalandadze, K.S. (1978). Archaeological Research of Karstic Caves in Tskaltubo region (in Georgian, with Russian summary). Caves of Georgia, 116–136.
64. Collin, T.C. (2019). Development and Application of a Metagenomic Ancient DNA Approach for the Identification and Assessment of Taxa from Anthropogenic Sediments: Reconstructing the Past. PhD thesis (University College Dublin).
65. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.
66. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, db.prot5448.
67. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders,

A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

68. Gunnarsdóttir, E.D., Li, M., Bauchet, M., Finstermeier, K., and Stoneking, M. (2011). High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. *21*, 1–11.

69. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE *7*, e34131.

70. Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpré, M.B., Sablin, M.V., López-Giráldez, F., Domingo-Roura, X., et al. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science *342*, 871–874.

71. Feuerborn, T.R., Palkopoulou, E., van der Valk, T., von Seth, J., Munters, A., Pečnerová, P., Dehasque, M., Ureña, I., Ersmark, E., Lagerholm, V.K., et al. (2020). Competitive mapping allows to identify and exclude human DNA contamination in ancient faunal genomic datasets. bioRxiv. https://doi.org/10.1186/s12864-020-07229-y.

72. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

73. Andrews, S. (2010). A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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

75. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

76. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.

77. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., and Fu, Q. (2017). 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. Curr. Biol. *27*, 3202–3208.e9.

78. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

79. van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., et al. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.

80. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.

81. Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W.K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225–228.

82. Yu, H., Spyrou, M.A., Karapetian, M., Shnaider, S., Radzeviciūtė, R., Nägele, K., Neumann, G.U., Penske, S., Zech, J., Lucas, M., et al. (2020). Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. Cell *181*, 1232–1245.e20.

83. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

84. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

85. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

86. Hajdinjak, M., Mafessoni, F., Skov, L., Vernot, B., Hübner, A., Fu, Q., Essel, E., Nagel, S., Nickel, B., Richter, J., et al. (2021). Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature *592*, 253–257.

87. Prüfer, K., Posth, C., Yu, H., Stoessel, A., Spyrou, M.A., Deviese, T., Mattonai, M., Ribechini, E., Higham, T., Velemínský, P., et al. (2021). A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia. Nat. Ecol. Evol. *5*, 820–825.

88. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

89. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture across inner Eurasia. Nat. Ecol. Evol. *3*, 966–976.

90. Renaud, G. (2018). libbam (Github).

91. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

92. Gilbert, M.T.P., Kivisild, T., Grønnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Götherström, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science *320*, 1787–1789.

93. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.

94. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.

95. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016). Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. Curr. Biol. *26*, 827–833.

96. Bollongino, R., Nehlich, O., Richards, M.P., Orschiedt, J., Thomas, M.G., Sell, C., Fajkosová, Z., Powell, A., and Burger, J. (2013). 2000 years of parallel societies in Stone Age Central Europe. Science *342*, 479–481.

97. Sánchez-Quinto, F., Schroeder, H., Ramirez, O., Avila-Arcos, M.C., Pybus, M., Olalde, I., Velazquez, A.M.V., Marcos, M.E.P., Encinas, J.M.V., Bertranpetit, J., et al. (2012). Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr. Biol. *22*, 1494–1499.

98. Benazzi, S., Slon, V., Talamo, S., Negrino, F., Peresani, M., Bailey, S.E., Sawyer, S., Panetta, D., Vicino, G., Starnini, E., et al. (2015). Archaeology. The makers of the Protoaurignacian and implications for Neandertal extinction. Science *348*, 793–796.

99. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., and Rollo, F. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. Curr. Biol. *18*, 1687–1693.

100. Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., and Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature *464*, 894–897.

101. Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınç, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol. *16*, e2003703.

102. Vai, S., Sarno, S., Lari, M., Luiselli, D., Manzi, G., Gallinaro, M., Mataich, S., Hübner, A., Modi, A., Pilli, E., et al. (2019). Ancestral mitochondrial N lineage from the Neolithic 'green' Sahara. Sci. Rep. *9*, 3530.

103. Hublin, J.-J., Sirakov, N., Aldeias, V., Bailey, S., Bard, E., Delvigne, V., Endarova, E., Fagault, Y., Fewlass, H., Hajdinjak, M., et al. (2020). Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. Nature *581*, 299–302.

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

105. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

106. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.

107. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

108. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534.

109. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650.

110. Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L.A., Warmuth, V., Ruiz-Linares, A., Manica, A., and Balloux, F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. Mol. Biol. Evol. *31*, 2780–2792.

111. Rambaut, A. (2018). FigTree.

112. Gopalakrishnan, S., Sinding, M.S., Ramos-Madrigal, J., Niemann, J., Samaniego Castruita, J.A., Vieira, F.G., Carøe, C., Montero, M.M., Kuderna, L., Serres, A., et al. (2018). Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr. Biol. *28*, 3441–3449.e5.

113. Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., and Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. Nat. Ecol. Evol. *2*, 124–131.

114. Sinding, M.S., Gopalakrishan, S., Vieira, F.G., Samaniego Castruita, J.A., Raundrup, K., Heide Jørgensen, M.P., Meldgaard, M., Petersen, B., Sicheritz-Ponten, T., Mikkelsen, J.B., et al. (2018). Population genomics of grey wolves and wolf-like canids in North America. PLoS Genet. *14*, e1007745.

115. Liu, Y.-H., Wang, L., Xu, T., Guo, X., Li, Y., Yin, T.-T., Yang, H.-C., Hu, Y., Adeola, A.C., Sanke, O.J., et al. (2018). Whole-Genome Sequencing of African Dogs Provides Insights into Adaptations against Tropical Parasites. Mol. Biol. Evol. *35*, 287–298.

116. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.

117. Leppälä, K., Nielsen, S.V., and Mailund, T. (2017). admixturegraph: an R package for admixture graph manipulation and fitting. Bioinformatics *33*, 1738–1740.

118. Skoglund, P., Northoff, B.H., and Shunkov, M.V. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

119. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

120. Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., Stephenson, J., Allendorf, F.W., Hogg, J.T., and Kijas, J. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol. Ecol. *24*, 5616–5632.

121. Yang, Y., Wang, Y., Zhao, Y., Zhang, X., Li, R., Chen, L., Zhang, G., Jiang, Y., Qiu, Q., Wang, W., et al. (2017). Draft genome of the Marco Polo Sheep (Ovis ammon polii). Gigascience *6*, 1–7.

122. Upadhyay, M., Hauser, A., Kunz, E., Krebs, S., Blum, H., Dotsev, A., Okhlopkov, I., Bagirov, V., Brem, G., Zinovieva, N., and Medugorac, I. (2020). The First Draft Genome Assembly of Snow Sheep (Ovis nivicola). Genome Biol. Evol. *12*, 1330–1336.

123. Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., et al. (2020). The origin of domestication genes in goats. Sci. Adv. *6*, eaaz5216.

124. Grossen, C., Guillaume, F., Keller, L.F., and Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat. Commun. *11*, 1001.

125. Alberto, F.J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. Nat. Commun. *9*, 813.

126. Martchenko, D., Chikhi, R., and Shafer, A.B.A. (2020). Genome Assembly and Analysis of the North American Mountain Goat (*Oreamnos americanus*) Reveals Species-Level Responses to Extreme Environments. G3 (Bethesda) *10*, 437–442.

127. The GIMP Development Team (2019). Gimp. https://www.gimp.org/.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Soil from Satsurblia cave	This Study	SAT22
Soil from Satsurblia cave	This Study	SAT23
Soil from Satsurblia cave	This Study	SAT26
Soil from Satsurblia cave	This Study	SAT27
Soil from Satsurblia cave	This Study	SAT28
Soil from Satsurblia cave	This Study	SAT29
Canis lupus bone from Satsurblia cave	This Study	Y7a
Canis lupus bone from Satsurblia cave	This Study	Y7d
Canis lupus bone from Satsurblia cave	This Study	AA8c
Canis lupus bone from Satsurblia cave	This Study	T24c
Canis lupus bone from Satsurblia cave	This Study	T20a
Critical commercial assays		
NextSeq 500/550 (75 cycle)	Illumina	TG-160-2005
DNeasy Blood and Tissue kit	QIAGEN	69506
Quick Blunting system	NEB	E1201S
Quick Ligation kit	NEB	M2200S
Expand Long Template PCR System	Roche	11681834001
AccuPrimePfx DNA Polymerase	Invitrogen	12344024
MinElute PCR Purification Kit	QIAGEN	28006
Qubit dsDNA HS Assay Kit	Invitrogen	Q32851
Agilent DNA 1000 Kit	Agilent	5067-1504
Deposited data		
Raw analyzed data and filtered genomes	N/A	ENA: PRJEB41420
Software and algorithms		
Cutadapt 2.7	[18]	https://cutadapt.readthedocs.io/en/stable/
FASTX-toolkit 0.0.1	[19]	http://hannonlab.cshl.edu/fastx_toolkit/
SGA	[44]	https://bioinformaticshome.com/tools/wga/descriptions/SGA.html
Centrifuge 1.0.3	[17]	https://ccb.jhu.edu/software/centrifuge/
Pavian	[45]	https://ccb.jhu.edu/software/pavian/
BWA 0.7.16	[46]	http://bio-bwa.sourceforge.net/bwa.shtml

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Samtools 1.10	46	http://samtools.github.io/bcftools/bcftools.html
Bedtools 2.29.2	47	https://bedtools.readthedocs.io/en/latest/
Qualimap 2.2.1	48	http://qualimap.conesalab.org/
Mapdamage 2.0.9	49	https://ginolhac.github.io/mapDamage/
SequenceTools	50	https://github.com/stschiff/sequenceTools
Admixtools 5.1	31	https://github.com/DReichLab/AdmixTools
Eigensoft 7.2.1	51	https://github.com/DReichLab/EIG
ADMIXTURE 1.3.0	52	https://bioinformaticshome.com/tools/descriptions/ADMIXTURE.html
PLINK 1.9	53	https://zzz.bwh.harvard.edu/plink/
PONG 1.4.9	54	https://github.com/ramachandran-lab/pong
ry_compute	55	https://github.com/pontussk/ry_compute/blob/master/ry_compute.py
MEGAN 6.19.9	56	https://software-ab.informatik.uni-tuebingen.de/download/megan6/welcome.html
BLAST+ 2.10	57	https://ncbiinsights.ncbi.nlm.nih.gov/2019/12/18/blast-2-10-0/
Schmutzi	21	https://github.com/grenaud/schmutzi/blob/master/.gitmodules
Bamutil 1.0.14	58	https://genome.sph.umich.edu/wiki/BamUtil
ContamMix v1.0.10	32	N/A
Calico 0.2	23	https://github.com/pontussk/calico
Geneious 8.1		http://assets.geneious.com/
Haplogrep 2.0	59	https://haplogrep.i-med.ac.at/
BEAST 1.8.4	60	https://beast.community/2016-06-17_BEAST_v1.8.4_released.html
Figtree v1.4.4	N/A	http://tree.bio.ed.ac.uk/software/figtree/
Picard Tools 2.21.4	N/A	https://broadinstitute.github.io/picard/
GATK 4.1.4.0	61	https://gatk.broadinstitute.org/hc/en-us
samtools 1.9	46	http://samtools.github.io/bcftools/bcftools.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
htsbox pileup r345	N/A	https://github.com/lh3/htsbox
admixturegraph R package	N/A	https://cran.r-project.org/web/packages/admixturegraph/index.html
PMDtools	N/A	https://github.com/pontussk/PMDtools
fastp	[62]	https://github.com/OpenGene/fastp

RESOURCE AVAILABILITY

Lead contact

Further information on materials, datasets, and protocols should be directed to and will be fulfilled by the Lead Contact, Pere Gelabert (pere.gelabert@univie.ac.at).

Materials availability

The raw genomic data used in all the analyses can be accessed at the European Nucleotide Archive (ENA) under ENA: PRJEB41420.

Data and code availability

Sequencing data and the filtered sequences are available at the European Nucleotide Archive (ENA) under ENA: PRJEB41420. All code used in this study and other previously published genomic data is available at the sources referenced in the key resource table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We have studied several genomic sequences from soil samples and *C. lupus* bone samples from Satsurblia cave in Georgia.

Archeological context

Satsurblia Cave is located in Georgia in the Southern Caucasus. The cave was discovered in 1975 by N. Kalandadze[63] and was excavated in: 1976, 1985–1988, 2008–2010, 2012–2013, and 2016–2017. Here we present results from those excavated during the 2016 campaign.

Fieldwork in Satsurblia cave focused on excavations in two areas: Area A in the north-western part of the cave, near the entrance, and Area B in the south-west, at the rear part of the cave. The stratigraphic sequence in Area A yielded a wealth of cultural and environmental remains from three main layers: A/I (Eneolithic, 5th millennium BCE), A/II (17.5-16.5 ka cal. BP), and A/III (25- 24 ka Cal. BP). Both areas yielded a series of *in situ* well preserved Upper Palaeolithic occupational layers with f *in situ* living surfaces (“floors”), some of them with preserved fireplaces and rich material culture assemblages. The sequence of Area B is divided into five main archaeological layers and encompasses deposits dated to several UP phases: 13 ka Cal. BP (upper part of Layer B/II); 19 ka Cal. BP (lower part of Layer B/II), 25-24 ka Cal. BP (lower B/II, B/III, B/IV); and 32-31 ka Cal. BP (Layer B/V).[16] A previous genome was published from remains located in square Y5, area A. The radiocarbon dating of this bone is 11,415 ± 50 uncal. bp (OxA-34632).[28] In addition to this bone, fieldwork only recovered one tooth and a fragmented ulna from layers dated between 13-15 ka Cal. BP. H findings were isolated finds and are not part of any burial contexts s. No human remains were found so far from any of the LGM and pre-LGM layers.

Sediment samples were removed with a knife from the exposed profile sections in association with micromorphological block samples and were then stored in zip-up bags and protected from light. Six samples from different layers were sequenced and examined (Data S1A). Here we report genomic data retrieved from sediment sample SAT29, which was taken from Area B, Layer BIII, square X4 (Figure 1B). Micromorphological analyses were performed on block sample SAT 15 14, taken next to SAT29, to investigate the formation processes and post-depositional alterations of layer B/III here, and to assess the integrity of the recovered aDNA and their potential source(s). The following presents preliminary results of these analyses (see Figure 1C). Natural processes include the weathering of the limestone bedrock with the deposition of limestone clasts and calcareous clay and silt; the deposition of rounded, cross-striated soil aggregates that originate from outside the cave; and the redistribution of clay and their deposition as coatings in large voids by percolating water. Both soil aggregates and clay coatings are connected to repeated water activity leaving the associated clay as an unlikely candidate for lasting DNA absorption in this context. Additionally, soil aggregates show variable color expressions, which result from exposure to heat at different temperatures, again making these aggregates an unlikely source for the recovered aDNA. Heating of soil aggregates and other sediment components results from UP people building combustion features at the site and distributing the combusting residues by trampling and dumping behaviors. Fire use and residues resulting from this

behavior - charcoal, ash, burnt sediments and bones - present the dominant anthropogenic component. However, it needs to be noted here that not all bones show heating traces and the heating traces are often limited to charring and low temperature heating. Microscopic bones are sand to gravel size and the most common anthropogenic component in this layer and present a potential source for the recovered aDNA. However, further research into the adsorption and preservation of DNA in archaeological sediments is needed.

METHOD DETAILS

DNA extraction

DNA extraction, library preparation, and indexing steps were undertaken in a dedicated aDNA facility within University College Dublin (UCD). All steps were undertaken within a grade B (EU) clean room under grade A unilateral air-flow hoods. Tyvek suits, hair nets, face masks, and nitrile gloves were used to limit contamination. Extraction of soil DNA was performed using 50 mg of soil in an extraction buffer[64] (to final concentration of 0.45M EDTA, 0.02M Tris-HCl (pH 8.0), 0.025% SDS, 0.5mg/mL Proteinase K and dH2O up to final volume). Samples were incubated at 37°C overnight within Matrix E lysing tubes (MPBIO116914) and using an Eppendorf Thermomixer® C with a rotational speed of 1600rpm. Samples were then cleaned according to the method outlined by Dabney et al.[65] and eluted using TET buffer. DNA libraries of the entire extract were prepared using the method outlined by Meyer and Kircher et al.[66] to produce 25uL of the library. Extraction and library negative controls were utilized using 50μl of deionised water.

PCR, quality control and next generation sequencing

Polymerase Chain Reaction (PCR) amplification and all subsequent steps were undertaken in a grade C laboratory due to increased sample stability. Amplification of 5uL of each library was performed using at a rate of 15 cycles and a single index was added onto the P7 end during amplification.[67] The amplified DNA was cleaned using PB and PE buffers (QIAGEN 28006). Concentration and molarity (nmol/L) of the working solution were ascertained through Agilent 2100 bioanalyzer and a Qubit4 for fluorometric quantification following manufacturer guidelines. Sequencing was undertaken at UCD Conway Institute of biomolecular and biomedical Research on an Illumina NextSeq 500/550 using the high output v2 (75 cycle) reagent kit (Illumina TG-160-2005). Further sequencing was performed on NovaSeq platforms.

Mitochondrial capture

Mitochondrial capture of both the human and canid mtDNA sequences were performed using the method outlined in Maricic et al.[20] Briefly, 50uL of modern human or dog blood were used to extract DNA using the QIAGEN Blood and Tissue kit. The modern DNA of each species was used for a long-range PCR (Sigma Aldrich Expand Long Template PCR System). Two primer pairs were used for the human mtDNA amplification[68] and for the bovine mtDNA amplification[69] and three primer pairs for the dog mtDNA amplification.[70] The long mtDNA fragments were sheared using a sonicator for eight 15-minute sessions and DNA was checked on a 2% agarose gel to make sure that the DNA was fragmented to below 1Kb in length. Next, fragmented DNA was blunt ended using NEB Quick Blunting system and the BioT/B adapters were ligated to the blunted fragments using the NEB Quick Ligation system to produce the bait for capture.

The single-indexed amplified SAT29 library was re-amplified using Accuprime pfx polymerase and the IS5/IS6 primer pairs[66] for 20 cycles and the concentration was measured on the Qubit 4.0. Before capture, the blocking oligonucleotides BO4, BO6, BO8 and BO10 were used to block the sequencing primer sites. Subsequently the bait and pool were combined and incubated at 65°C for two nights. The enriched DNA was melted off the baits using a 2% NaOH solution and the purified DNA was measured using qPCR to determine the ideal cycle number for amplification. The amplified capture libraries were measured on the Qubit 4.0 and Agilent 2100 bioanalyzer to determine concentration and subsequently sequenced on the Illumina Novaseq system.

Human DNA screening

We first explored the six libraries for the presence of human DNA. We obtained an average of 14,893,925 reads in each sequencing library. These reads were processed with the methodology described in Collin 2019.[64] This initial screening showed that five samples exhibited only residual presence of human DNA. The sixth sample, SAT29, had 0.03% of the total reads sequenced map to the human genome. Therefore, this sample was selected for further sequencing. Sequencing and classification results of these samples are presented in Data S1A.

Bioinformatic processing of sample SAT29

After sequencing the SAT29 library to saturation and merging the sequenced reads with the ones from the screening phase, we obtained a total of 561,263,536 reads. These were clipped using Cutadapt 2.7,[44] removing the sequencing adapters and the reads with poly-A tails (reads with more than four As).[44] We also removed reads with qualities below 30 of bases in at least 75% of the read bases with the FASTX-toolkit 0.0.1.[45] Clipped reads were processed with SGA[46] and redundant reads were removed disabling the kmer check. Finally, two bases per end were trimmed and reads shorter than 30bp were discarded using the FASTX-toolkit.[45] Once collapsed and filtered, we obtained a total of 226,880,778 reads that were used for further analyses. We used Centrifuge 1.0.3[17] with default parameters to classify the sequenced reads into taxa using the whole non-redundant nucleotide database from NCBI indexed following the Centrifuge manual and plotted using Pavian.[47] The classification showed the presence of four mammalian

taxa with more than 2% of the Eukaryotic classified reads, which we investigated in further analyses: *Ovis aries*, *Bos* taurus, *Homo sapiens and Canis lupus*. To separate the sequencing reads of the four major mammalian taxa we built a multi-fasta reference file with the genomes of: *H. sapiens* (GRCh37 Assembly GCA_000001405.1), O. aries (Oar_v3.1, assembly, GCA_000298735.1), B. taurus (ARS-UCD1.2 assembly, GCA_002263795.2) and C. lupus (CanFam3.1 assembly, GCF_000002285.3) following a similar strategy described in Feuerborn et al.[71] The filtered reads were aligned with bwa aln[48] disabling seeding, and with a gap open penalty of two. Only reads with mapping qualities above 30 were kept using Samtools 1.10.[72] Duplicated sequences were removed with picard 2.21.4. (Picard-tools., 2018) In total 4,956,676 reads were assigned to these species (661,765 *H. sapiens*, 2,378,237 *C. lupus*, 1,811,555 B. taurus and 72,100 *O. aries*) The characteristics and quality of the mapped reads was assessed with qualimap 2.2.1.[50] We determined the length distribution with fastqc[73] and assessed the level of damage with mapdamage 2.0.9.[51] Although two bases per end were clipped the deamination values are notably high: 3′ deamination of 23% in *Bos*, 28% in *Canis*, 20% in *Homo* and 19% in *Ovis* and 5′ deamination of 25% in *Bos*, 30% in *Canis*, 21% in *Homo* and 20% in Ovis. The distribution of these values along the sequence is presented in Figure 2B.

Damage pattern analysis

Reads mapping to all four species have fragment lengths and deamination patterns typical of ancient DNA (Figures 2B and 2C), but show some slight variability between species (Figure 2C). We examined the relationship between the read length and the deamination values. For doing that we have separated the reads by read length (from 30 to 70 bp) and from (70 to 100 bp) and we examined the deamination in these groups. We have seen that: Human short reads show a deamination of 21% and 19% in the short and long reads 3′ end respectively, that Canis reads show a deamination of 28% and 26% in the short and long reads 3′ end respectively and Bos shows a deamination of 23% and 20% in the short and long reads 3′ end respectively.

BLASTN analysis

In order to check the accuracy of the mapping results we used BLASTN+ from ncbiblastplus 2.11.0,[74] using the whole nt database from NCBI. We excluded the taxa from *Canis, Homo, Ovis and Bos* genera using the flag -negative_seqidlist and setting a minimum expected value (-evalue) of 1e-6 to check the accuracy of our assignment. The results were imported into MEGAN 6.21.1,[57] taxa identification was performed using a lowest common ancestor (LCA) value of 5% of assigned reads.

The 4,956,676 aligned reads were examined with BLAST+ and MEGAN. To prove the accuracy of the mapping, we excluded the four previous cited genera in the analysis. Out of the 4,956,676 aligned reads, 453,049 were assigned by MEGAN using a lowest common ancestor (LCA) score of 1 and a minimum support percent of 5%. 42,958 reads have been assigned to *Pan troglodytes,* representing the 9,5% of the assigned reads. The summed reads assigned to the *Simian* infraorder is 80,988 (18%) of the assigned reads. 79,470 reads (17,5%) have been assigned to the infraorder *Pecora,* which includes *Bison, Bos, Ovis and Capra genera.* 49,426 (11%) reads have been assigned to the suborder *Caniformia*. Other taxa, except these, are *Sus scrofa* with 6,745 reads and *Felis catus* with 10,931 reads. The other identified taxa are represented with less than 1,000 reads. These results show that the mapping process has been selective and the reads have been closely assigned as no species have been identified that could be important sources of the diversity.

Human population genetics

The final 661,765 filtered human reads were used for the following downstream analyses. We used sequenceTools[52] to call pseudo-haplotype genotypes of the 1240K dataset.[75] A total of 11,116 pseudo-haploid positions were recovered. These genotypes were combined with data from 82 ancient genomes[28,29,32,67,75–87] (Table S2) and 2,335 present-day individuals from 149 different populations[88,89] (Table S2) that were projected on a PCA using eigensoft 7.2.1,[53] using the 597,573 SNPs of the Human Origins (HO) dataset.[88] We used the option lsqproject in order to minimize the effect of the missing data on the distortion in the PCA location

Admixture analysis was run using ADMIXTURE 1.3.0[54] with all individuals from the Human Origins (HO) array and all the available sequences from the David Reich lab database (https://reich.hms.harvard.edu/). The HO dataset SNPs were pruned with option–indep-pairwise of PLINK 1.9[55] with parameters 250 50 0.4. The total number of remaining SNPs was 436,097. Figure 2B shows the 82 ancient individuals and SAT29 samples with PONG 1.4.9.[56]

To explore the genetic affinities and the amount of shared derived SNPs we used f_3-outgroup statistics using admixtools 5.1[31] in the form f_3(SAT29,X;Mbuti). X represents both the 82 ancient genomes (Table S2) and the 149 modern populations (Table S2). For the ancient individual comparisons we restricted the analysis to 2,000 shared SNPs and reduced the modern comparisons to 4,000 shared SNPs. We further explored the possible clusterization of SAT29 and Dzuzuana2 individuals with f_4 statistics in the form f_4(Dzuzuana2,X;SAT29,Mbuti), with X representing the ancient tested populations (Table S3). All these comparisons yielded no concluding results due to the lacking statistical significance due to the low coverage.

In addition, we used qpWave from admixtools 5.1 to test the possible single genetic pool for SAT29 and Dzuduana2. We assigned these two populations as left populations and used Chimp, Altai Neanderthal, Ju_hoan_North, Khomani_San and Vindija as the right populations, from the HO dataset. This yielded to non-significant results (tail probability of. of 0.38).

Sex determination of SAT29 human reads

For sex determination we used ry_compute.[19] The results show that the SAT29 soil sample is compatible with a female: R_y value of 0.0089 and a CI of: 0.0078-0.0099. To validate such estimation we have calculated the coverage of the in the mappable region of the

Y chromosome, extracted from 1000 genomes database (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/chrY/chrY_callable_regions.20130802.bed) using bedtools 2.29.2 and samtools 1.10. We identified 12,932 filtered reads aligning the mappable region of the human Y chromosome, these reads represent the 0.002% of the chromosome, while the mean value for the rest of chromosomes is 1%, with an SD of 2%.

Neanderthal ancestry in SAT29

We used an F4-ratio[31] to explore the Neanderthal ancestry of the SAT29 sample. We used the genotype data from the 1240k dataset available in https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V44/V44.3/SHARE/public.dir/v44.3_1240K_public.tar with the combination: (Chimp AltaiNeanderthal: X Mbuti:: Chimp Altai_Neanderthal: VindijaNeanderthal Mbuti). We estimate 1% Neanderthal ancestry in the SAT29 sample, although with large uncertainty due to the low amount of data (95% confidence intervals: 0%–6.6%). This point estimate is similar to that of Dzuzuana2 and likely lower than that of Palaeolithic Europeans due to dilution from Basal Eurasian ancestry.[29] The resolution of the data does not allow a high-precision estimate and this should thus be viewed more as a methodological question exploring the possibilities and limits of sediment DNA, rather than as providing novel and specific insights into the Neanderthal ancestry proportion of the SAT29 environmental genome.

Human mitochondrial analysis

Following the human mtDNA target enrichment step we sequenced 25,483,930 captured reads. After clipping and discarding reads with a base quality score below 30, we had a total of 24,448,710 reads. 2,183,282 reads mapped to the mtDNA human genome. To assure no non-human reads were left after mapping, we used MEGAN 6.19.9[57] and BLAST+ n 2.10[74] to remove non-human reads aligning all the reads against the whole nt database and selecting only the reads that MEGAN locates in the genus *Homo*. After removing duplicates, our final dataset contained 3,477 reads unique to *H. sapiens*, which represents 10.08-fold mtDNA genome coverage. The deamination rate of the mtDNA was 0.48 G > A at the 3′ end and 0.47 C > T at the 5′ end

mtDNA mixture estimate

We run contamMix (0.0.1),[32] Schmutzi,[21] and Calico 0.2[23] for the SAT29 human reads, while for the *Canis* sequences only two methods were used: ContamMix and Calico. The contamination levels in the *Bison* reads were based on ContamMix estimates alone. We also used the deamination values to check the presence of modern contamination by estimating these values in reads longer and shorter than 70 bp observing no difference in the obtained values between both ins (0.47 and 0.49).

Presence of multiple individuals identification

We used a similar strategy to the one described in Slon et al.[13] to explore the diversity in the SAT29 mitochondrial human reads. We filtered the reads that showed the presence of deaminated bases in the last ten positions on both ends with libbam,[90] and then we explored the allele distribution in a set of diagnostic and segregating positions through the base calling using samtools mpileup and the visual observation of reads through IGV. We also estimated the amount of deamination in the last 10 bases of the examined reads (Table S1). This data is also displayed graphically in Figure S2B.

We explored the allele frequency of the human environmental genome at all the 1000 genomes[91] variant positions of the mitochondrial sequence. First, we used GATK 3.7[62] Unified Genotyper to call all the mitochondrial variable positions from 1000 genomes on the filtered reads, a total of 3892 variants, using -L and with the out_mode EMIT_ALL_SITES, using the metagenomic filtered file as input. This has resulted in the identification of 130 sites with diverse alleles, 122 out of its being segregating. To estimate the effect of ancient damage in these sites we have plotted differentially the polymorphic and non polymorphic sites. The distribution of the allele frequencies of these sites is displayed in Figure S1C.

mtDNA consensus calling

To minimize the effect of low coverage, diversity and damage, we have used Geneouis to call the mtDNA genotypes based on the majority allele (calling the base supported by > 50% of the reads) for positions covered by at least five reads.

Human mitochondrial tip dating

The full mitochondrial dataset[28,84,88,92–103] (Table S2) was aligned using *MAFFT v7.309*,[104,105] Poly-C regions and mutation hotspots in positions 303-315, 515-522 and 16519 were masked in the consensus fasta.[106] The resulting alignment was used for assessing the nucleotide substitution model using IQ-tree.[107,108] The model TN+F+I+G4 had the lowest Bayesian information criterion. We used the alignment as input for BEAUti version 2.6.3, setting as priors the radiocarbon dates shown in Table S2. Tip dates were set to the years before the present. BEAST 2.6.3[109] was run with a 50,000,000 MCMC chain length with a strict clock model, Bayesian skyline tree prior and a mean clock rate of 2.2×10^{-8}.[110] The resulting log files were viewed with Tracer v1.7.1 and were checked for ESS above 200. The tree files were annotated with TreeAnnotator v2.6.3 and the resulting annotated trees were viewed with Figtree v1.4.4.[111]

To complement the Bayesian tree we inferred one Maximum Likelihood tree using the same alignment previously used for the Bayesian analyses. The tree was inferred with IQ-tree[108] with 100 bootstrap repetitions. Resulting annotated trees were viewed with Figtree v1.4.4.

Wolf genome: comparative dataset

To construct a dataset for ancestry analyses of the SAT29 reads of canid origin, we started from a previously compiled variant call set encompassing 722 dogs, wolves and other canid species.[33] We also incorporated a number of additional wolf and other canid genomes from other publications.[112–115] These additional genomes were mapped to the dog reference genome using bwa mem version 0.7.15,[116] marked for duplicates using Picard Tools version 2.21.4 (Picard-tools., 2018), genotyped at the sites present in the 722 canid variant call set using GATK HaplotypeCaller v3.6[62] with the "-gt_mode GENOTYPE_GIVEN_ALLELES" argument, and then merged into the variant call set using bcftools merge (http://www.htslib.org/).

The variants and genotypes were then filtered by excluding sites displaying excess heterozygosity ("ExcHet" annotation p value < 1x10-6, as computed using the bcftools fill-tags plugin), setting to missing any genotypes that included an indel allele or any allele with a frequency lower than the two most common alleles at the site and thereby removing such alleles (thus retaining only two SNP alleles overlapping any given position), setting genotypes to missing if the depth at the site (computed as the sum of the "AF" fields) was lower than one third of the genome-wide average coverage of the same, or lower than 5, or higher than twice the average, normalizing allele representation using bcftools norm, and finally excluding sites with missing genotypes for 130 or more individuals. This resulted in 65.5 million SNPs, of which 19.2 million are transversions with a minor allele count in the dataset of at least two.

We assigned genotypes to the SAT29 reads that mapped to the dog genome by sampling one random allele at each of these variants using htsbox pileup r345 (https://github.com/lh3/htsbox) and requiring a minimum read length of 35 ("-l 35"), mapping quality of 30 ("-q 30") and base quality of 30 ("-Q 30"). We also included data from two Pleistocene Siberian wolves, the 35,000 year old Taimyr-1 from the Taimyr peninsula[34] and the 33,000 year old CGG23 from the Yana RHS site in eastern Siberia,[35] and genotyped these in the same way as the SAT29 data. The SAT29 sample obtained genotype calls at 1,532,986 of the total set of SNPs (2.34%), and 439,426 of the transversions (2.28%).

To complement the bayesian tree we performed one Maximum Likelihood tree using the same alignment previously used for the bayesian analyses. The tree was performed with IQ-tree[108] with 100 bootstrap repetitions. Resulting annotated trees were viewed with Figtree v1.4.4.

Wolf genome: ancestry analyses

We calculated all possible f_4-statistics involving SAT29 and the publicly available canid genomes using AdmixTools v5.0,[31] using the qpDstat command with the "$f_4$mode: YES" and "numchrom: 38" arguments.

Using f_4-statistics of the form f_4(AndeanFox,SAT29;X,Wolf35Xinjiang) we find that the SAT29 canid data is closer to a member of *Canis lupus*, Wolf35 from Xinjiang, China, than to representatives of Coyote (Z = 31.37), Golden Jackal (Z = 37.58), African Golden Wolf (Z = 31.01) and Dhole (Z = 90.20. The strongly positive values in all tests shows that the data is clearly from a member of the wolf/dog species as opposed to any of these other canid species.

We used the admixturegraph R package[117] to systematically test admixture graphs by fitting them to the *f4*-statistics. We enumerated all possible graphs involving a coyote (Coyote01, California), a modern Eurasian wolf (Wolf35, Xinjiang), a modern dog (New Guinea singing Dog, pooling individuals NewGuineaSingingDog01, NGSD1, NGSD2 and NGSD3), the two Pleistocene Siberian wolves and the SAT29 sample without admixture events. To each of these admixture graphs, we then grafted on the dog reference genome as a clade with the New Guinea singing dog, and then a boxer individual (Boxer01) as a clade with the reference genome. Because the dog reference genome also derives from a boxer, but a different individual, the contrast between these two can serve to quantify reference bias in other genomes. We therefore introduced, in each graph, an admixture event from the reference genome into each of the ancient genomes – this "reference admixture" can then correct for any systematic shifts in allele frequency caused by reference bias in these genomes. Each graph was then fit five times, retaining the fit that achieved the lowest "best_error" score. Out of the 100 possible graphs, three provided good fits to the data, with the difference between them being only the placement of the SAT29 branch as being on, downstream of, or upstream of the Siberian Pleistocene wolf branch. They correctly predict 209 out of the 210 possible f_4-statistics ($|Z| < 3$) with the minor outlier statistic: f_4(CoyoteCalifornia,NewGuineaSingingDog ; Wolf35Xinjiang,CGG23). After these three, the next best graph has 26 outlier statistics.

Canid bone testing

We screened five samples from different layers and areas of Satsurblia cave (Data S1D) for two main reasons to A) compare the bone DNA and the canid DNA from soil, and B) determine the differential capacity to retrieve DNA from different sources of the same geological layer. In the last excavations no other human bones from the Satsurblia cave have been recovered, however several Canis bones have been identified. We extracted DNA and prepared libraries following the same procedure described for soil. We also captured the dog mtDNA with the same strategy previously described. Two libraries have shown enough DNA to be analyzed: T20a from layer IIa and sample Y7d from layer BIVa. We have used these samples in the *Canis* mt phylogenies canis molecular datings.

Canid mitochondrial tip dating

We generated a consensus sequence for the SAT29 canid mitochondrial capture data in the same way as we did for the human reads: using Geneious with the major (> 50%) allele for basecalling and mtDNA positions covered by at least five reads.

A full set of samples (107 as shown in Table S4) was aligned using MAFFT v7.309[105] and used for assessing the nucleotide substitution model using IQ-tree.[107,108] The model HKY+F+I+G4 had the lowest Bayesian information criterion. The resulting fasta

alignment was used as an input file for BEAUti version 2.6.3, setting as priors the radiocarbon dates shown in Table S5. Tip dates were set to the years before the present. A strict clock model and Bayesian skyline tree prior were used. The tree height prior was based on a normal distribution with a mean value of 125 kya according to estimates obtained by Loog et al.[37] SAT29 was given a prior age of 25,000 years BP based on a broad normal distribution with standard deviation of 15000 years. BEAST was run with a 50,000,000 MCMC chain length. The resulting log files were viewed with Tracer v1.7.1 and were checked for ESS above 200. The tree files were annotated with TreeAnnotator v2.6.3 and the resulting annotated trees were viewed with Figtree v1.4.4.

To complement the bayesian tree we performed one Maximum Likelihood tree using the same alignment previously used for the bayesian analyses. The tree was performed with IQ-tree[108] with 100 bootstrap repetitions. Resulting annotated trees were viewed with Figtree v1.4.4.

Presence of multiple individuals

We found evidence for polymorphism in the SAT29 wolf mitochondrial sequences, which could suggest that the retrieved DNA originates from more than one individual. The SAT29 consensus sequence falls on a branch together with two pre-LGM Armenian wolves (TU9 and TU10), but on many sites that define this branch SAT29 also displays observations of the reference allele. We summarized this evidence as follows:

1. We aligned the previously published ancient and modern wolf mitochondrial genomes to the dog mitochondrial reference genome using bwa mem 0.7.17[116] with the "-x intractg" argument, and obtained genotypes for them using htsbox pileup.
2. We merged the SAT29 mitochondrial reads obtained from the targeted capture experiment with those obtained from the shotgun sequencing experiment, to achieve a total coverage of 16.6x (mapping quality $\geq$ 30, base quality $\geq$ 30, read length $\geq$ 35).
3. To reduce the impact of ancient DNA damage and sequencing error on the assessment of polymorphism, we restricted the analysis to a set of polymorphic sites ascertained among the previously published wolf mitochondrial genomes. We first identified sites where the two samples CGG18 (Siberia, 41.7k BP) and TH10 (Alaska, 21k BP) carry the same nucleotide, as a rough approximation of the ancestral sequence of the "major clade" of wolf mitochondria to which these two samples, as well as the majority of ancient and present-day sequences, belong. Any sites containing indels were excluded. We then identified 79 sites on which the Armenian TU10 sample carries a different nucleotide from this major clade. This should constitute a set of variants where SAT29 often should carry the TU10 allele due to its shared phylogenetic history, but might carry the major clade allele if there are other sequences in the sample that carry alleles from that clade.
4. We then counted the number of alleles in the SAT29 sample matching the "Armenian" allele and the "major clade" allele at each of these ascertained sites, using htsbox pileup (-q30 -Q30 -l 35). Both alleles are observed at most sites (57 out of 75). A few sites (11 out of 75) display only the major clade allele, but this likely reflects more recent, private mutations in the history of TU10 after its divergence from the SAT29 sequence.
5. We restricted the SAT29 sample to reads displaying evidence of ancient DNA deamination damage, using PMDtools[118] with the "–threshold 3" argument. While the total read counts are reduced, most sites still display both alleles. This suggests that the additional mitochondrial sequences(s) in the sample are also of ancient origin, rather than representing modern contamination. These results are displayed in Figure S2A.

Bison genome: comparative dataset

To construct a dataset for ancestry analyses of the SAT29 reads of bovid origin, we downloaded raw sequence reads from the European Nucleotide Archive (ENA) from a number of previously sequenced bovid genomes: present-day gaur,[40,41] present-day gayal and banteng,[40] present-day and ancient domestic taurine and zebu domestic cattle,[41] ancient aurochs,[41] American bison,[40] present-day European bison from Poland,[40,42] and the historical (early 20th century) European bison from Poland and the Caucasus.[42]

We preprocessed the reads from all the samples using fastp,[119] filtering through the automatic adaptor detection and trimming that applied by default, as well as the "–low_complexity_filter" and "–length_required 30" arguments. For ancient genomes that had been sequenced paired-end, the "–merge" option was applied and only successfully merged read pairs were retained.

We mapped the filtered reads to the domestic cattle reference genome, using bwa mem v0.7.17[116] in paired-end mode for modern genomes and bwa aln v0.7.17[48] in single-end mode, with permissive parameters ("-l 16500 -n 0.01"), for the ancient genomes. We assigned read groups according to the library and run information specified in the ENA metadata for each of the studies, merged reads for each sample and sorted using samtools,[72] and marked duplicate reads using Picard MarkDuplicates v2.21.4 (Picard-tools., 2018).

To define a set of variants to use for ancestry analyses, we identified heterozygous sites in the genome of a single, high-coverage gaur, sample Ga5.[41] Ascertainment in the gaur outgroup species, which is estimated to have diverged from bison more than half a million years ago,[40] should result in variants that behave in an unbiased fashion in ancestry analyses. We called genotypes in Ga5 using GATK HaplotypeCaller v3.6.[62] We then filtered these genotype calls using bcftools (http://www.htslib.org/) to retain only those variants that were SNPs, were located on the 29 autosomal chromosomes, had a heterozygous genotype, had a genotype quality (GQ field) of > 30, a depth (sum of AD fields) of more than 15.04 and less than 49.63 (corresponding to 0.5 and 1.65 times the average autosomal coverage of 30.08, respectively). This resulted in 4,930,425 SNPs, of which 1,447,767 are transversions.

We assigned pseudo-haploid genotypes for all the bovid genomes, including Ga5 itself and the SAT29 reads that mapped to the cattle genome, by sampling one random allele at each of these Ga5 ascertained SNPs, using htsbox pileup r345 (https://github.com/lh3/htsbox) and requiring a minimum read length of 35 (''-l 35''), mapping quality of 30 (''-q 30'') and base quality of 30 (''-Q 30''). The SAT29 sample obtained genotype calls at 94,262 of the total set of SNPs (1.91%), and 27,724 of the transversions (1.91%).

Bison genome: ancestry analyses

We calculated all possible f_4-statistics involving the SAT29 sample and the publicly available bovid genomes using AdmixTools v5.0,[31] using the qpDstat command with the ''f_4 mode: YES'' and ''numchrom: 29'' arguments.

Using f_4-statistics of the form f_4(Ga5.Gaurus,SAT29;X,Wisent11) we find that the SAT29 bovid data is closer a bison individual, Wisent11 from Poland, than to representatives of aurochs (Gyu2, Armenia, Z = 20.59), taurine cattle (ScottishHighland, Z = 22.73), Zebu cattle (Tharparkar, Z = 24.75), banteng (ypt2230, Z = 35.96) and gayal (1107, Z = 43.76). The strongly positive values in all tests shows that the data is clearly from a member of the bison species as opposed to any of these other bovid species.

We used the admixturegraph R package[117] to systematically test admixture graphs by fitting them to the f_4-statistics. We enumerated and fit all possible graphs involving a gaur (Ga5), an American bison (mzc), an historical Polish bison (PLANTA), a historical Caucasian bison (Cc1) and the SAT29 sample with up to one admixture event. Each graph was fit five times, retaining the fit that achieved the lowest ''best_error'' score.

Among the 15 possible topologies that relate these five genomes without any admixture events, the best-fitting graph has the American bison as basal to the European bison and SAT29, and then SAT29 as basal to the historical Polish and Caucasian bison. This graph has just one outlier f_4-statistic ($|Z| < 3$), which fails to account for excess affinity between the American and the Polish bison (f_4(Gaur,-American bison;Polish bison,SAT29), Z = −4.03). After this, the next two best-fitting graphs differ from the best-fitting topology in that the position of SAT29 is swapped with that of the historical Polish wisent or the historical Caucasian wisent, respectively. These graphs both feature the same three outlier f_4-statistics, the first of which is shared by the best-fitting graph above, and the second and third of which fail to account for shared drift between the historical European bison to the exclusion of SAT29 (f_4(Gaur,Causasian bison;Polish bison,SAT29), Z = −4.51, f_4(Gaur,Polish bison;Caucasian bison, SAT29), Z = −3.43.f_4

Following these, all other graphs without admixture events have seven or more outlier f_4-statistics. When allowing for one admixture event, 10 out of the 315 possible graphs fit the data without any outlier statistics. Multiple solutions with quite variable topologies are thus possible, and with the limited data available we do not attempt to discriminate between these.

Bison mitochondria capture and analysis

We have captured a *B. bonasus* environmental mitochondrial genome through *Bos taurus* baits designed with the same methodology previously described. After aligning the reads against the *B. bonasus* mitochondria reference (NC_014044.1), we have filtered the reads with BLASTN and MEGAN as described previously. We have aligned the retrieved filtered reads with other 70 Bovid samples (Table S5) using MAFFT v7.309[105] and used it for assessing the nucleotide substitution model using IQ-tree.[107,108] We generated a bayesian tree and a calibrated tip-dating phylogeny with BEAST 2, setting as priors the radiocarbon dates shown in Figure S6.

Ovis genomic analysis

We explored the possible phylogenetic position of the reads that aligned to the Oar_v3.1 genome within the *Ovis* and *Capra* genus. In order to determine the SNPs to compare to SAT29, we built a dataset with individuals from all the available species of the genus *Ovis* and *Capra: Ovis vignei* (Nextgen project: https://projects.ensembl.org/nextgen/), *Ovis aries* (Nextgen project: https://projects.ensembl.org/nextgen/), *Ovis canadensis*,[120] *Ovis ammon*,[121] *Ovis orientalis* (Nextgen project: https://projects.ensembl.org/nextgen/), *Ovis nivicola*[122] Capra hircus (Nextgen Project: https://projects.ensembl.org/nextgen/), *C. caucasica*,[123] *C. ibex*,[124] *C. aegagrus*[125] and Oreamnos americanus as an outgroup.[126] We downloaded the available VCF files of the following genomes: 75 *Ovis aries* genomes, 4 *Ovis vignei* genomes, 14 *Ovis orientalis* genomes and one Capra hircus. This dataset consists of 48,870,177 SNPs in the autosomal chromosomes of the *Ovis aries* genome. After filtering SNPs for MAF < 0.05 and removing non-SNPs and no-biallelic SNPs and SNPs not located in autosomes, 22,553,044 SNPs were kept.

As the available VCF files does not cover all species we wanted to include, we downloaded the FASTQ files of one Ovis canadensis, one Ovis nivicola, one Ovis ammon, one C caucasica, one Capra ibex, one C sibirica, two C aegagrus and one Oreamnos americanus to produce additional VCF files for downstream analyses. The sequencing reads of the FASTQ files were aligned with BWA,[48] duplicate reads were removed with picard and low quality reads (< 30) were removed with Samtools.[72] These filtered reads were used for variant calling with GATK HaplotypeCaller v3.6 (67) by genotyping the positions of the filtered dataset with the ''-gt_mode GENOTYPE_GIVEN_ALLELES'' argument. The reads were then merged into the variant call set using bcftools merge (http://www.htslib.org/). Genotypes were filtered with bcftools for: MBQ > 30, depth of coverage below the half of the average coverage and more than double of the average coverage to eliminate possible misalignments in low and high complexity regions. These new VCF files were then merged with the downloaded VCF files. The final dataset was filtered for positions with more than 10% of missing sites and excess of Heterozygosity (pval < 1.10-6).

We called pseudo-haplotype genotypes using the 22 million positions of SAT29 Ovis reads using Sequence Tools[52] and recovered 19,469 SNPs of the SAT29 genome. We used f_4 statistics from admixtools[31] to determine the closest taxa to the SAT29 sample. The analysis did not yield any concluding result as the number of SNPS is likely too low.

All the displayed images in the publication have been edited with GIMP 2.10.24.[127]

3. Chapter 3: Methodological Advancements to improve Efficiency in sedaDNA Analyses

3.1. Publication II

Maximizing Efficiency in sedimentary ancient DNA Analysis: A Novel Extract Pooling Approach

Status: Published

Citation: Victoria Oberreiter, Pere Gelabert, Florian Brück, Stefan Franz, Evelyn Zelger, Sophie Szedlacsek, Olivia Cheronet, Fernanda Tenorio Cano, Florian Exler, Brina Zagorc, Ivor Karavanić, Marko Banda, Boris Gasparyan, Lawrence Guy Straus, Manuel R. Gonzalez Morales, John Kappelman, Mareike Stahlschmidt, Thomas Rattei, Stephan M. Kraemer, Susanna Sawyer & Ron Pinhasi. Maximizing Efficiency in sedimentary ancient DNA Analysis: A Novel Extract Pooling Approach. Scientific reports, 14(1), 19388. https://doi.org/10.1038/s41598-024-69741-5

scientific reports

Check for updates

OPEN

Maximizing efficiency in sedimentary ancient DNA analysis: a novel extract pooling approach

Victoria Oberreiter[1,2], Pere Gelabert[1,2,3]✉, Florian Brück[14], Stefan Franz[1], Evelyn Zelger[1], Sophie Szedlacsek[1], Olivia Cheronet[1,2], Fernanda Tenorio Cano[1], Florian Exler[1,2,11], Brina Zagorc[1,2], Ivor Karavanić[4], Marko Banda[4], Boris Gasparyan[5], Lawrence Guy Straus[6,7], Manuel R. Gonzalez Morales[8], John Kappelman[9], Mareike Stahlschmidt[1,2], Thomas Rattei[10], Stephan M. Kraemer[2,11,12,13], Susanna Sawyer[1,2]✉ & Ron Pinhasi[1,2]✉

In the last few decades, the field of ancient DNA has taken a new direction towards using sedimentary ancient DNA (sedaDNA) for studying human and mammalian population dynamics as well as past ecosystems. However, the screening of numerous sediment samples from archaeological sites remains a time-consuming and costly endeavor, particularly when targeting hominin DNA. Here, we present a novel high-throughput method that facilitates the fast and efficient analysis of sediment samples by applying a pooled testing approach. This method combines multiple extracts, enabling early parallelization of laboratory procedures and effective aDNA screening. Pooled samples with detectable aDNA signals undergo detailed analysis, while empty pools are discarded. We have successfully applied our method to multiple sediment samples from Middle and Upper Paleolithic sites in Europe, Asia, and Africa. Notably, our results reveal that an aDNA signal remains discernible even when pooled with four negative samples. We also demonstrate that the DNA yield of double-stranded libraries increases significantly when reducing the extract input, potentially mitigating the effects of inhibition. By embracing this innovative approach, researchers can analyze large numbers of sediment samples for aDNA preservation, achieving significant cost reductions of up to 70% and reducing hands-on laboratory time to one-fifth.

Genomic material can be obtained from various environmental sources, including soil[1,2], water[3,4], or air[5,6]. Hence, environmental DNA (eDNA) is a valuable tool for ecological surveys of species that are challenging to observe directly or in the absence of their remains. The initial applications of eDNA focused on retrieving microbial genetic material from environmental samples rather than culturing them prior to extraction[7–9]. Since then, numerous studies have been published, focusing on microbial taxa recovered from sediments and permafrost[10–14]. These studies have exemplified that the retrieval of DNA from archaeological sediments and permafrost could

[1]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria. [2]Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Vienna, Austria. [3]Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia, Universitat Autònoma de Barcelona, Bellaterra, Spain. [4]Department of Archaeology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia. [5]Institute of Archaeology and Ethnography, National Academy of Sciences of the Republic of Armenia, Yerevan, Armenia. [6]Department of Anthropology, University of New Mexico, Albuquerque, USA. [7]EvoAdapta Group Universidad de Cantabria, Santander, Spain. [8]Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Universidad de Cantabria, Gobierno de Cantabria, Banco Santander, Spain. [9]Department of Anthropology and Department of Earth and Planetary Sciences, The University of Texas, Austin, TX, USA. [10]Division of Computational Systems Biology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria. [11]Department of Environmental Geosciences, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria. [12]Institut für Analytische Chemie, University of Vienna, Vienna, Austria. [13]Forschungsverbund Umwelt und Klima, University of Vienna, Vienna, Austria. [14]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria. ✉email: pere.gelabert@univie.ac.at; susanna.sawyer@univie.ac.at; ron.pinhasi@univie.ac.at

be used to study our past[15,16]. In its early stages, the analysis of ancient sediments focused on simple presence/absence studies through metabarcoding quantification[17,18] involving the use of taxon-specific primers to amplify short gene regions, followed by high throughput sequencing (HTS)[19]. This approach has demonstrated great potential for the reconstruction of past ecosystems and the impacts of climate change[20] from various sources, including sediments[21,22], lake sediments[20,23–26], and frozen soils[27], even in the absence of macrofossils. Metabarcoding offers a cost-effective option. However, it does not allow further analyses such as phylogenies or population genetic tests. Alternatively, shotgun sequencing of sedaDNA provides a greater taxonomic resolution, as it analyzes all the genetic material contained within a sample, potentially allowing for the retrieval of full genomes[28–30]. As shotgun sequencing is usually ineffective in terms of endogenous DNA sequencing, studies have applied targeted hybridization capture, particularly focusing on mammalian and hominin mitochondrial DNA[31–33] and nuclear DNA in archaeological sediments[34]. These capture approaches, however, add significantly to the expense and time needed, particularly with rare hominin sedaDNA from archaeological sediments.

Even with the addition of targeted hybridization capture, sediment samples may not have any aDNA preservation. The preservation of sedaDNA in sediment is variable and depends on climatic conditions and the chemical composition of the sediment[31,33,35–37]. In Denisova Cave, for example, the majority of samples exhibited the presence of faunal and hominin mitochondrial (mt)DNA[33], while, in contrast, in Sefunim Cave, only four out of 33 tested samples showed traces of mammalian mtDNA[38]. In their analysis of Pleistocene sediments from sites across Europe, Asia, Africa, and North America, Massilani et al. were able to extract ancient mammalian mtDNA from approximately 50% of the analyzed, resin-impregnated sediment blocks[39]. Their study highlights the potential of resin-impregnated sediments to preserve ancient DNA, even from small particles like bone fragments and coprolites. This finding is crucial as it demonstrates that ancient DNA can remain stably localized within sediments over time, allowing for the recovery of genetic material in a microstratigraphic context. These studies provide valuable insights into the variable preservation and spatial distribution of ancient DNA within archaeological sites, emphasizing the need for efficient screening protocols.

The potentially low success rate of mammalian sedaDNA calls for a methodological screening optimization to allow for a faster and more cost-effective assessment of samples. Sample pooling is a common practice in the field of eDNA, primarily used to assess the biodiversity of ecological communities. The concept of pooling multiple samples for collective analysis was established in the 1940s when it was initially employed to detect positive cases of Syphilis during World War II[40]. Over the years, this approach has found applications in various fields, including the screening of infectious diseases such as hepatitis viruses B and C[41,42] human immunodeficiency virus (HIV)[43], and, more recently, during the SARS-CoV-2 pandemic[44,45]. In the context of sedimentary eDNA screening, several pooling strategies have been proposed to reduce costs, including pre-extraction[46–48], post-extraction[49], and post-PCR pooling[20,49]. A recent study demonstrated the successful pooling of multiple sample libraries into a single capture reaction while optimizing reagents costs[50]. However, there is currently no experimental framework examining how post-extraction sample pooling of ancient sediments affects the behavior of the deamination signal and the abundance of authentic ancient DNA fragments.

In this study, we hypothesize that deamination signals can be maintained and detected with increased post-extraction pooling of up to five sediment samples. We further hypothesize that the abundance of unique DNA reads, mapping to our reference panel, will not significantly decrease in the various extract pools compared to the individual stand-alone sample. To test these hypotheses, we applied multiple degrees of extract-pooling to Paleolithic sediment samples from European, Asian, and African sites and one burial site. Our findings demonstrate that up to five extracts can be pooled while maintaining a detectable signal of aDNA and achieving an average 1.36-fold increased mammalian target DNA yield across all pooling levels compared to the average unpooled state. This high-throughput pooling method will allow for large-scale sediment analyses for aDNA preservation with great reductions in costs and time.

Materials and methods

In the following wet lab experiments, we used sediment samples obtained from multiple Paleolithic cave sites, an open air site, and one burial site: El Mirón (Spain), Velika Pećina—Kličevica (Croatia), Hovk-1 (Armenia), Krems-Wachtberg (Austria), and Shinfa-Metema 1 (Ethiopia) (Supplementary Information S1).

In silico simulation of sample pooling

Following previous evidence for detectable sedaDNA after pooling post-PCR extracts[20], we first conducted an in-silico simulation to examine the impact of pooling on the characteristics of ancient mammalian mtDNA sequences using mitochondrial reference genomes of *Homo sapiens* and domestic goat (*Capra hircus*). To simulate the characteristics of ancient DNA we employed an online sequence manipulation tool (https://www.bioinformatics.org/sms2/split_fasta.html), fragmenting the original sequences into segments of 30, 35, and 40 base pair length. These fragmented sequences were then combined into composite files for each species, containing fragments of varying lengths. Next, we introduced deamination on the composite sequences with deamination rates set at 10, 30, and 50 percent per species. To model differential deamination levels between species within a single sediment sample, we combined human sequences with a 10 percent deamination rate with *Capra* sequences exhibiting deamination rates of 30 and 50 percent, respectively.

We carried out sequential pooling of the deaminated composite sequence files for human, *Capra*, and the combination of human and *Capra*. These pooling steps were simulated by adding random reads extracted from a sediment sample that showed no detectable aDNA reads after the initial screening process. The pooling was obtained up to a dilution of 1:10 (Supplementary Fig. S1). We assessed deamination and read count, which are the main measures of aDNA, at each pooling step using a custom bioinformatics pipeline detailed in the methods.

DNA extraction and library preparation

To mitigate the risk of modern contamination, we implemented strict measures such as the use of gloves, disposable lab suits, and other sterile equipment. All of the samples were prepared in dedicated clean room facilities at the University of Vienna. We included negative controls at each step (extractions, libraries, and PCR) to monitor potential reagent contamination. During the screening for candidate samples, amounts of 50–60 mg of sediment were used for DNA extraction following the protocol from Dabney[51] with adaptations from Korlević[52], and eluted in 50 µL TET buffer (10 mM Tris–HCl, 1 mM EDTA, 0.05% Tween 20, pH 8.0). Double-stranded libraries were prepared of half of the extract as described in[53], omitting the shearing of DNA into smaller fragments. Instead of SPRI beads, a MinElute PCR Purification kit from Qiagen was used for cleaning up the samples and eluted in 40 µL EBT buffer (1 mM EDTA, 0.05% Tween-20). We added a positive control using 24 µL of deionized water and 1 µL of a 1:250 dilution of CL104.

Mammalian mtDNA hybridization capture and sequencing

The number of PCR cycles for optimal amplification was determined individually for each library using a qPCR machine. We double-indexed[54] and amplified 25% of each library with PfuTurbo Cx HotStart DNA Polymerase. Amplification products were cleaned up using NGS clean-up magnetic beads, introducing a size selection through a concentration of 1.2 × beads per sample and eluted in a total amount of 25 µL EBT buffer (1 mM EDTA, 0.05% Tween-20). In preparation for the subsequent capture, we further amplified 3 × 2 µL of each indexed library using KAPA HiFi HotStart DNA Polymerase and the primer pairs IS5/IS6[53] at a rate of 20 cycles and cleaned up with magnetic beads. The concentration of the PCR product was measured on a Qubit.

We enriched for 51 mammalian mitogenomes, including human mitochondrial DNA, using custom-designed capture probes[55] following the TWIST capture protocol[56]. After a hybridization time of 16 h at 65 °C and four rounds of washing the enriched DNA was cooked off the probes in a PCR cycler at 95 °C for 5 min (including a heated lid at 110 °C). A qPCR was performed to determine the correct number of PCR cycles for each library. Half of the captured library was re-amplified using KAPA HiFi HotStart DNA Polymerase and the primer pairs IS5/IS6 and cleaned up with magnetic beads. Library concentration (ng/µL) and molarity (nmol/L) were determined through Qubit and Tapestation. For sequencing on an Illumina NovaSeq6000XP SP lane with a total number of 300 million (3,000,000,000) reads, libraries were pooled by including 20 ng of DNA for each (which roughly equals five million reads per sample) and the amount for blanks was calculated to approximate 200,000 target reads per blank (for numbers of sequenced reads per sample refer to Supplementary Table S1). The sequencing was performed at the Vienna BioCenter Core Facilities in Vienna.

Pooling experiment

To generate sufficient extracts for the pooling experiment, we employed a modified extraction protocol that involved processing 1 g of sediment for the negative samples (i.e., samples previously sequenced but without detectable aDNA, defined as < 1000 unique reads mapping to the mammalian reference panel and < 10% of deamination) and 250 mg of test samples (i.e., samples previously sequenced with detectable aDNA, defined as either > 1000 unique reads mapping to the mammalian reference panel or > 10% of deamination, or both). Following the protocol described in Dabney et al.[51] and Korlević et al.[52], we adjusted the amounts of buffers to the amount of used sediment and eluted the extracts in TET buffer, using 1 mL and 250 µL for negative and test samples, respectively. To ensure reproducibility of the initial sequencing results across all samples (Supplementary Table S1), we conducted downstream lab analyses using the same methods as previously described, and again carried out shallow sequencing on a NovaSeq SP lane, alongside other samples. Any sample for which the screening results were reproducible was included in the experimental set-up. In cases where the screening results could not be replicated, samples were substituted with new candidate samples. The extracts of the previously sequenced sediment samples which showed signs of aDNA (Table 1) were diluted with the extract of a negative sample (< 1000 unique reads mapping to the mammalian reference panel and < 10% of deamination) in increasing amounts (ranging from 1:1 to 1:4) up to a total volume of 25 µL. Each pool of extracts was prepared in three replicates and converted into double-stranded libraries according to[53]. The library preparation process for pooling levels incorporated different test extract input volumes, combined with increasing volumes of a negative sample up to a total volume of 50 µL; That equals to 25 µL of test extract for the unpooled state, 12.5 µL for a 1:1 pooling, 8.33 µL for a 1:2 pooling, 6.25 µL for a 1:3 pooling, and 5 µL for a 1:4 pooling state. We proceeded with the downstream analyses (PCR, mammalian capture, and quality controls) as described above.

Bioinformatic processing

First, we removed the adapter sequences and sequences shorter than 30 bp using Cutadapt 4.0[57]. The clipped reads were mapped to a composite reference file encompassing the 51 mammalian mt genomes included in the

State	Deamination 5′ C–T (%)	Reads
1	> 10	> 1000
2	> 10	< 1000
3	< 10	> 1000
4	< 10	< 1000

Table 1. Summary of the four states in which we group the screened sediment samples.

hybridization capture[55] using bwa aln version 0.7.17-r1188[58] with an edit distance of 0.01, a gap penalty opening of 2 and seeding disabled. Reads were filtered using samtools[59] retaining the ones with average mapping qualities above 30. We removed duplicate sequences with samtools rmdup and assessed damage levels through mapdamage 2-2.2.1[60]. The quantity of the accepted reads was examined using Qualimap v.2.3[61].

Statistical analyses

All statistical analyses and plots were performed in Rstudio Version 1.1.414[62] using R version 3.6.3[63]. Boxplots were produced using the default settings: The boxes represent the Interquartile Range (IQR) and span from the first quartile (25th percentile) to the third quartile (75th percentile) of the data. The midline of the box is the median. The whiskers extend up to 1.5 times the IQR from both ends of the box to the furthest datum within that distance. Data plotting beyond that distance are represented as individual points and considered outliers.

P-values were calculated in R studio using a paired Wilcoxon rank sum test with continuity correction and considered to be significant at a p-value below 0.05.

Results

In silico simulation

To evaluate the impact of sample pooling on ancient DNA detection, we conducted an in-silico simulation using fragmented mitochondrial sequences of *Homo sapiens* and *Capra hircus*, introducing varying levels of deamination and simulating pooling at different dilutions (Fig. 1). The simulation results (Supplementary Table S2) demonstrated stable read numbers and deamination levels across pooling steps with a non-significant decrease in read count ($p > 0.05$), indicating that an aDNA signal can be detected at a dilution up to 1:10-fold (Fig. 1).

Wet lab pooling of extracts does not decrease sensitivity

As part of the MINERVA project we screened archaeological sediments for ancient DNA employing a custom hybridization capture kit comprising 51 mammalian species (Supplementary Table S3)[55]. Due to the age of the sequences, aDNA displays distinct signs of degradation characterized by two main molecular features: strand breaks resulting in short DNA fragments and C to T substitutions caused by deamination[64–66]. We classified our samples using two parameters: (A) the quantity of DNA expressed as the number of unique reads mapping to the used mammalian reference panel and (B) the damage of these reads, measured by the occurrence of C to T transitions. We group our sediment samples into four states: (1) reads with a deamination level above 10%, and more than 1000 recovered unique reads, (2) reads with a deamination level above 10%, and fewer than 1000 recovered unique reads, (3) reads with a deamination level below 10% and more than 1000 recovered unique reads, and (4) reads with a deamination level below 10%, and fewer than 1000 recovered unique reads (Table 1). For convenience, all states are from here on referred to by their corresponding reference number (1–4) followed by the values for (a) deamination in percent and (b) the amounts of unique reads mapping to the mammalian reference panel. We further define state 1 as the threshold for a positive sediment sample (i.e., deamination level above 10% and > 1000 unique endogenous reads aligning to the mammalian reference genome panel used during the capture process). A failed sample is defined as a sample falling into states 2 (> 10%, < 1000), 3 (< 10%, > 1000),

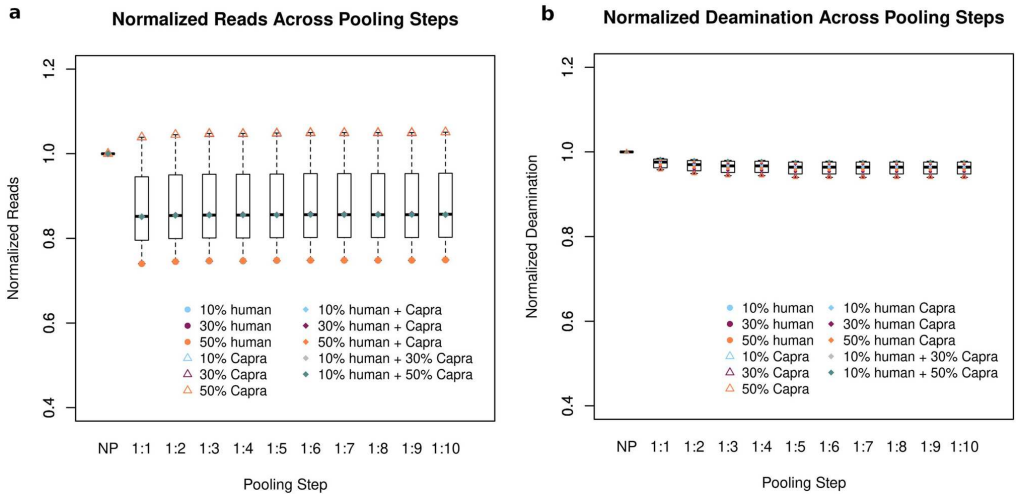


Figure 1. In silico simulation of pooling. (**a**) Read numbers normalized by the unpooled state for each simulated mitochondrial dataset at different pooling steps, relative to the non-pooled ("NP") condition. (**b**) Simulated deamination values normalized by the unpooled state for each simulated mitochondrial dataset at different pooling steps.

and 4 (< 10%, < 1000). In a different study on sedaDNA, a sample was considered to be positive for ancient hominin DNA if the number of DNA fragments assigned to hominins constituted at least 1% of the total identified fragments and if it had at least 10 fragments showing deamination which was significantly higher than 10%[33]. Another study also used a 10% deamination frequency at both ends as a cut-off (binomial CI ≥ 10%)[34].

To evaluate our pooling simulation in a laboratory setting, we selected eight test samples to represent these four states (two samples per state according to earlier screening analyses). However, two of the test samples had to be excluded from the study. One test sample falling into state 2 (> 10%, < 1000) was contaminated with 1.67 × modern human DNA. Modern human contamination can be introduced while handling the sample on site or during the wet lab process and will be captured alongside ancient DNA fragments. We identified modern contamination using the software mapdamage 2-2.2.1[60], which assesses DNA damage patterns characteristic of ancient DNA, such as C-to-T substitutions at the ends of DNA fragments[64–66]. Additionally, the screening results for a state 3 (> 10%, < 1000) sample could not be reproduced; the sample displayed higher deamination after re-extraction. A negative sample, devoid of aDNA after screening (< 1000 unique reads mapping to the mammalian reference panel and < 10% of deamination) was chosen for the pooling. This sample had an average deamination of 4.7% at both ends and 651 reads aligning to our mammalian mtDNA panel. The extract of each test sample was pooled with the extract of our chosen negative sediment sample in increasing amounts up to a dilution of 1:4 (Fig. 2). This process was performed in two sets for each state, with two distinct candidate samples meeting states 1–4. To ensure accuracy, each dilution level was replicated three times (Supplementary Table S1).

The experiment revealed that deamination remains stable with increased pooling in three states: state 1 (> 10%, > 1000, R2 = − 0.73), state 3 (< 10%, > 1000, R2 = 0.45), and state 4 (< 10%, < 1000, R2 = − 0.86), as shown in Fig. 3b. Notably, even in state 4, characterized by low deamination (< 10%) and low read levels (< 1000), the signal for the target aDNA could be maintained. In contrast, state 2 (> 10%, < 1000, R2 = − 0.94) exhibited a significant decrease in deamination level with increased pooling (p = 0.01789).

We observed an increase in normalized average read numbers when pooling was implemented (Fig. 3a). Despite considerable variation within each pooling level (SD = 0.12–0.23), there was a significant increase in average reads between the non-pooled state and the 1:3 dilution level ($p = 0.018$) as well as the 1:4 dilution level ($p = 0.011$). DNA yields were on average increased by 17, 47, 33, and 45 percent for pooling levels 1–4, respectively (Table 2). Overall, the DNA yield across all pooling levels was increased by 1.36-fold, compared to the average unpooled sample.

A cost-efficient method

The screening of sedaDNA samples is a laborious and expensive undertaking. We have modeled multiple situations in order to gain insights into the cost and labor reduction of this approach. We estimate that under conditions where 10% of the samples exhibit the characteristics previously defined as positive, the implementation of our approach would result in a 50% reduction in the number of libraries prepared. In the case of a 5% positivity rate, this would amount to a 75% decrease in the number of prepared libraries. Thus, we suggest that the pooling

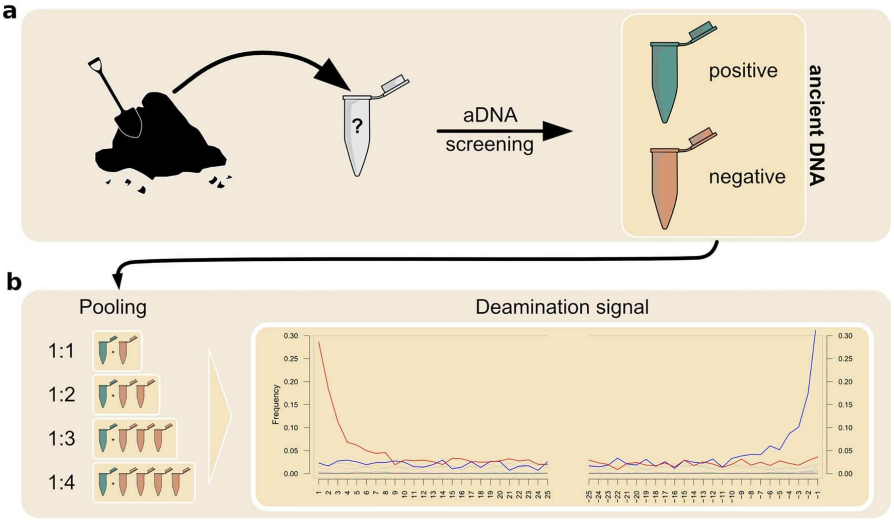


Figure 2. Streamlining of initial screening of samples followed by pooling. Sediment samples were individually screened to distinguish samples with preserved ancient DNA (positive) from samples devoid of aDNA (negative). The extracts of positive samples were pooled with up to 4 extracts of a negative sample to recreate a pool testing method. The presence of an aDNA signal was evaluated by the deamination signal and the number of unique reads.

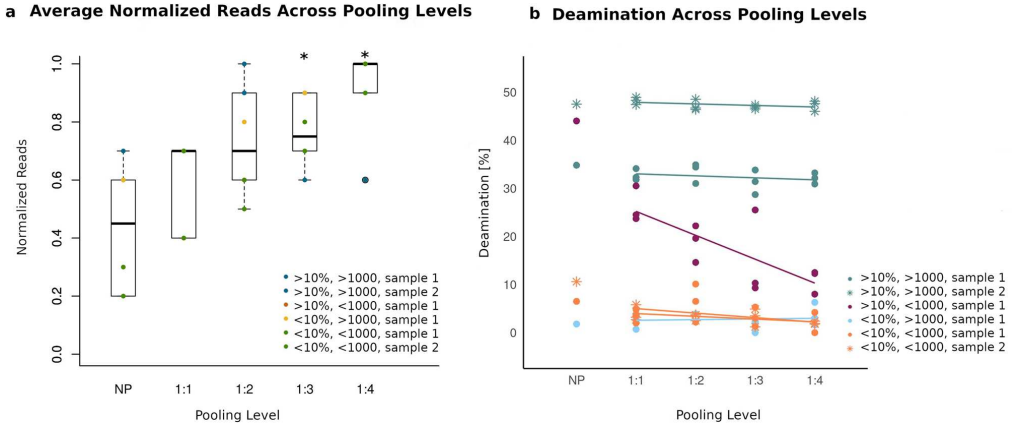


Figure 3. Average normalized read numbers percent of deamination across pooling levels. (**a**) Average read numbers of sediment samples normalized by the highest value across pooling levels. Each dataset is represented by a unique color in the legend. We calculated the differences between the unpooled state with each pooling step using a paired Wilcoxon rank sum test with continuity correction and considered it to be significant at a p-value below 0.05. Significance is indicated by a black asterisk. (**b**) Percent of deamination measured in sediment samples across increasing extract-pooling steps in comparison to the non-pooled ("NP") sample. The colors indicate the four different states that sedaDNA samples commonly fall into, specified in Table 1. Different shapes within the same states indicate different datasets.

Pooling level	Avg reads	% Increase in reads	Fold increase in reads
NP	542	N/A	N/A
1:1	634	17	1.17
1:2	798	47	1.47
1:3	722	33	1.33
1:4	786	45	1.45
Avg of all pooling levels	735	36	1.36

Table 2. Comparison of DNA yields in normalized, average read numbers across pooling levels.

of five extracts is a cost-effective strategy in scenarios where the prevalence of positive samples is less than 20% among the total.

Discussion

With our extract-pooling approach, we set out to define a cost-efficient and high-throughput method for screening archeological sediments for aDNA. When pooling was implemented, we observed a cost reduction of approximately 50–75%, including costs for reagents and sequencing. Hands-on time in the lab was reduced to one-fifth. We demonstrate that the pre-PCR pooling of Paleolithic sediment extracts does not reduce the number of aDNA reads detected in double-stranded libraries[53] across all pooling levels: extract-pooling, in fact, significantly ($p = 0.0001$) increases the efficiency of target aDNA recovery across all four states on average by 1.36-fold (Fig. 3a). While we observe an increase in unique DNA fragments mapping to the target panel with reduced extract input volume in double-stranded libraries, we acknowledge that understanding the underlying mechanisms requires experimental investigation beyond the scopes of this framework. One of the potential explanations for the observed increase could be the presence of substances in the sediment exerting a negative effect on PCR amplification. Inhibition is a prevalent obstacle in many PCR related applications such monitoring of foodborne diseases[67–70], water and environmental surveillance[71–73], plant pathogen detection[74–76], soil, and sediment analyses[77,78]. For a detailed review on the occurrence, properties, and methods to remove inhibitors see[79]. Common PCR inhibitors present in the organic component of sediment include plant-derived substances such as tannic, fulvic, and humic acids[77,78,80–82], as well as Phenols, Polysaccharides, Pectin, Polysaccharides, and Xylan[83–85]. Other substances known to impair PCR analysis include calcium ions, which compete with magnesium ions (a common component in commercial extraction kits) for binding to DNA polymerases, while humic acids interact directly with nucleic acids[81]. These inhibitors can be co-extracted with DNA from complex environmental samples like sediments and interfere with enzymatic reactions in qPCR and PCR amplification

due to inhibition of polymerase activity[86]. This can lead to reduced PCR signals or an elevated number of false negative results[87,88]. Besides the diverse nature and mechanisms of PCR inhibitors, their concentration, which can vary within a sample, plays an important role in PCR results[89]. Many methodological approaches have been introduced to minimize the adverse effects of inhibitors at different steps of the wet lab pipeline. There are several treatments to reduce the co-purified amount of inhibitors during DNA extraction, such as adding polyvinylpolypyrrolidone (PVPP)[90,91], hexadecyltrimethylammonium bromide (CTAB)[90,92,93], and aluminum ammonium sulfate ($AlNH_4(SO_4)_2$)[94]. Further methods include the introduction of hydroxypatite columns[95], column chromatography[90,96], cesium chloride density centrifugations[97–100], and DNA recovery from agarose after gel electrophoresis[90,92,101]. While most of these methods are laborious and and can effectively reduce the recovery of target DNA[90,94], the dilution of extracts has been found to effectively mitigate inhibition; A recent study investigating PCR inhibitor removal methods found that the pre-dilution of extracts led to maximum amplification results[71]. The success of dilution and commercial PCR removal kits were, however, dependent on the initial concentration of inhibitors.

Our results show that a reduction of extract input into double-stranded libraries increases DNA yield significantly when comparing read numbers of the unpooled sample (25 µL) with those of pooling level 1:3 with 6.25 µL ($p = 0.018$) and pooling level 1:4 with 5 µL input ($p = 0.011$) (Fig. 3a). A study conducted on bones and teeth demonstrated that the preparation of single-stranded libraries with minimal extract input reduces the effects of inhibition[102]. A reduction of sample input in qPCR analyses provides additional support for our hypothesis that inhibition can be mitigated through this approach, as demonstrated by previous studies[103,104]. The extensive dilution of a DNA extract (often up to 400-fold), is a common practice across different scientific disciplines to circumvent inhibition[87,103,105,106]. As a consequence, the target DNA will simultaneously be diluted along with the inhibitory substances in PCR based methods[107,108]. In cases of underrepresented taxa or genes of interest, a dilution may even result in failed detection[109,110]. Another study, therefore, urges a moderate extract dilution of 40–60-fold to effectively reduce qPCR inhibition[104].

Besides the reduction of PCR inhibiting agents, another potential explanation for the observed read increase is the improvement of PCR template availability in complex sediment samples: PCR is a highly sensitive process that can be affected by the availability and quality of endogenous DNA template and depends on the concentration and availability of PCR reagents, such as polymerase and dNTPs. In diverse environmental samples with high amounts of background DNA from non-target sources, endogenous DNA may be present in low quantities, thus reagents might be the limiting factor in a PCR reaction. Diluting the sample can reduce overall concentration, thereby improving the availability of PCR reagents for target template. This enhancement can lead to more efficient and consistent amplification, resulting in higher edogenous read numbers. We hypothesize that in our framework, the reduction of test sample input combined with increased amounts of a negative sample, containing very low amounts of target material mapping to the reference panel[55], results in elevated read numbers.

To ensure consistency across our experimental set-up, the same DNA extraction protocol[51,52] and library protocol[53] were used for all samples included. This standardization rules out the possibility of any protocol-specific enhancements or adaptations contributing to the increase in reads within the experiment. Additionally, the same extract was used as a negative pooling extract consistently across the dilution series, ensuring that intrinsic properties of the sediment samples remained constant. Overall, this uniform approach across all samples eliminates the likelihood of variations in DNA extraction techniques, sample handling, or intrinsic sediment properties affecting the results.

The main objective of our experimental set-up was to mimic extract pooling of sedaDNA samples; This required the use of an authentic negative sediment sample for the dilution of the test samples. Therefore, the effect of inhibition was not specifically controlled for in this framework and we acknowledge that this aspect requires further testing. Moreover, due to the diverse nature and mechanisms of inhibitory substances, no one combination of PCR reagents can be used to work uniformly well against all of them, as tested by[86]. Hence, a test for the individual identification of inhibitory agents in different types of sediment in order to adapt extraction, purification, and PCR protocols has yet to be developed.

The minimal extract volume converted into double-stranded libraries in our study was 5 µL, which equals a 1:4 dilution of the regular input. In our approach, we intentionally pooled a positive sediment extract with increasing volumes of a negative extract. However, it is important to note that when applying pooling to several previously unscreened extracts, in which case the extent of inhibitors is unknown, we recommend pooling at small volumes. Another unknown factor in unscreened extracts is the extent of DNA preservation. Therefore, we cannot make a generalized conclusion that dilution consistently leads to increased DNA yields, as it relies on the assumption that the concentration of humic substances corresponds to DNA content, which would require further testing.

Additional experiments will be necessary to test the optimal volume of extract converted into libraries to examine the potential for even smaller volumes than those employed in this study for enhanced DNA recovery. We recommend conducting similar experiments to assess the optimal input concentrations of both extracts and pre-PCR libraries in the context of ancient DNA.

The extent of deamination remains stable as pooling increases, with one exception observed in state 2 ($> 10\%, < 1000$) (Fig. 3b). For samples exhibiting the characteristics of state 2, it is plausible that the introduction of increased dilution with the negative sample containing a 2.4 × modern human contamination masks the impact of the limited aDNA portion within the positive sample. This aligns with the observations in Fig. 3, demonstrating a gradual reduction in C to T transitions with each increment in pooling. Notably, a parallel observation can be made for another sample categorized under state 2 which was ultimately excluded from the dataset due to a 1.67 × modern human contamination. This sample displayed a substantial reduction in its deamination signal as pooling intensified (Supplementary Fig. S2). This result supports the hypothesis that modern DNA can effectively mask the distinctive markers of aDNA: The various iterations of PCR amplification during the laboratory pipeline can introduce a technical bias towards the amplification of the more abundant and less fragmented modern

DNA. Characteristic aDNA damage patterns including miscoding lesions and single-strand breaks[111] can block polymerases, leading to chimeric sequences or the termination of amplification[112]. Consequently, this phenomenon leads to the prioritized sequencing of modern sequences over ancient ones during the screening process.

In this study, we employed equal-volume pooling of non-normalized extracts (according to concentration) to achieve increased pooling levels. Rather than merging different extracts, we utilized a single (test) extract and combined it with progressively larger volumes of another (negative) extract to attain higher pooling levels. This strategy was used to simulate the potential low success rate of retrieving ancient mammalian DNA from sedaDNA samples[38,39]. It is important to note that equal-volume pooling does not guarantee equal representation of each extract within the pool. When applying our extract-pooling method for the initial screening of different sediment samples, equal mass pooling might not circumvent heterogeneity in the extract pool. Quantifying the concentration of a sample's extract prior to pooling is reasonable and necessary when dealing with modern DNA, as the target (endogenous) DNA would typically approximate 100 percent. However, for ancient DNA, especially ancient sedimentary DNA, it is more complicated as the total amount of DNA within a sample does not reflect endogenous target DNA content (which, in our case is mostly human and faunal DNA); The vast majority of the DNA stems from microorganisms, plants, and faunal components[31,113,114]. Thus, a bias could also, arguably, arise from pooling by qubit quantification: A sample with overall high DNA concentration may have a very low amount of human or faunal DNA and would thus be underrepresented if corrected for the DNA concentration.

It is a widely used approach in the field of eDNA to pool extracts in the order of hundreds to determine the presence or absence of organisms[20]. However, when the objective is to identify individual positive samples for downstream analyses, an elevated number is suboptimal as a large (financial) effort is required to produce individual libraries from sample pools that tested positive. Therefore, it is clear that an equilibrium must be found. To date, no studies have specifically addressed the positivity rates of sedaDNA from archaeological contexts and there is yet no compelling data on the success rate and its determinants. Based on extensive screening efforts published so far[31,33,34], it appears that positivity—hitherto defined as the proportion of samples with sufficient DNA for classifying mammalian reads—may be around 10–15%. In our laboratory experiment, we established an upper threshold of five extracts per individual pool, assuming a more conservative 20% success rate for our samples. This decision was reached to ensure a balance between analytical efficiency and resource optimization, even if the success rate is high. One of the main objectives of sedaDNA analyses in the light of paleogenomics is to produce high-quality DNA data from hominins or fauna. When defining positivity as samples with enough reads to recover substantial information through population genetics, the success rate drops below 5%. The strength of our pooling method lies in its flexibility: the number of samples pooled can be adapted to specific research objectives. If the goal is to assess the overall preservation (presence or absence) of ancient DNA at a site, the number of samples within a pool can be increased. Conversely, if the objective is to identify hominins or specific mammals, the number of samples can be decreased (as pools with aDNA signals will have to be analyzed individually). Additionally, a subset of samples can first be pooled and pre-screened to determine the general state of aDNA preservation at a site, thereby concluding the optimal number of samples to pool.

Recently, there has been another notable methodological advancement aimed at enhancing the cost-effectiveness of large-scale sample screening, focusing on a step much closer to the end of the laboratory pipeline, referred to as multiplex capture[50]. In contrast to our method, multiplex capture involves individual library preparation, double-indexing, and initial amplification. The streamlining of lab processes is achieved at the step of hybridization capture, through the pooling of multiple double-indexed libraries into a single capture reaction. While reducing the costs for capture reactions considerably, the majority of lab processes have to be conducted individually. This method could additionally be tested in combination with our extract-pooling method.

In conclusion, we have successfully demonstrated two key findings that have important implications for the field of sedaDNA research: 1) Preservation of sensitivity: Our study shows that the pooling of up to five extracts did not compromise the sensitivity of ancient DNA detection. This finding implies that extract pooling can be employed as a cost-effective and efficient strategy for large-scale sediment sample screening without sacrificing the ability to detect ancient DNA. 2) Effective reduction of PCR drawbacks: Our results highlight that negative effects of undetermined nature in double-stranded libraries can be significantly reduced by using low extract input volumes. Although the nature of this observation requires further experimental testing, our approach allows researchers to optimize extract input volumes, leading to enhanced recovery of ancient DNA. Overall, this cost-effective and high-throughput approach has the potential to advance the field and facilitate the study of ancient ecosystems and populations.

Data availability

The datasets generated during the current study are available in the European Nucleotide Archive (ENA) through the accession number PRJEB67982.

Received: 12 January 2024; Accepted: 8 August 2024
Published online: 20 August 2024

References

1. Ariza, M. *et al.* Plant biodiversity assessment through soil eDNA reflects temporal and local diversity. *Methods Ecol. Evol.* **14**, 415–430 (2023).
2. Allen, M. C. *et al.* Sampling environmental DNA from trees and soil to detect cryptic arboreal mammals. *Sci. Rep.* **13**, 180 (2023).
3. Antich, A. *et al.* Marine biomonitoring with eDNA: Can metabarcoding of water samples cut it as a tool for surveying benthic communities?. *Mol. Ecol.* **30**, 3175–3188 (2021).
4. Blattner, L., Ebner, J. N., Zopfi, J. & von Fumetti, S. Targeted non-invasive bioindicator species detection in eDNA water samples to assess and monitor the integrity of vulnerable alpine freshwater environments. *Ecol. Indic.* **129**, 107916 (2021).

5. Lynggaard, C. *et al.* Airborne environmental DNA for terrestrial vertebrate community monitoring. *Curr. Biol.* **32**, 701-707.e5 (2022).
6. Johnson, M. D., Fokar, M., Cox, R. D. & Barnes, M. A. Airborne environmental DNA metabarcoding detects more diversity, with less sampling effort, than a traditional plant community survey. *BMC Ecol. Evol.* **21**, 218 (2021).
7. Ogram, A., Sayler, G. S. & Barkay, T. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Methods* **7**, 57–66 (1987).
8. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: A ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
9. Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. The analysis of natural microbial populations by ribosomal RNA sequences. In *Advances in Microbial Ecology* (ed. Marshall, K. C.) 1–55 (Springer US, 1986).
10. Shi, T., Reeves, R. H., Gilichinsky, D. A. & Friedmann, E. I. Characterization of viable bacteria from Siberian permafrost by 16S rDNA sequencing. *Microb. Ecol.* **33**, 169–179 (1997).
11. Coolen, M. J. & Overmann, J. Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment. *Appl. Environ. Microbiol.* **64**, 4513–4521 (1998).
12. Willerslev, E. *et al.* Long-term persistence of bacterial DNA. *Curr. Biol.* **14**, R9-10 (2004).
13. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 626–631 (2006).
14. Johnson, S. S. *et al.* Ancient bacteria show evidence of DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14401–14405 (2007).
15. Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. Molecular caving. *Curr. Biol.* **13**, R693–R695 (2003).
16. Willerslev, E. *et al.* Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791–795 (2003).
17. Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**, 110–117 (2009).
18. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050 (2012).
19. Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **11**, 759–769 (2011).
20. Rijal, D. P. *et al.* Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. *Sci. Adv.* **7**, eabf9557 (2021).
21. Jørgensen, T. *et al.* Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA meta-barcoding. *Mol. Ecol.* **21**, 1980–1988 (2012).
22. ter Schure, A. T. M. *et al.* Sedimentary ancient DNA metabarcoding as a tool for assessing prehistoric plant use at the Upper Paleolithic cave site Aghitu-3, Armenia. *J. Hum. Evol.* **172**, 103258 (2022).
23. Han, W. *et al.* Anthropogenic activities altering the ecosystem in Lake Yamzhog Yumco, southern Qinghai-Tibetan Plateau. *Sci. Total Environ.* **904**, 166715 (2023).
24. Seeber, P. A. *et al.* Evaluation of lake sedimentary ancient DNA metabarcoding to assess fungal biodiversity in Arctic paleoecosystems. *Environ. DNA* **4**, 1150–1163 (2022).
25. Jia, W. *et al.* Sedimentary ancient DNA reveals past ecosystem and biodiversity changes on the Tibetan Plateau: Overview and prospects. *Quat. Sci. Rev.* **293**, 107703 (2022).
26. Giguet-Covex, C. *et al.* Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* **5**, 1–7 (2014).
27. Bellemain, E. *et al.* Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. *Environ. Microbiol.* **15**, 1176–1189 (2013).
28. Parducci, L. *et al.* Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments From Southern Sweden. *Front. Ecol. Evol.* **7**, (2019).
29. Stahlschmidt, M. C. *et al.* Ancient mammalian and plant DNA from late quaternary stalagmite layers at Solkota Cave, Georgia. *Sci. Rep.* **9**, 6628 (2019).
30. Gelabert, P. *et al.* Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Curr. Biol.* **31**, 3564-3574.e9 (2021).
31. Slon, V., Hopfe, C., Weiß, C. L. & Mafessoni, F. Neandertal and denisovan DNA from pleistocene sediments. *Science* **356**, 605–608 (2017).
32. Zhang, D. *et al.* Denisovan DNA in late pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. *Science* **370**, 584–587 (2020).
33. Zavala, E. I. *et al.* Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature* **595**, 399–403 (2021).
34. Vernot, B. *et al.* Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science* **372**, (2021).
35. Pietramellara, G. *et al.* Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biol. Fertil. Soils* **45**, 219–235 (2009).
36. Parducci, L. *et al.* Ancient plant DNA in lake sediments. *New Phytol.* **214**, 924–942 (2017).
37. Nagler, M., Insam, H., Pietramellara, G. & Ascher-Jenull, J. Extracellular DNA in natural environments: Features, relevance and applications. *Appl. Microbiol. Biotechnol.* **102**, 6343–6356 (2018).
38. Slon, V. *et al.* Extended longevity of DNA preservation in Levantine Paleolithic sediments, Sefunim Cave, Israel. *Sci. Rep.* **12**, 14528 (2022).
39. Massilani, D. *et al.* Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
40. Dorfman, R. The detection of defective members of large populations. *Ann. Math. Stat.* **14**, 436–440 (1943).
41. García, Z. *et al.* Evaluation of a pooling method for routine anti-HCV screening of blood donors to lower the cost burden on blood banks in countries under development. *J. Med. Virol.* **49**, 218–222 (1996).
42. Roth, W. K., Weber, M. & Seifried, E. Feasibility and efficacy of routine PCR screening of blood donations for hepatitis C virus, hepatitis B virus, and HIV-1 in a blood-bank setting. *Lancet* **353**, 359–363 (1999).
43. Quinn, T. C. *et al.* Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence. *AIDS* **14**, 2751–2757 (2000).
44. Mutesa, L. *et al.* A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* **589**, 276–280 (2021).
45. Sawicki, R., Korona-Glowniak, I., Boguszewska, A., Stec, A. & Polz-Dacewicz, M. Sample pooling as a strategy for community monitoring for SARS-CoV-2. *Sci. Rep.* **11**, 3122 (2021).
46. Tveit, A., Schwacke, R., Svenning, M. M. & Urich, T. Organic carbon transformations in high-Arctic peat soils: Key functions and microorganisms. *ISME J.* **7**, 299–311 (2013).
47. Keshri, J., Yousuf, B., Mishra, A. & Jha, B. The abundance of functional genes, cbbL, nifH, amoA and apsA, and bacterial community structure of intertidal soil from Arabian Sea. *Microbiol. Res.* **175**, 57–66 (2015).
48. Song, Z. *et al.* Effort versus reward: Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS One* **10**, e0127234 (2015).

49. Hestetun, J. T., Lanzén, A., Skaar, K. S. & Dahlgren, T. G. The impact of DNA extract homogenization and replication on marine sediment metabarcoding diversity and heterogeneity. *Environ. DNA* **3**, 997–1006 (2021).
50. Zavala, E. I. *et al.* Quantifying and reducing cross-contamination in single- and multiplex hybridization capture of ancient DNA. *Mol. Ecol. Resour.* **22**, 2196–2207 (2022).
51. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).
52. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
53. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
54. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
55. Tejero, J.-M. *et al.* Cervidae antlers exploited to manufacture prehistoric tools and hunting implements as a reliable source of ancient DNA. *Heliyon* **10**, e31858 (2024).
56. Rohland, N. *et al.* Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
57. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
58. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
59. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
60. 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).
61. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
62. Team & R. S,. *RStudio: Integrated Development for R* (RStudio. Inc., 2015).
63. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013). http://www.R-project.org/.
64. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14616–14621 (2007).
65. Brotherton, P. *et al.* Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* **35**, 5717–5728 (2007).
66. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
67. Suther, C. & Moore, M. D. Quantification and discovery of PCR inhibitors found in food matrices commonly associated with foodborne viruses. *Food Sci. Hum. Wellness* **8**, 351–355 (2019).
68. Andersen, M. R. & Omiecinski, C. J. Direct extraction of bacterial plasmids from food for polymerase chain reaction amplification. *Appl. Environ. Microbiol.* **58**, 4080–4082 (1992).
69. Jaykus, L. A., De Leon, R. & Sobsey, M. D. A virion concentration method for detection of human enteric viruses in oysters by PCR and oligoprobe hybridization. *Appl. Environ. Microbiol.* **62**, 2074–2080 (1996).
70. Richards, G. P. Limitations of molecular biological techniques for assessing the virological safety of foods. *J. Food Prot.* **62**, 691–697 (1999).
71. Hamza, I. A. & Leifels, M. Assessment of PCR inhibitor removal methods to monitor viruses in environmental water samples: DAX-8 outperforms competitors. *Water Air Soil Pollut.* **235**, 20 (2024).
72. Haramoto, E. *et al.* A review on recent progress in the detection methods and prevalence of human enteric viruses in water. *Water Res.* **135**, 168–186 (2018).
73. Hamza, I. A. *et al.* Detection of human viruses in rivers of a densly-populated area in Germany using a virus adsorption elution method optimized for PCR analyses. *Water Res.* **43**, 2657–2668 (2009).
74. de Chaves, M.Q.-G. *et al.* A new and accurate qPCR protocol to detect plant pathogenic bacteria of the genus 'Candidatus Liberibacter' in plants and insects. *Sci. Rep.* **13**, 3338 (2023).
75. Kageyama, K., Komatsu, T. & Suga, H. Refined PCR protocol for detection of plant pathogens in soil. *J. Gen. Plant Pathol.* **69**, 153–160 (2003).
76. Approaches for eliminating PCR inhibitors and designing PCR primers for the detection of phytopathogenic fungi. *Crop Prot.* **26**, 145–161 (2007).
77. Albers, C. N., Jensen, A., Bælum, J. & Jacobsen, C. S. Inhibition of DNA polymerases used in Q-PCR by structurally different soil-derived humic substances. *Geomicrobiol. J.* **30**, 675–681 (2013).
78. Tebbe, C. C. & Vahjen, W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl. Environ. Microbiol.* **59**, 2657–2665 (1993).
79. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. PCR inhibitors—Occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
80. Hering, J. G. & Morel, F. M. Humic acid complexation of calcium and copper. *Environ. Sci. Technol.* **22**, 1234–1237 (1988).
81. Sutlovic, D., Gamulin, S., Definis-Gojanovic, M., Gugic, D. & Andjelinovic, S. Interaction of humic acids with human DNA: Proposed mechanisms and kinetics. *Electrophoresis* **29**, 1467–1472 (2008).
82. Matheson, C. D., Gurney, C., Esau, N. & Lehto, R. Assessing PCR inhibition from humic substances. *Open Enzym. Inhib. J.* **3**, 38–45 (2014).
83. Demeke, T. & Adams, R. P. The effects of plant polysaccharides and buffer additives on PCR. *Biotechniques* **12**, 332–334 (1992).
84. Henson, C., Truchot, D. & Canevello, A. PTSD and PTG in French and American firefighters: A comparative study. *Int. J. Environ. Res. Public Health* **19**, 11973 (2022).
85. Wei, T., Lu, G. & Clover, G. Novel approaches to mitigate primer interaction and eliminate inhibitors in multiplex PCR, demonstrated using an assay for detection of three strawberry viruses. *J. Virol. Methods* **151**, 132–139 (2008).
86. Uchii, K. *et al.* Comparison of inhibition resistance among PCR reagents for detection and quantification of environmental DNA. *Environ. DNA* **1**, 359–367 (2019).
87. 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).
88. 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).
89. Rossen, L., Nørskov, P., Holmstrøm, K. & Rasmussen, O. F. Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *Int. J. Food Microbiol.* **17**, 37–45 (1992).
90. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* https://doi.org/10.1128/aem.62.2.316-322.1996 (1996).
91. Frostegård, A. *et al.* Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* **65**, 5409–5420 (1999).

92. Malik, M., Kain, J., Pettigrew, C. & Ogram, A. Purification and molecular analysis of microbian DNA from compost. *J. Microbiol. Methods* **20**, 183–196 (1994).
93. Cho, J.-C., Lee, D.-H., Cho, Y., Cho, J.-C. & Kim, S.-J. Direct extraction of DNA from soil for amplification of 16S rRNA gene sequences by polymerase chain reaction. *J. Microbiol.* **34**, 229–235 (1996).
94. Braid, M. D., Daniels, L. M. & Kitts, C. L. Removal of PCR inhibitors from soil DNA by chemical flocculation. *J. Microbiol. Methods* **52**, 389–393 (2003).
95. Torsvik, V. L. Isolation of bacterial DNA from soil. *Soil Biol. Biochem.* **12**, 15–21 (1980).
96. Cullen, D. W. & Hirsch, P. R. Simple and rapid method for direct extraction of microbial DNA from soil for PCR. *Soil Biol. Biochem.* **30**, 983–993 (1998).
97. Holben, W. E., Jansson, J. K., Chelm, B. K. & Tiedje, J. M. DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbiol.* **54**, 703–711 (1988).
98. Leff, L. G., Dana, J. R., McArthur, J. V. & Shimkets, L. J. Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* **61**, 1141–1143 (1995).
99. Lovell, C. R. & Piceno, Y. Purification of DNA from estuarine sediments. *J. Microbiol. Methods* **20**, 161–174 (1994).
100. Walia, S., Khan, A. & Rosenthal, N. Construction and applications of DNA probes for detection of polychlorinated biphenyl-degrading genotypes in toxic organic-contaminated soil environments. *Appl. Environ. Microbiol.* **56**, 254–259 (1990).
101. Moré, M. I., Herrick, J. B., Silva, M. C., Ghiorse, W. C. & Madsen, E. L. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* **60**, 1572–1580 (1994).
102. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res.* **27**, 1230–1237 (2017).
103. Schneider, S., Enkerli, J. & Widmer, F. A generally applicable assay for the quantification of inhibitory effects on PCR. *J. Microbiol. Methods* **78**, 351–353 (2009).
104. Wang, H., Qi, J., Xiao, D., Wang, Z. & Tian, K. A re-evaluation of dilution for eliminating PCR inhibition in soil DNA samples. *Soil Biol. Biochem.* **106**, 109–118 (2017).
105. Shutler, G. G. *et al. Removal of a PCR Inhibitor and Resolution of DNA STR Types in Mixed Human-Canine Stains from a Five Year Old Case* (ASTM International, 1999).
106. Imaizumi, K. *et al.* DNA typing of bone specimens—The potential use of the profiler test as a tool for bone identification. *Leg. Med.* **7**, 31–41 (2005).
107. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
108. Lebuhn, M., Effenberger, M., Garcés, G., Gronauer, A. & Wilderer, P. A. Evaluating real-time PCR for the quantification of distinct pathogens and indicator organisms in environmental samples. *Water Sci. Technol.* **50**, 263–270 (2004).
109. Lindberg, E., Albrechtsen, H.-J. & Jacobsen, C. S. Inhibition of real-time PCR in DNA extracts from aquifer sediment. *Geomicrobiol. J.* **24**, 343–352 (2007).
110. Jane, S. F. *et al.* Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Mol. Ecol. Resour.* **15**, 216–227 (2015).
111. Lindahl, T. Instability and decay of the primary. *Nature* **362**, 22 (1993).
112. Pääbo, S., Higuchi, R. G. & Wilson, A. C. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J. Biol. Chem.* **264**, 9709–9712 (1989).
113. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. Biol. Sci.* **272**, 3–16 (2005).
114. Pedersen, M. W. *et al.* Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**, 45–49 (2016).

Acknowledgements

This study was supported by a grant from the internal Research Platform MINERVA (https://minerva.univie.ac.at). P.G. was funded through COST iNEAL STSM Grant CA19141-8d068698. S.Sa. was funded by the Austrian Science Fund (FWF) M3108-G. The excavation and sampling of Velika Pećina in Kličevica was financially supported by the Croatian Science Foundation (NECEM, HRZZ-IP-2019-04-6649) and M.S. received funding from the European Research Council (ERC) MicroStratDNA project under Grant Agreement No. 101042570. The authors would like to express their gratitude to the following individuals and institutions for their invaluable contributions to this research: Paul Knabl for the artistic expertise demonstrated in creating the figures and Natalija Čondić and Archaeological Museum Zadar for various forms of assistance during the sampling of Velika Pećina in Kličevica.

Author contributions

Study conceptualization: S.Sa., P.G., V.O., R.P. Methodological design: V.O., S.Sa., P.G., R.P., O.C., Experimental work: V.O., F.B., S.F., E.Z., S.Sz., F.TC., F.E., B.Z. Data analysis: V.O., P.G., and S.Sa. Archaeological conceptualization and sampling: I.K., M.B., L.G.S., M.R.G.M., B.G., M.B., J.K. Project administration and funding acquisition: R.P., T.R., S.M.K., M.S. Writing: V.O., S.Sa., P.G., R.P. with input from the rest of the authors.

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/s41598-024-69741-5.

Correspondence and requests for materials should be addressed to P.G., S.S. or R.P.

4. Chapter 4: New human and faunal sedaDNA genomes from Iberia, Spain

4.1. Publication III

A sedimentary ancient DNA perspective on human and carnivore persistence through the Last Glacial Maximum in El Mirón Cave, Spain

Status: Second round of revisions with Nature Communications

Citation: Pere Gelabert & Victoria Oberreiter, Lawrence Guy Straus, Manuel Ramón González Morales, Susanna Sawyer, Ana B. Marín-Arroyo, Jeanne Marie Geiling, Florian Exler, Florian Brueck, Stefan Franz, Fernanda Tenorio Cano, Sophie Szedlacsek, Evelyn Zelger, Michelle Hämmerle, Brina Zagorc, Alejandro Llanos-Lizcano, Olivia Cheronet, José-Miguel Tejero, Thomas Rattei, Stephan M. Kraemer, Ron Pinhasi. A sedimentary ancient DNA perspective on human and carnivore persistence through the Last Glacial Maximum in El Mirón Cave, Spain.

A sedimentary ancient DNA perspective on human and carnivore persistence through the Late Pleistocene in El Mirón Cave, Spain

Pere Gelabert [1,2,3,*], Victoria Oberreiter [1,2,3], Lawrence Guy Straus [4,5], Manuel Ramón González Morales [6], Susanna Sawyer [1,2], Ana B. Marín-Arroyo [5], Jeanne Marie Geiling [5], Florian Exler [1,2,7], Florian Brueck [1], Stefan Franz [1], Fernanda Tenorio Cano [1], Sophie Szedlacsek [1], Evelyn Zelger [1], Michelle Hämmerle [1,2], Brina Zagorc [1,2], Alejandro Llanos-Lizcano [1,2,8], Olivia Cheronet [1,2], José-Miguel Tejero [1,2,9,*], Thomas Rattei [10], Stephan M. Kraemer [2,7], Ron Pinhasi [1,2,*]

[1] Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

[2] Human Evolution and Archeological Sciences (HEAS), University of Vienna, Vienna, Austria

[3] These authors contributed equally to this work.

[4] Department of Anthropology, University of New Mexico, Albuquerque, NM, USA

[5] Grupo I+D+i EvoAdapta, Departamento de Ciencias Históricas, Universidad de Cantabria, Santander, Spain

[6] Instituto Internacional de Investigaciones Prehistóricas de Cantabria (Universidad de Cantabria, Gobierno de Cantabria, Santander), Santander, Spain

[7] Department of Environmental Geosciences, Centre for Microbiology and Environmental Systems Science, University of Vienna, Austria

[8] Facultad de Química y Farmacia, Universidad del Atlántico, Barranquilla, Colombia

[9] Seminari d'Estudis i Recerques Prehistòriques (SERP), University of Barcelona, Barcelona, Spain

[10] Division of Computational Systems Biology, Centre for Microbiology and Environmental Systems Science, University of Vienna, Austria

* Corresponding authors: pere.gelabert@univie.ac.at, jmtejero@ub.edu, ron.pinhasi@univie.ac.at

Abstract

Caves are primary sites for studying human and animal activity, occupation, genetic ancestry, and subsistence patterns throughout the Palaeolithic. Iberia served as a critical human and animal refugium in Europe during the Last Glacial Maximum (LGM), 26.5 to 19 thousand years before the present (cal kya), marked by significant climatic cooling and other environmental changes. Therefore, the Iberian Peninsula is a key location for understanding human and animal population dynamics before, during and after these climatic events. We recovered and analysed sedimentary ancient DNA (sedaDNA) data from the lower archaeological stratigraphic sequence of El Mirón cave (Cantabria, Spain), encompassing the (1) Late Mousterian period, associated with Neanderthals, and (2) the Gravettian (c. 31.5 cal kya), Solutrean (c. 24.5-22cal kya), and Initial Magdalenian (d. 21-20.5 cal kya) periods, associated with anatomically modern humans. We identified 28 animal taxa, including humans from levels spanning from the Mousterian to the Initial Magdalenian. Fifteen of these taxa had not been identified from the archaeozoological (i.e., faunal) record, including the presence of hyenas in the Magdalenian. Additionally, we provide phylogenetic analyses on 70 sedaDNA mtDNA genomes of fauna, including the densest Pleistocene sampling of *C. lupus* in Iberia. Finally, we recovered three human mtDNA sequences from the Solutrean levels. These sequences and published data suggest mtDNA haplogroup continuity in Iberia throughout the Solutrean/Last Glacial Maximum period.

Introduction

During the Late Pleistocene (Marine Isotope Stages 5-2), Neanderthals and early anatomically modern humans (AMH) were competing with large carnivores for the occupation of the same ecological niches[1,2]. This competition included access to vital resources such as food and shelter[3–9]. Notably, during the Late Middle and Early Upper Palaeolithic, the primary carnivore contenders for these resources were the cave lion (*Panthera spelaea*), leopard (*Panthera pardus*), cave hyena (*Crocuta crocuta spelaea*), wolf (*Canis lupus*) and dhole (*Cuon alpinus*)[10]. Many archaeological cave sites with Middle and Upper Palaeolithic occupational phases exhibit compelling evidence of alternating occupations between carnivores and humans, especially when these occupations were seasonal[11,12]. These caves are thus invaluable repositories for insights into the adaptations and interactions of species during the Upper Palaeolithic (c. 45 to 12 thousand years ago [kya])[13] while shedding light on the behaviour of AMH[3]. However, the physical presence of such animal evidence is very scarce, complicating the assessment of the co-occurrence of humans and particular animal taxa.

For the most part, archaeozoological studies rely on the morphological identification of animal remains and the taphonomic aspects affecting such remains (i.e. carnivore and/or anthropic and other biotic and abiotic agents of modification) to determine the presence of taxa in the available archaeological record[14]. These analyses have also shed light on the distribution of prey species, some potentially related to human subsistence and occupation[15,16]. Recently, palaeoproteomic methods, particularly Zooarchaeology by Mass Spectrometry (ZooMS)[17], enable the detection and charting of taxa from skeletal fragments lacking taxon-specific features[18–23]. However, ZooMS identification is restricted to preserved skeletal material and, as such, cannot reveal traces of vertebrate species that may have frequented the cave. Their remains are not necessarily recovered or part of the archaeozoological assemblage. Moreover, despite recent improvements, ZooMS only enables taxa identification without further phylogenetic inference and may lack resolution at the species level. Hence, the analyses of ancient DNA from sediment (sedaDNA) can provide additional insights by identifying the presence of animal species or human groups in archaeological sites and recovering their DNA even without visible skeletal remains[24–27].

Dated skeletal remains from Europe reveal a significant decline and eventual disappearance of multiple Pleistocene taxa at the end of the Late Pleistocene (Marine Isotope Stages 3 and 2, c. 60-10 kya). Cave hyenas (*Crocuta crocuta spelaea*) are believed to have disappeared from the European archaeological record around 31 kya[28], although there is potential evidence of their later persistence in southern Europe[29]. The faunal record shows cave lions (*Panthera spelaea*) persisted in Europe until 14 kya[30–34], potentially surviving longer in some areas. In northern Atlantic Iberia's Basque Country, the Armintxe cave paintings provide indirect evidence of lions' presence during the Magdalenian[35]. The leopard (*Panthera pardus*)[31–33,36,37] and dhole (*Cuon alpinus*)[10,14,38] eventually vanished at the end of the Pleistocene. Due to the limited amount of remains, the exact extirpation dates cannot be ascertained.

The Cantabrian region in Northern Spain was one of the main European human refugia during the Last Glacial Maximum (LGM, between 26.5-19 cal kya [39]) and has some of the best-preserved Upper Palaeolithic assemblages. Palaeoclimatic and palaeoenvironmental conditions during the LGM involved extreme temperature fluctuations and the expansion of ice sheets, which consequently limited the extent of human-inhabited areas across Europe[40,41]. Recent studies have revealed genomic homogeneity among European populations during the pre-LGM Gravettian period (33-24.5 kya)[42,43]. Later, part of this gene pool survived in the Franco-Cantabrian refugium with genetic continuity persisting during the Solutrean (24.5–21 kya)[44] and the succeeding post-LGM Magdalenian archaeological, cultural period (21-12 kya)[43,45].

Here, we present our results for the study of sedaDNA data from the lower archaeological stratigraphic sequence of the El Mirón Cave (Cantabria, Spain) vestibule rear, north section (Figure 1, Supplementary Figure 1, Supplementary Section 1). The site has a culture-stratigraphic sequence of levels dated by 101 radiocarbon determinations spanning the period from >46,000 to c. 4000 cal BP (Straus and González Morales 2018; Straus and González Morales 2012). The cave exhibits excellent organic preservation, and all the faunal remains recovered from the Gravettian and Solutrean levels have been studied, making it one of the best records of the Last Glacial Maximum (LGM) in the Iberian Peninsula. The results show 1) the identification of mammalian species through sedaDNA, which are absent or rare in the studies of the archaeofaunal assemblages, including the late persistence of carnivores

such as hyenas and leopards in Iberia, evidencing cooccurrence with humans, 2) the discovery of undocumented genetic phylogenies through the study of 70 mtDNA sequences from 11 fauna species recovered from El Mirón sediments, and 3) human mtDNA continuity through the LGM, pointing towards genetic stability.

Results

Patterns of DNA presence in El Mirón

We screened 32 sediment samples (Supplementary Table 1) from the rear vestibule of El Mirón cave (deep *sondage* meter-squares V, W, X10) (Figure 1, Supplementary Figure 1, Supplementary Section 1) spanning from the late Mousterian to the Initial Magdalenian (>46,000-21,000 cal BP). We used a mitochondrial (mtDNA) in-solution capture designed at the University of Vienna and manufactured by Twist[46], including 51 mammalian species for retrieving sedaDNA from a broad spectrum of taxa. After a strict classification process (Supplementary Sections 2-3), we detected the presence of 31 taxa of animals (including humans), 28 of which had congruent signals of aDNA based on length and deamination (Supplementary Tables 2-3, Supplementary Sections 2-3). The number of recovered species and DNA amounts varies across the profile (Figure 2). In the samples from this study, DNA preservation is related to archaeological level age, and lower levels show less preservation (Supplementary Figure 4). Based on a PCR assay, we tested the inhibitory effect of the purified extracts, and we found that the lower DNA yields of level 130 are not related to inhibition and are likely due to poor organic preservation (Supplementary Section 2, Supplementary Figures 3-4).

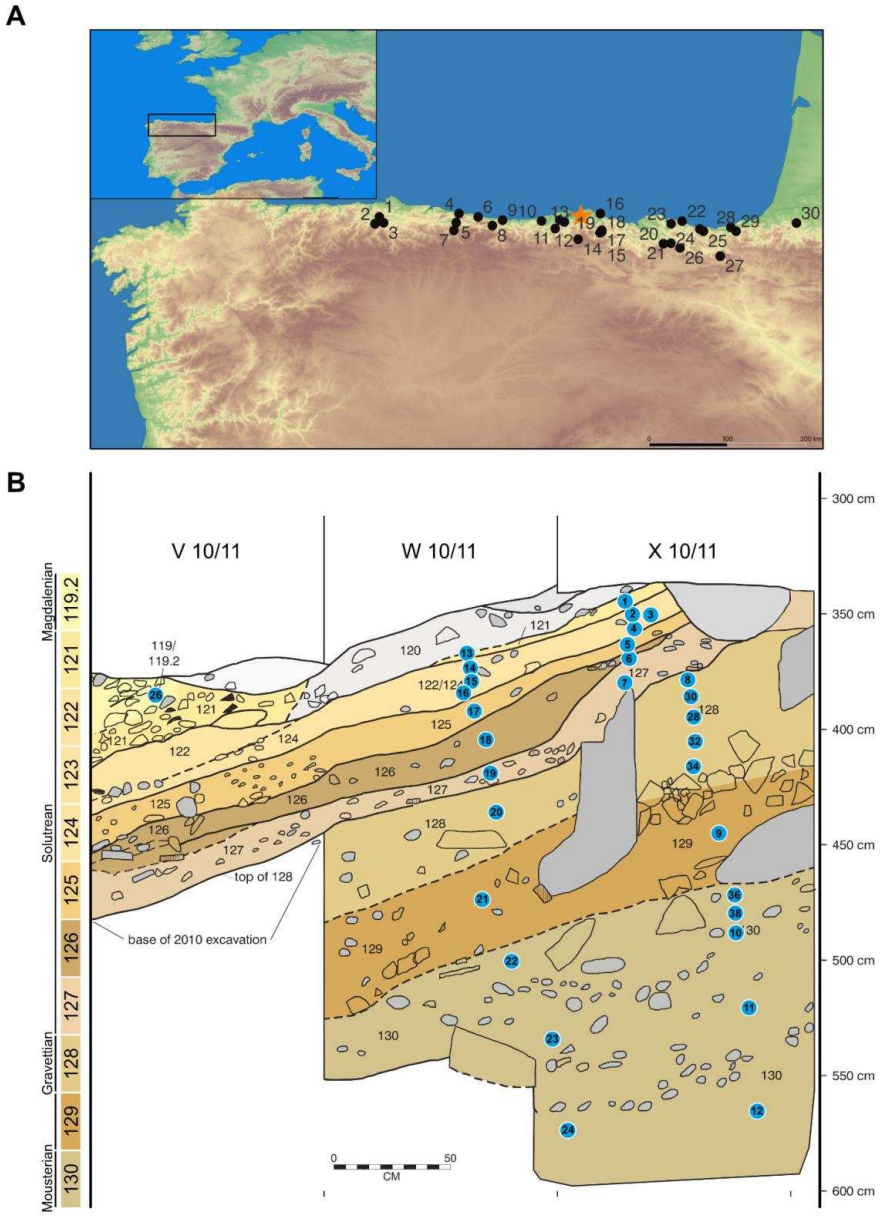


Figure 1: Sediment samples from El Mirón: *A) Location of El Mirón cave (denoted as an orange star) in the Cantabrian Region of northern Spain as well as some significant other caves located in this area: 1. La Paloma, 2. El Conde, 3. La Viña, 4. Tito Bustillo, 5. Los Azules, 6. La*

Riera, Cueto de la Mina 7. Collubil, 8. Coimbre, 9. Llonín, 10. Altamira, 11. Hornos de la Peña, El Castillo, 12. El Pendo, 13. Morín, 14. Rascaño, 15. La Garma, 16. La Fragua, 17. El Otero, 18. El Valle, 19. El Mirón, 20. Axlor, 21. Bolinkoba, 22. Cueva de Santa Catalina, 23. Santimamiñe, 24. Ekain, 25. Erralla, 26. Lezetxiki, 27. Coscobilo, 28. Ametzagaina, 29. Aitzbitarte III & IV, 30. Isturitz (French Basque Country) B) El Mirón Cave NE corner of the vestibule rear stratigraphic section with locations of the collected samples (blue dots) in the north face of the V-W-X/10 excavation units, including the W-X10 deep sondage, numbers inside dots are the sample numbers (modified from L.G. Straus & R. L. Stauber).

The archaeological evidence of human activity in the Mousterian (Level 130), Early Upper Paleolithic (Levels 129-128) or even Solutrean (Levels 127-121) periods is far less abundant than in the Late Upper Palaeolithic (Magdalenian). These Levels (130-121) are located at the rear of the cave (squares V-W-X10, about 2-3 m^2 of excavated area), (Figure 1B, Supplementary Figure 1): Level 130 yielded 115 lithic artefacts; in comparison, 573 lithic elements have been recovered from Level 128[47]. However, none of the levels 130, 129 and 128 yielded evidence of hearths or fire-cracked rocks that would have suggested intense human occupation in this part of the cave and during these periods[47,48]. The overlying Solutrean levels (127-121) show an increased density of archaeological remains and organic matter with seasonal, but somewhat more intense occupations[11,48]. In this study, we observe that the lowest level of El Mirón (130) presents the lowest amounts of aDNA (Supplementary Section 3) and also the lowest number of identified animal species through sedaDNA (Supplementary Table 3). While the average in our dataset is 10.6 species per sample, only one sample from level 130 yielded 10 identified species (Figure 2, Supplementary Table 3). Overall, the distribution of prey taxa is in agreement with the archaeofaunal descriptions. However, this is not the case for carnivores and uncommon taxa (Figure 2, Supplementary Section 3, Supplementary Table 16). We did not recover endogenous human DNA from level 130, but traces of contamination (Supplementary Table 3-9, Supplementary Section 3).

Detection of undocumented species at El Mirón through sedaDNA

Carnivores and some herbivores, such as reindeer (*Rangifer tarandus*), are present in relatively low numbers in Cantabrian archaeofaunal Upper Palaeolithic assemblages[11,44,48,49]. Many of the especially larger carnivores and bears are usually not observed during periods of

human occupation, suggesting alternating use of certain caves by humans, carnivores, and bears[5,11,14,18,44,48–52]. We successfully recovered sedaDNA of multiple species that are not identified among the faunal remains, namely *Crocuta crocuta, Lynx pardinus, Cuon alpinus, Rangifer tarandus, Falco sp., Columba livia, Pyrrhocorax sp., Sorex araneus, Talpa europea, Mustela nivalis, Coelodonta antiquitatis and Strix sp.* (Supplementary Tables 3-4, Figure 2). Furthermore, we discovered species at certain levels with no existing records of their physical remains, including those of *Panthera pardus, Mammuthus primigenius* and *Vulpes vulpes.* For example, archaeozoological remains of *Panthera pardus* have only been recovered from level 128, while we have genetic evidence of its presence in all the sampled levels except 122 (Supplementary Table 3). In addition, we also identified reindeer DNA in the undated and artefact-poor level 129 and the Gravettian-age level 128. Overall, its presence in both the sedaDNA and archaeofaunal registers is limited. Previously, it was only identified in the fossil record of El Mirón in the form of a grooved incisor pendant in the Lower Magdalenian level 17 (Cabin area, Supplementary Figure 1)[44]. However, although not abundant, its presence has been identified in several archaeological sites, mainly in the Basque region[53] but as far west as Asturias, dating from the Aurignacian to the Magdalenian[18,54,55].

We have identified an average of 3.4 species of carnivores per sample, with eight samples from levels 129 to 119.2 showing up to 5 different species in a single sample (Supplementary Tables 3 and 5). Previous archaeozoological research on El Mirón levels 130-119.2 identified *Canis lupus* fragments only in levels 128 and 130, and no *Cuon alpinus* fragments have been reported [48]. In contrast, here we identify the presence of both canids (dhole and wolf) in all the studied levels of El Mirón (Supplementary Tables 3-4, Figure 2). Finally, based on the study of sedaDNA, we determined that while no cave bear (*Ursus spelaeus*) is present in El Mirón, genetic traces of brown bears (*Ursus arctos*) are present throughout the lower profile (Mousterian-Initial Magdalenian) of the site sequence.

In summary, we expand the temporal range of several carnivore species to later strata than previously documented. Specifically, we push back the extinction timeline of spotted hyenas in Iberia to at least the Magdalenian/post-LGM period. Additionally, despite their limited representation or absence in the archaeological record, we provide evidence for the

continued presence of species such as dhole and leopard in El Mirón through the MIS2 (Figure 2, Supplementary Section 3).

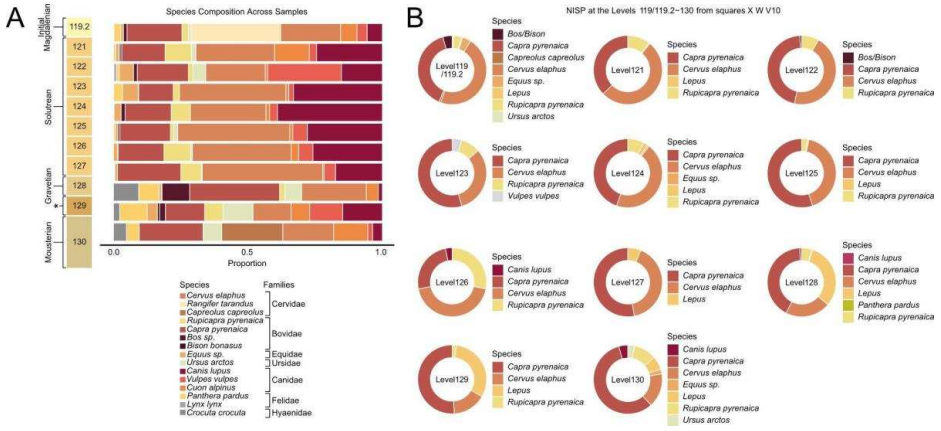


Figure 2. Faunal distribution at El Mirón: *A) Principal species identified with sedaDNA per archaeological level. The families with corresponding genera/species are displayed in the legend according to highest to lowest overall read presence and grouped by families. The X-axis displays the normalised read counts for each genus/species within each archaeological level, while the Y-axis shows the archaeological levels and associated human archaeological cultures. The asterisk denotes a level (129) with little human presence and undefinable archaeological culture. B) NISP according to level (Supplementary Table 4).*

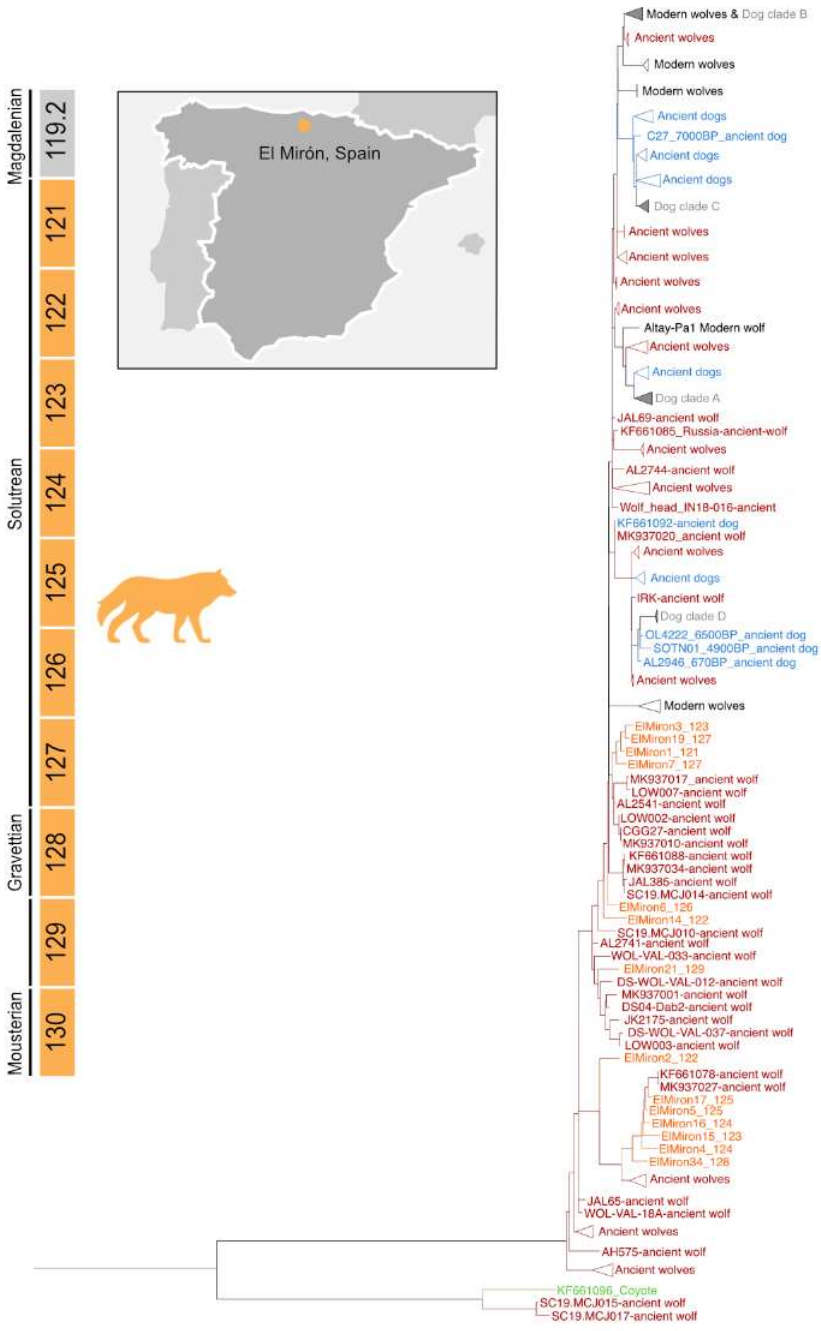
sedaDNA reveals multiple faunal mtDNA lineages at El Mirón

Our previous results focused on identifying the presence of mammalian mtDNA based on a classification of sedaDNA sequences. However, identifying the presence of specific species does not shed light on their genetic phylogenies. The remarkable DNA preservation at El Mirón enabled the assembly of 70 partial and complete mtDNA genomes to go beyond species identification and explore the phylogenetic relationships of 12 of the 28 identified species, including humans.

For the phylogenetic analyses, we focused on the samples with average coverage depths greater than 5X and signs of deamination (more than 0.4 at read end) (Supplementary Table 3, Supplementary Sections 2, 3 and 4), which allowed us to assemble mtDNA genomes of 21 *Cervus elaphus,* 17 *Capra pyrenaica,* 14 *Canis lupus,* 6 *Rupicapra pyrenaica*, 4 *Vulpes vulpes*, 3 *Cuon alpinus*, 2 *Panthera pardus,* 1 *Rangifer tarandus*, 1 *Ursus arctos*, and 1 *Equus* sp samples. (Supplementary Table 6). *These numbers indicate the individual samples for each species that had a coverage depth greater than 5X.* All these genomes have high deamination patterns and missing sites (Supplementary Table 3). The missing sites are likely due to the stringency of the classification that has reduced the coverage across mammal-conserved mtDNA regions (Supplementary Table 3). We also observed variability in the mtDNA sequences in each sample, both using all the substitutions or restricting it to variable transversions to minimise the effect of damage (Supplementary Table 7). This suggests that each genome originated from multiple individuals, a scenario that precludes the feasibility of conducting calibrated phylogenies for estimating split times.

Only seven bone fragments of *Canis lupus* have been identified, where sedaDNA levels have been studied (Supplementary Tables 4-5). In contrast, all the levels except 119.2 and 130 yielded enough mtDNA to reconstruct partial *C. lupus* mitogenomes with average coverages ranging from 5.3X to 35X, These partial mitogenomes represent the densest up-to-date sequencing of mtDNA Pleistocene *C. lupus* of Iberia. The mitogenomes were aligned with Palaeolithic and modern *C. lupus* mitogenomes[56–59]. First, we observe that wolf lineages are diverse in time and space, all showing phylogenetic relationships with Pleistocene wolves of Europe [56,58,60]. Some El Mirón sequences cluster close to Palaeolithic wolves from Goyet Cave, Belgium[59] (Figure 3, Supplementary Figure 11), showing temporal similarity. The oldest

255 available mtDNA sequences of Pleistocene European dogs fall close to or within dog clades A,
256 C and D [57,58,61,62]. All the newly reported sequences from El Mirón fall out of this diversity and
257 are close to the oldest wolves of the dataset (Figure 3). Therefore, we can only confirm the
258 presence of wolves in the Solutrean sequence and not the presence of domestic *C. lupus*
259 lineages[58].
260
261 From levels 129, 128 and 127 we also recovered the partial genomes of dholes
262 (Supplementary Figure 7), with remains previously found in northern Atlantic Spain only in
263 the Lower Magdalenian of nearby Rascaño[38] cave. The dhole mtDNA from El Mirón resembles
264 one Pleistocene mtDNA sequence from Bacho Kiro (Bulgaria) and expands the support of the
265 existence of at least two different haplotypes of *Cuon alpinus* in Europe[63] differentiated from
266 the modern dholes (Supplementary Figure 7). We further identified four *Vulpes vulpes*
267 genomes from levels 129, 126, 122, and 121. The *V. vulpes* sedaDNA mtDNA genomes all fall
268 within the range of *V. vulpes* modern diversity (Supplementary Figure 14).


Magdalenian
Solutrean
Gravettian
Mousterian
119.2
121
122
123
124
125
126
127
128
129
130
El Mirón, Spain
Modern wolves & Dog clade B
Ancient wolves
Modern wolves
Modern wolves
Ancient dogs
C27_7000BP_ancient dog
Ancient dogs
Ancient dogs
Dog clade C
Ancient wolves
Ancient wolves
Ancient wolves
Ancient wolves
Altay-Pa1 Modern wolf
Ancient wolves
Ancient dogs
Dog clade A
JAL69-ancient wolf
KF661085_Russia-ancient-wolf
Ancient wolves
AL2744-ancient wolf
Ancient wolves
Wolf_head_IN18-016-ancient
KF661092-ancient dog
MK937020_ancient wolf
Ancient wolves
Ancient dogs
IRK-ancient wolf
Dog clade D
OL4222_6500BP_ancient dog
SOTN01_4900BP_ancient dog
AL2946_670BP_ancient dog
Ancient wolves
Modern wolves
ElMiron3_123
ElMiron19_127
ElMiron1_121
ElMiron7_127
MK937017_ancient wolf
LOW007-ancient wolf
AL2541-ancient wolf
LOW002-ancient wolf
CGG27-ancient wolf
MK937010-ancient wolf
KF661088-ancient wolf
MK937034-ancient wolf
JAL385-ancient wolf
SC19.MCJ014-ancient wolf
ElMiron6_126
ElMiron14_122
SC19.MCJ010-ancient wolf
AL2741-ancient wolf
WOL-VAL-033-ancient wolf
ElMiron21_129
DS-WOL-VAL-012-ancient wolf
MK937001-ancient wolf
DS04-Dab2-ancient wolf
JK2175-ancient wolf
DS-WOL-VAL-037-ancient wolf
LOW003-ancient wolf
ElMiron2_122
KF661078-ancient wolf
MK937027-ancient wolf
ElMiron17_125
ElMiron5_125
ElMiron16_124
ElMiron15_123
ElMiron4_124
ElMiron34_126
Ancient wolves
JAL65-ancient wolf
WOL-VAL-18A-ancient wolf
Ancient wolves
AH575-ancient wolf
Ancient wolves
KF661096_Coyote
SC19.MCJ015-ancient wolf
SC19.MCJ017-ancient wolf

269

Figure 3: Maximum likelihood tree of the Canis lupus mtDNA genomes. *Orange denotes the Canis lupus mtDNA genomes from El Mirón sediment samples. Ancient wolf mtDNA genomes from other locations are coloured in red. Modern wolves are coloured light grey, Modern dog clades are coloured grey, and Ancient dogs are coloured blue. The green colour is used for a Coyote genome. The tree is rooted with dhole mtDNA sequences, including two from El Mirón (Supplementary Figure 11).*

P. pardus is the only species from the genus *Panthera* identified at El Mirón through archaeological faunal analyses. Our sedaDNA analysis confirmed this result. We recovered two partial genomes of *P. pardus* from levels 129 (46890–33160 cal BP) and 125 (22980–22240 cal BP), showing nucleotide diversity that suggests multiple individual origins. We generated a Maximum Likelihood tree that shows that both El Mirón sequences form a clade closely related to the BAR001 genome from Mezmaiskaya Cave (Russia, Northern Caucasus)[64] (Figure 4, Supplementary Figure 19). Hence, the El Mirón mtDNA consensus sequences are more similar to 35.000 YBP sequences from the Caucasus than the Baumannshöhle (Germany, ~ 40 kya) genomes, meaning that the *P. pardus* from Europe would not be a monophyletic clade providing further complexity to the scenario suggested in Paijmans et al. [64]

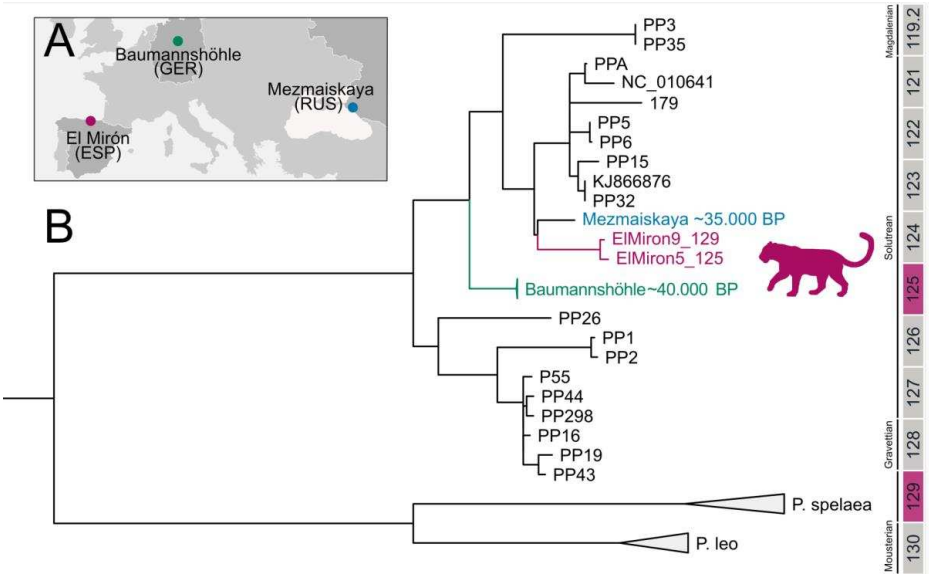


Figure 4: Maximum likelihood tree of P. pardus. *A genomic similarity between El Mirón leopards and Caucasus P. pardus BAR001 from the Pleistocene is observed. Circles represent the P. pardus mitogenomes: El Mirón (purple), Baumannshöhle (green), and Mezmaiskaya (blue).*

We recovered *Ursus arctos* DNA from all periods. However, only the El Miron_9 sample from level 129 has enough DNA to recover a 10X mtDNA genome. The *U. arctos* genome from El Mirón belongs to Clade 3a with other European genomes but is not similar to other Pleistocene genomes from Spain, which are present in Clade 1[65]. These published genomes are, however, much younger and could indicate recent changes in mtDNA diversity in Iberia (Extended Figure 1A, Supplementary Figure 22).

One species, *Crocuta crocuta spelaea,* did not have sufficient DNA preservation to produce mtDNA genomes using our 5X coverage filtering. However, we used the reads from ElMiron_10 from level 130 to assess its phylogenetic connection. The most recent fossils of *Crocuta crocuta spelaea* in Iberia are dated to 23.9 kya[66], with a single dated coprolite attributed to hyena (12,780 cal BP), suggesting potential persistence in Iberia until the late-glacial period [29]. The currently available data from Palaeolithic hyenas suggest diversity in mitochondrial haplotypes in Europe[67]. We thus performed a distance matrix analysis with the available mtDNA genomes from ancient and present-day *C. crocuta* (Supplementary Table 8) to understand the genetic relationship of the ElMiron_10 with other hyena populations. The Hyena mtDNA reads from level 130 of El Mirón and are most similar to *C. crocuta* specimens from Haplogroup A[67]. Haplogroup A is the same as CC8 and CC9 from France, dated to 22.6 kya[68].

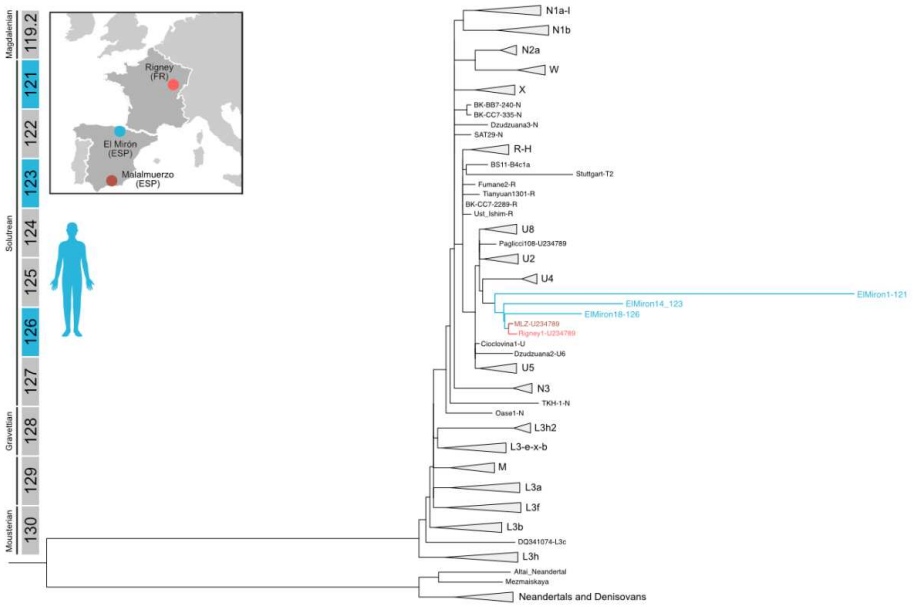
Finally, we recovered partial mtDNA genomes of multiple herbivores. *Cervus elaphu*s mtDNA genomes with coverages ranging from 5.5X to 63.68X could be reconstructed from levels 129 to 121. (Supplementary Tables 3 and 6). Phylogenetically, they form a clade with the Solutrean-age genome from another Palaeolithic cave site (Cueva Chufín) in western Cantabria[69] and other sequences from Denmark and Poland. A Maximum Likelihood tree (Extended Figure 1B, Supplementary Figure 26) shows that all the sequences from El Mirón are placed in Western Clade A[70] and closely related to the previously reported Iberian

Pleistocene *C. elaphus* with the exclusion of some Late Pleistocene *C. elaphus* from Northern Spain in the palaeontological site of Liñares cave[71] forms a separate clade. Five partial genomes of Spanish chamois (*R. rupicapra pyrenaica*) and 17 partial genomes of Spanish Ibex (*Capra pyrenaica)* were recovered. The sequences of both species from El Mirón sedaDNA samples show little population diversity at the mtDNA level (Supplementary Figure 34), and the sequences are congruent with those originating from multiple individuals. An mtDNA genome of *Rangifer tarandus* from level 129 falls close to *R. tarandus* sequences (Supplementary Figure 26). Lastly*, Equus* sp. from the El Mirón Solutrean level 125 mtDNA is close to Pleistocene sequences and falls as an outgroup of clade A, B, and D [72] (Supplementary Figure 37).

Human mtDNA at El Mirón resembles Solutrean Iberian lineages.

We recovered and validated human DNA from 10 sediment samples. Every library produced from these sediment samples shows the typical deamination pattern and congruent read length distribution (Supplementary Table 9). The number of recovered filtered reads varies, and only three samples from Solutrean levels presented enough reads to perform analyses: Sample ElMiron_1 from level 121 (2.5X), sample ElMiron_14 from level 122 (5.6X), and sample ElMiron_18 from level 126 (6.4X) (Supplementary Figures 39-42). The contamination estimates of Schmutzi are congruent with limited modern DNA contamination (Supplementary Table 9). Calico shows that the three sequences originate from multiple donors (Supplementary Table 9). The genotypes of all the relevant alleles are presented in Supplementary Table 10, suggesting multiple donor individuals but no modern contamination. Next, we called the majority sequence of these individuals using Schmutzi. The consensus sequences were analysed with Haplogrep 3 to determine the haplogroup of most calls. We successfully assigned two samples (levels 122 and 126) to haplogroup U2'3'4'7'8'9, and a sample from level 121 could not be assigned further than being identified as falling within R or U haplogroup diversity, typical for the Upper Pleistocene in Europe[43,73]. The sequences from levels 122 and 126 are closely related to the Solutrean-age Malalmuerzo individual (~23,000 cal yr BP) in Andalucia (southern Spain)[74], and they have the same haplogroup as the Solutrean individual from La Riera (21,011-20,725 yr calBP)[43], located ca. 125 km west of El Mirón in Asturias. All these Fournol-ancestry individuals have haplogroup U2'3'4'7'8'9, currently identified in individuals with genetic ancestry linked to Gravettian,

353 Solutrean or Magdalenian genetic groups and restricted to individuals that lived from 27,000
354 to around 13,000 years ago. [73]. The Fournol-ancestry refers to the Gravettian-like ancestry
355 defined by an individual from Southern France that is thought to represent genetically the
356 Iberian Solutrean genetic diversity(Posth et al. 2023). The mtDNA results point to
357 mitochondrial continuity in the site throughout the Solutrean. We produced a Maximum
358 Likelihood phylogeny with multiple modern and Pleistocene mtDNA sequences from Europe
359 that evidence the connections between El Mirón samples and other Solutrean individuals
360 (Figure 5, Supplementary Table 11). So far the Haplogroup U2'3'4'7'8'9 has been identified in
361 thirteen [42,43,74]. All these individuals belong the Gravettian, Solutrean or Magdlaenian cultures
362 from Poland to the Iberian Peninsula.
363



364

365 ***Figure 5: Maximum likelihood tree of Homo sapiens*** *performed with 100 bootstrap*
366 *replications. We observe that the three mtDNA sequences from El Mirón sediment samples*
367 *from levels 121,122, and 126 (in blue) are located together and close to the individual from la*
368 *Cueva de Malalmuerzo (brown)[74] and another individual from Rigney (France) (pink)[43]. El*

Mirón samples have long branches, suggesting an excess of substitutions added, a product of originating from multiple individuals.

Discussion

In this study, we present the first high-resolution sedaDNA analysis of faunal mitochondrial genomes through the Ice Age in a refugium. This pioneering approach is combined with extensive previous archaeological and archaeozoological research conducted at El Mirón Cave [11,75–79]. Our findings do not contradict the archaeological evidence, indicating relatively limited human activity and significant carnivore presence during the Middle to Upper Palaeolithic transition in the rear vestibule of El Mirón Cave [48]. The sedaDNA analyses identified 28 taxa (6 carnivores and 21 herbivores and humans). In contrast, the faunal analyses identified only 15 species, highlighting the capacity of sedaDNA to detect the presence of uncommon taxa throughout the archaeological profile. We detected genetic material indicating that uncommon taxa such as hyenas and leopards survived in Iberia until the end of the Solutrean or even into the Initial Magdalenian periods. Additionally, we revealed the extensive presence of dholes in the Pleistocene, currently rare in Pleistocene assemblages [61]. Overall, we provide key genetic data that extends the known occurrence dates of Pleistocene megafauna in Iberia. Future research combining sedaDNA with other molecular methods, such as palaeoproteomics, may help identify rare, taxonomically undiagnostic fragments of these species [18,19], shedding light on their temporal dynamics.

We report the presence of carnivores (*C. crocuta, L. pardinus, P. pardus, C. lupus, C. alpinus and V. vulpes*) across all stratigraphic levels, with no evident differences in the number of carnivore species per sample except for notably lower preservation in level 130 and 119.2. Notably, levels 128-130 are situated near the cave walls in the rear part of the vestibule, an area closer to the dark inner cave. This location likely attracted carnivores more frequently than humans, as indicated by the limited archaeological evidence of human occupation in this part of the cave as observed in Amalda cave.[7] In the future, denser sampling in space and time could provide substantial data to identify quantitative differences and resolve occupation hiatuses, and date alteration patterns. We observed that older samples exhibit reduced DNA preservation, younger samples however show more inhibition which indicates an overall better preservation of organic material (DNA and inhibitors). Additionally, our data suggest

401 that the archaeofaunal representation aligns more closely with sedaDNA findings for prey
402 animals rather than carnivores. The persistent presence of carnivore DNA, coupled with their
403 absence in the physical remains, indicates that carnivores frequently utilized the cave during
404 the Palaeolithic. They likely scavenged leftovers from humans and intermittently occupied the
405 cave in the absence of humans [14]. While carnivores may leave biological traces such as faeces
406 and urine, their physical remains, such as teeth or bones, are less commonly found unless
407 they perish in the cave. Future research linking sedaDNA to tissue origin might shed critical
408 light on cave occupation patterns.
409
410 Beyond the ability to prove the presence of animals without any previous record in the cave,
411 sedaDNA studies allowed the study of the genetic variability of the recovered animals, as
412 previously demonstrated[24,26,80]. Specifically, sedaDNA enables two unique types of analyses
413 that are not (or only very rarely) possible through the morphological study of remains: 1) The
414 determination of the phenotypic/genotypic affinities of the data in comparison with other
415 previously published specimens, 2) The identification of population or species replacements
416 that are not visible in the osseous remains. In this work, we extended the genetic datasets of
417 key species such as wolves, leopards, brown bears, and hyenas while describing previously
418 undocumented differences in their genetic ancestry. In this direction, we detected at least
419 two differentiated lineages of wolves living in the same area and a surprising mtDNA similarity
420 of the leopards from El Mirón to the one from Mezmaiskaya Cave in the Northern Caucasus,
421 proving that the mtDNA from leopards in Europe do not form a monophyletic clade. Finally,
422 we observe that the hyena mtDNA from El Mirón shows similarities with Pleistocene hyenas
423 from present-day France[68], suggesting regional connections at the mtDNA level at the
424 Pleistocene.
425
426 In our study, sedaDNA was also used to provide evidence of the genetic stability of human
427 mtDNA Solutrean lineages in Iberia. The human mtDNA data from three Solutrean levels at El
428 Mirón show little diversity, all similar to two recently published Iberian Solutrean fossils[43,74],
429 suggesting a stable refugium population in this region during and immediately after the
430 LGM[74]. In the future, nuclear genetic data and denser sampling can date the population
431 admixture that resulted in the Lower Magdalenian genetic pool represented by the Red Lady
432 (El Mirón Ancestry)[42].

We have identified several limitations in our study that could impact our findings. Firstly, our focus area within the vestibule of El Mirón is limited to 2-3 m², whereas the entire vestibule site spans over 300 m². This specific focus, aligned with previous studies on the lower sequence of El Mirón may bias our results towards this particular area and its materials. This limitation suggests that different regions within the vestibule could yield different outcomes. Moreover, the location of our study area at the rear of the vestibule might influence the higher presence of carnivore DNA. A larger cave sampling may produce more robust statistical data, especially in the front vestibule [49,81]. In addition, future research microcontextual studies should be carried out to integrate the current results with the microstratigraphic context of the sediments at El Mirón[82], delving into the specific origin of the DNA and resolving possible secondary alterations that have not been identified so far with the current sedimentological, taphonomical, geological and chronological research at El Mirón[48,49,83–86]. Secondly, linking the amount of DNA discovered to specific activities or animal presence is challenging due to uncertainties regarding the tissue or organic material origins (i.e., faeces, saliva, urine, or other fluids). Our data proves the presence of the notorious amount of DNA from species without remains. Further studies on DNA tissue origin can link animal activities and DNA presence. Third, we stretched the capacities to study ancient genomes due to the small amount of available mtDNA Pleistocene genomes of relevant species such as dholes or leopards. This limits our analyses and the inferences we can make. Finally, a potential limitation lies in the risk of DNA not contemporaneously deposited with the sediment. However, our rigorous validation methods and comprehensive analyses are designed to address and minimise these concerns effectively. We have evidenced the capacity to exclude the possible contaminant taxa and individually verify the identifications. The genetic similarity of our data to other Pleistocene sequences and the temporal dynamics strongly support the fact that the DNA is contemporaneous with the stratigraphic levels and available dates.

Our study provides a novel approach to understanding the complex relationships between archaeological and sedimentary aDNA data. Future research on DNA origins in sediments can complement these findings and archeozoological data. Pioneering studies have shown that most aDNA from sediments likely originates from faeces and skeletal elements.[87] This data can be cross-referenced with archaeological findings and taphonomic indicators to determine

whether the occurrence of certain species needs to be re-evaluated or if it indicates their absence in particular archaeological contexts. It can be regarded as a preliminary step towards new strategies for assessing faunal turnovers, identifying domestic forms of animals through sedaDNA, determining the extirpation dates of animal populations, analysing human activity patterns, and understanding the genetic affinities of both animal and human populations.

Methods

The complete methods are described in the Supplementary Material.

Sampling

Sampling was performed in the profile of the archaeological excavation trench at the rear of the El Mirón vestibule in excavation units V-W-X/10, and the W-X10 deep sondage (Figure 1)[86]

Experimental Procedures

Sediment samples were processed at the Palaeogenomics laboratory at the University of Vienna. 50 mg of sediment was processed per sample, following the protocol from Dabney[88] with adaptations from Korlević[89]. We enriched the indexed libraries for 51 animal mitogenomes, including human mitochondrial DNA, using custom-designed capture following the TWIST capture protocol[90]. Libraries were sequenced at the Vienna BioCenter Core Facilities (VBCF) on the Illumina NovaSeq 6000 platform.

Bioinformatics

Sequencing reads were demultiplexed. We removed adapters and reads shorter than 30 bases using Cutadapt 4.2[91]. These reads were processed with SGA[92] and FASTX-toolkit[93] to remove reads with low qualities and exact duplicates. Filtered reads were then classified at the family level using euka[94]. The classified reads were then converted to fastq reads with Samtools and aligned with BWA aln 0.7.17[95] against each reference sequence (Supplementary Section 3). Single animal genomes were studied, and we determined the deamination values, average read length, and the number of recovered sites. We explored the diversity within the genomes by determining the number of variable positions. We determined the presence of partial genomes when the average coverage of a taxon was >5X. We also included a partial

human genome with 2.6x coverage in our analyses because human DNA was rarely detected in this study. Consensus sequences were merged with present-day and ancient sequences and were aligned using MAFFT 7.52[96] and maximum likelihood phylogenies performed with MEGA[97].

Sequencing reads are available at ENA through accession code PRJEB74514

Author Contributions

PG, JMT, RP, LGS, and MRGM conceptualised the study. OC, MRGM, VO, and PG carried out the fieldwork (sampling). VO, FB, SF, FTC, SSz, EZ, BZ, FE, ALL, and SS performed the experiments. PG, VO, MH, ABMA, and JMG analysed the data. PG, VO, JMT, LGS, SS, ABMA, and RP wrote the text with input from all collaborators. LGS wrote the supplementary description of the site, its stratigraphy, industries and faunas.

Acknowledgements

The research was funded by the MINERVA Research Platform grant (RP, SMK, TR, VO) of the University of Vienna. A HEAS Seed Grant funded JMT, PG, and OC. PG is supported by INEAL through STMG grant CA19141-8d068698. Research of JMT is supported by the program Ramón y Cajal of the Spanish MCIN/AEI (MCIN/AEI/10.13039/501100011033. Project Number RYC2021-033759-I) and the European Community (NextGenerationEU»/PRTR). Excavations in El Mirón Cave directed by LGS and MRGM between 1996 and 2013 and continuing since 2022 with the additional direction of Igor Gutiérrez Zugasti and David Cuenca Solana, are authorised and partially funded by the Gobierno de Cantabria, with additional funding from the US National Science Foundation, Fundación M. Botín, Leakey Foundation, Ministerio de Educación y Ciencia, National Geographic Society, University of New Mexico, UNM Foundation Fund for Stone Age Research (J. and R. Auel, principal donors) and material support from the Town of Ramales de la Victoria and IIIPC Universidad de Cantabria. The campaigns from 2021 to 2023 have been financed through a nominative subsidy from the Department of Culture, Tourism and Sports of the Government of Cantabria within the framework of the project “The Prehistory of the Asón Basin: The caves of El Mirón (Ramales de la Victoria) and La Chora (Voto) ABMA is supported by SUBSILIENCE project (H2020-ERC-2018.CoG N. 818299). During his MSc, ALL was funded by Colfuturo (Fundación para el futuro

de Colombia). Thanks to Paul Knabl for illustrating Figures 1,2 and 4 and assisting with other figures.

References

1. Jones, E. L. & Carvalho, M. Ecospaces of the Middle to Upper Paleolithic transition: The archaeofaunal record of the Iberian Peninsula. *J. Hum. Evol.* **177**, 103331 (2023).
2. Lindly, J. Hominid and carnivore activity at Middle and Upper Paleolithic cave sites in eastern Spain. *Munibe Antropologia-Arkeologia* **40**, 45–70 (1988).
3. Rosell, J., Baquedano, E., Blasco, R. & Camarós, E. New insights on hominid-carnivore interactions during the Pleistocene. *Journal of taphonomy* **10**, 125–128 (2012).
4. Arilla, M., Rosell, J. & Blasco, R. A neo-taphonomic approach to human campsites modified by carnivores. *Sci. Rep.* (2020).
5. Pinto-Llona, A. C., Estaca, V., Grandal-d'Anglade, A., Romero, A. J. & Yravedra, J. Alternation between humans and carnivores in the occupations of the Mousterian site of Sopeña rock-shelter (Asturias, Spain). *Quat. Sci. Rev.* **328**, 108468 (2024).
6. Yravedra, J. A taphonomic perspective on the origins of the faunal remains from Amalda Cave (Spain). *Journal of Taphonomy* **8**, 301–334 (2010).
7. Sánchez-Romero, L. *et al.* New insights for understanding spatial patterning and formation processes of the Neanderthal occupation in theAmalda I cave (Gipuzkoa, Spain). *Sci. Rep.* **10**, 8733 (2020).
8. Marín-Arroyo, A. B. & Sanz-Royo, A. What Neanderthals and AMH ate: reassessment of the subsistence across the Middle–Upper Palaeolithic transition in the Vasco-Cantabrian region of SW Europe. *of Quaternary Science* (2022).
9. Yravedra, J. New Contributions on Subsistence Practices during the Middle-Upper

Paleolithic in Northern Spain. in *Zooarchaeology and Modern Human Origins: Human Hunting Behavior during the Later Pleistocene* (eds. Clark, J. L. & Speth, J. D.) 77–95 (Springer Netherlands, Dordrecht, 2013).

10. Ripoll, M. P., Morales Pérez, J. V., Sanchis Serra, A., Aura Tortosa, J. E. & Montañana, I. S. Presence of the genus Cuon in upper Pleistocene and initial Holocene sites of the Iberian Peninsula: new remains identified in archaeological contexts of the Mediterranean region. *J. Archaeol. Sci.* **37**, 437–450 (2010).
11. Marín-Arroyo, A. B. *et al.* Seasonality of Human Occupations in El Mirón Cave: Late Upper Paleolithic Hunter-Gatherer Settlement-Subsistence Systems in Cantabrian Spain. *Journal of Paleolithic Archaeology* **6**, 7 (2023).
12. Churchill, S. E. *Thin on the Ground: Neandertal Biology, Archeology, and Ecology*. (John Wiley & Sons, 2014).
13. Zilio, L. *et al.* Examining Neanderthal and carnivore occupations of Teixoneres Cave (Moià, Barcelona, Spain) using archaeostratigraphic and intra-site spatial analysis. *Sci. Rep.* **11**, 4339 (2021).
14. Sanchis, A. *et al.* Neanderthal and carnivore activities at Llonin Cave, Asturias, northern Iberian Peninsula: Faunal study of Mousterian levels (MIS 3). *C. R. Palevol* **18**, 113–141 (2019).
15. Vidal-Cordasco, M., Ocio, D., Hickler, T. & Marín-Arroyo, A. B. Ecosystem productivity affected the spatiotemporal disappearance of Neanderthals in Iberia. *Nat Ecol Evol* **6**, 1644–1657 (2022).
16. Vidal-Cordasco, M., Terlato, G., Ocio, D. & Marín-Arroyo, A. B. Neanderthal coexistence with *Homo sapiens* in Europe was affected by herbivore carrying capacity. *Sci Adv* **9**, eadi4099 (2023).

17. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854 (2009).
18. Torres-Iglesias, L., Marín-Arroyo, A. B., Welker, F. & de la Rasilla, M. Using ZooMS to assess archaeozoological insights and unravel human subsistence behaviour at La Viña rock shelter (northern Iberia). *J. Archaeol. Sci.* **161**, 105904 (2024).
19. Wang, N. *et al.* Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China. *Proc. Biol. Sci.* **290**, 20231129 (2023).
20. Ruebens, K. *et al.* Neanderthal subsistence, taphonomy and chronology at Salzgitter-Lebenstedt (Germany): a multifaceted analysis of morphologically unidentifiable bone. *J. Quat. Sci.* **38**, 471–487 (2023).
21. Sinet-Mathiot, V. *et al.* Identifying the unidentified fauna enhances insights into hominin subsistence strategies during the Middle to Upper Palaeolithic transition. *Archaeol. Anthropol. Sci.* **15**, 139 (2023).
22. Pothier Bouchard, G., Riel-Salvatore, J., Negrino, F. & Buckley, M. Archaeozoological, taphonomic and ZooMS insights into The Protoaurignacian faunal record from Riparo Bombrini. *Quat. Int.* **551**, 243–263 (2020).
23. Arenas-Sorriqueta, E. *et al.* Subsistence strategies during the Gravettian in the rock shelter of La Viña (Asturias, N Spain). *Quaternary Science Advances* **12**, 100113 (2023).
24. Gelabert, P. *et al.* Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Curr. Biol.* (2021) doi:10.1016/j.cub.2021.06.023.
25. Vernot, B. *et al.* Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science* (2021) doi:10.1126/science.abf1667.

26. Zavala, E. I. *et al.* Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature* **595**, 399–403 (2021).
27. Özdoğan, K. T. *et al.* Archaeology meets environmental genomics: implementing sedaDNA in the study of the human past. *Archaeol. Anthropol. Sci.* **16**, 108 (2024).
28. Stuart, A. J. & Lister, A. M. New radiocarbon evidence on the extirpation of the spotted hyaena (Crocuta crocuta (Erxl.)) in northern Eurasia. *Quat. Sci. Rev.* **96**, 108–116 (2014).
29. Sauqué, V. *et al.* Pleistocene cave hyenas in the Iberian Peninsula: New insights from Los Aprendices cave (Moncayo, Zaragoza). *Palaeontol. Electronica* (2017).
30. Stuart, A. J. & Lister, A. M. Extinction chronology of the cave lion Panthera spelaea. *Quat. Sci. Rev.* **30**, 2329–2340 (2011).
31. Echegaray, J. G. & Freeman, L. G. *Excavando La Cueva de El Juyo: Un Santuario de Hace 14000 Años*. (Ministerio de Educación, Cultura y Deporte, 2015).
32. Castaños, P. La Macrofauna de la Cueva de la Paloma, Pleistoceno terminal de Asturias. in *La Cueva de la Paloma.* (ed. Hoyos, M.) (Ministerio de Cultura., 1980).
33. Jerjotoma-Ortín, V., Cuenca-Bescós, G. & Mazo, C. The Mark of the Beast: a bone assemblage assessment from the North of the Iberian Peninsula (MIS 3). *Journal of Archaeological Science: Reports* **54**, 104409 (2024).
34. Cueto, M., Camarós, E., Castaños, P., Ontañón, R. & Arias, P. Under the Skin of a Lion: Unique Evidence of Upper Paleolithic Exploitation and Use of Cave Lion (Panthera spelaea) from the Lower Gallery of La Garma (Spain). *PLoS One* **11**, e0163591 (2016).
35. Garate, D. New Insights into the Study of Paleolithic Rock Art: Dismantling the 'Basque Country Void'. *J. Anthropol. Res.* **74**, 168–200 (2018).
36. Sommer, R. S. & Benecke, N. Late Pleistocene and Holocene development of the felid fauna (Felidae) of Europe: a review. **269**, 7–19 (2006).

37. Altuna, J. Fauna de mamiferos de los yacimientos prehistoricos de Guipuzcoa. Con catalogo de los mamiferos cuaternarios del Cantabrico y del Pirineo Occidental. *San Sebastian, Sociedad de ciencias naturales Aranzadi* **24**, 1–465 (1972).
38. Altuna, J. Restos óseos del yacimiento prehistórico del Rascaño. in *El Paleolítico superior de la cueva del Rascaño (Santander)* 223–269 (dialnet.unirioja.es, 1981).
39. Straus, L. G. The Human Occupation of Southwestern Europe during the Last Glacial Maximum. *J. Anthropol. Res.* **71**, 465–492 (2015).
40. Straus, L. G. Pointes solutréennes et l'hypothèse de territorialisme. *Bulletin de la Société préhistorique française. Comptes rendus des séances mensuelles* **74**, 206–212 (1977).
41. Straus, L. G. El Solutrense: 40 Anos de Reflexiones por un Arqueologo Norteamericano. *ETFI* (2012) doi:10.5944/etfi.5.2012.4768.
42. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
43. Posth, C. *et al.* Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature* **615**, 117–126 (2023).
44. Carvalho, M. *et al.* Initial and lower magdalenian large mammal faunas and human subsistence at El mirón cave (Cantabria, Spain). *J. Paleolit. Archaeol.* **4**, (2021).
45. Straus, L. G., Morales, M. R. G. & Carretero, J. M. *The Red Lady of El Mirón Cave: Lower Magdalenian Human Burial in Cantabrian Spain*. (Elsevier, Limited, 2015).
46. Tejero, J.-M. *et al.* A minimally-invasive method for ancient DNA sampling of Prehistoric bone and antler tools and hunting weapons. *bioRxiv* 2023.04.02.535282 (2023) doi:10.1101/2023.04.02.535282.
47. González-Morales, M. & Straus, L. G. La ocupación gravetiense de la cueva de El Mirón (Ramales de la Victoria, Cantabria) y el contexto del arte paleolítico temprano de la

cuenca del Asón. in *Pensando el Gravetiense: nuevos datos para la Región Cantábrica, en su contexto peninsular y pirenaico* (eds. de las Heras Martín, C., Lasheras Corruchaga, J. A., Arrizabalaga Valbuena, A. & de la Rasilla Vives, M.) 289–299 (Ministerio de CulturaEditor: Monografías del Museo y Centro de Investigación de Altamira, 2012).

48. Marín-Arroyo, A. B. *et al.* The Middle to Upper Palaeolithic transition at El Mirón Cave (Cantabria, Spain). *Quat. Int.* **544**, 23–31 (2020).

49. Geiling, J. M. Human ecodynamics in the Late Upper Pleistocene of Northern Spain: an archeozoological study of ungulate remains from the Lower Magdalenian and other periods …. (University of Cantabria, 2020).

50. Sauqué, V. & Cuenca-Bescós, G. The Iberian Peninsula, the last European refugium of panthera pardus linnaeus 1758 during the Upper Pleistocene. *Quaternaire, vol. 24/1 | 2013 Volume 24 Numéro 1* 13–24 (2013).

51. Castaños, J. *et al.* Carbon and nitrogen stable isotopes of bone collagen of large herbivores from the Late Pleistocene Kiputz IX cave site (Gipuzkoa, north Iberian Peninsula) for palaeoenvironmental reconstruction. *Quat. Int.* **339-340**, 131–138 (2014).

52. Straus, L. G. Carnivores and cave sites in Cantabrian Spain. *J. Anthropol. Res.* (1982).

53. Gómez-Olivencia, A. *et al.* New evidence for the presence of reindeer (*Rangifer tarandus*) on the Iberian Peninsula in the Pleistocene: an archaeopalaeontological and chronological reassessment. *Boreas* **43**, 286–308 (2014).

54. Tejero, J.-M. Towards complexity in osseous raw material exploitation by the first anatomically modern humans in Europe: Aurignacian antler working. *Journal of Anthropological Archaeology* **36**, 72–92 (2014).

673 55. Lefebvre, A. *et al.* New insights into the use and circulation of reindeer antler in
674 northern Iberia during the Magdalenian (ca. 21-13 cal ka BP). *J. Archaeol. Sci.* **150**,
675 105708 (2023).
676 56. Bergström, A. *et al.* Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*
677 **607**, 313–320 (2022).
678 57. Bergström, A. *et al.* Origins and genetic legacy of prehistoric dogs. *Science* **370**, 557–
679 564 (2020).
680 58. Thalmann, O. *et al.* Complete mitochondrial genomes of ancient canids suggest a
681 European origin of domestic dogs. *Science* **342**, 871–874 (2013).
682 59. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. Ancient wolf genome reveals an
683 early divergence of domestic dog ancestors and admixture into high-latitude breeds.
684 *Curr. Biol.* **25**, 1515–1519 (2015).
685 60. Loog, L. *et al.* Ancient DNA suggests modern wolves trace their origin to a Late
686 Pleistocene expansion from Beringia. *Mol. Ecol.* **29**, 1596–1610 (2020).
687 61. Hervella, M. *et al.* The domestic dog that lived ~17,000 years ago in the Lower
688 Magdalenian of Erralla site (Basque Country): A radiometric and genetic analysis.
689 *Journal of Archaeological Science: Reports* **46**, 103706 (2022).
690 62. Boschin, F. *et al.* The first evidence for Late Pleistocene dogs in Italy. *Sci. Rep.* **10**, 13313
691 (2020).
692 63. Taron, U. H. *et al.* Ancient DNA from the Asiatic Wild Dog (Cuon alpinus) from Europe.
693 **12**, (2021).
694 64. Paijmans, J. L. A. *et al.* Historical biogeography of the leopard (Panthera pardus) and its
695 extinct Eurasian populations. *BMC Evol. Biol.* **18**, 1–12 (2018).
696 65. Fortes, G. G. *et al.* Ancient DNA reveals differences in behaviour and sociality between

brown bears and extinct cave bears. *Mol. Ecol.* **25**, 4907–4918 (2016).

66. Altuna, J. Los mamíferos del yacimiento prehistórico de Morín (Santander). *Cueva Morín. Excavaciones* (1966).

67. Westbury, M. V. *et al.* Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Sci Adv* **6**, eaay0456 (2020).

68. Bon, C. *et al.* Coprolites as a source of information on the genome and diet of the cave hyena. *Proc. Biol. Sci.* **279**, 2825–2830 (2012).

69. Tejero, J.-M. *et al.* Cervidae antlers exploited to manufacture Prehistoric tools and hunting implements as a reliable source of ancient DNA. *Heliyon* e31858 (2024).

70. Mackiewicz, P. *et al.* Phylogeny and evolution of the genus Cervus (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Sci. Rep.* **12**, 16381 (2022).

71. Rey-Iglesia, A., Grandal-d'Anglade, A., Campos, P. F. & Hansen, A. J. Mitochondrial DNA of pre-last glacial maximum red deer from NW Spain suggests a more complex phylogeographical history for the species. *Ecol. Evol.* **7**, 10690–10700 (2017).

72. Cieslak, M. *et al.* Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* **5**, e15311 (2010).

73. Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016).

74. Villalba-Mouco, V. *et al.* A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. *Nat Ecol Evol* **7**, 597–609 (2023).

75. Marín Arroyo, A. B., Fosse, P. & Vigne, J.-D. Probable evidences of bone accumulation by Pleistocene bearded vulture at the archaeological site of El Mirón Cave (Spain). *J. Archaeol. Sci.* **36**, 284–296 (2009).
76. Marín-Arroyo, A. B. & Geiling, J. M. Archeozoological study of the macromammal remains stratigraphically associated with the Magdalenian human burial in El Mirón Cave (Cantabria, Spain). *J. Archaeol. Sci.* **60**, 75–83 (2015).
77. Alfaro-Ibáñez, M. P., Cuenca-Bescós, G., Bover, P., González Morales, M. & Straus, L. G. Implications of population changes among the Arvicolinae (Rodentia, Mammalia) in El Mirón Cave (Cantabria, Spain) for the climate of the last c. 50,000 years. *Quat. Sci. Rev.* **315**, 108234 (2023).
78. Cuenca-Bescós, G., Marín-Arroyo, A. B. & Martínez, I. Relationship between Magdalenian subsistence and environmental change: The mammalian evidence from El Mirón (Spain). *Quaternary* (2012).
79. Straus, L. G. & González Morales, M. R. *El Mirón Cave, Cantabrian Spain: The Site and Its Holocene Archaeological Record*. (University of New Mexico Press, 2012).
80. Pedersen, M. W. *et al.* Environmental genomics of Late Pleistocene black bears and giant short-faced bears. *Curr. Biol.* (2021) doi:10.1016/j.cub.2021.04.027.
81. Marín-Arroyo, A. B. Exploitation of the Montane Zone of Cantabrian Spain during the Late Glacial: Faunal Evidence from El Mirón Cave. *J. Anthropol. Res.* **65**, 69–102 (2009).
82. Aldeias, V. & Stahlschmidt, M. C. Sediment DNA can revolutionize archaeology-if it is used the right way. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2317042121 (2024).
83. Straus, L. G., González Morales, M., Farrand, W. R. & Hubbard, W. J. Sedimentological and stratigraphic observations in El Mirón, a late Quaternary cave site in the Cantabrian Cordillera, Northern Spain. *Geoarchaeology* **16**, 603–630 (2001).

84. Farrand, W. Sedimentology of el Mirón cave. in *El Mirón Cave, Cantabrian Spain: The Site and Its Holocene Archaeological Record* (eds. Straus, L. G. & González Morales, M. R.) 60–94 (University of New Mexico Press, 2012).
85. Hopkins, R. J. A., Straus, L. G. & González Morales, M. R. Assessing the chronostratigraphy of el Mirón cave, Cantabrian Spain. *Radiocarbon* **63**, 821–852 (2021).
86. Straus, L. G. & González Morales, M. R. New Dates for the Solutrean and Magdalenian of Cantabrian Spain: El Miron and La Riera Caves. *Radiocarbon* **60**, 1013–1016 (2018).
87. Massilani, D. *et al.* Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
88. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).
89. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
90. Rohland, N. *et al.* Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
91. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
92. Myers, E. W. The fragment assembly string graph. *Bioinformatics* **1;21 Suppl 2**, 79–85 (2005).
93. Hannon, G. J. *FASTX-Toolkit*. (2010).
94. 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. (2023) doi:10.1111/2041-210x.14214.

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

96. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

97. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

5. Contributions to the Field

5.1. Chapter 2

“Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment”

The study of Satsurblia Cave was one of the pioneering research papers demonstrating the successful retrieval of genome-wide human and faunal data directly from archeological sediments. The retrieved amounts of sedaDNA were comparable to low-coverage sequencing data from skeletal remains.
We applied metagenomic sequencing to sediments dating to approx. 25 ka (Upper Paleolithic period) and retrieved nuclear and mitochondrial DNA from humans, wolves, and bison. This data allowed for phylogenetic and ancestry analyses of the species above. One of the main results was the identification of a previously unknown pre-LGM human lineage from the Caucasus, contributing their DNA to modern populations across West Eurasia, the Near East, and North Africa. Moreover, we identified previously undocumented and extinct lineages, including a Caucasian wolf population and a European bison lineage. This research illustrates the potential of sedaDNA to contribute to our understanding of ancient mammalian and hominin populations.

5.2. Chapter 3

“Maximizing Efficiency in sedimentary ancient DNA Analysis: A Novel Extract Pooling Approach”

Since wet lab processes and sequencing of numerous samples can be costly, we developed a novel, cost-efficient approach that reduces financial and time burdens by up to 70%. This high-throughput pooled testing approach does not require any additional equipment beyond what is typically available in a paleogenomics lab, making it accessible and easy to implement. We streamline the wet lab workflow by pooling up to five extracts early in the process and discarding pools devoid of DNA, reducing hands-on laboratory time to one-fifth. This method will be especially useful during the initial screening of archaeological sites with unknown DNA preservation.
Moreover, our results demonstrate that aDNA signals remain detectable up to a pooling level of 4:1 (four negative extracts to one positive), maintaining the integrity of aDNA signals in extract pools. Considering

the low success rate of sediment samples, this parallelization enables a reliable and quick screening without the need for immediate individual sample testing. Notably, our approach enhances the efficiency of aDNA recovery, increasing DNA yields by 1.36-fold on average. By adopting this technique, researchers can screen large numbers of sediment samples for aDNA with reduced costs, resources, and time.

5.3. Chapter 4

"A sedimentary ancient DNA perspective on human and carnivore persistence through the Last Glacial Maximum in El Mirón Cave, Spain"

In our study of El Mirón Cave, we corrected and refined the previous zooarcheological records by applying targeted sedaDNA techniques. We identified 70 faunal and three human mitochondrial genomes, a total of 28 different taxa, including 15 not previously identified through classical zooarchaeological methods based on morphological differences. This represents almost double the number of taxa that would have been detected using traditional methods alone. Among these taxa were *Mammuthus primigenius*, *Crocuta crocuta, Cuon alpinus, Rangifer tarandus, Lynx pardinus*, and *Vulpes vulpes*. The dhole (*Cuon alpinus*) was undetected or misassigned in the faunal record (Marín-Arroyo et al., 2020) but identified through sedaDNA analysis in all levels of the cave. Our analyses also extended the presence of *Panthera pardus* from one level (identified via physical remains) to all the screened levels, except for level 122.

Moreover, our results extended the extinction dates for spotted hyenas (*Crocuta crocuta spelaea*) in Iberia into the post-LGM Magdalenian period (21-12 kya), which is much later than previously believed (Stuart & Lister, 2014). This finding revises our understanding of Iberia's Late Pleistocene ecosystems and carnivore extinctions. At the same time, the improved data resolution through sedaDNA provides essential insights into the faunal diversity of the region and the ecological dynamics of the LGM and beyond.

Furthermore, we recovered three human mitochondrial DNA sequences from Solutrean levels, supporting mtDNA continuity in Iberia throughout the LGM. Hitherto, the only human skeletal record from El Mirón is a burial dated to the Lower Magdalenian, referred to as the Red Lady (L. G. Straus, Morales, & Carretero, 2015). Our finding adds new human mtDNA data to the sparse record of genomes from the Gravattian and Solutrean periods when Iberia served as a crucial refugium during the LGM.

Overall, this study shows how sedaDNA can uncover previously hidden aspects of past ecosystems, correct potential misinterpretations in the zooarchaeological register, and push extinction timelines.

6. Discussion

As the above results demonstrate, sedaDNA has become a valuable method in paleogenomics, enabling a new level of resolution of past biodiversity, environmental changes, and population dynamics of ancient hominins and fauna. The sections below elaborate on some of the advantages and limitations of sedaDNA research based on the results obtained from my thesis:

6.1. Benefits of sedaDNA

6.1.1. Filling Knowledge Gaps in Classical aDNA Research through sedaDNA

Archaeological sediments as a repository for aDNA offer a significant advantage: a continuous timescale of genomic information. When DNA is preserved at a site with a complete stratigraphic profile, samples can be densely collected in a grid-like pattern (**Figure 1**), covering all represented time periods (Vernot et al., 2021; Zavala et al., 2021). Adhering to this sampling approach can help close genomic knowledge gaps in periods lacking fossil finds and other organic sources of DNA. For instance, in our study of El Mirón Cave, we added human mitochondrial data from three samples to the sparse genome records from the Solutrean period in Iberia (Chapter 2).

It is commonly known that DNA is rarely preserved in warm environments, which applies to substantial parts of the world, including Africa, South America, and tropical areas (Allentoft et al., 2012; Kistler, Ware, Smith, Collins, & Allaby, 2017; Smith et al., 2003). To date, only a single study has successfully retrieved sedaDNA from warm, arid climates. (Viviane Slon et al., 2022) reported extended preservation of sedaDNA molecules in sediments from Sefunim Cave, Israel, contrasting the otherwise sparse record of >30 ka old DNA from Levantine fossils. This prolonged survival of DNA in sediment may partially be attributed to the adsorption of nucleic acids to protective minerals outside the bone, slowing down their decay (Cai et al., 2006). Hitherto, it has not been specifically tested whether increased temperatures lead to stronger DNA adsorption to calcite, one of the predominant minerals at Sefunim Cave (Viviane Slon et al., 2022). Notably, Sarhan et al. demonstrated that the quality of extracellular DNA preserved in calcite crusts is of better quality than the DNA retrieved from the skeletal remains covered by this crust, although present in smaller quantities (Sarhan et al., 2021). Nevertheless, it has to be mentioned that the study by Sarhan et al. was conducted on a relatively young specimen and is a single case study. The absence of similar

sedaDNA results from sites with poor DNA preservation in bones currently suggests little optimism for the field of sedaDNA to close the geographical gap in paleogenomics.

6.1.2. Elevated Taxa Diversity in sedaDNA

Sediments can reveal the presence of organisms that are difficult to identify through zooarchaeological records based on morphological characteristics. There are taxa that i) do not preserve in the fossil record (Lammers, Heintzman, & Alsos, 2021), ii) have been previously misidentified and wrongly assigned to other taxa in zooarchaeological studies (Chapter 4), and iii) were thought to be extinct due to the absence of zooarchaeological evidence (Graham et al., 2016; Haile et al., 2009). This was demonstrated by our study of El Mirón Cave (Chapter 4), by confirming the presence of 28 taxa, including humans, an additional 15 of which had not previously been recorded through classic archeozoological methods. Our study underscored that certain species or taxa can be underrepresented in the morphological record, while their genomic signatures can still be detected in sediment. For example, only seven skeletal fragments of *Canis lupus* were identified at studied levels at El Mirón, while all but two levels produced partial mtDNA genomes above 5x. The zooarcheological record of herbivores was largely congruent with our sedaDNA results. Interestingly, this was not the case for carnivores or rare species.

The potential of sedaDNA to identify a vast diversity of organisms can add a previously undetectable layer of information to zooarcheological, archeobotanical, and geoarcheological research, thereby enhancing our understanding of past ecosystems.

6.2. Limitations of sedaDNA

DNA yields from bones are generally higher than those obtained from sediments. While DNA from skeletal remains is limited by the scarcity of their recovery, higher quantities of DNA, particularly nuclear DNA, are generally obtained from organic remains. For instance, a study comparing skeletal samples with their covering calcite deposit found nuclear DNA in quantities two orders of magnitude higher in the skeletal remains (Sarhan et al., 2021). Despite methodological advancements in the field, mineralogical sources have not yet been able to completely replace organic remains as the primary source for DNA extraction in paleogenomics. The genomic resolution of our sedaDNA results in Chapter 2 (comparable to low-coverage genomes from skeletal material) underscores the limitations of metagenomic studies on loose sediment regarding genomic depth (Gelabert et al., 2021). However, the study provides valuable insights into the ancestry and population histories of Paleolithic wolves, bison, and humans, supported by similar findings from another study (Lazaridis et al., 2018). In contrast, our capture-based study on El Mirón Cave

achieved 70 mitochondrial genomes (including two human genomes) exceeding a 5x coverage. This includes high-coverage mtDNA genomes, for instance, 63.78x for *Cervus elaphus,* 58.51x for *Canis lupus*, and 44.05x for *Vulpes vulpes* (Chapter 4).

A common trait of complex sediment samples is that DNA of the same taxa commonly stems from multiple donors rather than from a single individual. This can be computationally tested by assessing sequence polymorphisms (i.e. mutations) of the mitochondrial data. The haploid status of individual mtDNA genomes precludes the possibility of sequence divergence. If there are multiple alleles detected, more than one individual must have contributed its DNA to the mix (Renaud, Slon, Duggan, & Kelso, 2015). The mitochondrial genomes (faunal and human) we retrieved from Satsurblia (Chapter 2) and El Mirón Cave (Chapter 4) displayed nucleotide polymorphisms that indicate a mosaic of individuals as a source of DNA. These substitutions complicated the phylogenetic analysis of the data and prevented the calculation of split times. (V. Slon et al., 2017) identified a single individual from three samples taken from Denisova Cave and at least two Neanderthals that contributed to the human mtDNA pool extracted from El Sidrón. Another study employed a combination of sequence polymorphisms and sex determination to infer that samples from Galería de las Estatuas and Denisova Cave possibly stem from single individuals, whereas most samples from Chagyrskaya Cave are probably composed of individuals of different sexes (Vernot et al., 2021).

6.2.1. Microbial Dominance in sedaDNA

While the ability of sedaDNA to identify a broad spectrum of organisms is one of its strengths, it simultaneously is one of its challenges: the dominance of microbiota in environmental samples can often impair the amplification and detection of underrepresented species. Non-targeted metagenomic studies will result in high taxonomic diversity in the sequencing results and require deeper sequencing (i.e., decoding more unique DNA fragments from the extract). To minimize this issue, targeted DNA enrichment approaches (also called "hybridization capture") have been developed (V. Slon et al., 2017; Tejero et al., 2024; Vernot et al., 2021). In the study of El Mirón, we successfully captured 10.6 species per sample using the custom mammalian panel by (Tejero et al., 2024).

In a study of 2023, our lab developed a pre-extraction treatment that significantly improved the retrieval of endogenous DNA in petrous bones, simultaneously filtering out fractions rich in contaminants via density separation (Fernandes et al., 2023). We showed that bone can be separated into fractions that display different success rates in DNA yields based on their level of mineralization (i.e., the concentration of osteocytes with high endogenous DNA levels). The implementation of this method showed a 5.38x

increase in endogenous DNA compared to the homogenized bone powder. This approach could be adapted to sediment samples to segregate bone particles from the environmental fractions and, hence, the high proportion of microbiota.

6.2.2. Inhibition in sedaDNA

Another issue that comes with analyzing environmental samples, especially loose sediments, is that co-extracted substances, such as humic, fulvic, and tannic acids, can negatively influence the potential to amplify target DNA (Hering & Morel, 1988; Matheson, Gurney, Esau, & Lehto, 2014; Sutlovic, Gamulin, Definis-Gojanovic, Gugic, & Andjelinovic, 2008; Uchii et al., 2019). These so-called inhibitors can interfere with DNA polymerase activity and impair downstream wet lab processes, introducing a bias toward false negative results (Opel, Chung, & McCord, 2010). Although protocols have been developed to minimize the effect of inhibitors (e.g., Braid, Daniels, & Kitts, 2003; Cullen & Hirsch, 1998; Frostegård et al., 1999; Holben, Jansson, Chelm, & Tiedje, 1988; Leff, Dana, McArthur, & Shimkets, 1995; Lovell & Piceno, 1994; Malik, Kain, Pettigrew, & Ogram, 1994; Torsvik, 1980; Walia, Khan, & Rosenthal, 1990; Zhou, Bruns, & Tiedje, 1996), inhibition hitherto remains a significant issue in sedaDNA pipelines. As discussed in Chapter 3, we found that inhibition may be significantly mitigated through sample extract dilution (Oberreiter et al., 2024). This theory is supported by other studies that use dilution or reduced extract input to circumvent the effects of inhibition (Glocke & Meyer, 2017; Imaizumi et al., 2005; McKee, Spear, & Pierson, 2015; Schneider, Enkerli, & Widmer, 2009; Shutler et al., 1999).

Recent advancements have significantly enhanced the potential of sedaDNA analysis. A breakthrough was the discovery that resin-impregnated sediment blocks are highly effective repositories for aDNA retrieval (Massilani et al., 2022). While targeted microsampling yielded significantly more DNA and reduced the number of identified taxa, it also reduced inhibition. The authors suggest that this reduction was likely due to the lower amount of sediment in the sample, which was partially composed of resin. This composition led to fewer inhibitory substances in the extract, further supporting our hypothesis that a reduced sample input or dilution can improve the efficiency of wet lab processing (Oberreiter et al., 2024). It also demonstrates that loose sediment (i.e., a mix of organic and inorganic components) contains more inhibitors than organic inclusions (i.e., bones or coprolites) alone. We acknowledge that protocols to specifically identify, quantify, and remove inhibitors must be installed and that experimental testing is required in the future.

(Massilani et al., 2022) also found that microfeatures are more efficient targets for sedaDNA retrieval and observed that taxa diversity varies significantly within millimeters between sampling locations. This

suggests that DNA concentrates within these inclusions. However, the exact source of DNA in sediments remains to be determined. Their results do not exclude the possibility that DNA is additionally adsorbed to mineral particles, as the matrix was also found to contain aDNA while showing elevated inhibition rates. In our study of Satsurblia Cave (Chapter 2, (Gelabert et al., 2021)) we also encountered significant heterogeneity within the sediment, as only one sample (SAT29) from layer BIII of the profile produced authentic ancient DNA. This finding supports the notion that ancient DNA preservation occurs rather locally than evenly distributed and that inhibition may have played a pivotal role in the failing of the other sediment samples we analyzed from Satsurblia Cave. However, the micro-environmental conditions that led to the very local DNA preservation in SAT29 were not investigated in this framework. In contrast to Satsurblia Cave, El Mirón Cave exhibited exceptional aDNA preservation across a broad stratigraphy, spanning 25 ka from the Initial Magdalenian to the Mousterian period. The cave has a stable microclimate and an undisturbed stratigraphy, which likely played a vital role in preserving DNA in all the studied levels (Hopkins, Straus, & González Morales, 2021; Marín Arroyo, Fosse, & Vigne, 2009; Lawrence G. Straus & González Morales, 2018).

6.2.3. Bioinformatic Challenges in sedaDNA: Fragmentation, Misalignments, and Filtering

In addition to the discussed wet lab challenges, another hurdle in sedaDNA analysis is the bioinformatics pipeline. Ancient DNA is commonly highly fragmented due to degradation over time (Willerslev & Cooper, 2005). This results in very short DNA fragments that are difficult to assign to the corresponding locations in reference genomes. Furthermore, the extracted DNA from environmental sources is a diverse mix of organisms, often with large quantities of microbial DNA. The bioinformatic analysis ancient genomes involves several strict filtering steps, including filtering for authenticity to reduce modern contaminants and assigning fragments to reference genomes based on sequence similarity. This filtering process can result in losing a high proportion of the sequences, limiting the resolution of population or phylogenetic studies. Misalignments are one of the main bioinformatic challenges in sedaDNA studies. The short, degraded fragments typical of ancient DNA (Willerslev & Cooper, 2005) are often difficult to map accurately to reference genomes, and the presence of closely related mammalian species enhances this problem. (Vernot et al., 2021), in particular, reduced misalignments by designing a 1.6 million probe set to target genome regions with high divergence between hominins and other mammals, ensuring that hominin sequences could be more reliably distinguished.

Hybridization capture methods are naturally biased towards the genomic regions targeted by the probes, which means that regions of the genome outside of these targets are underrepresented. The

mitochondrial genomes we recovered from El Mirón Cave showed missing sites at regions conserved in mammals, which we attributed to the strict nature of the taxa classification (Chapter 4).

While capture efficiently reduces microbial signals, the procedure is not capable of eliminating modern contaminants. To ensure the authenticity of aDNA, specific tools such as mapdamage (Jónsson, Ginolhac, Schubert, Johnson, & Orlando, 2013) and schmutzi (Renaud et al., 2015) are used to determine the amount of DNA that does not show C to T damage patterns at their ends. This strict computational filtering is necessary to distinguish between endogenous hominin DNA and contaminants while further reducing the reads that remain for high-resolution analyses. To retrieve authentic results free from contaminant distortion, strict computational filtering was applied in (Gelabert et al., 2021, Chapter 2). However, the untargeted metagenomic approach we used resulted in low coverage data, particularly for humans and bison. Based on current knowledge, employing a targeted nuclear capture approach, such as the one used by (Vernot et al., 2021) for humans and the capture panel used in Chapter 3 and 4 by (Tejero et al., 2024), would have likely increased genomic coverage and benefited the robustness of population genetics analyses.

6.2.4. Mitigation, Leaching, and post-depositional Movement of sedaDNA

It has been demonstrated that DNA can be transported vertically with water (Andersen et al., 2012; Haile et al., 2007). Similar to (Haile et al., 2007), our study of El Mirón Cave also found modern sheep DNA in six Solutrean samples. This contamination was likely introduced because the cave was used as a sheep staple in modern times. Congruent with Haile et al.'s findings, which detected leaching as deep as 70 cm in an area of high urine deposition at the surface level, we also did not find evidence of DNA leaching below Solutrean layers (Chapter 4).

We identified the modern sheep DNA computationally through low deamination values and elevated read lengths. However, this did not influence the results of the other identified taxa, which displayed authentic damage and read length patterns.

6.3. Outlook and Future Directions

As paleogenomics research continues to evolve, mineralogical sources as archives for aDNA retrieval will likely play an increasingly important role in archaeological research. While recent advances, such as resin-impregnated sediment blocks and targeted sampling methods, have improved DNA recovery and reduced inhibition, there are still significant challenges to overcome, particularly regarding wet lab and computational approaches.

A very prominent issue that impairs the analysis of a range of sediment contexts, especially lake sediments and other sediments with high proportions of organic material, is inhibition. The identification and removal of inhibitory substances very early on in the pipeline (preferably even before extraction) will require extensive experimental efforts but must be addressed since it is the source of many false negative results in sedaDNA and limits our understanding of DNA preservation at numerous sites. During my PhD, I analyzed hundreds of sediments and other mineralogical samples from over 40 sites and different contexts. Notably, the success rate of retrieving authentic aDNA was less than 5%. I partially attribute this low yield to false negatives resulting from inhibition (as well as to overall bad DNA preservation at certain sites). Our samples undergo several quantitative assessment steps in the wet lab pipeline to determine the presence of DNA, including quantitative real-time PCR and subsequent PCR analysis. Since we do not perform any form of composition analysis prior to the processing of the sediment samples, the amount and impact of inhibitors such as humic acids remains undetermined. Humic acids do not only negatively affect PCR reactions by inhibiting polymerase activity but also interfere with fluorescence in real-time PCR, causing detection inhibition and skewing results toward false negatives (Sidstedt et al., 2015). To account for the effects of inhibition, I suggest assessing the inhibition potential of single extracts by comparing the amplification efficiency of a controlled template oligonucleotide with added extracts, as demonstrated in Chapter 4. Based on these results, the extract input into downstream laboratory steps can be reduced accordingly.

Future studies will also need to further investigate the primary source of genomic material in sediments and the details and extent of mineral-bound DNA. If a significant amount of extracellular DNA indeed is adsorbed to minerals, as indicated by previous research under laboratory conditions (Cai et al., 2006; Ogram et al., 1988), specific desorption protocols (to release DNA from minerals) will have to be developed to minimize the loss of informative fragments to extraction protocols.

Since (Massilani et al., 2022) showed that microsampling of polymerized sediment blocks enhances endogenous DNA yields significantly, I propose to produce these blocks at every excavation in addition to sampling loose sediment. From my experience in sampling loose sediment from stratigraphies, adhering to the boundaries of narrow sediment strata and preventing sediment movement during the process is a constant challenge. Even with the assistance of archaeologists and experts on-site, it remains difficult to sample without introducing material from adjacent layers. Polymerized blocks would help preserve the integrity of the stratigraphy and minimize contamination risks from neighboring levels during sampling.

For (archived) loose sediment, I propose conducting density separation experiments. The promising results from applying a density separation step prior to extraction in petrous bone samples suggest this method could be useful and adapted for environmental samples (Fernandes et al., 2023). Implementing this approach could enhance the overall yield of authentic aDNA by filtering out inorganic components, reducing microbial contamination, and retaining fractions of organic inclusions (e.g., bones and coprolites) for subsequent extraction.

Besides wet lab advances, the computational pipelines of environmental aDNA samples will require refining. As already outlined by (Vernot et al., 2021), misalignments are a fundamental challenge in the analysis of sedaDNA. Vernot et al. substantially reduced misalignments focussing on areas where the human genome displays high diversity compared to mammalian genomes. They developed a set of informative 1.6 million single nucleotide polymorphisms (SNP) to enrich for hominin sequences around these areas via hybridization capture. Especially in regards to hominin evolution and population dynamics, I encourage further research efforts at refining the genomic target regions of capture panels to minimize mammalian misalignments, given the complex genomic mosaics of sediment samples.

7. Conclusions

In this thesis, I have demonstrated the importance and potential of sedimentary archives for aDNA recovery to advance our knowledge of human evolution, mammalian diversity, ancestry and phylogenetic relationships.

In Chapter 2, I presented the genome-wide results of a single Upper Paleolithic sediment sample from Satsurblia Cave in the Caucasus. The sedaDNA revealed a previously undocumented pre-LGM human lineage and extinct wolf and bison lineages, offering new insights into their population histories.

Chapter 3 introduced a novel, high-throughput methodological advancement that is easy to implement in the wet lab pipeline and takes a significant amount of financial burden off of archeological sedaDNA surveys while increasing DNA yields.

Lastly, in Chapter 4, I presented crucial mitochondrial DNA results from El Mirón Cave, Spain. This study expanded the archeozoological record of the site by identifying a total of 72 mtDNA genomes (>5x coverage), including 15 undocumented taxa through sedaDNA. We added three Solutrean human mtDNA genomes to the spare genome record of that time. Further, we dated back the extinction of spotted hyenas (*Crocuta crocuta spelaea*) in Iberia to the Magdalenian and expanded the known presence of crucial carnivores such as the leopard *(Panthera pardus)* and the dhole (*Cuon alpinus*) at the site. The

computational classification approach used in this study entails several filtering steps to enable robust fragment assignments, accurately identifying DNA sequences down to the species level.

As science continues to evolve, especially with advances in computational tools and high-throughput sequencing technologies, the field of sedaDNA will continue to push the boundaries of paleogenomics. The techniques presented in this thesis lay the groundwork for further experiments to improve DNA recovery from mineralogical archives. The contributions of this research to the field, combined with other recent advances will help to further refine our understanding of the past and open up new possibilities for uncovering genomic information from regions and time periods previously thought inaccessible. Overall, the future of sedaDNA research holds great promise for improving our understanding of the past.

8. References

Allen, M. C., Kwait, R., Vastano, A., Kisurin, A., Zoccolo, I., Jaffe, B. D., Angle, J. C., et al. (2023). Sampling environmental DNA from trees and soil to detect cryptic arboreal mammals. *Scientific reports*, *13*(1), 180. nature.com.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., et al. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological sciences / The Royal Society*, *279*(1748), 4724–4733.

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature*, *522*(7555), 167–172.

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H. E., Kjaer, K. H., Orlando, L., et al. (2012). Meta-barcoding of "dirt"DNA from soil reflects vertebrate biodiversity. Molecular ecology, *21*, 1966–1979.

Antich, A., Palacín, C., Cebrian, E., Golo, R., Wangensteen, O. S., & Turon, X. (2021). Marine biomonitoring with eDNA: Can metabarcoding of water samples cut it as a tool for surveying benthic communities? *Molecular ecology*, *30*(13), 3175–3188. Wiley.

Araujo, A. G. M. (2013). Bioturbation and the upward movement of sediment particles and archaeological materials: comments on Bueno et al. *Journal of archaeological science*, *40*(4), 2124–2127.

Ariza, M., Fouks, B., Mauvisseau, Q., Halvorsen, R., Alsos, I. G., & de Boer, H. J. (2023). Plant biodiversity assessment through soil eDNA reflects temporal and local diversity. *Methods in ecology and evolution / British Ecological Society*, *14*(2), 415–430. Wiley.

Bergström, A., Stanton, D. W. G., Taron, U. H., Frantz, L., Sinding, M.-H. S., Ersmark, E., Pfrengle, S., et al. (2022). Grey wolf genomic history reveals a dual ancestry of dogs. *Nature*, *607*(7918), 313–320.

Blattner, L., Ebner, J. N., Zopfi, J., & von Fumetti, S. (2021). Targeted non-invasive bioindicator species detection in eDNA water samples to assess and monitor the integrity of vulnerable alpine freshwater environments. *Ecological indicators*, *129*, 107916. Elsevier.

Braid, M. D., Daniels, L. M., & Kitts, C. L. (2003). Removal of PCR inhibitors from soil DNA by chemical flocculation. *Journal of microbiological methods*, *52*(3), 389–393. Elsevier.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14616–14621.

Brundin, M., Figdor, D., Sundqvist, G., & Sjögren, U. (2013). DNA binding to hydroxyapatite: a potential mechanism for preservation of microbial DNA. *Journal of endodontia*, *39*(2), 211–216.

Cai, P., Huang, Q.-Y., & Zhang, X.-W. (2006). Interactions of DNA with clay minerals and soil colloidal particles and protection against degradation by DNase. *Environmental science & technology*, *40*(9), 2971–2976.

Clarke, C. L., Alsos, I. G., Edwards, M. E., Paus, A., Gielly, L., Haflidason, H., Mangerud, J., et al. (2020). A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. *Quaternary science reviews*, *247*, 106564.

Clarke, C. L., Heintzman, P. D., Lammers, Y., Monteath, A. J., Bigelow, N. H., Reuther, J. D., Potter, B. A., et al. (2024). Steppe-tundra composition and deglacial floristic turnover in interior Alaska revealed by sedimentary ancient DNA (sedaDNA). *Quaternary science reviews*, *334*, 108672.

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., et al. (2009). The Last Glacial Maximum. *Science*, *325*(5941), 710–714.

Crecchio, C., & Stotzky, G. (1998). Binding of DNA on humic acids: Effect on transformation of Bacillus subtilis and resistance to DNase. *Soil biology & biochemistry*, *30*(8), 1061–1067.

Cullen, D. W., & Hirsch, P. R. (1998). Simple and rapid method fordirect extraction of microbial DNA fromsoil for PCR. *Soil biology & biochemistry*, *30*(8-9), 983–993. Elsevier BV.

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor perspectives in biology*, *5*(7). Retrieved from http://dx.doi.org/10.1101/cshperspect.a012567

Dalén, L., Nyström, V., Valdiosera, C., Germonpré, M., Sablin, M., Turner, E., Angerbjörn, A., et al. (2007). Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(16), 6726–6729.

Essel, E., Zavala, E. I., Schulz-Kornas, E., Kozlikin, M. B., Fewlass, H., Vernot, B., Shunkov, M. V., et al. (2023). Ancient human DNA recovered from a Palaeolithic pendant. *Nature*, *618*(7964), 328–332.

Fernandes, D. M., Sirak, K. A., Cheronet, O., Novak, M., Brück, F., Zelger, E., Llanos-Lizcano, A., et al. (2023). Density separation of petrous bone powders for optimized ancient DNA yields. *Genome research*, *33*(4), 622–631. Cold Spring Harbor Lab.

Freeman, C. L., Dieudonné, L., Collins, M. J., & Sand, K. K. (2020). Survival of environmental DNA in natural environments: Surface charge and topography of minerals as driver for DNA storage. *BioRxiv*.

Frostegård, A., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., et al. (1999). Quantification of bias related to the extraction of DNA directly from soils. *Applied and environmental microbiology*, *65*(12), 5409–5420. American Society for Microbiology.

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, *524*(7564), 216–219.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., Johnson, P. L. F., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, *514*(7523), 445–449.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., et al. (2016). The genetic history of Ice Age Europe. *Nature*, *534*(7606), 200–205.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, *5*, 5257.

Gelabert, P., Sawyer, S., Bergström, A., Margaryan, A., Collin, T. C., Meshveliani, T., Belfer-Cohen, A., et al. (2021). Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Current biology: CB*, *31*(16), 3564–3574.e9.

Gilbert, M. T. P., Tomsho, L. P., Rendulic, S., Packard, M., Drautz, D. I., Sher, A., Tikhonov, A., et al. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science*, *317*(5846), 1927–1930.

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* , *27*(15), 2153–2155. Oxford University Press (OUP).

Glocke, I., & Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome research*, *27*(7), 1230–1237.

Graham, R. W., Belmecheri, S., Choy, K., Culleton, B. J., Davies, L. J., Froese, D., Heintzman, P. D., et al. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(33), 9310–9314.

Grave, P., & Kealhofer, L. (1999). Assessing Bioturbation in Archaeological Sediments using Soil Morphology and Phytolith Analysis. *Journal of archaeological science*, *26*(10), 1239–1248.

Greaves, M. P., & Wilson, M. J. (1969). The adsorption of nucleic acids by montmorillonite. *Soil biology & biochemistry*, *1*(4), 317–323.

Green, R. E., Briggs, A. W., Krause, J., Prüfer, K., Burbano, H. A., Siebauer, M., Lachmann, M., et al. (2009). The Neandertal genome and ancient DNA authenticity. *The EMBO journal*, *28*(17), 2494–2502. Wiley.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., et al. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211.

Hagelberg, E., Sykes, B., & Hedges, R. (1989). Ancient bone DNA amplified. *Nature*, *342*(6249), 485.

Haile, J., Froese, D. G., Macphee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. V., Rasmussen, M., et al. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22352–22357.

Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M. T. P., Nielsen, R., Munch, K., et al. (2007). Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular biology and evolution*, *24*(4), 982–989. Oxford University Press (OUP).

Hajdinjak, M., Mafessoni, F., Skov, L., Vernot, B., Hübner, A., Fu, Q., Essel, E., et al. (2021). Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature*, *592*(7853), 253–257.

Harney, É., Cheronet, O., Fernandes, D. M., Sirak, K., Mah, M., Bernardos, R., Adamski, N., et al. (2021). A minimally destructive protocol for DNA extraction from ancient teeth. *Genome research*, *31*(3), 472–483. genome.cshlp.org.

Heintzman, P. D., Froese, D., Ives, J. W., Soares, A. E. R., Zazula, G. D., Letts, B., Andrews, T. D., et al. (2016). Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8057–8063.

Hering, J. G., & Morel, F. M. (1988). Humic acid complexation of calcium and copper. *Environmental science & technology*, *22*(10), 1234–1237. ACS Publications.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*(5991), 282–284.

Hofreiter, M., Betancourt, J. L., Sbriller, A. P., Markgraf, V., & McDonald, H. G. (2003). Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Curá, Neuquén Province, southwest Argentina. *Quaternary Research*, *59*(3), 364–378. Cambridge University Press (CUP).

Hofreiter, M., Mead, J. I., Martin, P., & Poinar, H. N. (2003). Molecular caving. *Current biology: CB*, *13*(18), R693–5.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Pääbo, S. (2001). Ancient DNA. *Nature reviews. Genetics*, *2*(5), 353–359. Nature Publishing Group.

Holben, W. E., Jansson, J. K., Chelm, B. K., & Tiedje, J. M. (1988). DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. *Applied and environmental microbiology*, *54*(3), 703–711.

Hopkins, R. J. A., Straus, L. G., & González Morales, M. R. (2021). ASSESSING THE CHRONOSTRATIGRAPHY OF EL MIRÓN CAVE, CANTABRIAN SPAIN. *Radiocarbon*, *63*(3), 821–852. Cambridge University Press.

Höss, M., Dilling, A., Currant, A., & Pääbo, S. (1996). Molecular phylogeny of the extinct ground sloth Mylodon darwinii. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(1), 181–185.

Imaizumi, K., Noguchi, K., Shiraishi, T., Sekiguchi, K., Senju, H., Fujii, K., Yoshida, K., et al. (2005). DNA typing of bone specimens—the potential use of the profiler test as a tool for bone identification. *Legal medicine*, *7*(1), 31–41.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945.

Ives, J. W., Froese, D., Supernant, K., & Yanicki, G. M. (2014). Paleoamerican Odyssey 149–169. *Paleoamerican Odyssey*, 149–169. Texas A & M Univ. Press.

Johnson, M. D., Fokar, M., Cox, R. D., & Barnes, M. A. (2021). Airborne environmental DNA metabarcoding detects more diversity, with less sampling effort, than a traditional plant community survey. *BMC ecology and evolution*, *21*(1), 218. bmcecolevol.biomedcentral.com.

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R. L., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications*, *6*, 8912. nature.com.

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*. Retrieved from http://dx.doi.org/10.1093/bioinformatics/btt193

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., et al. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular ecology*, *21*(8), 1989–2003.

Khanna, M., Yoder, M., Calamai, L., & Stotzky, G. (1998). X-ray diffractometry and electron microscopy of DNA from Bacillus subtilis bound on clay minerals. *Sciences of Soils*, *3*(1), 1–10.

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic acids research*, *45*(11), 6310–6320.

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., et al. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, *612*(7939), 283–291.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, *464*(7290), 894–897.

Lammers, Y., Heintzman, P. D., & Alsos, I. G. (2021). Environmental palaeogenomic reconstruction of an Ice Age algal population. *Communications biology*, *4*(1), 220.

Lazaridis, I., Belfer-Cohen, A., Mallick, S., Patterson, N., Cheronet, O., Rohland, N., Bar-Oz, G., et al. (2018, September 21). *Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry*. *bioRxiv*. Retrieved September 6, 2024, from https://www.biorxiv.org/content/10.1101/423079

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature*, *536*(7617), 419–424.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413.

Leff, L. G., Dana, J. R., McArthur, J. V., & Shimkets, L. J. (1995). Comparison of methods of DNA extraction from stream sediments. *Applied and environmental microbiology*, *61*(3), 1141–1143.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*(6422), 709–715.

Liu, B., & Liu, J. (2014). DNA adsorption by magnetic iron oxide nanoparticles and its application for arsenate detection. *Chemical communications* , *50*(62), 8568–8570.

Liu, S., Kruse, S., Scherler, D., Ree, R. H., Zimmermann, H. H., Stoof-Leichsenring, K. R., Epp, L. S., et al. (2021). Sedimentary ancient DNA reveals a threat of warming-induced alpine habitat loss to Tibetan Plateau plant diversity. *Nature communications*, *12*(1), 2995.

Loreille, O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P., & Hänni, C. (2001). Ancient DNA analysis reveals divergence of the cave bear, Ursus spelaeus, and brown bear, Ursus arctos, lineages. *Current biology: CB*, *11*(3), 200–203.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., et al. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

Lorenz, M. G., & Wackernagel, W. (1987). Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Applied and environmental microbiology*, *53*(12), 2948–2952.

Lovell, C. R., & Piceno, Y. (1994). Purification of DNA from estuarine sediments. *Journal of microbiological methods*, *20*(3), 161–174.

Lydolph, M. C., Jacobsen, J., Arctander, P., Gilbert, M. T. P., Gilichinsky, D. A., Hansen, A. J., Willerslev, E., et al. (2005). Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and environmental microbiology*, *71*(2), 1012–1017. American Society for Microbiology.

Lynggaard, C., Bertelsen, M. F., Jensen, C. V., Johnson, M. S., Frøslev, T. G., Olsen, M. T., & Bohmann, K. (2022). Airborne environmental DNA for terrestrial vertebrate community monitoring. *Current biology: CB*, *32*(3), 701–707.e5.

Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K. A., Viola, B., Markin, S. V., et al. (2020). A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(26), 15132–15136.

Malik, M., Kain, J., Pettigrew, C., & Ogram, A. (1994). Purification and molecular analysis of microbian DNA from compost. *Journal of microbiological methods*, *20*(3), 183–196.

Marín Arroyo, A. B., Fosse, P., & Vigne, J.-D. (2009). Probable evidences of bone accumulation by Pleistocene bearded vulture at the archaeological site of El Mirón Cave (Spain). *Journal of archaeological science*, *36*(2), 284–296.

Marín-Arroyo, A. B., Geiling, J.-M., Jones, J. R., González Morales, M. R., Straus, L. G., & Richards, M. P. (2020). The middle to upper Palaeolithic transition at El mirón cave (Cantabria, Spain). *Quaternary international: the journal of the International Union for Quaternary Research*, *544*, 23–31. Elsevier BV.

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E. A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., et al. (2016). Past climate changes, population dynamics and the origin of Bison in Europe. *BMC biology*, *14*(1), 93.

Massilani, D., Morley, M. W., Mentzer, S. M., Aldeias, V., Vernot, B., Miller, C., Stahlschmidt, M., et al. (2022). Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1). National Acad Sciences. Retrieved from http://dx.doi.org/10.1073/pnas.2113666118

Matheson, C. D., Gurney, C., Esau, N., & Lehto, R. (2014). Assessing PCR Inhibition from Humic Substances. *The open enzyme inhibition journal*, *3*(1), 38–45. Bentham Science Publishers Ltd.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., et al. (2018). The genomic history of southeastern Europe. *Nature*, *555*(7695), 197–203.

McKee, A. M., Spear, S. F., & Pierson, T. W. (2015). 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. *Biological conservation*, *183*, 70–76. Elsevier.

Meiri, M., Lister, A. M., Collins, M. J., Tuross, N., Goebel, T., Blockley, S., Zazula, G. D., et al. (2014). Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proceedings. Biological sciences / The Royal Society*, *281*(1776), 20132167. The Royal Society.

Melzak, K. A., Sherwood, C. S., Turner, R. F. B., & Haynes, C. A. (1996). Driving Forces for DNA Adsorption to Silica in Perchlorate Solutions. *Journal of colloid and interface science*, *181*(2), 635–644.

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., et al. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, *531*(7595), 504–507.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, *338*(6104), 222–226.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, *456*(7220), 387–390.

Nguyen, T. H., & Elimelech, M. (2007). Plasmid DNA adsorption on silica: kinetics and conformational changes in monovalent and divalent salts. *Biomacromolecules*, *8*(1), 24–32.

Oberreiter, V., Gelabert, P., Brück, F., Franz, S., Zelger, E., Szedlacsek, S., Cheronet, O., et al. (2024). Maximizing efficiency in sedimentary ancient DNA analysis: a novel extract pooling approach. *Scientific reports*, *14*(1), 19388.

Ogram, A., Sayler, G. S., Gustin, D., & Lewis, R. J. (1988). DNA adsorption to soils and sediments. *Environmental science & technology*, *22*(8), 982–984.

Opel, K. L., Chung, D., & McCord, B. R. (2010). A study of PCR inhibition mechanisms using real time PCR. *Journal of forensic sciences*, *55*(1), 25–33. Wiley.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Özdoğan, K. T., Gelabert, P., Hammers, N., Altınışık, N. E., de Groot, A., & Plets, G. (2024). Archaeology meets environmental genomics: implementing sedaDNA in the study of the human past. *Archaeological and anthropological sciences*, *16*(7), 108.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, *314*(6012), 644–645.

Parducci, L., Alsos, I. G., Unneberg, P., Pedersen, M. W., Han, L., Lammers, Y., Salonen, J. S., et al. (2019). Shotgun Environmental DNA, Pollen, and Macrofossil Analysis of Lateglacial Lake Sediments From Southern Sweden. *Frontiers in Ecology and Evolution*, *7*. Retrieved from https://www.frontiersin.org/articles/10.3389/fevo.2019.00189

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. *The New phytologist*, *214*(3), 924–942. Wiley Online Library.

Parducci, L., Suyama, Y., Lascoux, M., & Bennett, K. D. (2005). Ancient DNA from pollen: a genetic record of population history in Scots pine. *Molecular ecology*, *14*(9), 2873–2882.

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., et al. (2013). A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary science reviews*, *75*, 161–168.

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., et al. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., et al. (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PloS one*, *10*(6), e0129102.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D. E., Buigues, B., Tikhonov, A., et al. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*, *311*(5759), 392–394.

Posth, C., Yu, H., Ghalichi, A., Rougier, H., Crevecoeur, I., Huang, Y., Ringbauer, H., et al. (2023). Palaeogenomics of upper Palaeolithic to neolithic European hunter-gatherers. *Nature*, *615*(7950), 117–126. Springer Science and Business Media LLC.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060.

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome biology*, *16*(1), 224. Springer Science and Business Media LLC.

Sanchez, G., Holliday, V. T., Gaines, E. P., Arroyo-Cabrales, J., Martínez-Tagüeña, N., Kowler, A., Lange, T., et al. (2014). Human (Clovis)-gomphothere (Cuvieronius sp.) association ~ 13,390 calibrated yBP in Sonora, Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(30), 10972–10977. Proceedings of the National Academy of Sciences.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. Proceedings of the National Academy of Sciences.

Sarhan, M. S., Lehmkuhl, A., Straub, R., Tett, A., Wieland, G., Francken, M., Zink, A., et al. (2021). Ancient DNA diffuses from human bones to cave stones. *iScience*, *24*(12), 103397.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one*, *7*(3), e34131. Public Library of Science (PLoS).

Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M. V., Derevianko, A. P., et al. (2015). Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(51), 15696–15700.

Schneider, S., Enkerli, J., & Widmer, F. (2009). A generally applicable assay for the quantification of inhibitory effects on PCR. *Journal of microbiological methods*, *78*(3), 351–353. Elsevier.

Seguin-Orlando, A., Korneliussen, T. S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., Albrechtsen, A., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science*, *346*(6213), 1113–1118.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., et al. (2004). Rise and fall of the Beringian steppe bison. *Science*, *306*(5701), 1561–1565.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, *26*(10), 1135–1145. Springer Science and Business Media LLC.

Shutler, G. G., Gagnon, P., Verret, G., Kalyn, H., Korkosh, S., Johnston, E., & Halverson, J. (1999). *Removal of a PCR Inhibitor and Resolution of DNA STR Types in Mixed Human-Canine Stains from a Five Year Old Case*. ASTM International.

Sidstedt, M., Jansson, L., Nilsson, E., Noppa, L., Forsman, M., Rådström, P., & Hedman, J. (2015). Humic substances cause fluorescence inhibition in real-time polymerase chain reaction. *Analytical biochemistry*, *487*, 30–37. Elsevier BV.

Slon, V., Clark, J. L., Friesem, D. E., Orbach, M., Porat, N., Meyer, M., Kandel, A. W., et al. (2022). Extended longevity of DNA preservation in Levantine Paleolithic sediments, Sefunim Cave, Israel. *Scientific reports*, *12*(1), 14528.

Slon, V., Hopfe, C., Weiß, C. L., & Mafessoni, F. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*. science.org. Retrieved from https://www.science.org/doi/abs/10.1126/science.aam9695

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., et al. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, *561*(7721), 113–116.

Smith, C. I., Chamberlain, A. T., Riley, M. S., Cooper, A., Stringer, C. B., & Collins, M. J. (2001). Not just old but old and cold? *Nature*, *410*(6830), 771–772. Nature Publishing Group.

Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C., & Collins, M. J. (2003). The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of human evolution*, *45*(3), 203–217.

Sodnikar, K., Parker, K. M., Stump, S. R., ThomasArrigo, L. K., & Sander, M. (2021). Adsorption of double-stranded ribonucleic acids (dsRNA) to iron (oxyhydr-)oxide surfaces: comparative analysis of model dsRNA molecules and deoxyribonucleic acids (DNA). *Environmental science. Processes & impacts*, *23*(4), 605–620.

Søe, M. J., Nejsum, P., Seersholm, F. V., Fredensborg, B. L., Habraken, R., Haase, K., Hald, M. M., et al. (2018). Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. *PloS one*, *13*(4), e0195481. Public Library of Science.

Stahlschmidt, M. C., Collin, T. C., Fernandes, D. M., Bar-Oz, G., Belfer-Cohen, A., Gao, Z., Jakeli, N., et al. (2019). Ancient Mammalian and Plant DNA from Late Quaternary Stalagmite Layers at Solkota Cave, Georgia. *Scientific reports*, *9*(1), 6628. nature.com.

Straus, L. G., & González Morales, M. R. (2018). New Dates for the Solutrean and Magdalenian of Cantabrian Spain: El Miron and La Riera Caves. *Radiocarbon*, *60*(3), 1013–1016. Cambridge University Press.

Straus, L. G., Morales, M. R. G., & Carretero, J. M. (2015). *The Red Lady of El Mirón Cave: Lower Magdalenian Human Burial in Cantabrian Spain*. Elsevier.

Stuart, A. J., & Lister, A. M. (2014). New radiocarbon evidence on the extirpation of the spotted hyaena (Crocuta crocuta (Erxl.)) in northern Eurasia. *Quaternary science reviews*, *96*, 108–116.

Sutlovic, D., Gamulin, S., Definis-Gojanovic, M., Gugic, D., & Andjelinovic, S. (2008). Interaction of humic acids with human DNA: proposed mechanisms and kinetics. *Electrophoresis*, *29*(7), 1467–1472.

Suyama, Y., Kawamuro, K., Kinoshita, I., Yoshimura, K., Tsumura, Y., & Takahara, H. (1996). DNA sequence from a fossil pollen of Abies spp. from Pleistocene peat. *Genes & genetic systems*, *71*(3), 145–149.

Tejero, J.-M., Cheronet, O., Gelabert, P., Zagorc, B., Álvarez-Fernández, E., Arias, P., Averbouh, A., et al. (2024). Cervidae antlers exploited to manufacture prehistoric tools and hunting implements as a reliable source of ancient DNA. *Heliyon*, *10*(11), e31858.

Thomson-Laing, G., Howarth, J. D., Atalah, J., Vandergoes, M. J., Li, X., Pearman, J. K., Fitzsimons, S., et al. (2024). Sedimentary ancient DNA reveals the impact of anthropogenic land use disturbance and ecological

shifts on fish community structure in small lowland lake. *The Science of the total environment*, *922*, 171266.

Torsvik, V. L. (1980). Isolation of bacterial DNA from soil. *Soil biology & biochemistry*, *12*(1), 15–21.

Uchii, K., Doi, H., Okahashi, T., Katano, I., Yamanaka, H., Sakata, M. K., & Minamoto, T. (2019). Comparison of inhibition resistance among PCR reagents for detection and quantification of environmental DNA. *Environmental DNA (Hoboken, N.J.)*, *1*(4), 359–367. Wiley.

van der Valk, T., Pečnerová, P., Díez-Del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., et al. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, *591*(7849), 265–269.

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., et al. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science*, *372*(6542). Retrieved from http://dx.doi.org/10.1126/science.abf1667

Walia, S., Khan, A., & Rosenthal, N. (1990). Construction and applications of DNA probes for detection of polychlorinated biphenyl-degrading genotypes in toxic organic-contaminated soil environments. *Applied and environmental microbiology*, *56*(1), 254–259.

Weiner, S., & Wagner, H. D. (1998). THE MATERIAL BONE: Structure-Mechanical Function Relations. *Annual review of materials research*, *28*(Volume 28, 1998), 271–298. Annual Reviews.

Whittle, A., Pollard, J., & Greaney, S. (2023). *Ancient DNA and the European Neolithic: Relations and Descent*. Oxbow Books.

Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M. B., Brand, T. B., Hofreiter, M., et al. (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, *317*(5834), 111–114.

Willerslev, E., & Cooper, A. (2005). Ancient DNA. *Proceedings. Biological sciences / The Royal Society*, *272*(1558), 3–16.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486), 47–51.

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., et al. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, *300*(5620), 791–795.

Willerslev, E., Hansen, A. J., & Poinar, H. N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in ecology & evolution*, *19*(3), 141–147.

Zavala, E. I., Jacobs, Z., Vernot, B., Shunkov, M. V., Kozlikin, M. B., Derevianko, A. P., Essel, E., et al. (2021). Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature*, *595*(7867), 399–403. nature.com.

Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., et al. (2020). Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. *Science*, *370*(6516), 584–587.

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and environmental microbiology*, *62*(2), 316–322.

9. Appendix

The Supplementary documents of the unpublished manuscript (Chapter 4) are attached here.

A sedimentary ancient DNA perspective on human and carnivore persistence through the Late Pleistocene in El Mirón Cave, Spain

Supplementary Information

Supplementary Section 1: Site Description

1. El Mirón cave

The archaeological site of El Mirón is located on the eastern edge of the northern Spanish region of Cantabria, about 20 km from the present Bay of Biscay shore. The site is 260 m a.s.l. on a mountainside dominating the broad upper Asón River valley, which here is at c. 100 m a.s.l. It is a limestone cave excavated since 1996 under the direction of L.G. Straus and M.R. González Morales, and recently I. Gutiérrez and D. Cuenca [1,2]. The cave has a prominent mouth and a spacious, mostly dry, sunlit vestibule (30 m deep to 16 m wide and 13 to 20 m high) (Supplementary Figure 1A). In contrast, the 100 m deep inner cave is narrow, low-ceilinged and completely dark.

El Mirón was occupied by humans from the late Middle Palaeolithic until modern times, including much of the Upper Palaeolithic (Gravettian, Solutrean, Magdalenian and the Epipalaeolithic Azilian period, as well as the Neolithic, Chalcolithic and Bronze Age). It is one of the best-dated late Quaternary sites of the Cantabrian, with 101 radiocarbon assays, major assemblages of artefacts and faunal remains, and numerous palaeoenvironmental studies [3–6]. With excellent preservation of osseous material and large-scale excavations, the site has yielded numerous human remains, notably those of the Lower Magdalenian "Red Lady" burial (18898-18747 cal BP) [7]. The genome of this individual has played a key role in defining the genetic characteristics of the Magdalenian population of Europe [8]. Recent genetic research has defined the Magdalenian population represented by the human buried remains called "Red Lady" as the result of the Gravettian-like population that survived the LGM during the Solutrean (represented by the genome of Fournol) plus an incoming gene flow represented by the genome of Arene-Candide 16. The genetic ancestry of the "Red Lady" can be modelled with these two sources [9]. Other human remains (loose teeth) have been found in other Lower Magdalenian levels. Surface deposits have been disturbed pertaining in part to the Chalcolithic and Bronze Age.

The archaeological evidence found at the lowermost sequence of El Mirón (Mousterian, Early Upper Palaeolithic: >46-28 cal kya BP), excavated below a series of Solutrean

levels only in a 2m^2 *sondage* (squares W-X10) dug below the base of a large looters' pit at the rear of the cave vestibule, but tied in stratigraphically to the overlying sequence of Initial, Lower, Middle and Upper Magdalenian levels in the contiguous area not affected by looting, indicates an alternating occupation of the cave by humans and carnivores. Human visits to the cave during the Mousterian and early Upper Palaeolithic were sporadic, ephemeral, and probably seasonal, with limited evidence showing that occupations took place in winter or spring based on the ages of ibex individuals [1].

The Lower Magdalenian levels (c. 20-18 cal kya), in particular, are extraordinarily rich in artefacts, including lithic and osseous manufacturing debris and finished products, namely domestic tools, weapons, personal ornaments, and works of portable art (some both characteristic of the regional archaeological culture at this time and others indicative of social connections with human groups in southern France), and anthropogenic features items such as hearths, pits and a possible wall. The ritual human burial documented during this period at El Mirón is associated with (indeed marked by) non-local red ochre and rock art and is unique for this period in Iberia. The classic Cantabrian Lower Magdalenian horizon was preceded in El Mirón by a series of rich Initial Magdalenian levels (c. 21-20 cal kya) extremely rare for Iberia. They overlie Solutrean levels (c. 25-21 cal kya) that attest to several brief visits to this montane site (possibly from base camps near the now-inundated LGM coast) by hunting parties who left behind numerous fragmentary stone projectile points and elements of personal adornment. The pre-LGM humans (presumably *Neanderthal* and *or* Anatomically Modern Humans (AMH)) who inhabited the cave in preceding times (> 47, and up to c. 31.5 cal kya) were even more ephemeral. All these archaeological levels (Mousterian, Early Upper Palaeolithic, Gravettian, Solutrean and Initial Magdalenian) in El Mirón provide a unique archive for high-resolution studies on the alternating patterns of human and animal presence within Pleistocene archaeological cave sites[10]. The genetic study of a Magdalenian burial (The Red Lady) was crucial to reveal the arrival of Villabruna genes into the former refugium of Iberia during the Magdalenian[8].

2. Sedimentology and sample stratigraphic and taphonomic context

The sedimentology of El Mirón is published by the late W.R. Farrand in a *Geoarchaeology* article and in a chapter of the El Mirón monograph [11,12]. The performed analyses include study on the particle size, the roudness, Calcium Carbonate and Organic Matter, Lithology and Clay Mineralogy.

Levels 130-121 are a series of colluvial, sandy loam deposits. Among these are finer grain levels, others more gravelly, with varying amounts of alluvial pebbles and cobbles and limestone rocks spalled rocks from the cave walls and ceiling, generally yellowish-brown or brownish-yellow, but a few browner or olive-brown. The lay-sized sediment fraction ranges from 30 to 50%, the rest is coarse fraction. There are no visible changes in this ratio through the sondage.

The pH ranges between 8.67-8.08, except Levels 125-123, which vary between 7.74-7.91. These levels are generally low in organic matter, although some Solutrean levels are richer therein [11]. These basal levels (130-121) slope down (westward) toward the cave mouth, c. 15-20°. Their faunal and cultural remains were deposited atop the eroded face of alluvial sediments that fill the inner cave (Supplementary Figure 1), which had been voided out of the cave vestibule by running water. Hence, these levels are at the interface between the alluvial in-filling of the inner cave and the vestibule, which had later been re-filled with highly anthropogenic Upper Palaeolithic and post-Palaeolithic deposits (coarse material). Levels 121-130 are described by Farrand [11] as “colluvial terrace” material. The Solutrean levels 121-127 are relatively rather thin, very clearly stratified and without macroscopic observations of diagenesis, with concentrated bones and stone artefacts. The bones in even levels 128-130 are well-preserved. There is no evidence of gullying or significant water disturbance, but the materials are quite scarce and dispersed through quite thick levels (especially 130)[13] reveal that the bones (including ones of very small animals—birds, lagomorphs) are in fine condition and showing not signs of transport (i.e., surfaces not altered by running water or other diagenetic processes, etc.) in these levels. So, neither the sedimentology nor the taphonomy would lead us to conclude disturbance. We do not have any evidence indicating that the deposits, particularly those from the Solutrean and Initial

Magdalenian levels, are in a secondary position. Similar studies at other Cantabrian archaeological sites have revealed secondary depositions and disturbances. For instance, in La Cueva de El Pendo[14,15], a taphonomic reappraisal of the sequences suggests that the interstratifications do not correspond to a primary archaeological sequence but may result from post-depositional processes. A similar situation has been reported at La Güelga cave, where micromorphological observations and new radiocarbon dating strongly suggest that the few presumably Châtelperronian finds were transported [16].

Description of the sampled levels

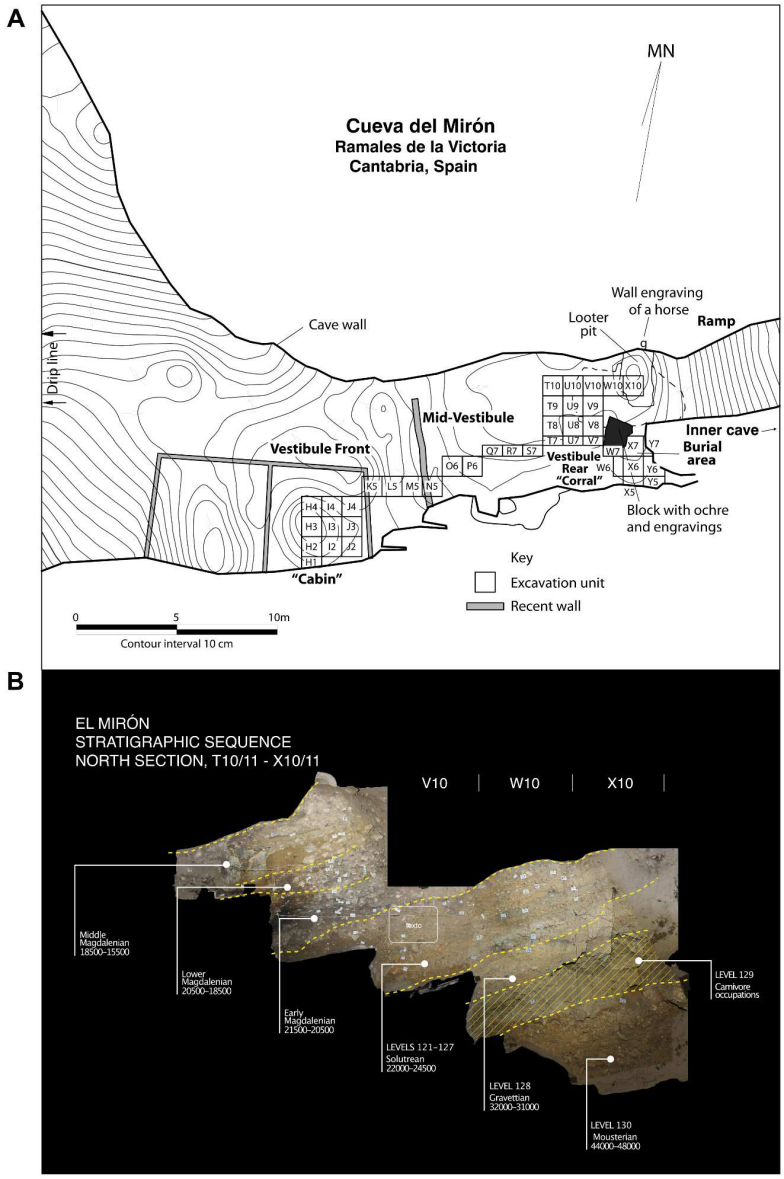
Level 130, which is a massive (> 1 m), undifferentiated horizon whose base was not reached, yielded 115 stone artefacts (2 denticulates, a burin, a flake core, 110 items of debitage and a hammerstone) that were not concentrated either horizontally or vertically. Undated, 35-50 cm-thick Level 129 produced only 14 items of debitage—including blades and bladelets [10]. Level 128—30-50 cm-thick—is artifactually much richer (550 items of debitage, including 66 blades and bladelets, a bladelet core and a mixed core, plus 23 retouched tools including many classic "Upper Palaeolithic" types none of which, however, is diagnostic of the Gravettian despite the 31-32 cal ky BP radiocarbon age[17 13].

Levels 127-122 are generally 10-20 cm thick each and some are discontinuous in the area excavated. All but one yielded Solutrean points (laurel, willow, shouldered or unifacial types), while artefact-poor Level 121 had none, although overlying, disturbed "Level" 120 (hard-packed, mixed sediments at the very base of the looter pit) had a shouldered point. Radiocarbon dates from Levels 127, 126, 125, 122 and 121 all fall within the Solutrean period. In total, these levels yielded 13,673 items of lithic debitage (ranging between 451-3739 per level) and 223 retouched tools/weapons (between 10-55 per level), and several perforated animal teeth. bones and mollusc shells, small numbers of bone/antler artefacts (points, needles, awls, and blanks) and a possibly engraved antler tine [3 14].

None of these basal levels sampled in the deep pit (130-121) had apparent hearths, concentrations of charcoal, or black staining from fires. Human occupations---at least in

this area at the rear of the cave vestibule—were relatively minor and probably short, although the Solutrean occupations—possibly by hunting parties—were more significant than the visits during late Middle Palaeolithic, possibly Early Upper Palaeolithic and "Gravettian" periods whose traces are quite ephemeral.

Atop this sequence of light-colour, organic- and artefact-poor levels, there lies the long, artifactually and faunally extraordinarily rich, dark (variegated, but often blackish-brown, organic-rich, loam, with angular, white limestone fragments and cobbles) series of early (Initial and Lower) Magdalenian levels. The one sampled for this study, Level 119.2 in square V10, pertains to the Initial Magdalenian. There is a clear break (unconformity) between Level 121—the uppermost Solutrean—and the c. 20-30 cm-thick Level 119-119.3 series. This 119-119.3 series yielded 62,087 items of lithic debitage and 639 retouched tools/weapon elements. Osseous artefacts—notably large, round-section points (*sagaies*)-- are relatively numerous and often engraved. There are also bone needles, blanks, and—in Level 19.2—a perforated, schist-like plaque with the engraving of a horse head. The contrasts between this stratigraphic unit (and the other overlying Initial and Lower Magdalenian levels) and the underlying Mousterian, Early Upper Palaeolithic and even Solutrean levels in terms of sediment colour, organic matter, faunal, charcoal and artefact contents are stark. Human occupations of the cave vestibule during the Initial Magdalenian were massive, long-duration, repeated (spring and fall) and multi-functional, as judged by the wide variety of artefacts (used for diverse manufacture/maintenance activities, hunting, decoration, etc.) and evidence of fires [18].



Supplementary Figure 1: A) Plan of El Mirón Cave (by L.G. Straus and R.L. Stauber, based on cave topography by E. Torres). B) Stratigraphic sequence of the North Section with corresponding periods. (photos and montage by M.R. González Morales).

3. Archaeozoological record at El Mirón

The Middle (Level 130, ≥46 cal ky BP), and/or Early Upper Palaeolithic (Levels 129-128, Level 128: 31-32 cal ky BP), Solutrean (Levels 127-121: c. 24.5-21.5 cal ky BP) and Initial Magdalenian (Levels 119-119.2: c.21-20.5 cal ky BP) mammalian faunas from El Mirón (which have been analysed taxonomically and taphonomically) principally include red deer, and ibex, in addition to chamois, roe deer, leporids, plus traces of horse, mammoth, wolf, cave lion, and bear in some of the levels (see Supplementary Tables 4 and 5) [13,19]. The summary of all the fauna identifications in levels 130-119.2 is presented in Supplementary Tables 3 and 5. The small amount of carnivore remains identified is compatible with the roaming activities of these taxa that did not leave evident remains. The archaeological evidence found at the lower sequence of El Mirón (Mousterian, Early Upper Paleolithic: >46-28 cal kya BP), clearly indicates an alternating occupation of the cave by humans and carnivores. Human visits to the cave during the Mousterian and early Upper Palaeolithic were sporadic, ephemeral, and probably seasonal, with limited evidence showing that visit occupations took place in winter or spring based on the ages of ibex individuals [1]. No hominin remains have been discovered in the deep but small (2x1 m) pit (squares W-X/10) that samples these basal levels or in the archaeologically richer overlying sequence of Solutrean and Initial Magdalenian levels (c. 24.5-20.5 cal kya BP) excavated over a somewhat larger area also at the vestibule rear (1m^2 more for the Solutrean levels in square V10); 2 m^2 for the Initial Magdalenian ones in squares V10 and V8)). Despite the evidence of albeit scanty human occupation, no human DNA has been recovered from the Mousterian and EUP levels.

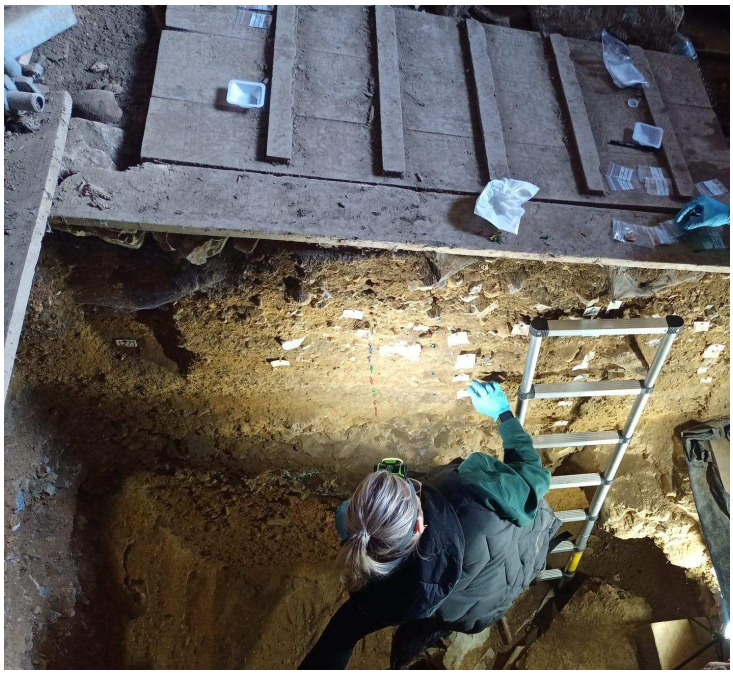
Supplementary Table 5: Ungulate & Leporid remains from the Middle and Palaeolithic, Early Upper Palaeolithic, Solutrean and Initial Magdalenian Levels in El Mirón Cave studied in this project (NISP/MNI).

Level	*Equus* sp.	Bos/Bison sp.	*Cervus elaphus*	*Capreolus capreolus*	*Capra pyrenaica*	*Rupicapra pyrenaica*	*Lepus* sp.
119.2	2/1	9/2	151/8	2/1	122/12	12/1	
121			107/4		77/4	22/33	1/1
122		2/1	106/3		107/5	19/2	1/1
123					12/1	2/1	
124	1/1		59/3		59/4	10/3	
125			73/4		97/4	5/2	1/1
126			14/2		8/3	9/1	1/1
127			20/3		26/2		
128			44/3		86/3	10/1	64/6
129			8/2		26/2	1/1	15/2
130	2/1		15/2		54/3	10/2	6/2

Supplementary Section 2: Supplementary Methods

1. Sampling

Sampling was performed in the profile of the archaeological excavation trench at the rear of the El Mirón vestibule [5] in 2022 and 2023 (Figure 1, Supplementary Table 1, Supplementary Figure 2). After assessing the cave's stratigraphy, samples were taken with documentation of the locations of the excavation's 3D grid system and the archaeological context. 100mg of sediment was taken for each sample, with control of contamination by using gloves and washing the sampling instruments with 5% bleach solution. Sediment was stored in sterilised bags and preserved in cold conditions until transported to the University of Vienna (Austria).



Supplementary Figure 2: Sampling procedure on the rear vestibule of the Cave, February 2023.

2. Laboratory protocols:

All samples were processed at the Palaeogenomics laboratory of the University of Vienna. We applied contamination control measures to mitigate the effect of modern DNA contamination. The samples were prepared in dedicated clean room facilities. We included contamination controls at each step of the wet lab pipeline to control for potential contamination of reagents. 50 mg of sediment was extracted per sample following the protocol from Dabney [20] with adaptations from Korlević [21]. Sample El Miron_18 was extracted twice from two different 50 mg aliquots to increase the chances of retrieving human DNA. The DNA was eluted it in 50 µL TET buffer (10 mM Tris-HCl, 1 mM EDTA, 0.05% Tween 20, pH 8.0) and double-stranded libraries of 25 µL of the extract, as described in Meyer and Kircher 2010[22], were prepared without shearing the DNA into smaller fragments. We used a MinElute PCR Purification kit from Qiagen to clean up the samples instead of SPRI beads and eluted in 40 µL EBT buffer (1 mM EDTA, 0.05% Tween-20). Positive control was added to each library batch using 24 µL of deionised water and 1 µL of a 1:250 dilution of CL104.

The number of PCR cycles per sample was determined through real-time PCR. Libraries were double-indexed [23] and amplified with PfuTurbo Cx HotStart DNA Polymerase from Agilent. A subsequent clean-up was performed with 1.2x NGS clean-up magnetic beads per sample, introducing a size selection. We eluted in 25 µL EBT buffer (1 mM EDTA, 0.05% Tween-20). Each indexed library was amplified further in preparation for solution capture using KAPA HiFi HotStart DNA Polymerase and IS5/IS6 as primer pairs[22]. We enriched the amplified libraries employing a custom-designed mitochondrial capture including human and 50 mammalian mtDNA sequences[24], following the TWIST capture protocol [25]. Following 16 hours of hybridisation at 65°C and four rounds of washing, we mobilised the target DNA from the probes in a PCR cycler at 95°C for 5 minutes. Another qPCR was performed before amplifying half of the captured library using KAPA HiFi HotStart DNA Polymerase, and the primer pairs IS5/IS6. We cleaned up with magnetic beads. Quality controls were performed by applying Qubit and Tapestation. The captured libraries were sequenced at the Vienna BioCenter Core Facilities (VBCF) on an Illumina NovaSeq 6000 platform.

3. Bioinformatics

Sequencing reads were demultiplexed. We removed adapters and reads shorter 30 bases and poly A tails were removed using Cutadapt 4.2 [26]. Filtered reads were processed with SGA [27] and FASTX-toolkit [28] to remove reads with low qualities (base quality lower than 30 in 25% or more of the read length) and exact duplicates (using SGA preprocess with --dust-threshold=1, followed by SGA index with -a ropebwt option, and SGA filter with --no-kmer-check option). Filtered reads were then classified at the family level using euka [29], with a minimum cutoff of 50 fragments per taxa, an entropy value of 1.17, and a minBins value of 6. We also used the general deamination values to set as priors of the euka classification, minimising the contamination. This was set to 0.2.

We then processed each of the individual families separately. Classified reads by euka were then converted to fasta reads using Samtools 1.3 [30] and analysed with BLASTn [31] using the whole nt database from NCBI (2023-08-17) and -outfmt 6 option. Classified reads were then analysed with MEGAN 6.25.6 [32], and reads were classified at the genus level using the LCA approach and the taxonomy file from NCBI, setting a minimum support of 100 reads to validate the presence. In cases where the number of 100 was clear at the species level, we also classified the reads at the species level and identified the species. This is common in taxa where only one taxon per genus is present. Only reads that were classified at the selected level were kept.

The classified reads at species/genus level were then converted to fastq reads with Samtools 1.3 [30] and aligned with BWA aln 0.7.17 [33] disabling seeding, setting a gap open penalty of 2 and an edit distance of 0.01 using the mtDNA reference sequence of the identified hit, if we could not resolve the species, we selected the most likely species reference sequence from that genus according to the literature (Supplementary Section 4). Reads were later filtered by read quality > 30, and duplicates were removed using Samtools and Picard tools 3.0.0 [34]. We used MapDamage 2.0 [35] and Qualimap 2.0 [87] to assess the mapping statistics. We then defined an identification at the species level when 50 reads were found, and the deamination values were >20% on both ends. To perform the read length distribution, we employed samtools [36], and the results were later plotted with R 4.1.2 [37].

4. Single taxa Analyses

After alignment, individual animal genomes were studied, and we determined the deamination values, average read length, and the number of recovered reads. We explored the diversity within the genomes by determining the number of variable positions (Supplementary Table 7). This was restricted to positions with coverage greater than 5 using samtools 1.9 mpileup and bcftools 1.19[38]. The analyses were performed in all the sites and only on the transversions, to exclude the effect of damage. To further minimise the effect of damage in such calculation, only positions with MAF>0.21 were considered, excluding positions in which only a single read was carrying the derived position.

We determined the presence of partial genomes when the average coverage of a taxon was >5X. In these cases, we called a consensus sequence using ANGSD 0.941[39] (calling the majority allele and a minimum coverage of 5). We also studied the sequence visually using IGV 2.16.1 [40] and through variant calling using bcftools [41] to observe the presence of discordant alignments.

5. Detection of inhibition in the DNA extracts

To evaluate the inhibition potential of each extract, 5 µL of an extract diluted in EBT by a factor of 5 (equivalent to 1 µL of pure extract) was added to a PCR reaction designed to amplify the CL104 template oligo (see Supplementary Table 12). The inhibition potential was determined by measuring the concentration of amplified DNA using the Qubit 1x dsDNA HS Assay Kit and comparing it to a control that contained EBT instead of the extract. The efficiency of the PCR reaction is calculated based on the concentration values obtained for the control (4,47±0,10 ng/µL; n=3) (Supplementary Table 13, Supplementary Figure 3).

Supplementary Table 12: Inhibition PCR conditions used in the experiment.

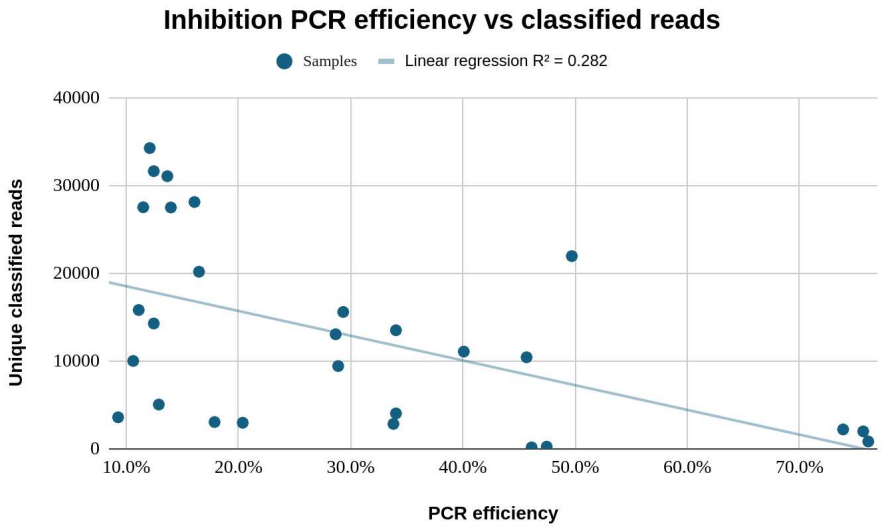
	Volume / µL
10X Standard Taq Reaction Buffer	2.5
dNTPs (10 mM)	0.5
CL107 (5 µM) 42	1
CL108 (5 µM) 42	1
CL 104 (0,4 nM) 42	1
Taq DNA Polymerase	0.125
Nuclease-free water	13.875
Diluted extract	5
Total	25

Step	Temp	Time
Initital denaturation	95°C	30 s
35 cycles	95°C	30 s
	52°C	30 s
	68°C	30s
Final Extension	68°C	5 min
Hold	4°C	

Supplementary Table 13: inhibition PCR results.

Sample	Level	Concentration of amplified template / ng*µL^{-1}	Efficiency	Classified reads (Table S3)
El Miron_26	119.2	2.78	62.2%	3629
El Miron_01	121	0.478	10.7%	26754
El Miron_13	121	0.532	11.9%	10934
El Miron_02	122	0.556	12.4%	14281
El Miron_14	122	0.514	11.5%	27521
El Miron_03	123	0.718	16.1%	28119
El Miron_15	123	0.496	11.1%	15819
El Miron_04	124	1.28	28.7%	13056
El Miron_16	124	0.54	12.1%	34262
El Miron_05	125	0.61	13.7%	31060
El Miron_17	125	0.624	14.0%	27497
El Miron_06	126	0.556	12.4%	31630
El Miron_18	126	1.29	28.9%	9436
El Miron_07	127	0.736	16.5%	20177
El Miron_19	127	0.474	10.6%	10017
El Miron_08	128	1.31	29.3%	15594
El Miron_20	128	1.52	34.0%	13515
El Miron_28	128	0.798	17.9%	3057
El Miron_30	128	0.576	12.9%	5054
El Miron_32	128	0.414	9.3%	3598
El Miron_34	128	2.04	45.7%	10434
El Miron_09	129	2.22	49.7%	21958

El Miron_21	129	1.79	40.1%	11082
El Miron_10	130.1	1.52	34.0%	4040
El Miron_22	130.1	3.3	73.9%	2219
El Miron_36	130.1	3.4	76.1%	845
El Miron_38	130.1	0.91	20.4%	2976
El Miron_11	130.2	1.51	33.8%	2848
El Miron_23	130.2	2.06	46.1%	177
El Miron_12	130.3	3.38	75.7%	1996
El Miron_24	130.3	2.12	47.5%	251



Supplementary Figure 3: Inhibition PCR efficiency vs classified sequencing reads.

We tested the presence of a linear regression (using the PCR efficieny and the sequencing reads) and observed that the R^2 value is 0.282. This indicates that approximately 28.2% of the variance in the dependent variable (classified reads) is explained by the independent variable (Efficiency). The *p-value* is 0.00127, showing statistical significance.

Interpretation: Our regression analysis indicates that inhibition is higher in younger samples compared to older ones, with statistical significance. The lower levels yielded fewer classified reads, suggesting that organic material (DNA and inhibitory substances) is less preserved in the older levels. Therefore, the reduced number of classified reads obtained for level 130 is attributed to low DNA preservation rather than inhibition.

Supplementary Section 3: The Classification of sedaDNA at the species level

1. Classification process

To classify the recovered eukaryotic mtDNA reads at a family-order level, we used the euka pipeline[29]. All 32 samples showed the presence of mammalian aDNA. The average number of reads classified per sample using euka is 41,701. The data was classified by euka into 26 taxa, including three genera, 12 families, and 11 other taxonomic groups from suprafamily to order (Supplementary Table 2). The number of recovered DNA reads varied for each archaeological level and sample (Supplementary Table 2), ranging from 2,641 to 124,874 classified unique reads. While we retrieved aDNA from all the levels, there are vertical differences in the section: the average classified amount of reads per sample from the Mousterian (level 130) is 10,518 classified unique reads. This value is five times higher in the Solutrean levels, with an average of 53,745 classified unique reads (Supplementary Table 2). This classification process enabled us to distinguish the animal families present at El Mirón and prepare the data for the single-taxa analyses (Supplementary Table 2). We do not use this data to assess the relative presence of taxa in the samples due to classification limitations and the inability to specifically validate these reads.

We then refined the euka identifications with BLASTn[31], and family-order level reads were classified into genera and species using the LCA algorithm implemented in MEGAN[32]. The following taxonomical levels were identified in euka and resolved at the genus or species level through the BLASTn process:

The reads are classified at the ***Bovidae*** family level and were classified into genera. We determined the presence of genera: *Bos, Bison, Ovis, Capra and Rupicapra.* Each of the reads aligning to these genera was later attributed to *Bos sp. (most likely Bos primigenius), Bison bonasus, Ovis aries, Capra pyrenaica, and Rupicapra pyrenaica.* In the case of *Rupicapra* and *Capra*, as the classification through BLASTn was not conclusive, we agreed on the assignations based on the phylogeny, which shows that the reads

belong to the diversity of *Capra pyrenaica* and *Rupicapra pyrenaica.* (Supplementary Figure 34).

The reads classified at the ***Canidae*** family level were later classified in genera. We determined the presence of the following genera: *Cuon, Canis and Vulpes.* Each of the reads aligning to these genera was later attributed to *Cuon alpinus, Canis lupus* and *Vulpes vulpes. Cuon alpinus* is an uncommon taxa in Iberia, with only some identifications in the archaeofauna register [43]. *Vulpes vulpes* and *Canis lupus* are however much more common, both have already been detected in El Mirón [10].

The reads belonging to ***Carnivora*** yielded only *Crocuta* identifications; all the reads classified as *Crocuta* genus were determined to belong to *Crocuta crocuta.*

The reads belonging to the ***Cervidae*** family are later classified in genera. We only identified reads belonging to the genus *Cervus.* All reads classified as *Cervus* were identified as *Cervus elaphus.*

No further attributions were possible for reads classified at the ***Charadriiformes*** level.

The reads belonging to the ***Corvoidea*** family are classified in genera. We successfully identified reads belonging to the genus *Pyrrhocorax.* All reads classified as *Pyrrhocorax* were identified as *Pyrrhocorax graculus, which is* common in Pleistocene sites in Iberia[44] although not identified in El Mirón remains yet.

The reads belonging to ***Corvus*** are classified into genera, and only the reads belonging to the genus *Corvus are classified.* All reads classified as *Corvus* were identified as *Corvus corax. Corvus corax* has been identified as a human exploited animal during the Pleistocene in Iberia [45].

The reads belonging to the ***Soricidae*** were later classified into genera. We successfully identified reads belonging to the genus *Sorex* and attributed them to *Sorex araneus.*

The reads belonging to the ***Suina*** level were later classified into the genera *Sus*, and those belonging to the genera *Sus* were classified as *Sus scrofa.*

The reads belonging to the ***Ursidae*** family were later classified in genera. We successfully identified reads belonging to the genus *Ursus.* All the reads belonging to *Ursus* were determined to belong to *Ursus arctos.*

The reads belonging to the ***Equidae*** family were later classified in genera, we determined the presence of the genus *Equus*. All the reads belonging to the genus *Equus* were determined to belong to *Equus* sp.

The reads belonging to the ***Lepus*** genus were determined to belong to *Lepus* sp. A common species in El Mirón fauna.

The reads belonging to the ***Passeroidea*** parvorder were not sufficient for further classifications.

The reads belonging to the ***Eulipotyphla*** order were later classified into genera, we determined the presence of the genus *Talpa*. All the reads belonging to the genus *Talpa* were classified as *Talpa europaea.*

The reads belonging to the ***Felidae*** family were later classified into genera. We determined the presence of the following genera: *Panthera, Felis, and Lynx.* The reads aligning to these genera were later attributed to *Panthera pardus*, *Felis catus,* and *Lynx pardinus.* The iberian Lynx was common in Iberia during the Pleistocene [46]

The reads belonging to the ***Neognathae*** infraclass were later classified into genera, and we determined the presence of the following genera: *Columba* and *Falco*. The reads aligning to these genera was later attributed to *Columba livia* and *Falco* sp. *Columba livia* has also descrived as a human-exploited ressource during the Pleistocene in Iberia [47]

The reads belonging to the ***Pecora*** infraorder were later classified into genera, and we determined the presence of the following genera: *Capreolus* and *Rangifer. The* reads aligning to these genera was later attributed to *Capreolus capreolus* and *Rangifer tarandus. Rangifer tarandus* is rare in the North of Iberia [48]. *Caperolus capreolus* is much more common [49].

The reads belonging to the ***Muridae*** family were not sufficient for further classifications.

The reads belonging to the ***Glires*** order were not sufficient for further classifications.

The reads belonging to the ***Galliformes*** *o*rder were too few for further classifications.

The reads belonging to the ***Microtus*** genus were classified as *Microtus sp.*

The reads belonging to the ***Cricetidae*** family were not sufficient for further classifications.

All the reads belonging to the genus ***Mustela*** were classified as *Mustela nivalis.*

All reads belonging to order ***Proboscidea*** were classified as *Mammuthus primigenius.* Wholly Mammoths have been identified in more than 20 sites from the MIS 3-2 in Iberia [50]

All reads belonging to order ***Perissodactyla*** were classified as *Coelodonta antiquitatis.* Wholly rhino is not common but has been identified in several sites of the North of Iberia. With documented presence until 20 kya [51].

All the reads ***Homininae*** subfamily were classified as *Homo sapiens.*

The unique classified reads were finally aligned to their reference sequence (as mentioned before either the proper identified sequence or the selected sequence based on literature) using BWA aln[33]. This process enabled a strict classification and the discarding of spurious assignments as each sequencing read was placed in a single taxon as well as removing PCR duplicates. To claim a specific species/genus identification in a given sample, we set a cut-off value of 100 reads, an average read length below 75 bp, and a 20% deamination signal (C>T changes to the reference) in the read ends.

This process successfully identified 31 mammal and avian species/genera (Figure 2, Supplementary Table 3). Multiple samples yielded more than 1,000 unique reads assigned to a specific mammalian species. The read length distributions of mapped reads show a general average read length between 50 and 60 bp. However, some taxa,

such as *O. aries,* do not have short reads, pointing towards modern origin. Following these observations (the read-end deamination and length estimates), we determined the presence of modern contaminants in some samples. We detected the presence of modern *Ovis aries* in samples El Miron_13, Mirón_1, El Miron_2, El Miron_14, El Miron_16 and especially in El Miron_3 where 9,460 unique *O. aries* reads were identified as modern. *Felis catus* DNA was identified in sample El Miron_3 (Supplementary Table 3). In addition, the read length distribution of *M. arvalis* (Supplementary Figure 4) shows the presence of elevated long reads despite a clear deamination signal (Supplementary Table 3), we decided to not study these reads due to the uncongruent length. Therefore, *Ovis' and Felis'* and *Microtus's* identifications were removed from the results, and these taxa were not further analysed. Following this identification we also removed the *Capra* sp. from ElMiron_3 from the analyses because of damage (10% lower than the average 55%) and long read distribution (average of 75 vs average arround 60 for the rest). The origin of the high numbers of *O. aries* DNA reads in some of the samples is congruent with the usage of the cave as a stable during recent times. This contamination is primarily present in the samples closest to the pre-excavation surface of the W-X10 sondage (at the base of a large looter pit)(Figure 1), probably from ovid faecal sources. This contamination, however, does not affect the rest of the results, as the deamination values of the other taxa are noticeably different (Supplementary Table 3). We also detected the presence of contaminant modern human DNA in more than half of the samples (Supplementary Table 4). Only human mtDNA samples with more than 30% of C>T changes relative to the reference at read ends have been further examined in downstream analyses and identified as specific ancient human identifications.

We studied the edit distances of the individual alignments observing that most of the sample edit distances averages are between 1 and 2, which is consistent with metagenomic ancient DNA [52]. The edit distance is plotted together with coverage and damage for all the discussed individual mitochondrial sequences.

2. Statistical analyses

Efficiency and DNA preservation

We used the total numbers of classified reads at the species level, focusing on the ones with compelling evidence of ancient origin (Supplementary Table 3), to elaborate on patterns of DNA preservation. The average of classified reads per sample is 12,904. All the Mousterian samples are below this number. We then corrected the total amount of present classified reads by the number of sequenced reads (Supplementary Table 1) and we calculated both the Average (0.001966016248) and the Median (0.001509933178). All the Mousterian samples (level 130) are below these numbers (Supplementary Table 14), the temporal trend shows less preservation in the older samples (Supplementary Figure 4).

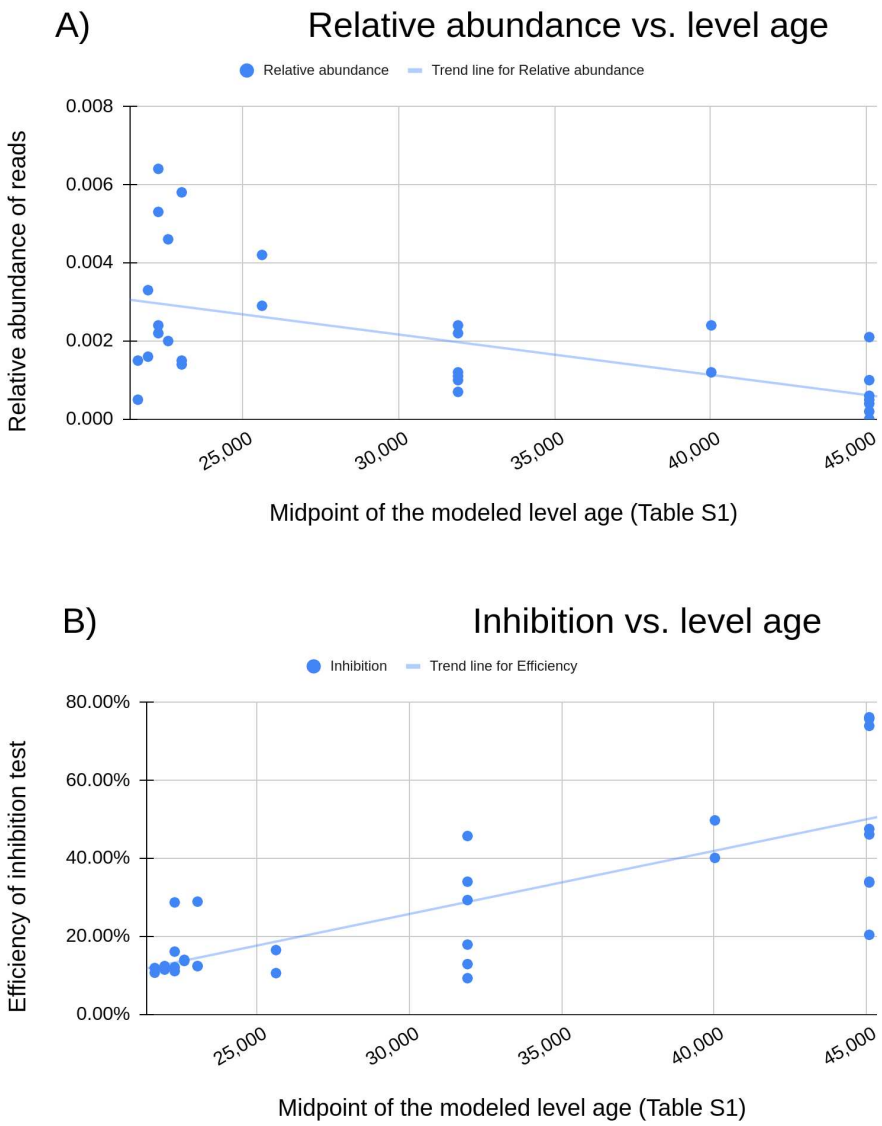
Supplementary Table 14: Preservation of sedaDNA. Description of the A) sequenced reads, B) Classified reads at the species level and C) relative abundance of these reads over the sequenced ones.

Sample	Level	Sequenced reads	Classified reads	Relative abundance
El Miron_26	119.2_Initial Magdalenian	7806511	3629	0.0005
El Miron_13	121_Solutrean	7127339	10934	0.0015
El Miron_1	121_Solutrean	8123344	26754	0.0033
El Miron_2	122_Solutrean	8688768	14281	0.0016
El Miron_14	122_Solutrean	11262225	27521	0.0024
El Miron_15	123_Solutrean	7159133	15819	0.0022
El Miron_3	123_Solutrean	4406322	28119	0.0064
El Miron_16	124_Solutrean	6453916	34262	0.0053
El Miron_4	124_Solutrean	6674901	13056	0.0020
El Miron_17	125_Solutrean	6015536	27497	0.0046
El Miron_5	125_Solutrean	5321453	31060	0.0058
El Miron_18	126_Solutrean	6205522	9220	0.0015
El Miron_18	126_Solutrean	6875295	9339	0.0014
El Miron_6	126_Solutrean	7586805	31630	0.0042
El Miron_7	127_Solutrean	7043110	20177	0.0029
El Miron_19	127_Solutrean	10007474	10017	0.0010
El Miron_8	128_Gravettian	7248214	15594	0.0022

El Miron_20	128_Gravettian	5529719	13495	0.0024
El Miron_30	128_Gravettian	4112118	5054	0.0012
El Miron_32	128_Gravettian	3165761	3598	0.0011
El Miron_28	128_Gravettian	4271627	3057	0.0007
El Miron_34	128_Gravettian	4410579	10434	0.0024
El Miron_21	129_unidentified	9166513	11082	0.0012
El Miron_9	129_unidentified	10322029	21958	0.0021
El Miron_36	130.1_Mousterian	4870182	845	0.0002
El Miron_38	130.1_Mousterian	3024452	2976	0.0010
El Miron_22	130.1_Mousterian	4982782	2219	0.0004
El Miron_10	130.1_Mousterian	6482939	4040	0.0006
El Miron_23	130.2_Mousterian	5747802	177	0.0000
El Miron_11	130.2_Mousterian	6099622	2848	0.0005
El Miron_24	130.3_Mousterian	7882363	251	0.0000
El Miron_12	130.3_Mousterian	7685749	1996	0.0003

We tested the regression of the relative abundance of DNA reads against the midpoint of the modeled age for each sampled level to assess the relationship between time and DNA preservation (Supplementary Table 1). The relationship is statistically significant, with an R^2 value of 0.3849 and a *p-value* of 0.000152.

Additionally, we evaluated the regression of inhibition test efficiency against the midpoint of the modeled age for each sampled level to further explore the relationship between time and DNA preservation. This relationship is also statistically significant, with an R^2 value of 0.4475 and a *p-value* of 0.000028. Both relationships are presented in Supplementary Figure 4.



Supplementary Figure 4: DNA preservation and PCR efficiency vs. sample age. A) Relative abundance of classified reads vs. level age B) Efficiency of the PCR reaction vs. level age. The regression values show that the older samples have less relative DNA presence and less inhibition.

The previous results suggest that the DNA preservation is negatively related to sample age and the efficiency of the DNA amplification is positively related to sample age.

sedaDNA and Archaeofauna relationship

Prey animals are the most present in the faunal assemblages and are relevant to assessing subsistence patterns among human populations[53]. At the species level, it is clear that most of the genetic material belongs to Spanish ibex (*Capra pyrenaica)* and red deer (*Cervus elaphus).* Previous archaeozoological taxa identifications reported that 47% of the Upper Palaeolithic ungulate faunal remains from El Mirón correspond to Spanish ibex, followed by red deer (19%) and chamois (6%)[10] (Supplementary Table 4). Our findings indicate that the sedaDNA of red deer represents an average 27% of the reads, followed by Spanish ibex with 23% and Chamois with 6% (Supplementary Table 3, Figure 2).

We next tested the possibility that the observations of sedaDNA and archeozoofauna show the same trend. We first converted Supplementary Table 4 into relative values according to the total sample per level (total (NISP)[10,13,54,55] or total classified reads at the species level). This enables us to compare the values across the sample profile. The obtained relative values are presented in Supplementary Table 15.

Supplementary Table 15: Relative presence of the taxa classified at the individual level. The series labelled as sedaDNA refer to sedaDNA observations, the others to NISP observations. The results are extracted from Supplementary Table 4.

Level	130	129	128	127	126	125	124	123	122	121	119
Canis-DNA	0.36	4.63	3.55	7.93	9.41	20.9	20.7	11.11	12.59	8.59	0.14
Canis	57.1	0.0	28.5	0.0	14.2	0.0	0.0	0.0	0.00	0.00	0.00
Vulpes-DNA	0.00	11.89	6.86	2.65	46.89	3.66	5.18	7.45	8.48	6.70	0.22
Vulpes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.0	0.00	0.00	0.00
Cuon-DNA	6.78	15.73	27.9	22.1	3.38	4.07	4.57	2.87	7.50	1.88	3.14
Cuon	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Cervus-DNA	2.76	4.08	9.05	8.01	8.14	19.3	11.1	10.94	12.66	13.31	0.54
Cervus	2.45	1.31	7.19	3.27	2.29	11.9	9.64	1.14	17.32	17.48	25.8
Capreolus-DNA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Capreolus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100
Rangifer-DNA	0.00	0.00	100.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rangifer	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Capra-DNA	6.36	6.47	12.6	6.56	14.38	10.2	10.6	10.34	9.94	11.28	1.13
Capra	7.86	3.78	12.5	3.78	1.16	14.1	8.59	1.75	15.57	11.21	19.6
Rupicapra-DNA	1.18	12.28	8.08	13.7	4.08	6.58	18.4	2.16	19.23	14.27	0.00
Rupicapra	9.90	0.99	9.90	0.00	8.91	4.95	9.90	1.98	18.81	21.78	12.8
Bos-DNA	19.7	15.34	21.4	2.67	8.90	1.49	14.6	4.98	2.59	5.98	2.25
Bos	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.76	0.00	88.2
Bison-DNA	7.00	65.01	0.00	23.0	4.97	0.00	0.00	0.00	0.00	0.00	0.00
Bison	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Equus-DNA	8.00	24.76	10.5	1.44	13.85	23.2	8.43	2.38	4.70	2.71	0.00
Equus	11.7	0.00	0.00	0.00	0.00	0.00	5.88	0.00	0.00	0.00	82.3
Panthera-DNA	18.9	5.84	24.8	3.39	7.67	29.5	1.87	3.38	0.00	1.79	2.63
Panthera	0.00	0.00	100.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lynx-DNA	0.00	0.00	13.8	0.00	86.13	0.00	0.00	0.00	0.00	0.00	0.00
Lynx	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ursus-DNA	10.9	40.13	9.29	4.18	24.17	0.00	0.00	4.67	2.53	1.58	2.51
Ursus	60.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	40.0
Crocuta-DNA	59.2	24.15	4.32	0.99	1.17	1.60	0.00	0.00	0.00	3.58	4.97
Crocuta	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lepus-DNA	0.00	6.78	0.00	8.47	19.92	8.26	16.7	14.92	7.98	16.90	0.00
Lepus	6.25	16.67	66.6	3.13	0.00	1.04	3.13	0.00	0.00	1.04	2.08

We then tested the possibility that the two observations (NISP and sedaDNA) describe the same underlying phenomenon. The data consists of discrete values, and we assume that these two observations are paired and that the distribution does not follow normality. The Kolmogorov-Smirnov (KS) test is a non-parametric statistical test used to compare two datasets' cumulative distribution functions (CDFs). It's particularly useful for determining whether two samples are likely to come from the same distribution. The KS test compares the shape and spread of two distributions, here, we use it to compare if sedaDNA and NISP observations come from the same distribution. We used Rstudio [37] to test the relationship. The results are presented in Supplementary Table 16.

Supplementary Table 16: *Kolmogorov-Smirnov test results* regarding the possible relationship between NISP and sedaDNA at El Mirón.

Taxa	**D value**	***p-value***
Equus	0.63636	0.02325
Rangifer	Not Tested	Not Tested
Cervus	0.27273	0.8326
Capreolus	Not Tested	Not Tested
Bos	0.81818	0.001268
Bison	Not Tested	Not Tested
Capra	0.27273	0.8079
Rupicapra	0.18182	0.9934
Lepus	0.54545	0.0758
Crocuta	Not Tested	Not Tested
Canis	0.72727	0.005946
Vulpes	0.81818	0.001268

Cuon	Not Tested	Not Tested
Panthera	Not Tested	Not Tested
Lynx	Not Tested	Not Tested
Ursus	0.63636	0.02325

Results interpretation: Four taxa present no significant *p-value* ($p<0.05$): *Lepus, Capra, Rupicapra* and *Cervus*. In summary, according to the KS test results, there is insufficient evidence to conclude that, in these four taxa, the DNA and Zooarcheological results come from different distributions. The high *p-value* suggests that any observed differences in the distributions of relative presence are likely due to chance. These results suggest that the main animals of prey, *Capra, Rupicapa,* and *Cervus,* do not have significant differences in relative presence in both datasets. Surprisingly, this is not observed for *Canis,* which presents significant sedaDNA across the profile. Overall, The results suggest that sedaDNA matches the zooarcheological distribution results when the taxa have large numbers of representation in the archeofaunal and sedaDNA observations.

We next tested the possibility of correlation between both datasets. Spearman's correlation coefficient assesses the strength and direction of association between two ranked variables. It is suitable when the data does not meet the assumptions of normality or when there is not an assumption of linearity. We used Rstudio 2023.12.1 and R version 4.1.2 [37] to test the relationship. The results are presented in Supplementary Table 17.

Supplementary Table 17: Spearman's correlation test results regarding the possible relationship between NISP and sedaDNA at El Mirón.

Taxa	**Spearman's correlation coefficient (rho)**	***p-value***
Equus	-0.3237481	0.3314398
Rangifer	Not Tested	Not Tested

Cervus	0.2909091	0.3864346
Capreolus	Not Tested	Not Tested
Bos	-0.5258759	0.09660888
Bison	Not Tested	Not Tested
Capra	-0.277905	0.407993
Rupicapra	0.1743193	0.608212
Lepus	-0.5953553	0.05330582
Crocuta	Not Tested	Not Tested
Canis	-0.4393724	0.1763393
Vulpes	0.2	0.5554454
Cuon	Not Tested	Not Tested
Panthera	Not Tested	Not Tested
Lynx	Not Tested	Not Tested
Ursus	0.108118	0.7516816

Results interpretation: The results show the presence of no significant correlation between both datasets at a statistical threshold of *p-value<0.05*.

Supplementary Section 4: Individual genomes and phylogenies analyses

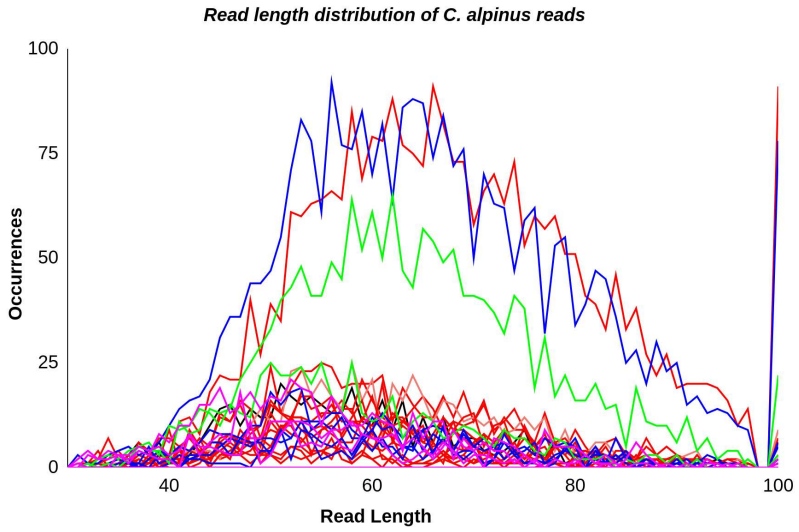
In this section, we describe the analyses of the individual mtDNA sequences from the sediments and their comparisons with present-day and ancient diversity. Consensus sequences were merged with present-day and ancient sequences and were aligned using MAFFT 7.52 [56]. Maximum Likelihood trees were produced using partial deletions of 95%, GTR substitution model and 100 bootstrap replications using MEGA 6 [57]. We also estimated the presence of missing sites on the consensus sequences, counting the number of "N" with a tailor-made script. Trees were produced with Fig Tree 1.4.4 [58] and TreeViewer[59]. For classifying the *C. crocuta* sequences from sample ElMiron_10 we performed a pairwise distance matrix in MEGA[57] 10.2.4 using 100 bootstrap replications, 95% partial deletion and the Maximum composite method. The datasets used for each of the analyses are presented in Supplementary Table 11. We have used more than 750 mtDNA sequences from present-day and ancient genomes from the following publications: [8,9,24,60–155]. The read length distributions of the taxa from the recovered partial genomes (Supplementary Table 6) are described individually in this section.

1. Cuon alpinus

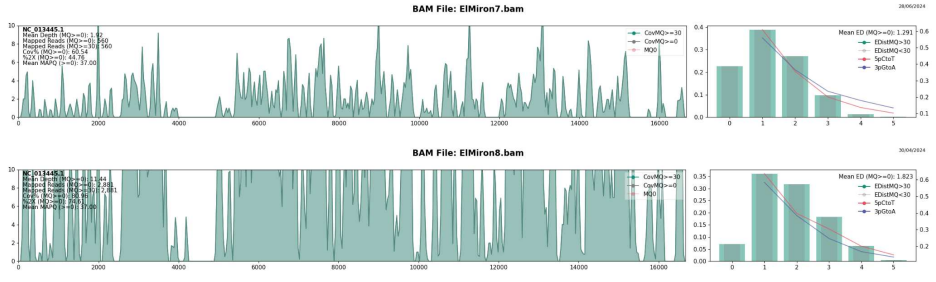
We analyzed three mitochondrial DNA (mtDNA) sequences from *Cuon alpinus*: ElMiron_8, ElMiron_9, and ElMiron_7 (Supplementary Figures 5-7). For reconstructing the consensus sequences and subsequent alignments, we used the reference sequence NC_013445.1 (*Cuon alpinus*). These sequences exhibit 31 variable sites across their length. Coverage plots indicate that the sequences are incomplete; indeed, *Cuon alpinus* shows one of the highest rates of missing sites among all animals examined in this study. The average across the 70 recovered genomes is 3,877 missing sites. Notably, all three *Cuon* sequences have nearly double this number of missing sites (Supplementary Figure 6, Supplementary Table 3).

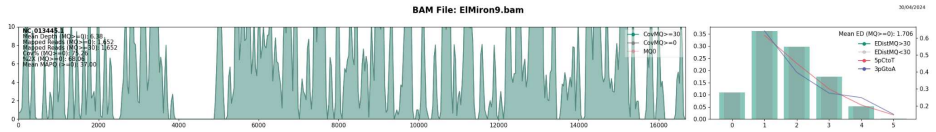
The mapping plot analysis (Supplementary Figure 6) shows that the recovered sequences present coverage gaps compared to the reference (NC_013445.1), representing an Asian individual. The reads aligning to *Cuon* sequences exhibit a lower

edit distance (average = 1.3, SD = 0.19) compared to those aligning to *Canis* sequences (average = 1.6, SD = 0.27). This difference suggests that we are only recovering more conserved regions of the *Cuon* sequence. This could be explained for the fact that the sequence of *Cuon alpinus* was not present in the capture design and therefore we only recovered more conserved regions.

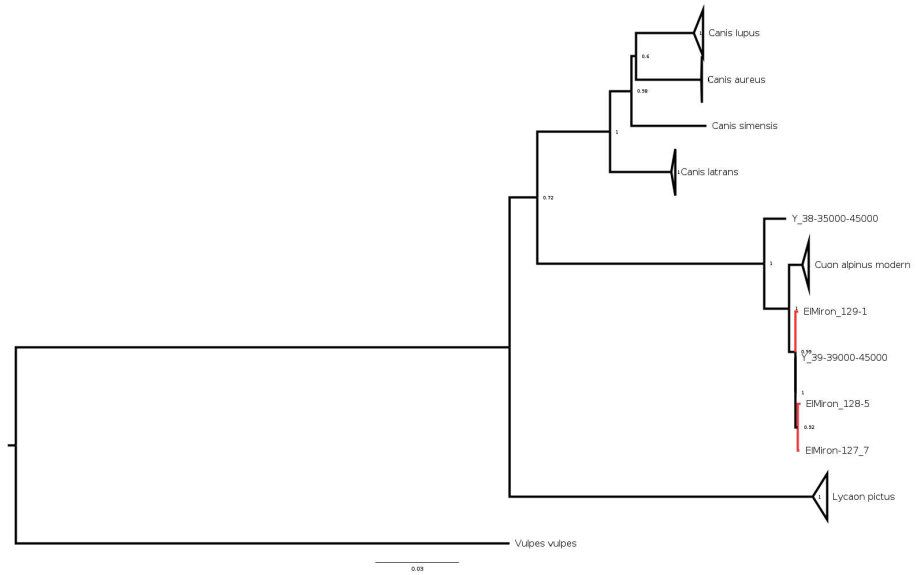


Supplementary Figure 5: read length distribution of El Mirón *C. alpinus* sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.





Supplementary Figure 6*:* Coverage plots and damage plots of *C. alpinus* sedaDNA mtDNA genomes.



Supplementary Figure 7: Maximum likelihood tree of the *Cuon mtDNA* genomes. Red denotes the *Cuon alpinus* mtDNA genomes from El Mirón soil samples*.* Node numbers denote bootstrap values. The tree is rooted with *Vulpes vulpes* mtDNA. All the sequences from El Mirón are located close to sample Y-39 from Bacho Kiro in Bulgaria dated to the late Pleistocene.

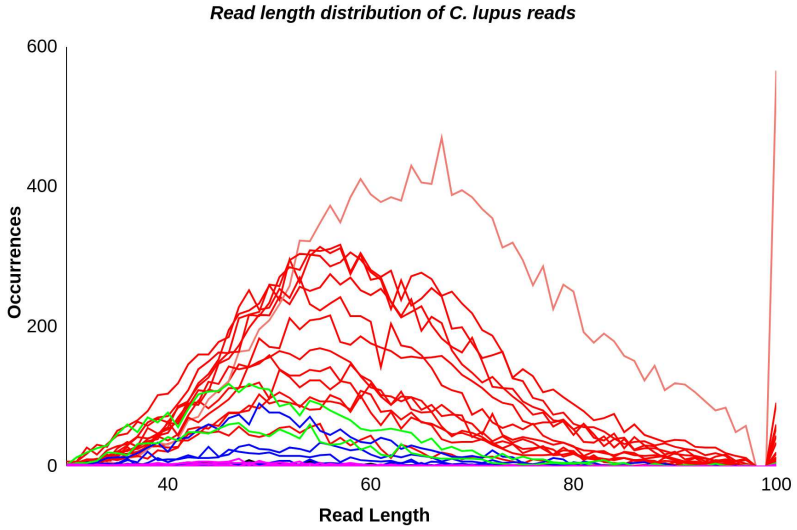
2. Canis lupus

For the *C. lupus* analyses*,* We used samples: ElMiron_1, ElMiron_2, ElMiron_14, ElMiron_15, ElMiron_3, ElMiron_16, ElMiron_4, ElMiron_17, ElMiron_5, ElMiron_6, ElMiron_7, ElMiron_19, ElMiron_20, ElMiron_21 (Supplementary Figures 8-11). For reconstructing the consensus sequences as well as the individual alignments we used the reference sequence NC_008092.1. Most of the sequences are almost complete, with several with only a few hundred missing sites, being one of the species with the highest

values of covered sequences (Supplementary Figure 10, Supplementary Table 3). This shows the similarity between the reference sequence and the El Mirón Pleistocene sequences. As *C. lupus* is one of the taxa with more recovered genomes from El Mirón sediments, we have tested the presence of a correlation between the coverage and the missing sites. We observe that with higher coverage, the amount of missing sites is decreased in a relationship that is not linear (Supplementary Figure 9). We tested the significance of the correlation with Sperman's correlation test and the relationship is highly significant rho=-0.996 and *p-value*=1.3e-13.

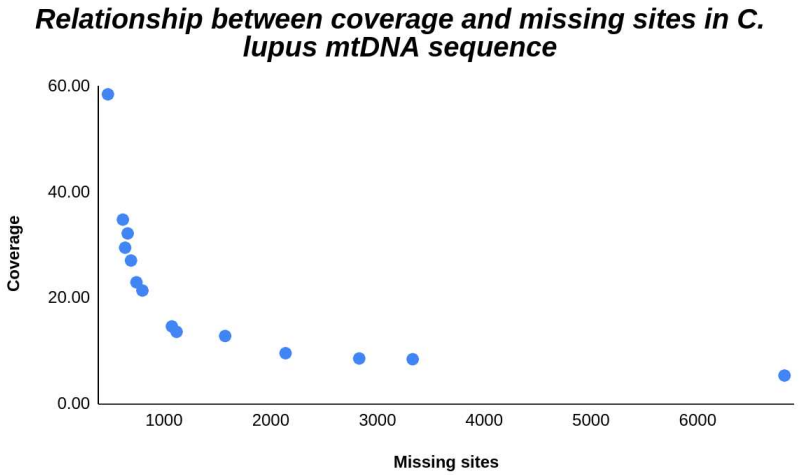
We constructed the multiple sequence alignment with our 14 samples plus 215 other sequences of *C. lupus and* four outliers (NC_013445.1 plus El_Mirón8, ElMiron_7 and ElMiron_8 *C. alpinus* sequence)

The alignment, excluding outgroups, has 2039 variable positions and 194 variable positions within El Mirón 14 sequences.

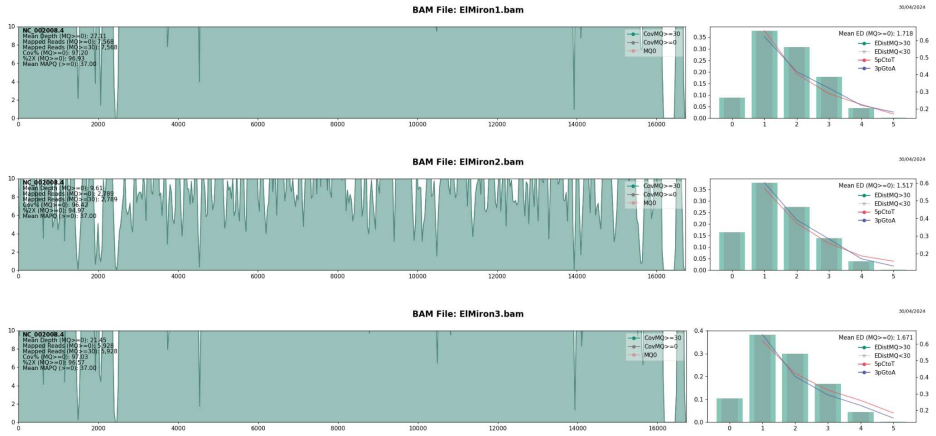


Supplementary Figure 8: read length distribution of El Mirón *C. lupus* sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue

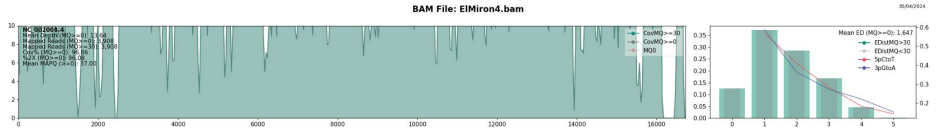
depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



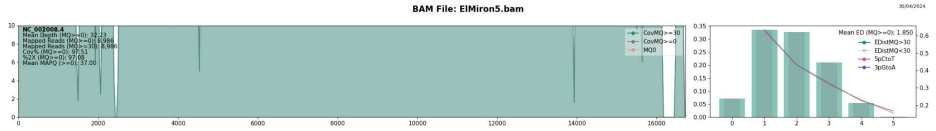
Supplementary Figure 9: Relationship between missing sites and coverage in the reads attributed to *C. lupus* from El Mirón. The distribution suggests that significant yields are necessary to increase the coverage across the sequence with the selected methodology.



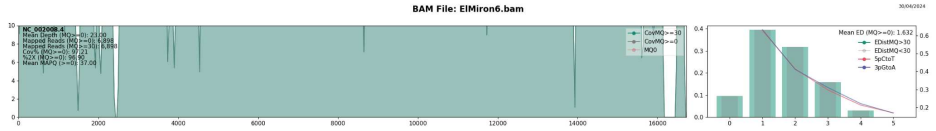
BAM File: ElMiron4.bam



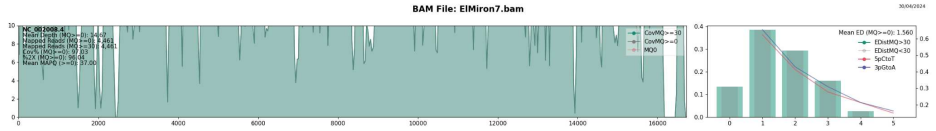
BAM File: ElMiron5.bam



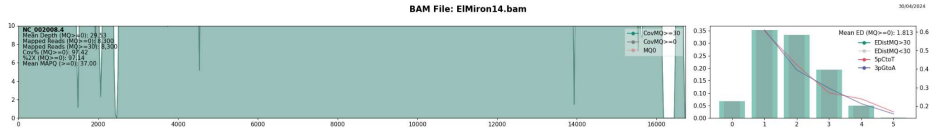
BAM File: ElMiron6.bam



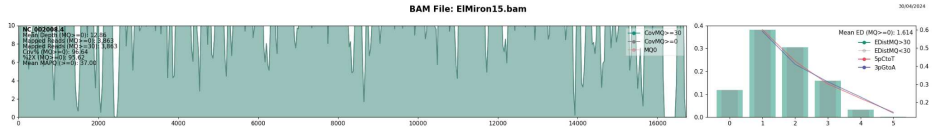
BAM File: ElMiron7.bam



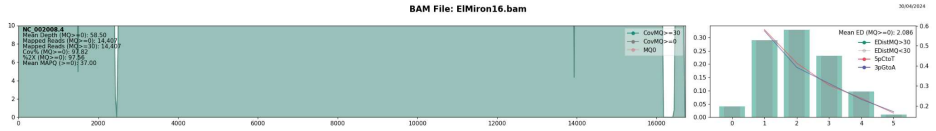
BAM File: ElMiron14.bam



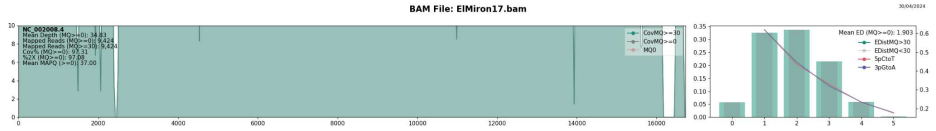
BAM File: ElMiron15.bam



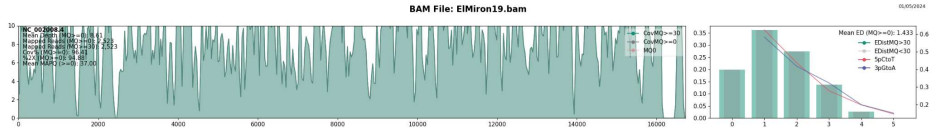
BAM File: ElMiron16.bam

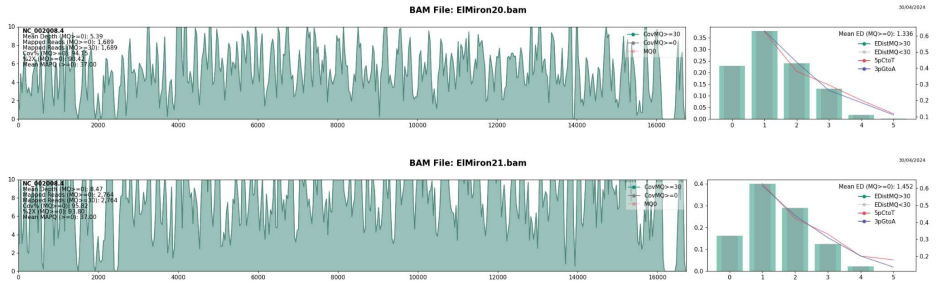


BAM File: ElMiron17.bam



BAM File: ElMiron19.bam





Supplementary Figure 10: Coverage and damage plots of *C. lupus* sedaDNA sedaDNA mtDNA genomes.

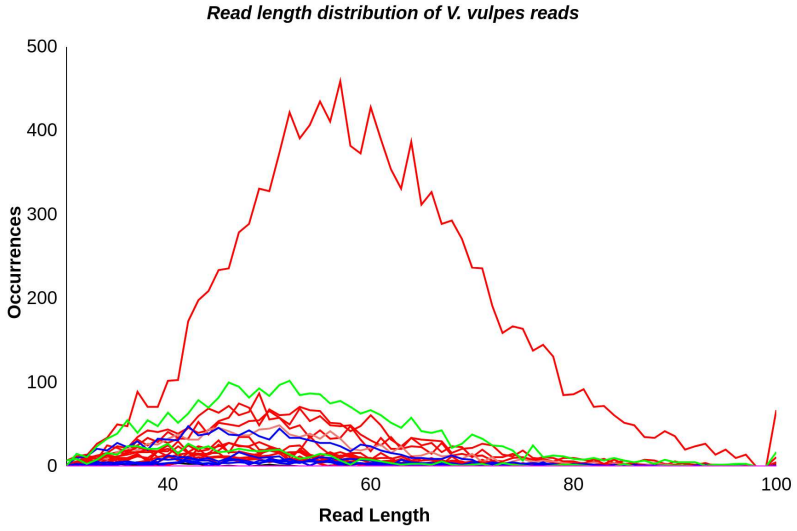

Modern wolves
Modern wolves
TU148_3200BP_wolf
TU144_3200BP_wolf
Modern wolves
Modern wolves
ALAS016_15542ca BP_wolf
DQ480505_Russia-Bij3
Modern wolves
HMNH007_3525calBP_wolf
Modern wolves
Dog_Clade B
KF661046_Russia3
Modern wolves
AL3185_5657calBP_wolf
TU4_6160calBP_wolf
AL2350_28345calBP_wolf
Modern wolves
Pleistocene wolves
Modern wolves
Modern wolves
TU11_14130 BP_ancientwolf
Pagliccci_14000calBP_ancientdog
C27_7000BP_ancientdog
CTC_5000BP_ancientdog
C89_4800BP_ancientdog
HXH_7000BP_ancientdog
C90_4800_ancientdog
AL2397_4400BBP_ancientdog
TU14_2861calBP_ancientwolf
Germany_12500BP_ancientdog
Germany_14700BP_ancientdog
Dog_Clade C
Pleistocene wolves
Pleistocene wolves
Pleistocene wolves
PA1_modernwolf
Swiss2_14500BP_ancientdog
TH11_14341BP_wolf
JK2181_15447calBP_wolf
AL3284_21435BP_wolf
IN18-005_18148BP_wolf
USA_1000BP_ancientdog
USA_8500BP_ancientdog
Argentina_1000BP_ancientdog
C26_7000BP_ancientdog
OL4223_7000BP_ancientdog
OL4061_10900BP_ancientdog
Dog_Clade A
JAL65_89816BP_wolf
Russia_22000BP_wolf
Modern wolves
Pleistocene wolves
IN18-016_32020BP_wolf
Russia_33500_wolf
TH12_34048calBP_wolf
Pleistocene wolves
TU13_14120 calBP_wolf
TH15_14542 BP_wolf
Swiss14500_14500BP_ancientdog
IRK_5657calBP_wolf
AL2946_670BP_ancientdog
SOTN01_4900BP_ancientdog
OL4222_6500BP_ancientdog
Dog_Clade D
Ancient wolves
Modern wolves
ElMiron3_123
ElMiron19_127
ElMiron1_121
ElMiron7_127
Pleistocene wolves
Pleistocene wolves
ElMiron6_126
ElMiron14_122
SC19_29943calBP_wolf
AL2741_81793BP_wolf
WOL-VAL-033_63922BP_wolf
ElMiron21_129
VAL-012_56376BP_wolf
CGG18_41738calBP_wolf
Dab2_51102BP_wolf
JK175_33163calBP_wolf
DS-WOL-VAL-037_40801BP_wolf
LOW003_44450calBP_wolf
ElMiron2_122
Belgium_26000BP_wolf
TH1_26036calBP_wolf
ElMiron17_125
ElMiron5_125
ElMiron16_124
ElMiron15_123
ElMiron4_124
ElMiron34_128
AH577_37616BP_wolf
TU7_46348BP_wolf
TU6_38481_wolf
AH574_58002BP_wolf
TU2_33318BP_wolf
TU!_35000BP_wolf
Belgium_36000BP_wolf
TU3_32723BP_wolf
JK2174_31364BP_wolf
JK2183_32366BP_wolf
JAL65_89816BP_wolf
WOL-VAL-18A_100706BP_wolf
JAL48_89816BP_wolf
Russia_18000BP_wolf
ALAS-024_82674BP_wolf
LOW008_48900BP_wolf
AH575_57233BP_wolf
Belgium_33000BP_wolf
TH3_28834calBP_wolf
TU11_29957BP_wolf
TU9_31109BP_wolf
KF661096_Coyote
SC19.MCJ015
SC19.MCJ017
Cuon_alpinus
C_alpinus-ElMiron9_129
C_alpinus-ElMiron8_128
C_alpinus-ElMiron-127_7

Supplementary Figure 11: Maximum likelihood tree of the *Canis lupus mtDNA* genomes. Pink denotes the *Canis lupus* mtDNA genomes from El Mirón soil samples. Node numbers denote bootstrap values. The tree is rooted with dhole mtDNA sequences, including three from El Mirón. All the sequences from El Mirón are located within the Pleistocene *C. lupus* diversity of Eurasia.

3. Vulpes vulpes

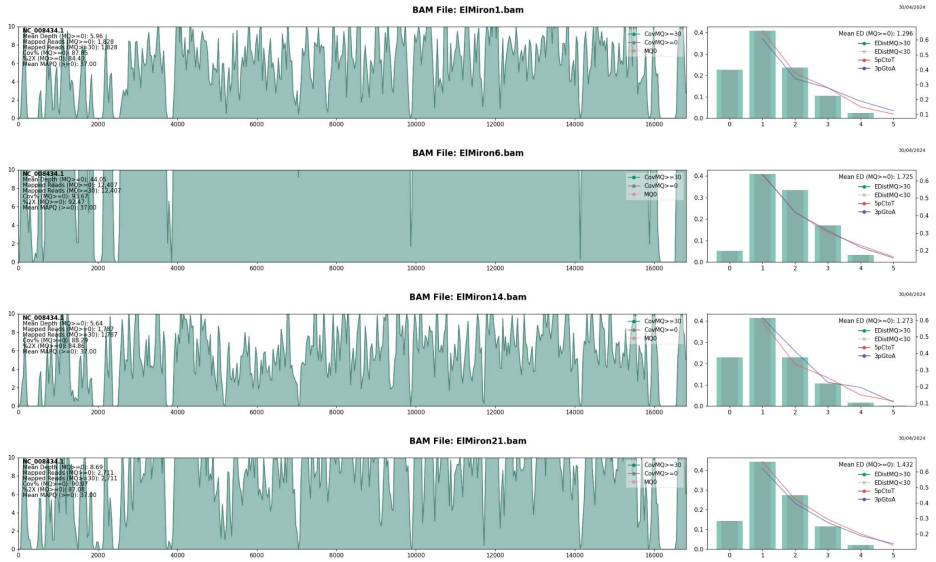
We analysed four mtDNA sedaDNA genomes: ElMiron_1, ElMiron_6, ElMiron_14 and El_Mirón21 (Supplementary Figures 12-14). We used the reference sequence NC_008434.1 to reconstruct the consensus sequences. The Mirón sedaDNA mtDNA sequences were aligned with other 12 *Vulpes spp.* sequences and 2 outgroup. One sequence is almost complete, only presenting 1,653 missing sites (Supplementary Figure 13, Supplementary Table 3).

The alignment, excluding outgroups, has 3280 variable sites. Within the four El Mirón samples, there are 39 variable sites.

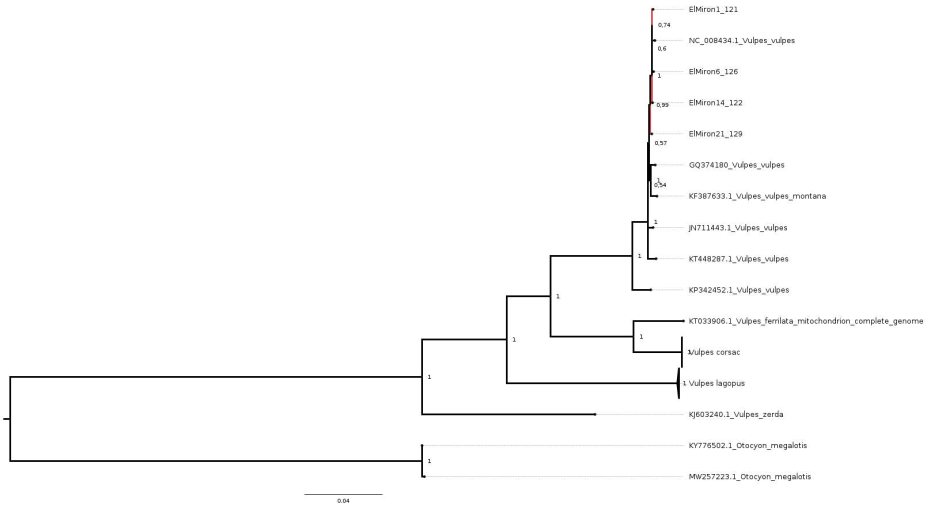


Supplementary Figure 12: Read length distribution of El Mirón *V. vulpes* sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue

depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



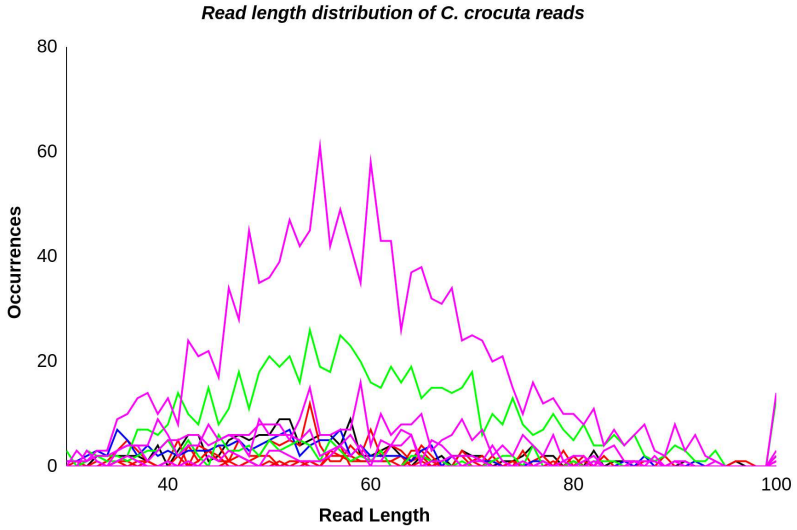
Supplementary Figure 13: Coverage plots and damage plots of *V. vulpes* sedaDNA mtDNA genomes.



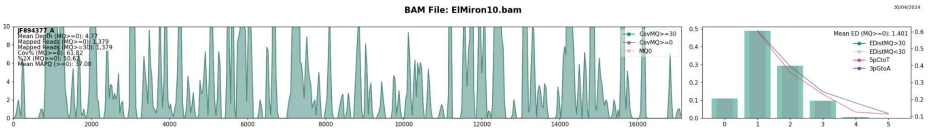
Supplementary Figure 14: Maximum likelihood tree of the *Vulpes vulpes* seda mtDNA genomes. The red colour denotes the sedaDNA sequences from El Mirón soil samples. Node numbers denote bootstrap values. The tree is rooted in *Otocyon megalotis*. The sequences from El Mirón are located within the diversity of *V. vulpes.*

4. Crocuta crocuta

We recovered a partial genome of *Crocuta crocuta* from sample El Miron_10 (Supplementary Figures 15-16). For reconstructing the consensus sequences as well as the individual alignments we used the reference sequence NC_020670.1. We compared the sequence with other 23 present day and ancient *Crocuta crocuta* diversity using a distance matrix.



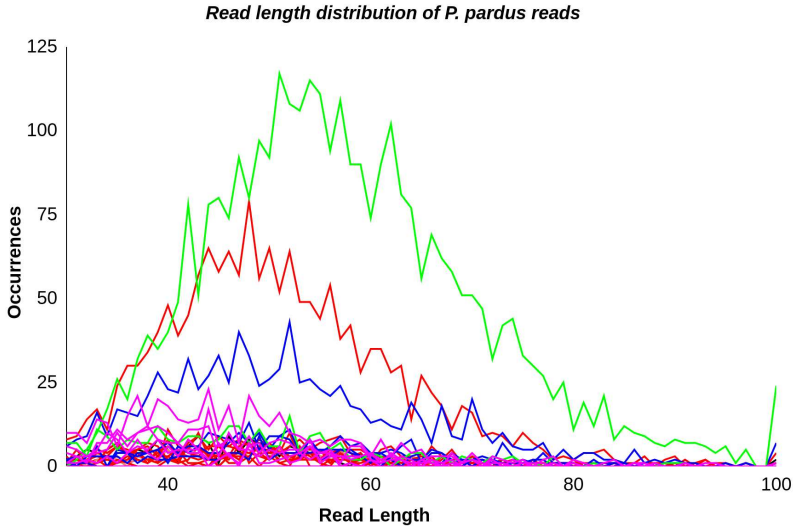
Supplementary Figure 15: Read length distribution of El Mirón *C. crocuta.* Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



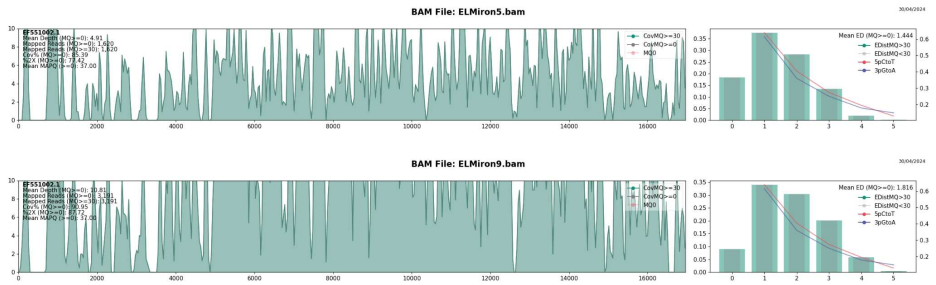
Supplementary Figure 16: Coverage and damage plot of *C. crocuta* mtDNA sedaDNA sequence.

5. *Panthera pardus*

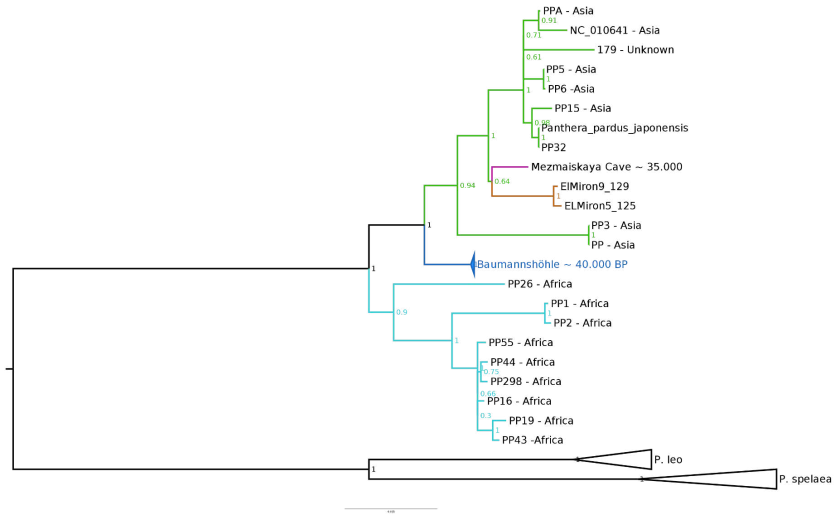
We analysed two mtDNA sequences ElMiron_5 and ElMiron_9 (Supplementary Figures 17-19). For reconstructing the consensus sequences as well as the individual alignments we used the reference sequence NC_010641.1. We aligned our 2 new reported sequences together with 25 present-day and ancient *P. pardus* mtDNA genomes *and* 5 other *Panthera* genus sequences. The two sequences from El Mirón show 21 variable sites across the sequence.



Supplementary Figure 17: Read length distribution of El Mirón *P. pardus* mtDNA sedaDNA sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



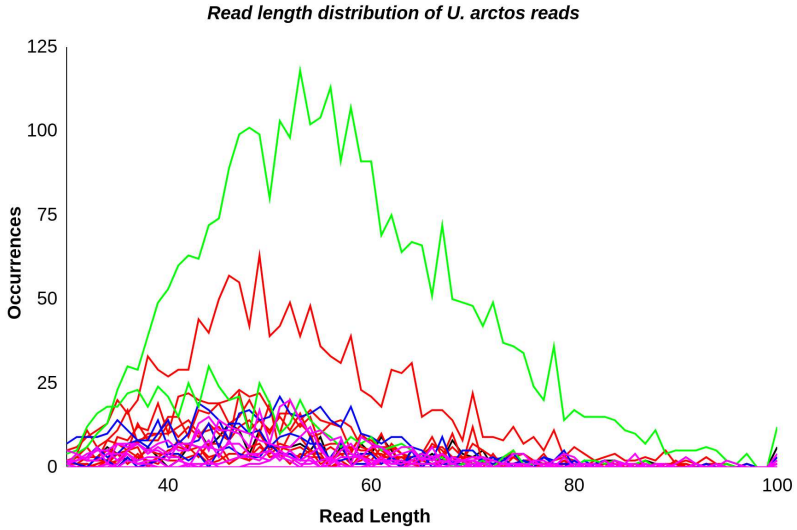
Supplementary Figure 18: Coverage and damage plots of *P. pardus* sedaDNA mtDNA sequences.



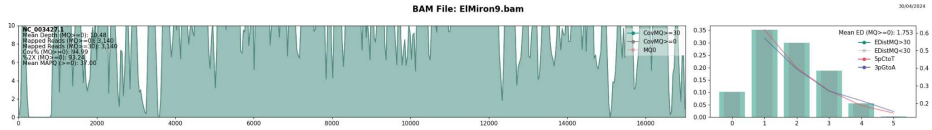
Supplementary Figure 19: Maximum likelihood tree of *Panthera* mtDNA genomes: A genomic similarity between El Mirón leopards and Caucasus *P. pardus* BAR001 from the Pleistocene is observed. Light blue corresponds to African leopards, Blue to Pleistocene European leopards, Purple for caucasian Pleistocene leopards and green for Asian leopards. El Mirón leopards are in brown.

6. Ursus arctos

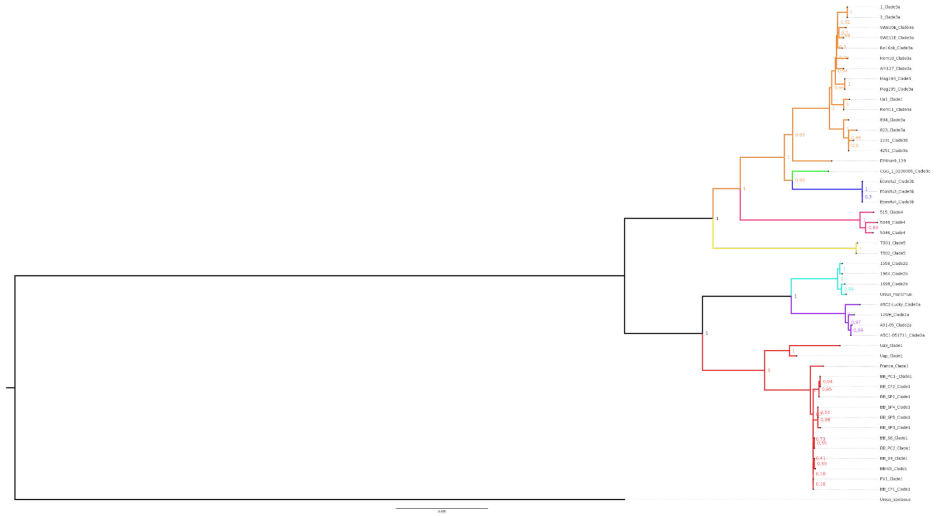
We analysed one mtDNA partial genome, ElMiron_9 (Supplementary Figure 20-22). For reconstructing the consensus sequences and individual alignments, we used the reference sequence NC_003427.1. The sequence is almost fully covered, showing the similarity of *Ursus arctos* from El Mirón and the reference sequence (Supplementary Figure 21).



Supplementary Figure 20: Read length distribution of El Mirón *U. arctos* sequences. Black depicts the Initial Magdalenian 119.2 level, Red depicts the Solutrean 121-127 levels, Blue depicts the Gravettian 128 level, Green depicts the semi-sterile 129 level, and Purple depicts the Mousterian 130 level.



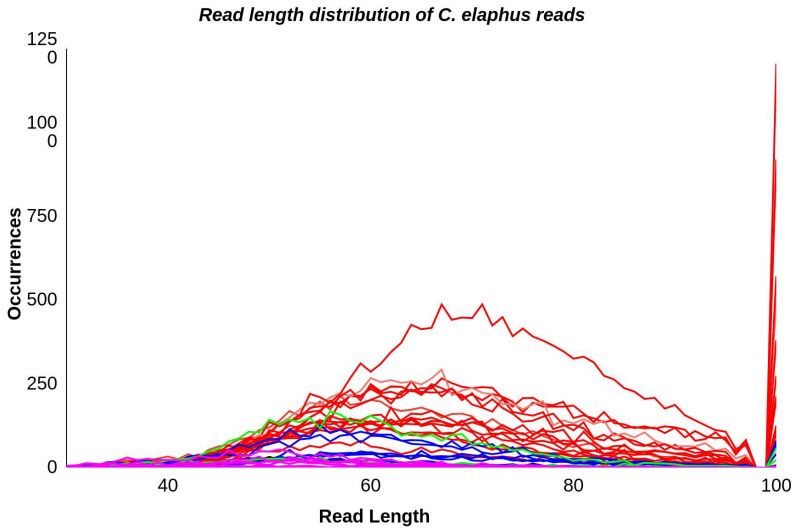
Supplementary Figure 21: *U. arctos sedaDNA mtDNA genomes coverage plots and damage plots.*



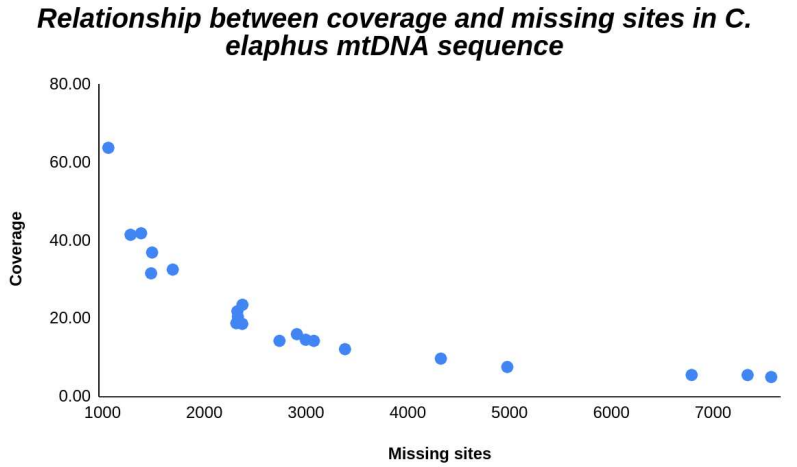
Supplementary Figure 22: Maximum likelihood tree of the *Ursidae* mtDNA genomes: The tree is rooted in *the U. spelaeus mtDNA genome. Colours denote clades. Orange denotes clade 3a, Dark blue clade 3b, Green clade 3C, Pink clade 4, Light blue clade 2b, Purple clade 2a, and red clade 1. ElMiron_9 from level 129 is placed in clade 3a, separated from other Pleistocene bears from Iberia,* which are placed in clade 1.

7. Cervus elaphus

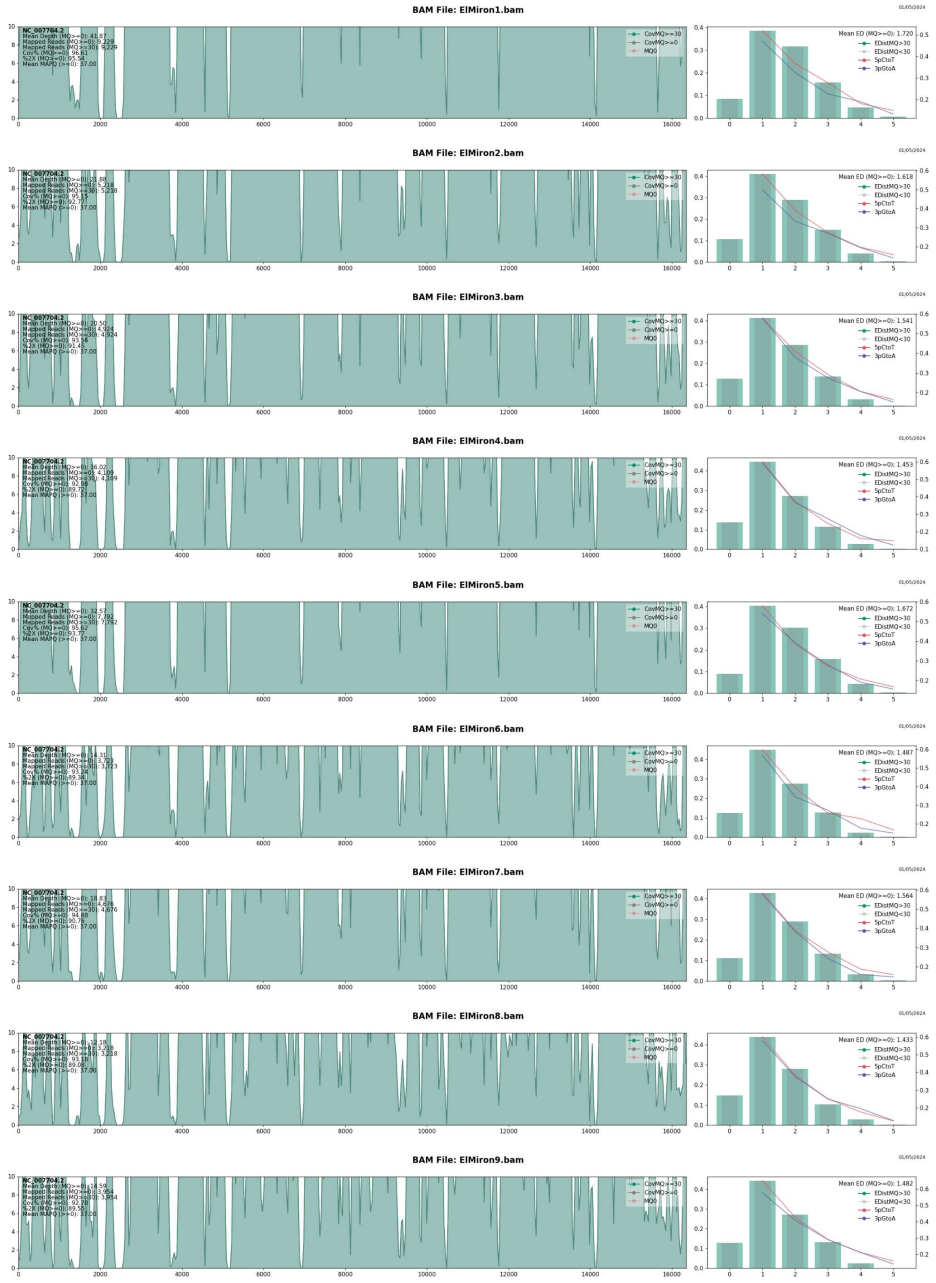
We analysed 21 mtDNA partial genomes of *C. elaphus,* including: ElMiron_1, ElMiron_2, ElMiron_3, ElMiron_4, ElMiron_5, ElMiron_6, ElMiron_7, ElMiron_8, ElMiron_9, ElMiron_13, ElMiron_14, ElMiron_15, ElMiron_16, ElMiron_17, ElMiron_18, ElMiron_18 extract 2, ElMiron_19, ElMiron_20, ElMiron_28, ElMiron_30, ElMiron_34 (Supplementary Figures 23-26). For reconstructing the consensus sequences as well as the individual alignments we used the reference sequence NC_007704.2. Many sequences show the presence of limited missing sites, but others show elevated numbers (Supplementary Table 3). As *C. elaphus* is one of the taxa with more genomes, we have tested the presence of a relationship between the coverage and the missing sites. We observe that with higher coverage, the amount of missing sites is reduced (Supplementary Figure 24), in a relationship that is nonlinear. We tested the significance of the correlation with Sperman's tests, and the relationship is highly significant rho=-0.972 and *p-value*=$1.8e^{-10}$. The 21 sequences show 139 variable sites when aligned.



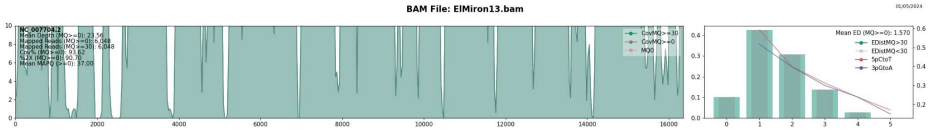
Supplementary Figure 23: Read length distribution of El Mirón *C. elaphus* sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



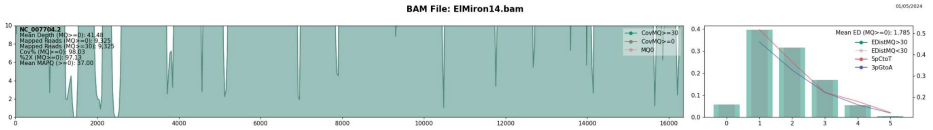
Supplementary Figure 24: Relationship between missing sites and coverage in the reads attributed to *C. elaphus* from El Mirón.



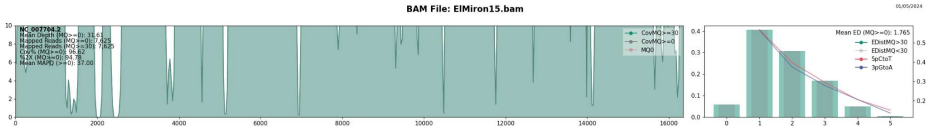
BAM File: ElMiron13.bam



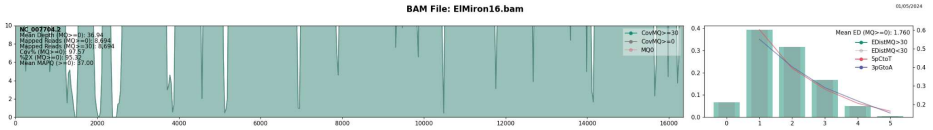
BAM File: ElMiron14.bam



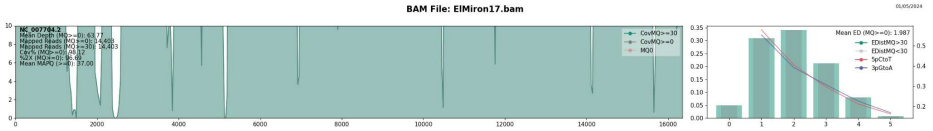
BAM File: ElMiron15.bam



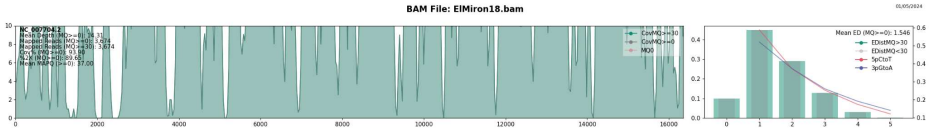
BAM File: ElMiron16.bam



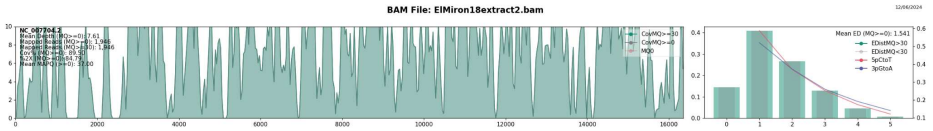
BAM File: ElMiron17.bam



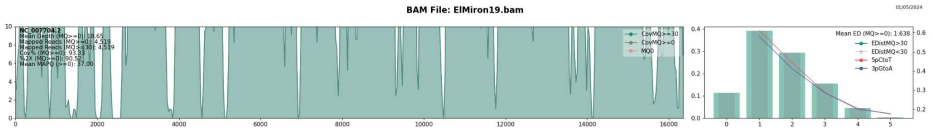
BAM File: ElMiron18.bam



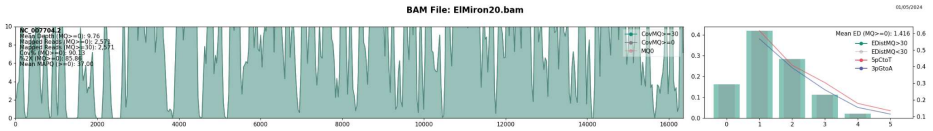
BAM File: ElMiron18extract2.bam

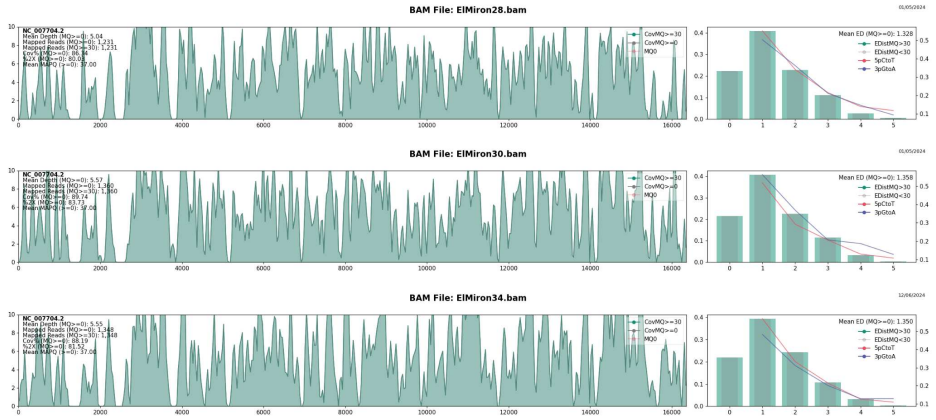


BAM File: ElMiron19.bam

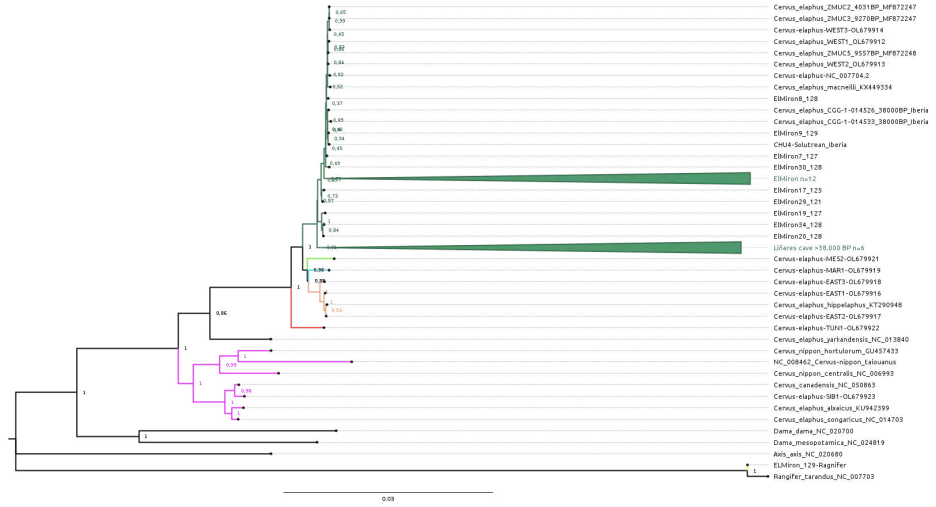


BAM File: ElMiron20.bam





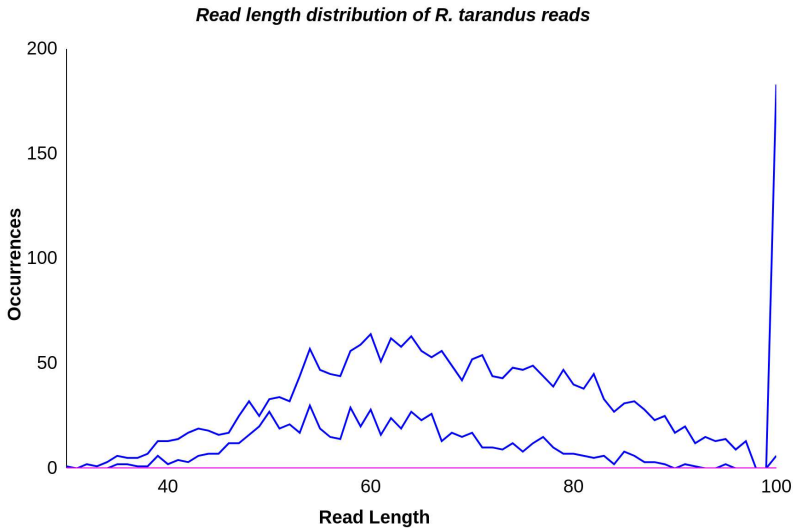
Supplementary Figure 25: Coverage plots and damage plots of *C. elaphus sedaDNA* mtDNA genomes.



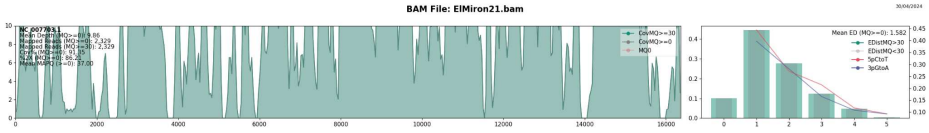
Supplementary Figure 26: Maximum likelihood tree of the Cervidae mtDNA genomes. The tree is rooted with *R. tarandus* mtDNA genomes. Colours denote clades. Dark green represents Western Clade A; pink represents the Eastern *Cervus* diversity; red represents Western Clade B; orange Western Clade C; light green Western Clade D, and light blue Western Clade E according to [129]. All the sequences from El Mirón are related to the rest of Clade A's diversity. It also shows the genetic proximity of El Mirón deer mtDNA sequences to the Pleistocene sequences from Liñares cave in Spain and the Solutrean bone tool from Chufín. Numbers in the nodes depict bootstrap values.

8. Rangifer tarandus

We analysed one mtDNA sedaDNA genome of *Rangifer tarandus* corresponding to ElMiron_21 (Supplementary Figure 27-28). We used the reference sequence NC_007703.1 to reconstruct the consensus sequence and individual alignments. The sequence is included in the *Cervidae* phylogeny (Supplementary Figure 26).



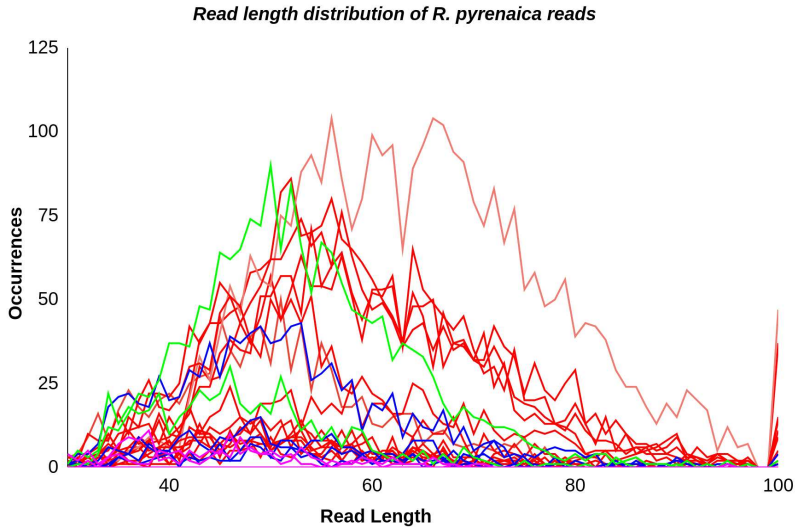
Supplementary Figure 27: Read length distribution of El Mirón *R. tarandus* mtDNA sedaDNA sequences. Black depicts the Initial Magdalenian 119.2 level, Red depicts the Solutrean 121-127 levels, Blue depicts the Gravettian 128 level, Green depicts the semi-sterile 129 level, and Purple depicts the Mousterian 130 level.



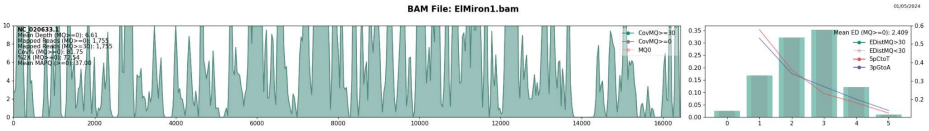
Supplementary Figure 28: Coverage and damage plots of *R. tarandus* sedaDNA mtDNA genome.

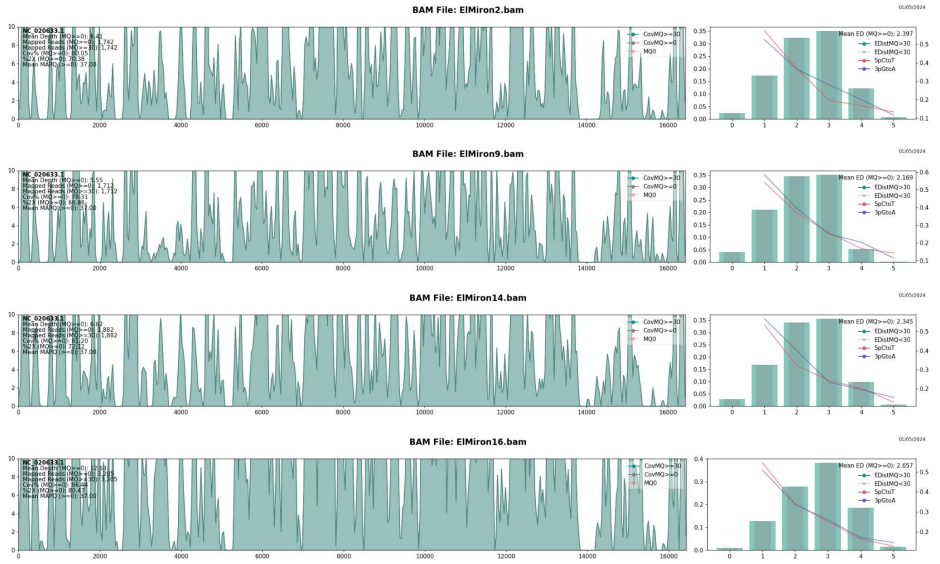
9. Rupicapra pyrenaica

We recovered five partial mtDNA genomes: ElMiron_1, ElMiron_2, ElMiron_14, ElMiron_7, and ElMiron_16 (Supplementary Figures 29-30). We used the reference sequence NC_0206331 to reconstruct the consensus sequences, we aligned the sequences from El Mirón with 35 other samples. When aligned, the five sequences from El Mirón show 37 variable sites.



Supplementary Figure 29: Read length distribution of El Mirón *R. pyrenaica* genomes. Black depicts the Initial Magdalenian 119.2 level, Red depicts the Solutrean 121-127 levels, Blue depicts the Gravettian 128 level, Green depicts the semi-sterile 129 level, and Purple depicts the Mousterian 130 level.



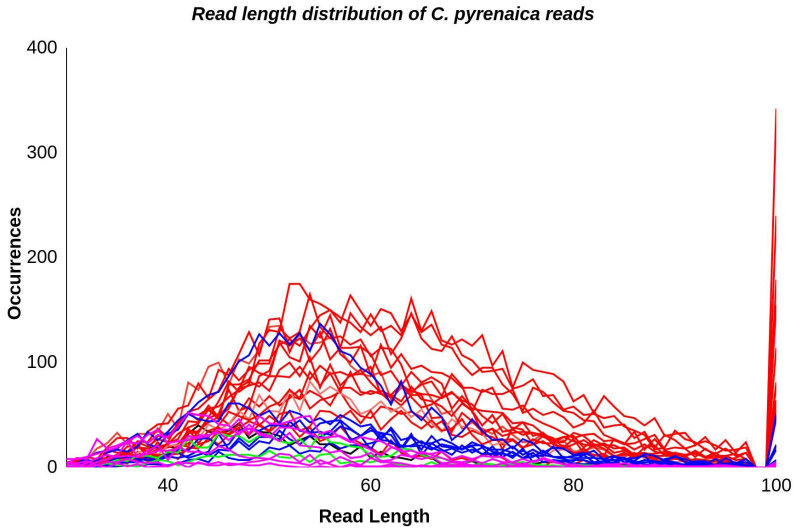


Supplementary Figure 30: read length distribution of El Mirón *R. pyrenaica* sedaDNA sequences.

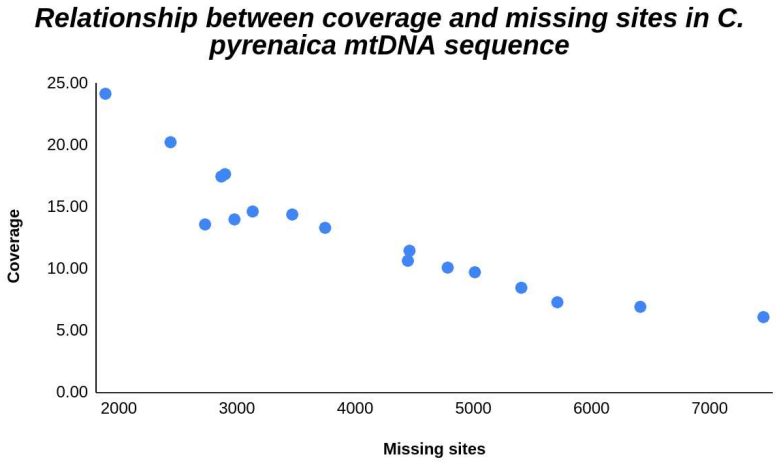
10. Capra pyrenaica

We analysed 17 mtDNA sequences ElMiron_2, ElMiron_4, ElMiron_5, ElMiron_6, ElMiron_7, ElMiron_8, ElMiron_14, ElMiron_15 ,ElMiron_16, ElMiron_17, ElMiron_18, ElMiron_18 extracat 2, ElMiron_19, ElMiron_20, ElMiron_30, ElMiron_32, ElMiron_34 (Supplementary Figures 31-34). For reconstructing the consensus sequences and the individual alignments we used the reference sequence NC_020623.1 (alpine chamoix *Capra ibex)*. These sequences were aligned with the other 16 *Caprinae* sequences. The 17 sequences from El Mirón show 297 variable sites when aligned. As *C. pyrenaica* is one of the taxa with more genomes, we have tested the presence of a relationship between the coverage and the missing sites. We observe that with higher coverage, the amount of missing sites decreases (Supplementary Figure 32); in a relationship that is not lineal, we tested the significance of the correlation with Sperman's tests and the relationship is highly significant rho=-0.94 and *p-value*=4.79e−9. We can observe that there are regions of the genome not covered, which could indicate that the Pleistocene chamois present in Cantabria was differentiated from the reference used in this current analysis or that the capture baits designed with *Capra hircus* did not capture these regions. Despite these gaps, we are confident that these sequences resemble *C. pyrenaica* based on the

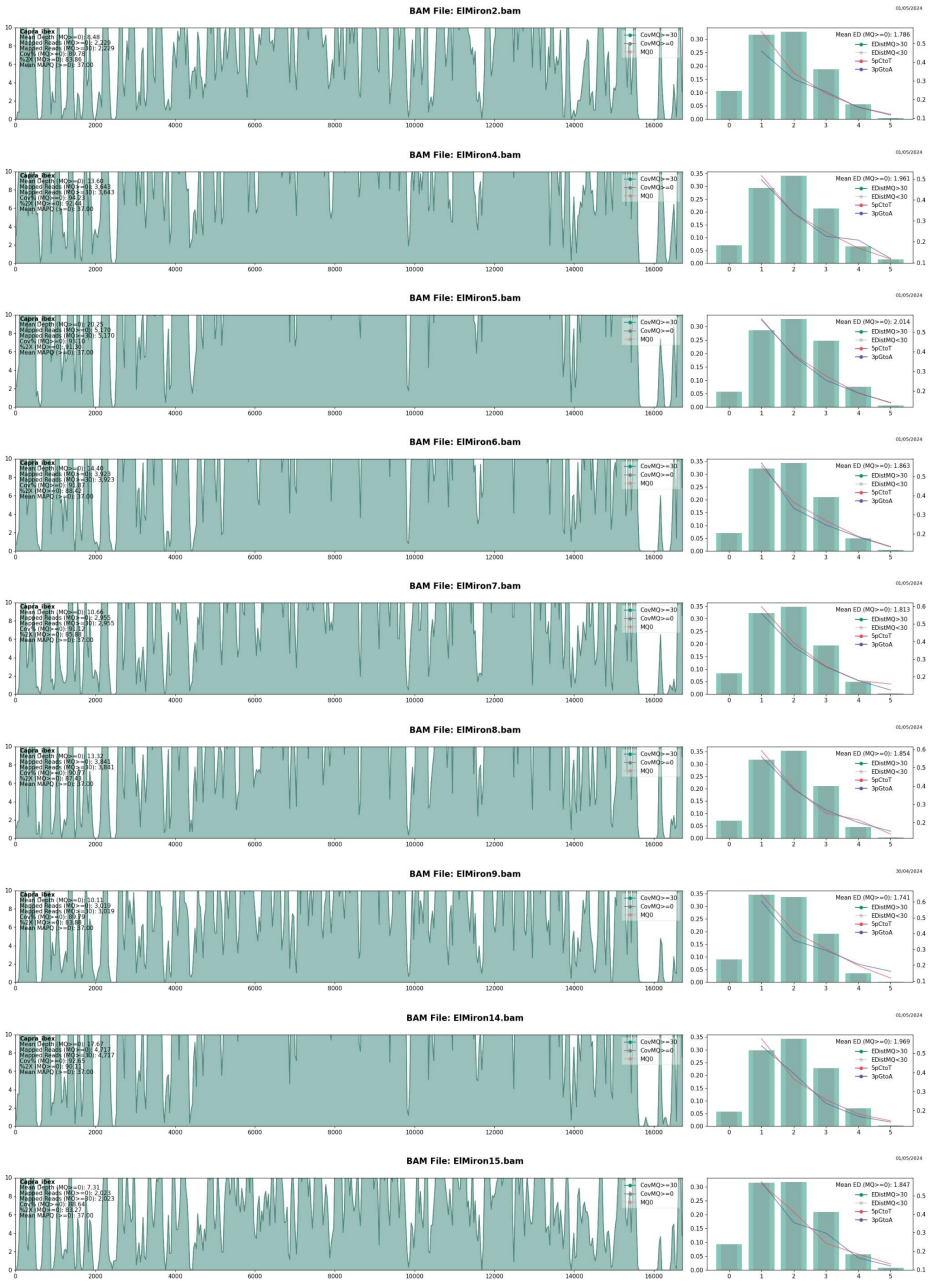
unequivocally placing in the phylogeny (Supplementary Figure 33). The 21 sequences show 139 variable sites when aligned.



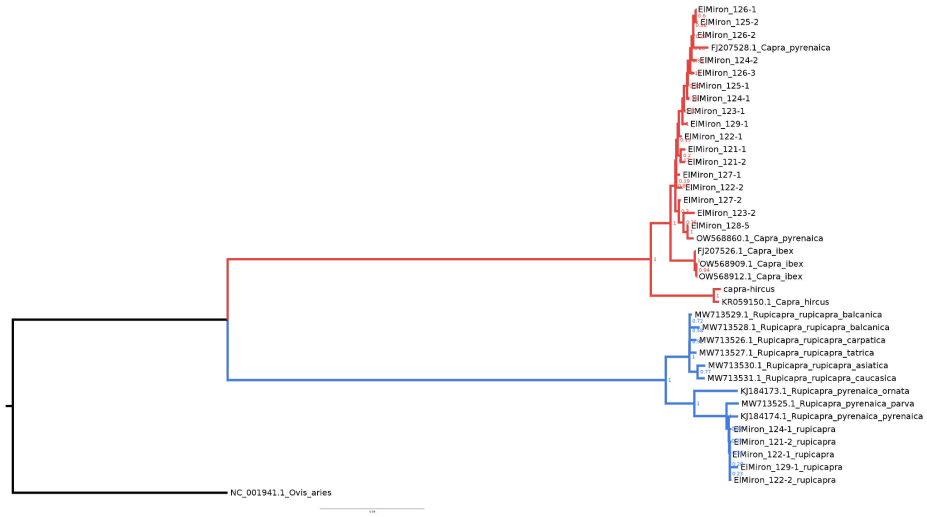
Supplementary Figure 31: Read length distribution of El Mirón *C. pyrenaica* sequences. Black depicts Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts Gravettian 128 level, Green depicts semi-sterile 129 level, Purple depicts Mousterian 130 level.



Supplementary Figure 32: Relationship between missing sites and coverage in the reads attributed to *C. pyrenaica* from El Mirón.



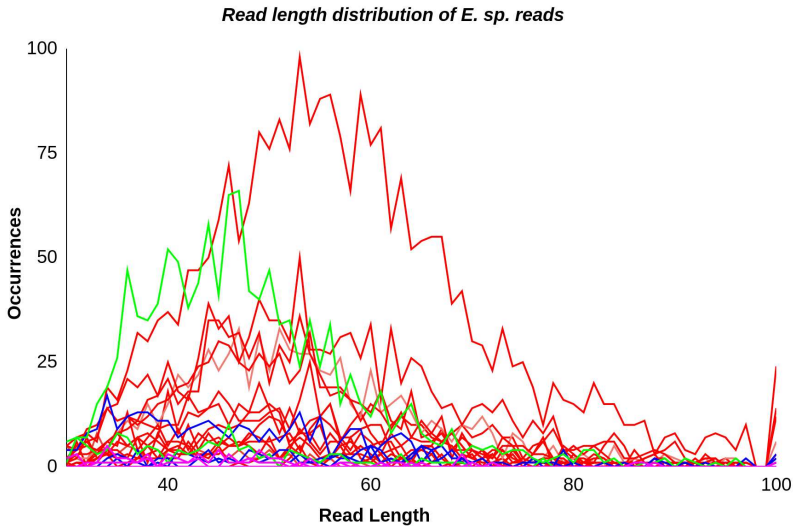
Supplementary Figure 33: Coverage plots and damage plots of *C. pyrenaica* sedaDNA genomes.



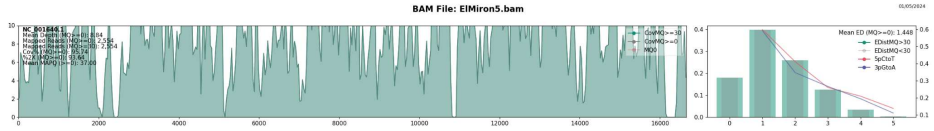
Supplementary Figure 34: Maximum likelihood tree of the *Caprinae mtDNA* genomes. The red colour denotes the *Capra* mtDNA genomes, and the blue ones represent the *Rupicapra* mtDNA genomes. The phylogeny has been rooted with *O. aries* mtDNA sequence. The Phylogeny confirms that the mtDNA sequences from El Mirón belong to *C. pyrenaica and R. pyrenaica.* Node numbers denote bootstrap values.

11. Equus sp.

We analysed one mtDNA partial genome ElMiron_9 (Supplementary Figures 35-37). For reconstructing the consensus sequences as well as the individual alignments we used the reference sequence NC_001640.1. We aligned the sequence from ElMiron_9 sample with other 106 ancient and modern horse sequences plus an outgroup.



Supplementary Figure 35: read length distribution of El Mirón *E.* sp. sequences.



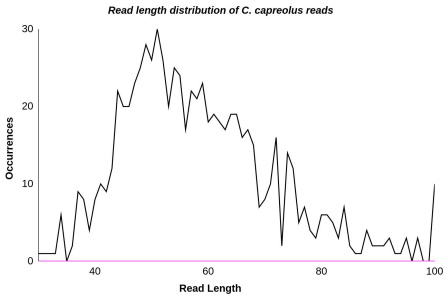
Supplementary Figure 36: read length distribution of El Mirón *Equus.* sp. sedaDNA mtDNA genomes.

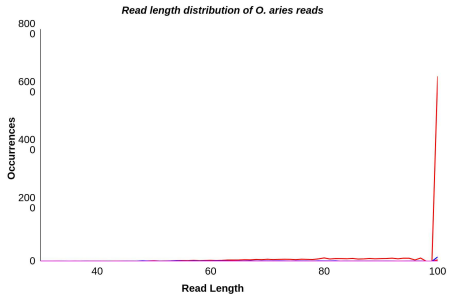
Supplementary Figure 37: Maximum likelihood tree of the *Equus* sp. *mtDNA* genomes. Purple denotes the Mirón5_125 mtDNA sedaDNA sequence. This sequence appears within the present-day and ancient Eurasian diversity as a separate clade.

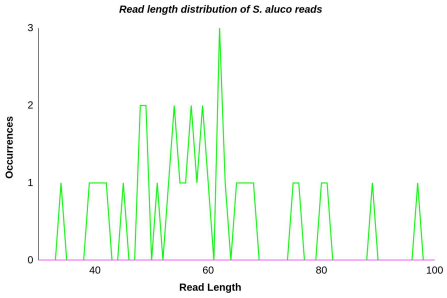
We have not analysed the individual genomes for the rest of the species we identified in Supplementary Section 3. Still, we have aligned the reads against the reference sequence (Supplementary Table 18), and we present the read length distribution in Supplementary Figure 38. The deamination values are present in Supplementary Table 3.

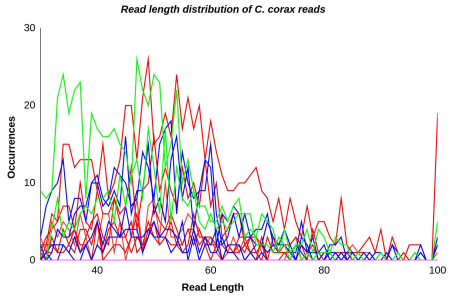
Supplementary Table 18: *Reference mtDNA genomes used for the species not described before*

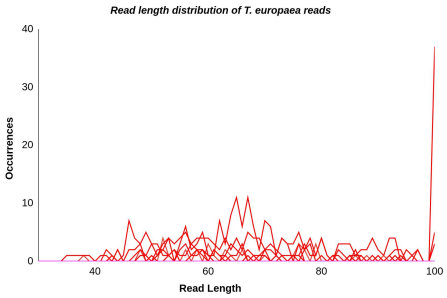
Taxa	Reference used
C. capreolus	NC_020684.1
O. aries	NC_001941.1
S. aluco	NC_072569.1
C. corax	NC_034838.1
T. europaea	NC_002391.1
C. livia	NC_013978.1
S. scrofa	NC_000845.1
M. arvalis	NC_038176.1
F. catus	NC_001700.1
B. bonasus	NC_006853.1
B. primigenius	NC_013996.1
L. europaeus	NC_004028.1
S. araneus	NC_027963.1
P. graculus	NC_025927.1
F. tinnunculus	NC_011307.1
M. nivalis	NC_020639.1
L. pardinus	NC_028319.1
C. antiquitatis	FJ905813.1
M. primigenius	NC_007596.2

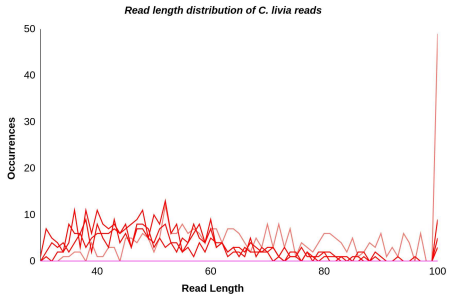

Read length distribution of C. capreolus reads
Occurrences
Read Length

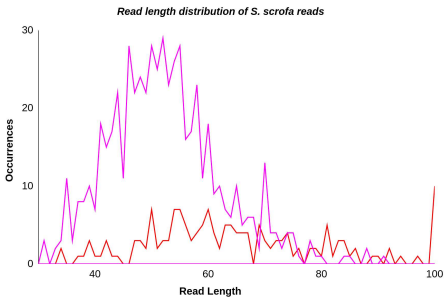

Read length distribution of O. aries reads
Occurrences
Read Length

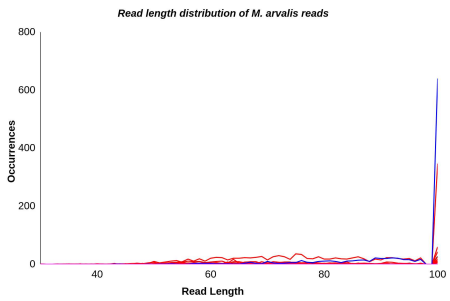

Read length distribution of S. aluco reads
Occurrences
Read Length

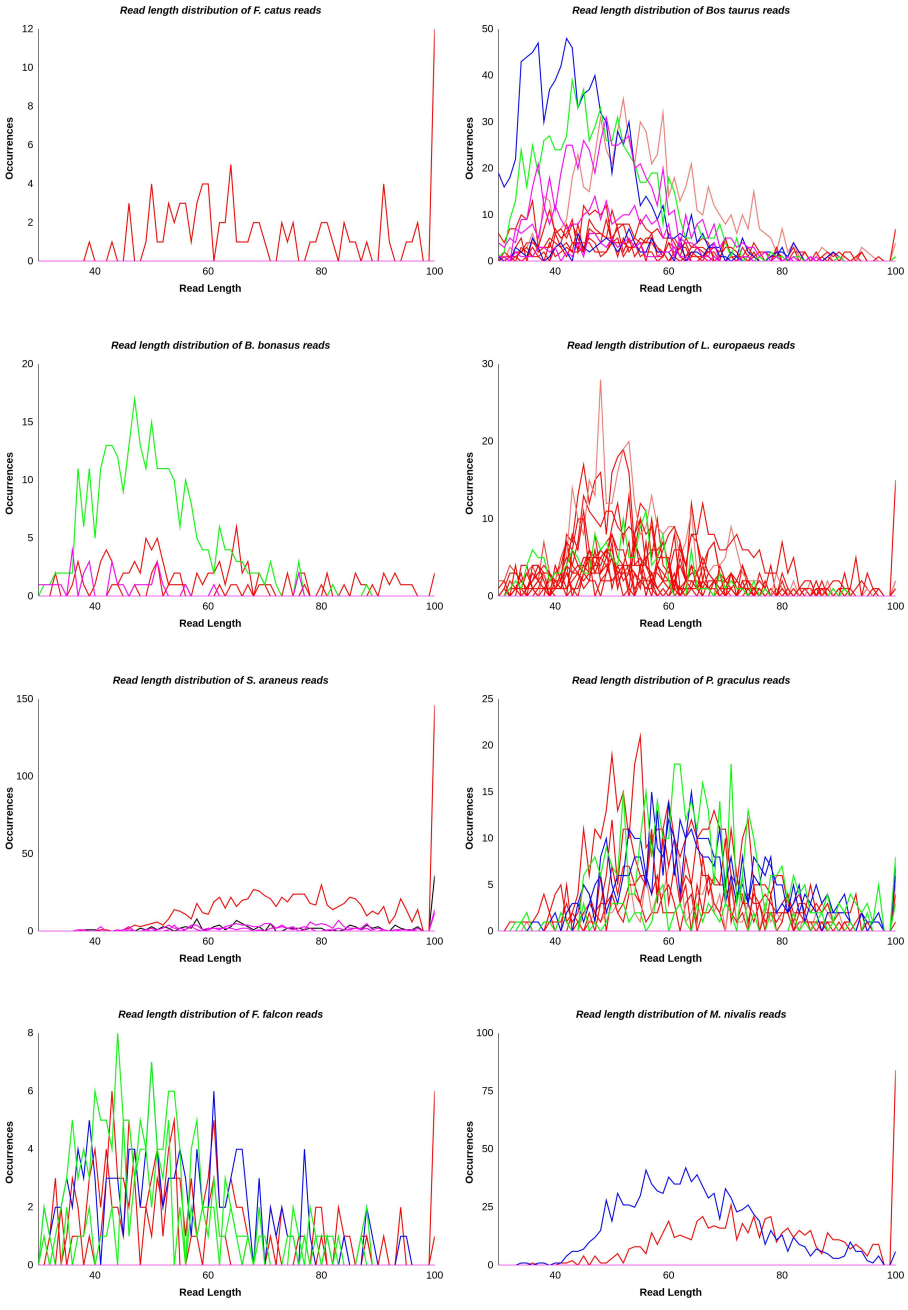

Read length distribution of C. corax reads
Occurrences
Read Length

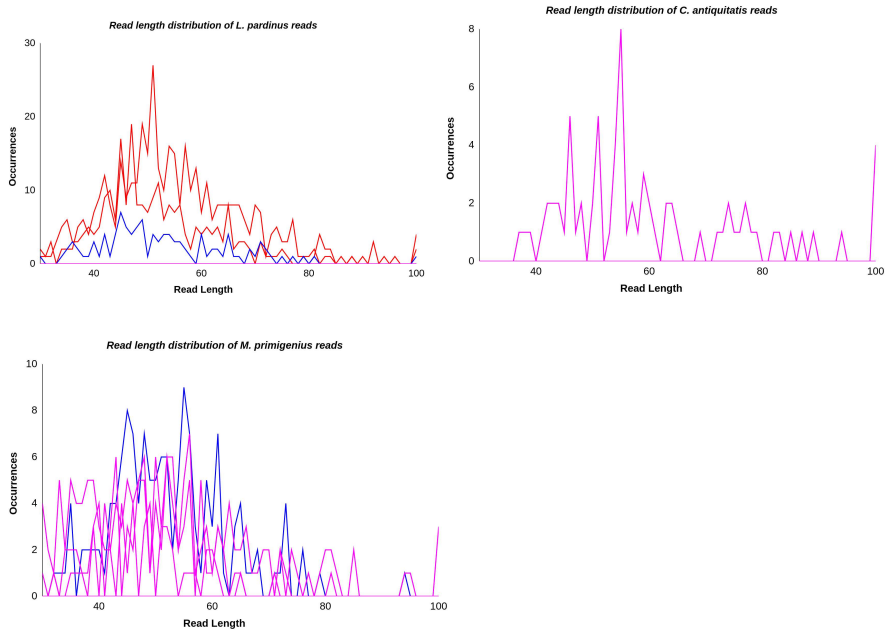

Read length distribution of T. europaea reads
Occurrences
Read Length


Read length distribution of C. livia reads
Occurrences
Read Length


Read length distribution of S. scrofa reads
Occurrences
Read Length


Read length distribution of M. arvalis reads
Occurrences
Read Length


Read length distribution of F. catus reads
Occurrences
Read Length
Read length distribution of Bos taurus reads
Occurrences
Read Length
Read length distribution of B. bonasus reads
Occurrences
Read Length
Read length distribution of L. europaeus reads
Occurrences
Read Length
Read length distribution of S. araneus reads
Occurrences
Read Length
Read length distribution of P. graculus reads
Occurrences
Read Length
Read length distribution of F. falcon reads
Occurrences
Read Length
Read length distribution of M. nivalis reads
Occurrences
Read Length

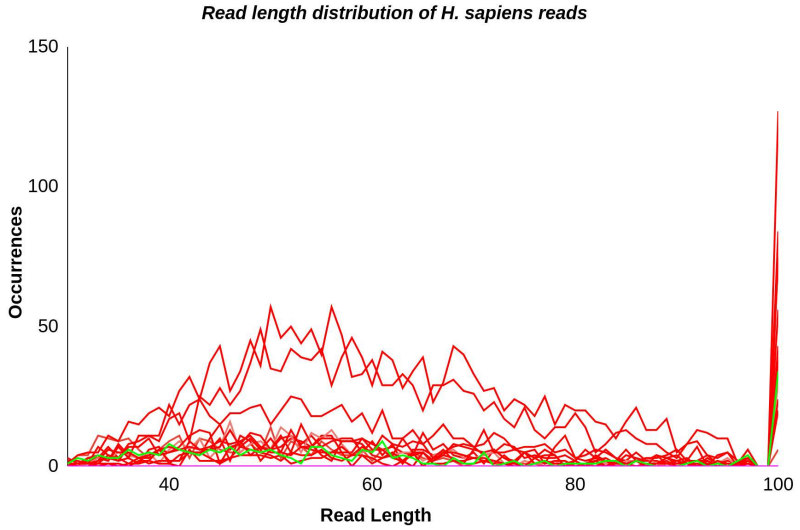


Supplementary Figure 38: Read length distribution of the taxa that have been individually aligned. Black depicts the Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts the Gravettian 128 level, Green depicts the semi-sterile 129 level, and Purple depicts the Mousterian 130 level.

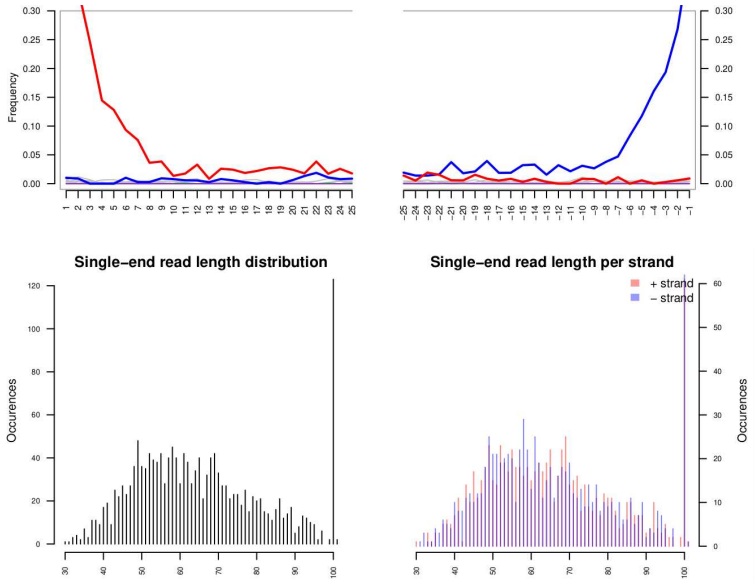
12. Human analyses

Human DNA reads were analysed with mapDamage2-2.2.2 to check for deamination as we did for the rest of analyzed taxa. We later calculated the amount of contamination and determined the presence of multiple sequences using Schmutzi 2.0 [156] (with and without prediction of contamination) and Calico [157]. Calico does not distinguish between modern or ancient reads but enables us to determine the proportions of the sequence donors, assessing the presence of multiple individual sequences mixed in our reads. We later generated the mtDNA sedaDNA genome consensus sequences with Schmutzi and checked the positions individually using IGV 2.16.1 genome viewer. We also determined the mtDNA haplogroup with haplogrep 3.0 [158]. We aligned the resulting sequences with a dataset comprising modern and ancient sequences (Supplementary Table 11) using MAFFT 7.52 [56]. We produced Maximum Likelihood trees using MEGA 10.2.4 with a partial deletion of 95%, GTR substitution model and 500 bootstrap replicates[57].

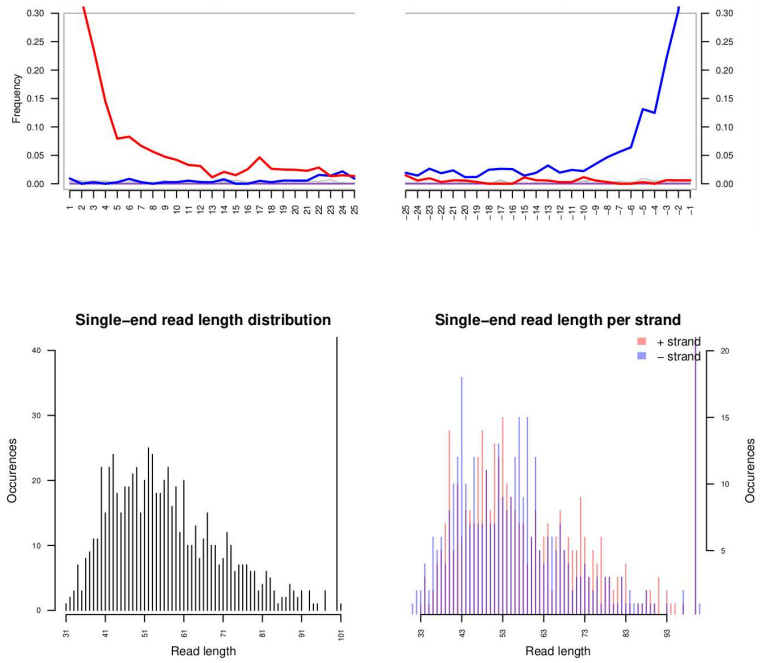
We detected the presence of human reads in all samples but only 10 with deamination values greater than 30%, which we consider the ones showing real Pleistocene DNA. Only ElMiron_1, ElMiron_14 and ElMiron_18 showed enough reads to assess the relationship of the mtDNA sequences with other ancient and modern ones. None of these samples corresponds to level 130, not detecting the presence of human DNA linked to Mousterian archaeological culture in El Mirón.



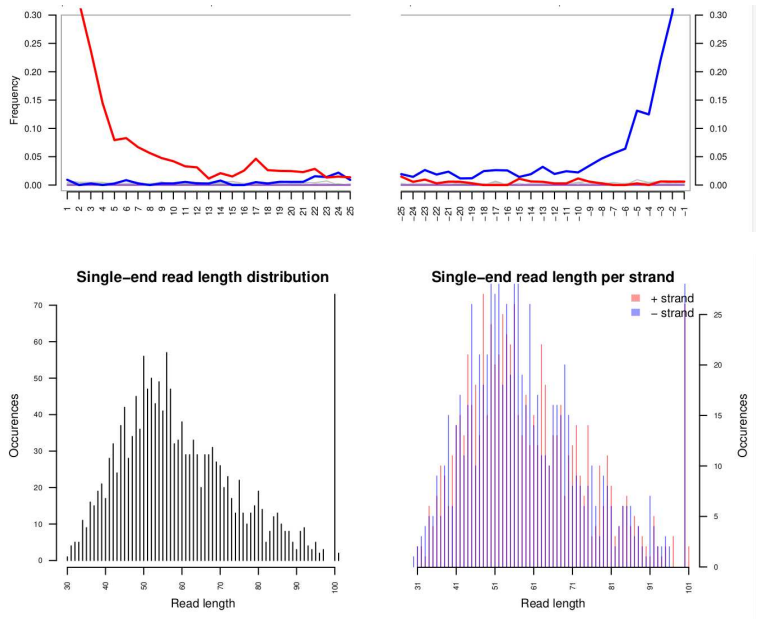
Supplementary Figure 39: read length distribution of El Mirón *H. sapiens* sequences. Black depicts the Initial Magdalenian 119.2 level, Red depicts Solutrean 121-127 levels, Blue depicts the Gravettian 128 level, Green depicts the semi-sterile 129 level, and Purple depicts the Mousterian 130 level.



Supplementary Figure 40: Read length distribution and deamination pattern of El Miron18 (Level 126) sample human reads.



Supplementary Figure 41: Read length distribution and deamination pattern of El Miron1 (Level 121) sample human reads.



Supplementary Figure 42: Read length distribution and deamination pattern of El Miron14 (Level 122) sample human reads.

Supplementary Section 5: Supplementary References

1. Marín-Arroyo, A. B. *et al.* Seasonality of Human Occupations in El Mirón Cave: Late Upper Paleolithic Hunter-Gatherer Settlement-Subsistence Systems in Cantabrian Spain. *Journal of Paleolithic Archaeology* **6**, 7 (2023).
2. Cuenca-Bescós, G., Marín-Arroyo, A. B. & Martínez, I. Relationship between Magdalenian subsistence and environmental change: The mammalian evidence from El Mirón (Spain). *Quaternary* (2012).
3. Hopkins, R. J. A., Straus, L. G. & González Morales, M. R. Assessing the chronostratigraphy of el Mirón cave, Cantabrian Spain. *Radiocarbon* **63**, 821–852 (2021).
4. Straus, L. G. & González Morales, M. R. *El Mirón Cave, Cantabrian Spain: The Site and Its Holocene Archaeological Record*. (University of New Mexico Press, 2012).
5. Straus, L. G. & González Morales, M. R. New Dates for the Solutrean and Magdalenian of Cantabrian Spain: El Miron and La Riera Caves. *Radiocarbon* **60**, 1013–1016 (2018).
6. Straus, L. G. & González Morales, M. R. The Upper Paleolithic sequence in el Mirón Cave (Ramales de la Victoria, Cantabria, Spain): An overview. *Journal of Archaeological Science: Reports* **27**, 101998 (2019).
7. Straus, L. G., González Morales, M. R., Carretero, J. M. & Marín-Arroyo, A. B. 'The Red Lady of El Mirón'. Lower Magdalenian life and death in Oldest Dryas Cantabrian Spain: an overview. *J. Archaeol. Sci.* **60**, 134–137 (2015).
8. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
9. Posth, C. *et al.* Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature* **615**, 117–126 (2023).
10. Marín-Arroyo, A. B. *et al.* The Middle to Upper Palaeolithic transition at El Mirón

Cave (Cantabria, Spain). *Quat. Int.* **544**, 23–31 (2020).

11. Farrand, W. Sedimentology of el Mirón cave. in *El Mirón Cave, Cantabrian Spain: The Site and Its Holocene Archaeological Record* (eds. Straus, L. G. & González Morales, M. R.) 60–94 (University of New Mexico Press, 2012).
12. Straus, L. G., González Morales, M., Farrand, W. R. & Hubbard, W. J. Sedimentological and stratigraphic observations in El Mirón, a late Quaternary cave site in the Cantabrian Cordillera, Northern Spain. *Geoarchaeology* **16**, 603–630 (2001).
13. Geiling, J. M. Human ecodynamics in the Late Upper Pleistocene of Northern Spain: an archeozoological study of ungulate remains from the Lower Magdalenian and other periods …. (University of Cantabria, 2020).
14. Sanguino, J. & Montes, R. *La cueva del 'El Pendo' : actuaciones arqueológicas 1994-2000*. (Consejería de Cultura, Educación y Deporte, 2001).
15. Montes, R., Sanguino, J., Martín, P., Gómez, A. J. & Morcillo, C. La secuencia estratigráfica de la cueva de El Pendo (Escobedo de Camargo, Cantabria): problemas geoarqueológicos de un referente cronocultural. *Geoarqueología y patrimonio en la Península Ibérica y el entorno mediterráneo, ADEMA, Almazán (Soria)* 127–138 (2005).
16. Kehl, M. *et al.* Towards a revised stratigraphy for the Middle to Upper Palaeolithic boundary at La Güelga (Narciandi, Asturias, Spain). Soil micromorphology and new radiocarbon data. *Bol. geol. min.* **129**, 183–206 (2018).
17. González-Morales, M. & Straus, L. G. La ocupación gravetiense de la cueva de El Mirón (Ramales de la Victoria, Cantabria) y el contexto del arte paleolítico temprano de la cuenca del Asón. in *Pensando el Gravetiense: nuevos datos para la Región Cantábrica, en su contexto peninsular y pirenaico* (eds. de las Heras Martín, C., Lasheras Corruchaga, J. A., Arrizabalaga Valbuena, A. & de la Rasilla Vives, M.)

289–299 (Ministerio de CulturaEditor: Monografías del Museo y Centro de Investigación de Altamira, 2012).

18. Straus, L. G. & González Morales, M. R. *El Mirón Cave, Cantabrian Spain: The Site and Its Holocene Archaeological Record*. (University of New Mexico Press, 2012).
19. Marín-Arroyo, A. B. *Arqueozoología En El Cantábrico Oriental Durante La Transición Pleistoceno/holoceno: La Cueva Del Mirón*. (PUbliCan, Ediciones de la Universidad de Cantabria, 2010).
20. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).
21. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
22. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
23. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
24. Tejero, J.-M. *et al.* A minimally-invasive method for ancient DNA sampling of Prehistoric bone and antler tools and hunting weapons. *bioRxiv* 2023.04.02.535282 (2023) doi:10.1101/2023.04.02.535282.
25. Rohland, N. *et al.* Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
26. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
27. Myers, E. W. The fragment assembly string graph. *Bioinformatics* **1;21 Suppl 2**,

79–85 (2005).

28. Hannon, G. J. *FASTX-Toolkit*. (2010).
29. 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.* (2023) doi:10.1111/2041-210x.14214.
30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
31. 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).
32. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Picard-tools. http://broadinstitute.github.io/picard.
35. 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).
36. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
37. Team, R. S. RStudio: integrated development for R. RStudio. *Inc., Boston, MA* (2015).
38. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).
39. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

40. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
41. Danecek, P., McCarthy, S. & Li, H. bcftools—utilities for variant calling and manipulating vcfs and bcfs. *The MIT/Expat License or GPL License, see the* (2015).
42. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
43. Ripoll, M. P., Morales Pérez, J. V., Sanchis Serra, A., Aura Tortosa, J. E. & Montañana, I. S. Presence of the genus Cuon in upper Pleistocene and initial Holocene sites of the Iberian Peninsula: new remains identified in archaeological contexts of the Mediterranean region. *J. Archaeol. Sci.* **37**, 437–450 (2010).
44. Lloveras, L. *et al.* The role of birds in Upper Palaeolithic sites: Zooarchaeological and taphonomic analysis of the avian remains from Arbreda Cave (Serinyà, northeast Iberia). *Quat. Int.* **626-627**, 22–32 (2022).
45. Nabais, M., Pimenta, C. & Zilhão, J. Human-bird interaction in last Interglacial Iberia: A combined approach using skeletal part analysis, bone surface modification, bird ethology and ethnography. *Journal of Archaeological Science: Reports* **49**, 104023 (2023).
46. Sommer, R. S. & Benecke, N. Late Pleistocene and Holocene development of the felid fauna (Felidae) of Europe: a review. *J. Zool.* **269**, 7–19 (2006).
47. Finlayson, S. & Finlayson, C. The birdmen of the Pleistocene: On the relationship between Neanderthals and scavenging birds. *Quat. Int.* **421**, 78–84 (2016).
48. Gómez-Olivencia, A. *et al.* New evidence for the presence of reindeer (*Rangifer tarandus*) on the Iberian Peninsula in the Pleistocene: an archaeopalaeontological and chronological reassessment. *Boreas* **43**, 286–308 (2014).
49. Sanz-Royo, A., Terlato, G. & Marín-Arroyo, A. B. Taphonomic data from the transitional Aurignacian of El Castillo cave (Spain) reveals the role of carnivores at

the Aurignacian Delta level. *Quaternary Science Advances* **13**, 100147 (2024).

50. Álvarez-Lao, D. J. & García, N. Comparative revision of the Iberian woolly mammoth (Mammuthus primigenius) record into a European context. *Quat. Sci. Rev.* **32**, 64–74 (2012).
51. Álvarez-Lao, D. J. & García, N. Southern dispersal and Palaeoecological implications of woolly rhinoceros (Coelodonta antiquitatis): review of the Iberian occurrences. *Quat. Sci. Rev.* **30**, 2002–2017 (2011).
52. Pochon, Z. *et al.* aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* **24**, 242 (2023).
53. Vidal-Cordasco, M., Terlato, G., Ocio, D. & Marín-Arroyo, A. B. Neanderthal coexistence with *Homo sapiens* in Europe was affected by herbivore carrying capacity. *Sci Adv* **9**, eadi4099 (2023).
54. Carvalho, M. *et al.* Initial and lower magdalenian large mammal faunas and human subsistence at El mirón cave (Cantabria, Spain). *J. Paleolit. Archaeol.* **4**, (2021).
55. Straus, L. G., Morales, M. G., Arroyo, A. B. M. & Chiapusso, M. J. I. Las ocupaciones humanas de la cueva del Mirón (Ramales de la Victoria, Cantabria, España) durante el Último Máximo Glacial y el periodo Solutrense. *Espac. Tiempo Forma Ser. Prehist. Arqueol.* 413–426 (2012).
56. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
57. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
58. Rambaut, A. & Drummond, A. FigTree v1. 3.1 Institute of Evolutionary Biology. *Univ. Edinb. J.* (2010).
59. Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: Flexible, modular software to

visualise and manipulate phylogenetic trees. *Ecol. Evol.* **14**, e10873 (2024).

60. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
61. Yang, M. A. *et al.* 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017).
62. Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
63. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
64. Hajdinjak, M. *et al.* Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature* **592**, 253–257 (2021).
65. Bollongino, R. *et al.* 2000 years of parallel societies in Stone Age Central Europe. *Science* **342**, 479–481 (2013).
66. Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016).
67. Benazzi, S. *et al.* Archaeology. The makers of the Protoaurignacian and implications for Neandertal extinction. *Science* **348**, 793–796 (2015).
68. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
69. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
70. Gilbert, M. T. P. *et al.* DNA from pre-Clovis human coprolites in Oregon, North America. *Science* **320**, 786–789 (2008).
71. Günther, T. *et al.* Population genomics of Mesolithic Scandinavia: Investigating early

postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).

72. Vai, S. *et al.* Ancestral mitochondrial N lineage from the Neolithic 'green' Sahara. *Sci. Rep.* **9**, 3530 (2019).
73. Hublin, J.-J. *et al.* Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. *Nature* 1–4 (2020).
74. Lazaridis, I. *et al.* Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry. *bioRxiv* 423079 (2018).
75. Gelabert, P. *et al.* Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Curr. Biol.* (2021) doi:10.1016/j.cub.2021.06.023.
76. Villalba-Mouco, V. *et al.* A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. *Nat Ecol Evol* **7**, 597–609 (2023).
77. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).
78. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
79. Soares, P. *et al.* The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
80. Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol.* **24**, 757–768 (2007).
81. Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H.-J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339–345 (2006).

82. Behar, D. M. *et al.* The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140 (2008).

83. Barbieri, E. & Sestili, P. Reactive oxygen species in skeletal muscle signaling. *J. Signal Transduct.* **2012**, 982794 (2012).

84. Behar, D. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).

85. Harich, N. *et al.* The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evol. Biol.* **10**, 138 (2010).

86. Cerný, V. *et al.* Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol. Biol.* **9**, 63 (2009).

87. Batini, C. *et al.* Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol. Biol. Evol.* **28**, 1099–1110 (2011).

88. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).

89. Podgorná, E., Soares, P., Pereira, L. & Cerný, V. The genetic impact of the lake chad basin population in North Africa as documented by mitochondrial diversity and internal variation of the L3e5 haplogroup. *Ann. Hum. Genet.* **77**, 513–523 (2013).

90. Vyas, S., Zaganjor, E. & Haigis, M. C. Mitochondria and Cancer. *Cell* **166**, 555–566 (2016).

91. Pennarun, E. *et al.* Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. *BMC Evol. Biol.* **12**, 234 (2012).

92. Olivieri, A. *et al.* The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* **314**, 1767–1770 (2006).

93. González, A. M. *et al.* Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* **8**, 223 (2007).
94. Fernandes, V. *et al.* The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *Am. J. Hum. Genet.* **90**, 347–355 (2012).
95. Greenspan, B. Family Tree DNA. Preprint at (2008).
96. Derenko, M. *et al.* Complete mitochondrial DNA diversity in Iranians. *PLoS One* **8**, e80673 (2013).
97. Pereira, L. *et al.* No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. *Mol. Biol. Evol.* **24**, 868–874 (2007).
98. Achilli, A. *et al.* Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PLoS One* **6**, e21029 (2011).
99. Costa, M. D. *et al.* Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the 'golden mean' to longevity. *Mech. Ageing Dev.* **130**, 222–226 (2009).
100. Fraumene, C. *et al.* High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. *Mol. Biol. Evol.* **23**, 2101–2111 (2006).
101. Derbeneva, O. A., Starikovskaya, E. B., Wallace, D. C. & Sukernik, R. I. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am. J. Hum. Genet.* **70**, 1009–1014 (2002).
102. Thalmann, O. *et al.* Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874 (2013).
103. Loog, L. *et al.* Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. *Mol. Ecol.* **29**, 1596–1610 (2020).

104. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* **25**, 1515–1519 (2015).

105. Matsumura, S., Inoshima, Y. & Ishiguro, N. Reconstructing the colonization history of lost wolf lineages by the analysis of the mitochondrial genome. *Mol. Phylogenet. Evol.* **80**, 105–112 (2014).

106. Björnerfeldt, S., Webster, M. T. & Vilà, C. Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* **16**, 990–994 (2006).

107. Chen, L. & Zhang, H. H. Canis lupus chanco mitochondrion, complete genome. Preprint at (2009).

108. Zhang, H., Zhang, J., Chen, L. & Liu, G. The complete mitochondrial genome of Chinese Shinjang wolf. *Mitochondrial DNA* **25**, 106–108 (2014).

109. Pang, J.-F. *et al.* mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* **26**, 2849–2864 (2009).

110. Arnason, U., Gullberg, A., Janke, A. & Kullberg, M. Mitogenomic analyses of caniform relationships. *Mol. Phylogenet. Evol.* **45**, 863–874 (2007).

111. Gopalakrishnan, S. *et al.* Interspecific Gene Flow Shaped the Evolution of the Genus Canis. *Curr. Biol.* **28**, 3441–3449.e5 (2018).

112. Bergström, A. *et al.* Grey wolf genomic history reveals a dual ancestry of dogs. *Nature* **607**, 313–320 (2022).

113. Webb, K. M. & Allard, M. W. Mitochondrial genome DNA analysis of the domestic dog: identifying informative SNPs outside of the control region. *J. Forensic Sci.* **54**, 275–288 (2009).

114. Bon, C. *et al.* Coprolites as a source of information on the genome and diet of the

cave hyena. *Proc. Biol. Sci.* **279**, 2825–2830 (2012).

115. Westbury, M. V. *et al.* Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Sci Adv* **6**, eaay0456 (2020).

116. Westbury, M. V. *et al.* Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. *Mol. Biol. Evol.* **35**, 1225–1237 (2018).

117. Westbury, M. V., De Cahsan, B., Dalerum, F., Norén, K. & Hofreiter, M. Aardwolf Population Diversity and Phylogenetic Positioning Inferred Using Complete Mitochondrial Genomes. *sawr.1* **49**, 27–33 (2019).

118. Hiendleder, S., Lewalski, H., Wassmuth, R. & Janke, A. The complete mitochondrial DNA sequence of the domestic sheep (Ovis aries) and comparison with the other major ovine haplotype. *J. Mol. Evol.* **47**, 441–448 (1998).

119. Hassanin, A., Ropiquet, A., Couloux, A. & Cruaud, C. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *J. Mol. Evol.* **68**, 293–310 (2009).

120. Zhong, H.-M., Zhang, H.-H., Sha, W.-L., Zhang, C.-D. & Chen, Y.-C. Complete Mitochondrial Genome of the Red Fox (Vuples vuples) and Phylogenetic Analysis with Other Canid Species. *Dongwuxue Yanjiu* **31**, 122–130 (2010).

121. Zhang, Z. *et al.* High-Quality Chromosome-Level Genome Assembly of the Corsac Fox (Vulpes corsac) Reveals Adaptation to Semiarid and Harsh Environments. *Int. J. Mol. Sci.* **24**, (2023).

122. Zhao, C., Zhang, H., Liu, G., Yang, X. & Zhang, J. The complete mitochondrial genome of the Tibetan fox (Vulpes ferrilata) and implications for the phylogeny of Canidae. *C. R. Biol.* **339**, 68–77 (2016).

123. Yu, J.-N., Kim, S., Oh, K. & Kwak, M. Complete mitochondiral genome of the Korean red fox Vulpes vulpes (Carnivora, Canidae). *Mitochondrial DNA* **23**, 118–119 (2012).
124. Frank, K. *et al.* Complete mitochondrial genome sequence of a Hungarian red deer (Cervus elaphus hippelaphus) from high-throughput sequencing data and its phylogenetic position within the family Cervidae. *Acta Biol. Hung.* **67**, 133–147 (2016).
125. Kim, H.-J. *et al.* The complete mitochondrial genome of Cervus canadensis (Erxleben, 1777), as a model species of Chronic Wasting Disease (CWD). *Mitochondrial DNA B Resour* **5**, 2621–2623 (2020).
126. Hassanin, A. *et al.* Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335**, 32–50 (2012).
127. Rey-Iglesia, A., Grandal-d'Anglade, A., Campos, P. F. & Hansen, A. J. Mitochondrial DNA of pre-last glacial maximum red deer from NW Spain suggests a more complex phylogeographical history for the species. *Ecol. Evol.* **7**, 10690–10700 (2017).
128. Wada, K., Nishibori, M. & Yokohama, M. The complete nucleotide sequence of mitochondrial genome in the Japanese Sika deer (Cervus nippon), and a phylogenetic analysis between Cervidae and Bovidae. *Small Rumin. Res.* **69**, 46–54 (2007).
129. Mackiewicz, P. *et al.* Phylogeny and evolution of the genus Cervus (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Sci. Rep.* **12**, 16381 (2022).
130. Li, Y., Ba, H. & Yang, F. Complete mitochondrial genome of Cervus elaphus songaricus (Cetartiodactyla: Cervinae) and a phylogenetic analysis with related species. *Mitochondrial DNA A DNA Mapp Seq Anal* **27**, 620–621 (2016).

131. Wada, K., Okumura, K., Nishibori, M., Kikkawa, Y. & Yokohama, M. The complete mitochondrial genome of the domestic red deer (Cervus elaphus) of New Zealand and its phylogenic position within the family Cervidae. *Anim. Sci. J.* **81**, 551–557 (2010).

132. Hirata, D. *et al.* Molecular phylogeography of the brown bear (Ursus arctos) in Northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Mol. Biol. Evol.* **30**, 1644–1652 (2013).

133. Miller, W. *et al.* Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2382–90 (2012).

134. Keis, M. *et al.* Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia. *J. Biogeogr.* **40**, 915–927 (2013).

135. Anijalg, P. *et al.* Ongoing recovery of a brown bear population from a century-old severe bottleneck: insights from population genetic and spatially explicit analyses. *Conserv. Genet.* **21**, 27–40 (2020).

136. Fortes, G. G. *et al.* Ancient DNA reveals differences in behaviour and sociality between brown bears and extinct cave bears. *Mol. Ecol.* **25**, 4907–4918 (2016).

137. Korsten, M. *et al.* Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Mol. Ecol.* **18**, 1963–1979 (2009).

138. Delisle, I. & Strobeck, C. Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Mol. Biol. Evol.* **19**, 357–361 (2002).

139. Krause, J. *et al.* Mitochondrial genomes reveal an explosive radiation of extinct and

extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.* **8**, 220 (2008).

140. Rey-Iglesia, A. *et al.* Evolutionary history and palaeoecology of brown bear in North-East Siberia re-examined using ancient DNA and stable isotopes from skeletal remains. *Sci. Rep.* **9**, 4462 (2019).
141. Botigué, L. R. *et al.* Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8**, 16082 (2017).
142. Bergström, A. *et al.* Origins and genetic legacy of prehistoric dogs. *Science* **370**, 557–564 (2020).
143. Boschin, F. *et al.* The first evidence for Late Pleistocene dogs in Italy. *Sci. Rep.* **10**, 13313 (2020).
144. Achilli, A. *et al.* Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2449–2454 (2012).
145. Xu, S. *et al.* High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J. Genet. Genomics* **34**, 720–729 (2007).
146. Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
147. Lippold, S. *et al.* Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* **2**, 450 (2011).
148. Der Sarkissian, C. *et al.* Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr. Biol.* **25**, 2577–2583 (2015).
149. Librado, P. *et al.* Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6889–97 (2015).
150. Yoon, S. H. *et al.* Origin and spread of Thoroughbred racehorses inferred from

complete mitochondrial genome sequences: Phylogenomic and Bayesian coalescent perspectives. *PLoS One* **13**, e0203917 (2018).

151. Xu, X., Gullberg, A. & Arnason, U. The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *J. Mol. Evol.* **43**, 438–446 (1996).
152. Taron, U. H. *et al.* Ancient DNA from the Asiatic Wild Dog (Cuon alpinus) from Europe. *Genes* **12**, (2021).
153. Campana, M. G. *et al.* Genome sequence, population history, and pelage genetics of the endangered African wild dog (Lycaon pictus). *BMC Genomics* **17**, 1013 (2016).
154. Sosale, M. S. *et al.* The complete mitochondrial genome and phylogenetic characterization of two putative subspecies of golden jackal (Canis aureus cruesemanni and Canis aureus moreotica). *Gene* **866**, 147303 (2023).
155. Koepfli, K.-P. *et al.* Genome-wide Evidence Reveals that African and Eurasian Golden Jackals Are Distinct Species. *Curr. Biol.* **25**, 2158–2165 (2015).
156. 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).
157. Skoglund, P. *et al.* Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).
158. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–63 (2016).