UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA

# Ancient DNA Screening of Rhinoceros Species at Denisova Cave

Raquel Alexandra Leal Teixeira

**Mestrado em Biologia Molecular e Genética**

Dissertação oriendada por:
Professor Bruno Miguel Santos de Almeida Nevado
Professor Love Dalén

2025

# Acknowledgments

Undertaking this project would not have been possible without my supervising committee. To Professor Love Dalén I would like to thank the very opportunity to carry out this work, for entrusting me with this task, as well as the feedback provided throughout. I also want to thank Professor Bruno Nevado, who provided invaluable insights which allowed me to refine the final product. Additionally, this work was made possible with the support of ULisboa that, under the ERASMUS+ program, funded my internship.

I am extremely grateful to all the incredible people I met at CPG. I find myself privileged for the time spent working with such a conscientious, kind and very brilliant group of people. To one and all, thank you for making me feel so welcome, assured, and safe. I extend a special thanks to Alexandre Gilardet, who took upon himself the task of my mentor. Not only did he provide me with guidance and the tools to carry out my project, but it was also his work of trail blazing the methods I would employ that allowed me to carry out such a monumental task.

Finally, I want to extend my deepest and fondest gratitude to my family and friends. To my father, thank you for every, near-daily call, so that the *saudade* I felt was of the most gentle sort. To my siblings, Ricardo and Rodrigo, for being an incredible source of inspiration and motivation, and my anchor home. I am very grateful to my dear friends, those who accompanied me in this journey and provided an endless source of solace and joy. Thank you, Inês, and thank you to the Genomigas, all of whom supported me even from afar, even without quite understanding what I was doing, confident only that it had something to do with rhinos.

# Resumo

Entre as diversas famílias da megafauna do Cenozoico, a Rhinocerotidae destaca-se como uma das mais impressionantes. Ao longo do Pleistoceno (2.6 milhões a 12 mil anos atrás), os rinocerontes terão ocupado um território vasto na região da Eurásia. Porém, a maioria destas espécies extinguiu-se durante o Pleistoceno Tardio (115 – 11.7 mil anos atrás). Atualmente, apenas três espécies persistem no continente asiático, e outras duas em África, todas as quais enfrentam o risco de extinção.

No foro de investigação genética, o estudo dos rinocerontes extintos da Eurásia tem por principal foco o rinoceronte-lanudo (*Coleodonta antiquitatis*), e concentra-se nas dinâmicas populacionais que imediatamente antecedem a sua extinção, assim como a resposta e adaptação da espécie a ecossistemas impactados pelas alterações climáticas do final do Pleistoceno. Em contraste, existe pouca informação genética disponível para o estudo de outras espécies, assim como escassa amostragem proveniente do Pleistoceno Médio (774 – 129 mil anos atrás). O estudo comparativo das demais espécies de rinocerontes extintos – sendo estes o rinoceronte-de-Merck (*Stephanorhinus kirchbergensis*), o rinoceronte-da-estepe (*Stephanorhinus hemitoechus*) e o "unicórnio" da Sibéria (*Elasmotherium sibiricum*) –, especialmente durante este período, possuí particular interesse científico. O estudo das características morfológicas e ecologia destas espécies extintas indica que as espécies do género *Stephanorhinus* e *E. sibiricum* estariam mais aptas a prosperar em ambientes moderados, ao passo que o rinoceronte-lanudo estaria adaptado a ambientes secos e frios.

Recentemente, foi realizado um estudo de espectrometria de massa em fragmentos ósseos petrificados (ZooMS) que foram encontrados da caverna Denisova (Montes Altai, Sibéria). Através deste método, foi identificado um vasto conjunto de restos de rinocerontes. Devido a limitações no método, desconhece-se, porém, a sua espécie. Estes fragmentos provêm de diversas camadas estratigráficas consecutivas, abrangendo períodos de transição climática do Pleistoceno Médio e Tardio. Neste projeto, rastreámos a presença de DNA antigo (*ancient DNA* ou aDNA) nestes fragmentos, com o objetivo de identificar amostras que apresentassem DNA endógeno preservado. Reconstruimos o genoma mitocondrial (mitogenoma) para identificação taxonómica ao nível da espécie e alinhámos os mitogenomas gerados com os da literatura, de modo a considerar estes espécimenes na filogenia dos rinocerontes. Analisámos também os padrões de distribuição das espécies pelas camadas estratigráficas da caverna.

Atendendo às diferenças morfológicas e ecológicas das espécies, tomamos por hipótese que um maior número de rinocerontes-lanudos seria identificado nas camadas correspondentes aos períodos glaciares, enquanto as outras espécies deveriam figurar principalmente das camadas

correspondentes a períodos interglaciares. Este trabalho visa aprofundar o conhecimento acerca das espécies de rinocerontes presentes na região do Altai durante o Pleistoceno Médio, e acrescentar ao catálogo de informação genética de modo a potenciar estudos sobre a história evolutiva mais ancestral não só do rinoceronte-lanudo como das demais espécies eurasiáticas extintas.

Um total de 141 amostras foram obtidas de fragmentos ósseos da câmara leste da caverna Denisova, apresentando uma distribuição estratigráfica com repartição uniforme pelos estágios isotópicos (*marine-oxygen isotope stages* ou MIS) 5, 6 e 7. Pretender-se-ia deste modo identificar padrões na distribuição das espécies pelos períodos glaciares e interglaciares do Pleistoceno Médio. Procedeu-se à extração de DNA e criação de bibliotecas genéticas para NGS (*Next Generation Sequencing*). Após sequenciação, foram empregues ferramentas bioinformáticas para processar as sequências, e estas mapeadas ao genoma do rinoceronte-de-Sumatra (o parente vivo mais próximo dos rinocerontes extintos da eurásia, em termos evolutivos) para obter estatísticas referentes à presença e estado de preservação do DNA. De entre os valores estimados, proporção de conteúdo endógeno > 5% e cobertura média > 0.1X foram usados como critério no rastreio inicial para identificação das amostras mais promissoras, que foram novamente extraídas com o objetivo de maximizar a quantidade de informação genética obtida.

Para a identificação taxonómica, as sequências com cobertura média > 0.1X foram mapeadas competitivamente, primeiro a um painel de diversas espécie-engodo para despistar contaminações e depois novamente a um painel com três espécies de rinocerontes extintos. Desta análise, identificamos 12 amostras pertencentes à espécie rinoceronte-lanudo e uma amostra de rinoceronte-de-Merck. Em dez destas amostras, das quais se destaca o rinoceronte-de-Merck, foi possível validar a identificação taxonómica, uma vez que o valor de cobertura média do genoma mitocondrial (> 1X) permitiu reconstruir o respetivo mitogenoma, e deste modo analisar o posicionamento destas amostras na árvore filogenética mitocondrial das espécies. Dada a distribuição das amostras pelas camadas estratigráficas da caverna, os resultados evidenciam uma presença contínua do rinoceronte-lanudo nos Montes Altai durante os estágios isotópicos MIS 6 e 5. O único espécimen de rinoceronte-de-Merck identificado provém do MIS 7, o que poderá indicar que ocorreu uma transição de espécies após este período interglacial. Contudo, tratando-se de somente uma amostra, não é possível validar esta hipótese, nem identificar padrões de coexistência interespécies que relacionem os sucessivos períodos de transição climática à predominância da espécie mais adaptada ao clima em cada fase na região. O padrão de distribuição de amostras de rinoceronte-lanudo corresponde de modo geral ao proposto pela hipótese, havendo, porém, falta de resolução na estratigrafia que permita determinar se a espécie permaneceu na região durante o primeiro sub-estágio do interglacial MIS 5 (MIS 5e), ou se terá ocorrido uma extinção local temporária.

Reconstruimos sete mitogenomas de rinocerontes-lanudos com elevada cobertura média (> 3X). Estes novos mitogenomas permitiram expandir a árvore filogenética mitocondrial da espécie, quer de um ponto de vista temporal como geográfico. A árvore filogenética gerada neste estudo apresenta suporte sólido para a existência de quatro clados, corroborando a existência do clado D proposta num estudo prévio. Porém, não foi possível resolver de forma conclusiva a relação entre os clados. Verificou-se que a amostra RT087 contribuía, em parte, para a incerteza na resolução da relação entre os clados B e D. A rede de haplótipos revelou que RT087 se posiciona num ramo que diverge do mesmo nó ancestral correspondente à divergência entre B e D. Como a diferença em passos mutacionais entre RT087 e os clados B e D é pequena e quase equivalente, esta amostra posiciona-se num ramo curto e com baixo valor de suporte. Este padrão poderá indicar que este espécimen viveu num período próximo da divergência dos dois clados, o que contrasta com a idade estimada para a amostra com base no contexto estratigráfico. Assim, propõe-se que, em futuros estudos, estas amostras sejam novamente datadas com recurso a métodos filogenéticos que utilizam o relógio molecular, de modo a validar ou corrigir as idades atribuídas a estes fragmentos.

A árvore filogenética indica também a coexistência de múltiplas linhagens de rinoceronte-lanudo na região, evidenciando uma ampla dispersão e conectividade entre as populações de rinocerontes-lanudos da Eurásia, bem como o papel dos Montes Altai como um ponto de contacto entre elas.

Neste estudo, foi ainda investigada a presença de DNA obtido de forma incidental (*by-catch*) em fragmentos ósseos de rinocerontes, assim como de amostras de bovídeos provenientes de outro estudo, explorando as aplicações e relevância biológica desta metodologia. Foi investigada a presença de DNA de hiena (*Crocuta crocuta*) em quatro amostras de bovídeo e uma amostra de rinoceronte, assim como de lobo (*Canis lupus*) e ovelha (*Ovis aries*) noutras duas amostras de bovídeo. Os mesmos métodos de mapeamento e reconstrução do mitogenoma foram empregues em bibliotecas previamente geradas, usando o genoma mitocondrial das espécies *by-catch*. Em duas amostras, uma vez que a cobertura média do mitogenoma foi superior a 1X, foi possível reconstruir dois mitogenomas de hiena. Os mitogenomas gerados pertencem ao haplogrupo A1, que terá sido previamente identificado como o principal haplogrupo de hiena presente na caverna Denisova no final do Plesitoceno Médio (MIS 5). Juntamente com os resultados da identificação taxonómica dos rinocerontes, este estudo contribui para a análise e interpretação dos conjuntos faunísticos do sul da Sibéria. No entanto, sublinha também o risco de basear a identificação taxonómica deste tipo de fragmentos exclusivamente na informação genética, evidenciando a utilidade da complementaridade da mesma com os resultados de ZooMS.

Ao preencher lacunas no registo fóssil e genético, este trabalho contribui para a compreensão da história evolutiva da família Rhinocerotidae. Introduz perspetivas relevantes acerca da dispersão das linhas maternais de rinoceronte-lanudo e, particularmente, do papel dos Montes Altai como um corredor ecológico para esta espécie e ponto de contacto entre as múltiplas linhagens. Paralelamente, o registo obtido para o rinoceronte-de-Merck acrescenta dados valiosos para avaliar possíveis padrões de substituição entre espécies com nichos ecológicos distintos. O rastreio realizado neste projeto permitiu ainda gerar bibliotecas de DNA com conteúdo genético não só a nível mitocondrial (que foi alvo de análise neste estudo), mas também a nível nuclear. A informação genética do genoma nuclear possibilita estudos futuros sobre a adaptação das espécies ao seu ambiente, assim como sobre as respostas na estrutura populacional face a transições climáticas abruptas do Pleistoceno Médio.

Palavras-chave: DNA antigo, paleogenética, filogenia mitocondrial, rinoceronte-lanudo, rinoceronte-de-Merck

# Abstract

The Rhinocerotidae family is among some of the most impressive megafauna of the Cenozoic. During the Pleistocene, rhinoceroses were widely distributed across Eurasia, with most species becoming extinct during the Late Pleistocene, 129 – 11.7 thousand years ago (ka).

Research into the population dynamics of rhinoceroses and their responses to climate-driven environmental changes has primarily focused on the woolly rhinoceros (*Coleodonta antiquitatis*) and the period immediately preceding their extinction. While the Late Pleistocene has been extensively studied, the Middle Pleistocene (774 – 129 ka) remains poorly sampled.

Mass spectrometry analysis of bone fragments found at Denisova Cave has revealed a vast archive of rhinoceros remains, deposited across stratigraphic layers spanning climatic transitions from the Middle to Late Pleistocene. In this project, we screened these rhinoceros bone fragments for ancient DNA (aDNA). Through the processing and analysis of genomic data, we taxonomically identified bone fragments from 12 woolly rhinoceros and one Merck's rhinoceros (*Stephanorhinus kirchbergensis*). From this data, seven high coverage woolly rhinoceros mitogenomes from the late Middle Pleistocene were reconstructed and used to broaden the temporal and geographical reach of the species' phylogeny. We found evidence that woolly rhinoceroses were present in the Altai region continually throughout marine-oxygen isotope stages (MIS) 5 and 6. Multiple maternal lineages co-existed in the region, highlighting a wide dispersal and connectivity between the populations of woolly rhinoceros across Eurasia, as well as the role of Altai as a point of contact between them.

This study also investigated the presence of by-catch DNA in Rhinocerontidae and bovid bone fragments, exploring its applications and biological implications. Two hyena (*Crocuta crocuta*) mitogenomes were generated, belonging to haplogroup A1, which is prevalent in Europe. Together with the previous results, our research contributes to the study and interpretation of faunal assemblages from southern Siberia.



Key-words: ancient DNA, paleogenetics, mitochondrial phylogeny, woolly rhinoceros, Merck's rhinoceros

# Table of Contents

**Acknowledgments** ..... **ii**
**Resumo** ..... **iii**
**Abstract** ..... **vii**
**Table of Contents** ..... **viii**
**List of Figures** ..... **x**
**List of Tables** ..... **x**
**List of Abbreviations** ..... **xi**
**1. Introduction** ..... **1**
1.1. Emergence of the Rhinocerotidae family ..... 1
1.2. Ecology of the extinct Pleistocene Eurasian rhinoceros ..... 1
1.3. Population dynamics in response to climatic changes ..... 3
1.4. Paleontological context at Denisova Cave ..... 5
1.5. Ancient DNA: Challenges and Approaches ..... 6
1.6. Main Goals ..... 8
**2. Materials and Methods** ..... **9**
2.1. Sample information ..... 9
2.2. Bone fragment sampling ..... 10
2.3. Lysis and extraction ..... 10
2.4. Library preparation ..... 11
2.5. Sequencing data processing and analysis ..... 13
2.6. Mitogenomes and phylogenetic analysis ..... 14
2.7. Investigation of by-catch genomic data ..... 15
**3. Results** ..... **16**
3.1. Report on the screening and taxonomical assignment ..... 16
3.2. Denisova Cave rhinoceros placement in the woolly rhinoceros mitochondrial phylogeny ..... 18
3.3. Mitogenomes generated from by-catch genomic data ..... 20
**4. Discussion** ..... **21**
4.1. Temporal patterns of rhinoceros species diversity in Altai ..... 21
4.2. Geographical range of woolly rhinoceros clades ..... 23
4.3. Resolution of clade divergence ..... 24
4.4. Implications of by-catch genomic data recovery ..... 25
**5. Conclusion** ..... **26**
**References** ..... **28**
**Supplementary Material** ..... **36**
Annex I. Remarks on RTBL5 contamination ..... 36

Annex II. Supplementary Figures ..... 37
Annex III. Supplementary Tables ..... 41

# List of Figures

Figure 1.1. Ranges of rhinoceros species surviving into the Late Quaternary..... 2
Figure 1.2. Whole-genome phylogeny of the Rhinocerotidae ..... 3
Figure 1.3. Temporal distribution of ancient DNA studies..... 4
Figure 1.4. Location and structure of Denisova Cave..... 5
Figure 3.1. Competitive mapping results for samples with mitochondrial depth >0.1X .. 18
Figure 3.2. Consensus phylogenetic tree of woolly rhinoceros mitogenomes..... 19
Figure 3.3. Median-joining haplotype network of woolly rhinoceros mitogenomes..... 20
Figure 4.1. Distribution of taxonomically identified samples through the layers of the East Chamber ..... 22

Supplementary Figure 1. Competitive mapping results of the full screening data ..... 37
Supplementary Figure 2. Consensus phylogenetic tree of low and high depth of coverage mitogenomes ..... 38
Supplementary Figure 3. Consensus phylogenetic tree generated from the alignment of high depth of coverage mitogenomes and European woolly rhinoceros from hyena coprolite .. 39
Supplementary Figure 4. Hyena phylogenetic tree ..... 40
Supplementary Figure 5. AMBER plots of extraction blank RTBL5..... 40

# List of Tables

Table 2.1. Distribution of the fragments sampled in this study through layers of the East Chamber and corresponding marine-oxygen isotope stage..... 10
Table 3.1. Summary of mapping statistics of deep sequenced samples..... 16

Supplementary Table 1. Record of samples used in this study..... 41
Supplementary Table 2. Indexing PCR reaction and cycle programs for libraries prepared with the Santa Cruz Reaction (SC) and Meyer & Kircher (MK) protocols ..... 44
Supplementary Table 3. Screening mapping statistics..... 45
Supplementary Table 4. Bait species on the panel for the first round of competitive mapping ..... 48
Supplementary Table 5. R packages used in data visualization..... 49
Supplementary Table 6. Mapping statistics of the libraries investigated for by-catch mitogenome content ..... 50

# List of Abbreviations

| | |
|---|---|
| aDNA | Ancient DNA |
| Cq | Quantification cycle |
| ka | Kiloannum, thousand years |
| Ma | Milliannum, million years |
| MIS | Marine-oxygen isotope stages |
| Mitogenome(s) | Mitochondrial genome(s) |
| qPCR | Quantitative polymerase chain reaction |
| USER | Uracil-DNA-glycosylase and endonuclease VIII |
| ZooMS | Zooarchaeology by Mass Spectrometry |

# 1. Introduction

## 1.1. Emergence of the Rhinocerotidae family

The Cenozoic landscape served as a stage for the development of some of the largest mammals that ever lived. Changes in the climate occurring throughout the Quaternary, and particularly during the onset of the Holocene (the current geological age) (Lorenzen et al., 2011) would render extinct a significant number of megafauna species. Among those species that have survived to modern day we count a number of species of rhinoceros.

The Rhinocerotidae family, also referred to as "true rhinoceroses", traces its ancestry to the early Eocene, 55 - 60 million years ago (Ma) (Bai et al., 2020), when the Rhinocerotoidea superfamily diverged from Tapiroidea. During this time, Eurasia was still connected to North America, which allowed Rhinocerotoidea to spread across the northern territories alongside other large mammal species (Prothero, 1993). Before the end of the Eocene, this connection was severed, and the ensuing geographical isolation would steer the divergence of the superfamily. Rhinocerotidae radiated from the superfamily and further diversified into over 100 species, distributed across North America, Eurasia and Africa (Cerdeño, 1998; Prothero, 1993).

Following the extinction of other species that had radiated from the superfamily (Cerdeño, 1998), only Rhinocerotidae lasted into the Pleistocene, which spanned from 2.6 Ma to 12 thousand years ago (ka). During and following this epoch more extinctions occurred, and in the present day only five rhinoceros species remain. According to the Rhino Recovery Fund, all five extant species can be found in either Africa or Asia, making up a sum of *ca.* 27,000 individuals. These are the white and black rhinoceroses in Africa, and the Sumatran, Javan, and Greater one-horned rhinoceroses from Asia. The species are categorized from near threatened to critically endangered according to the IUCN Red List.

## 1.2. Ecology of the extinct Pleistocene Eurasian rhinoceros

Fossil records have indicated that, during the Middle Pleistocene (774 – 129 ka) and lasting into the early stages of the Holocene, there used to be rhinoceroses inhabiting the region of Eurasia. At this time, there were four rhinoceros species inhabiting the territories of northern Eurasia (Figure 1.1) (Cerdeño, 1998), belonging to two subfamilies of Rhinocerotidae: Rhinocerotinae and Elasmotheriinae. From the extant Rhinocerotinae subfamily, the species present in the region were *Coleodonta antiquitatis*, *Stephanorhinus hemitoechus* and *S. kirchbergensis*. From the extinct Elasmotheriinae subfamily, *Elasmotherium sibiricum* inhabited in the area of Siberia during the Late Pleistocene (129 – 11.7 ka), until its extinction 35 – 36 ka (Kosintsev et al., 2018).

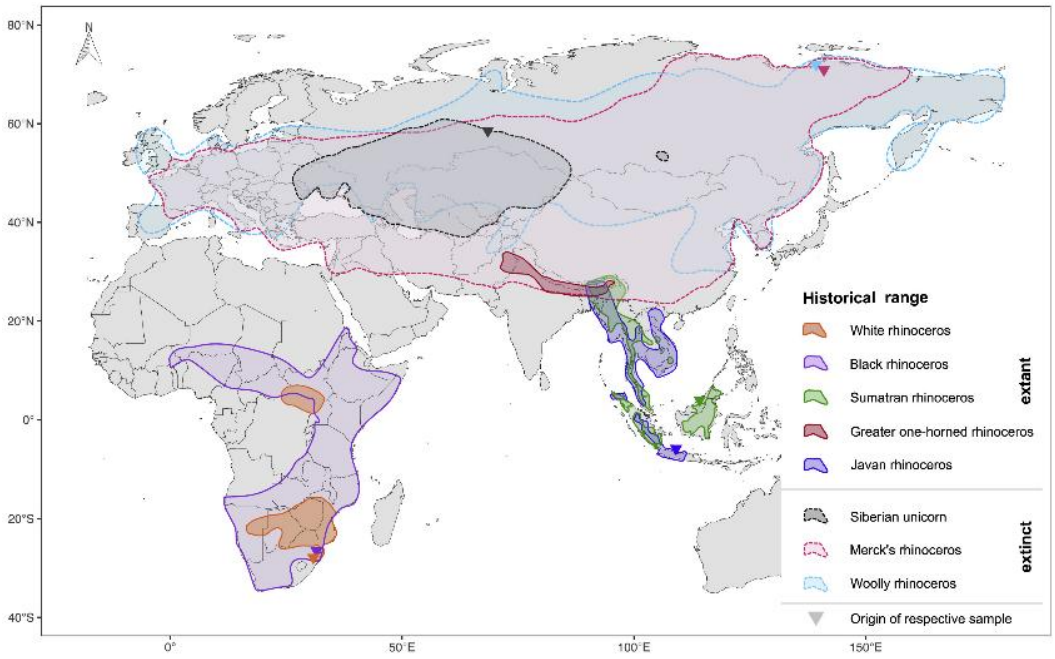



**Figure 1.1. Ranges of rhinoceros species surviving into the Late Quaternary,** (Liu et al., 2021). All ranges are approximated to the distribution, rather than detailed. Historical distribution range of extant species (white rhinoceros, *Ceratotherium simum*; black rhinoceros, *Diceros bicornis*; Sumatran rhinoceros, *Dicerorhinus sumatrensis*; Greater one-horned rhinoceros, *Rhinoceros unicornis*; Javan rhinoceros, *R. sondaicus*) and extinct species (Siberian unicorn, *E. sibiricum*; Merck's rhinoceros; *S. kirchbergensis*; woolly rhinoceros, *C. antiquitatis*) is shown. *S. hemitoechus* was not addressed by Liu et al. (2021).

The study of these ancient rhinoceroses' remains has informed the palaeoecological contexts they inhabited. The Elasmotheriinae and Rhinocerotinae subfamilies diverged around 36 Ma (Figure 1.2) (Liu et al., 2021). *E. sibiricum* was the largest of the Quaternary rhinoceros, characterized by a single large horn (Kosintsev et al., 2018) – a feature that earned the common name "Siberian unicorn". This species had a high degree of dietary specialization, relative to other rhinoceroses of the time. Morphological and stable isotope analysis has suggested hypergrazing adaptation, meaning a diet composed primarily of $C_4$ plants, particularly the consumption of non-photosynthetic plant parts like roots (Kosintsev et al., 2018). Its habitat would have consisted of areas of dry steppe or desert biomes.

The genera *Stephanorhinus* and *Coleodonta* form a monophyletic clade with the Sumatran rhinoceros (*Dicerorhinus sumatrensis*), with their last common ancestor dating to 9.38 Ma. Subsequently, the two genera diverged 5.5 Ma (Liu et al., 2021) (Figure 1.2). *Stephanorhinus* were mixed feeders, browsing or grazing as the environment permitted (Boeskorov, 2012; Stefaniak et al., 2021). Adapted to moderate/warm climates, the genus was thus both dietarily and ecologically flexible. Though *S. kirchbergensis* (also referred to as Merck's rhinoceros) was mainly present in forests, and *S. hemitoechus* in steppes, the two were often sympatric (van Asperen & Kahlke, 2015).

*Coleodonta antiquitatis*, commonly known as the woolly rhinoceros, possessed thick wool and skin, features that would have been beneficial in colder, dry environments (Boeskorov, 2012). *C. antiquitatis* would have primarily grazed on bush vegetation in steppe and tundra-steppe biomes (Agadjanian & Shunkov, 2018), though isotope analysis by Stefaniak et al. (2021) indicates some degree of seasonal dietary adjustment.

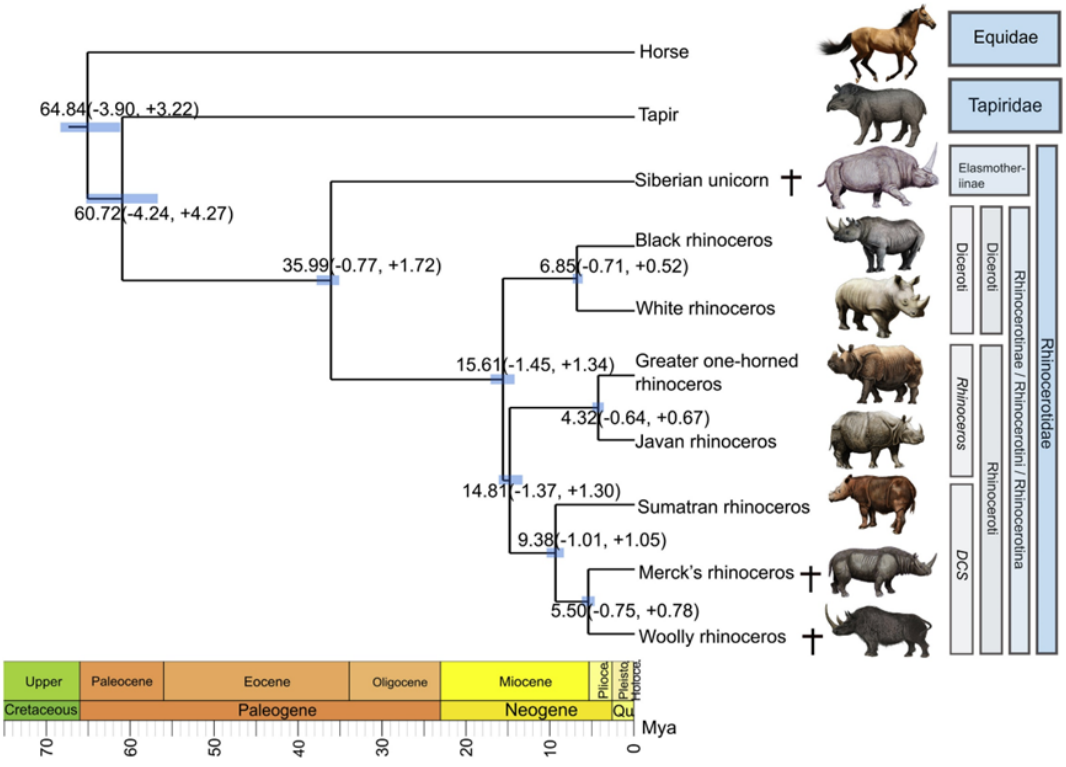



**Figure 1.2. Whole-genome phylogeny of the Rhinocerotidae,** based on a consensus of trees generated every 100 kb, modified from Liu et al. (2021). The consensus species tree of Rhinocerotidae is dated and shows the estimated divergence dates between lineages at the nodes. Extant species (white rhinoceros, *C. simum*; black rhinoceros, *D. bicornis*; Sumatran rhinoceros, *D. sumatrensis*; Greater one-horned rhinoceros, *R. unicornis*; Javan rhinoceros, *R. sondaicus*) and extinct species (Siberian unicorn, *E. sibiricum*; Merck's rhinoceros, *S. kirchbergensis*; woolly rhinoceros, *C. antiquitatis*) are shown, as well as outgroups *Equus caballus* (horse) and *Tapirus terrestris* (tapir). *S. hemitoechus* was not addressed by Liu et al. (2021).

The diversity in adaptive characteristics among rhinoceros species reflects the complex evolutionary history of the family and is crucial to understanding the success of each species in different habitats. When viewed in the context of the climate-driven environmental shifts of the Quaternary, these characteristics offer clues to infer probable responses to changing pressures in the rhinoceros habitat. However, these observations do not constitute the sum of information necessary to comprehend the intricate population dynamics of the species' past.

## 1.3. Population dynamics in response to climatic changes

The Pleistocene was an epoch marked by several events of massive climatic turnover, seeing alternating glacial and interglacial periods (Anderson et al., 2013). The shifting climate

and associated geological changes had an impact on ecosystems across the globe, shaping the distribution and diversity of fauna and flora alike (Hofreiter & Stewart, 2009).

The genetic divergence of rhinoceros lineages that survived to the end of the Pleistocene hints at a complex, unresolved population dynamic in their recent past. Through the study of mitochondrial phylogenies, Lord et al. (2020) identified deep genetic divergence between two Late Pleistocene woolly rhinoceros mitochondrial DNA lineages in northern Siberia. This divergence, along with the demographic patterns inferred from the dataset, led the authors to suggest that these lineages emerged due to population fragmentation and subsequent merging occurring during the Middle Pleistocene. The addition of samples from northern China (Yuan et al., 2023) to the woolly rhinoceros phylogeny reinforces the weak geographic and temporal structuring between the two lineages. To study the population events that gave rise to the lineages of the Late Pleistocene would require investigating population structure further back in time, by broadening the temporal reach of this phylogeny with older samples.

Despite ancient DNA having been recovered from samples dating as far back as the Early Pleistocene, the Middle Pleistocene remains significantly underrepresented (Figure 1.3) (Dalén et al., 2023), limiting the study of this particular time period.

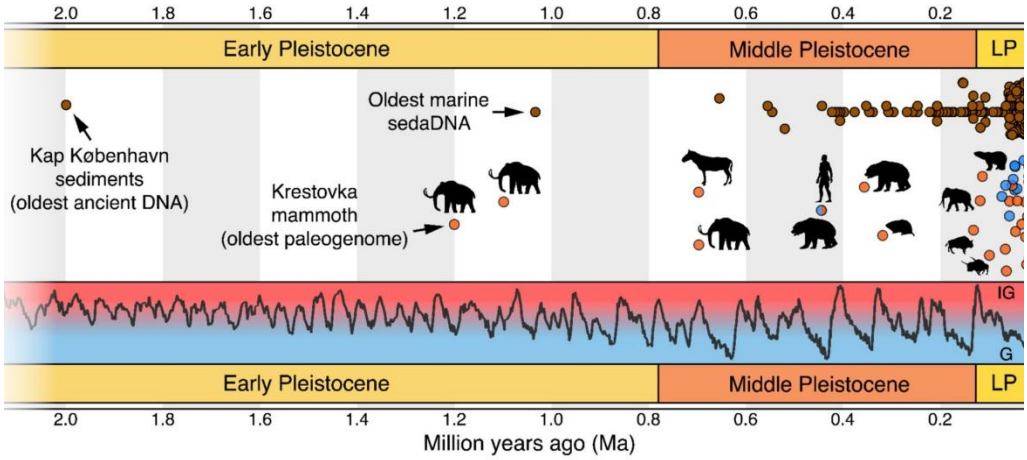



**Figure 1.3. Temporal distribution of ancient DNA studies**, in which circles in orange are non-human animal paleogenomes, blue are hominin, and brown are sedimentary ancient DNA records (Dalén et al., 2023).

Understanding organisms from this period could provide valuable context for the lineage divergence first observed by Lord et al. (2020), offering insight into how climate change affected ancient rhinoceroses and their evolutionary trajectories. With this goal in mind, this project sought to screen a large archive of Rhinocerontidae remains from Denisova Cave and dated to the Middle and Late Pleistocene.

## 1.4. Paleontological context at Denisova Cave

Found along the range of the Altai mountains in southern Siberia (Figure 1.4 A), Denisova Cave is an archeological and paleontological site providing a context of great significance to the study of the Paleolithic period of the region (Kuzmin et al., 2022). In 2008, a phalanx of a juvenile hominin was excavated, which would later be determined to belong to a new unidentified lineage of archaic human, the Denisovans (Briggs et al., 2009; Reich et al., 2010). It is recognized as the oldest inhabited cave in Siberia, possessing evidence of Neanderthal, Denisovan, and modern human occupation (Jacobs et al., 2019, 2025; Zavala et al., 2021).

Denisova Cave has one of the longest stratigraphical records in Eurasia, dating as far back as the Middle Pleistocene (Douka et al., 2019; Jacobs et al., 2019). The cave is made up of three sections (Figure 1.4 B): the main, south, and east chambers. In the East chamber, the stratigraphic sequence spans from around 50 to 300 ka, preserving a record of sediments and remains from the late Middle to Late Pleistocene. The state of fragmentation of most of the remains recovered from the cave prevents their identification and taxonomic classification through morphological analysis. Other methods have thus been employed to investigate these fragments so they may be considered in broader paleontological studies.

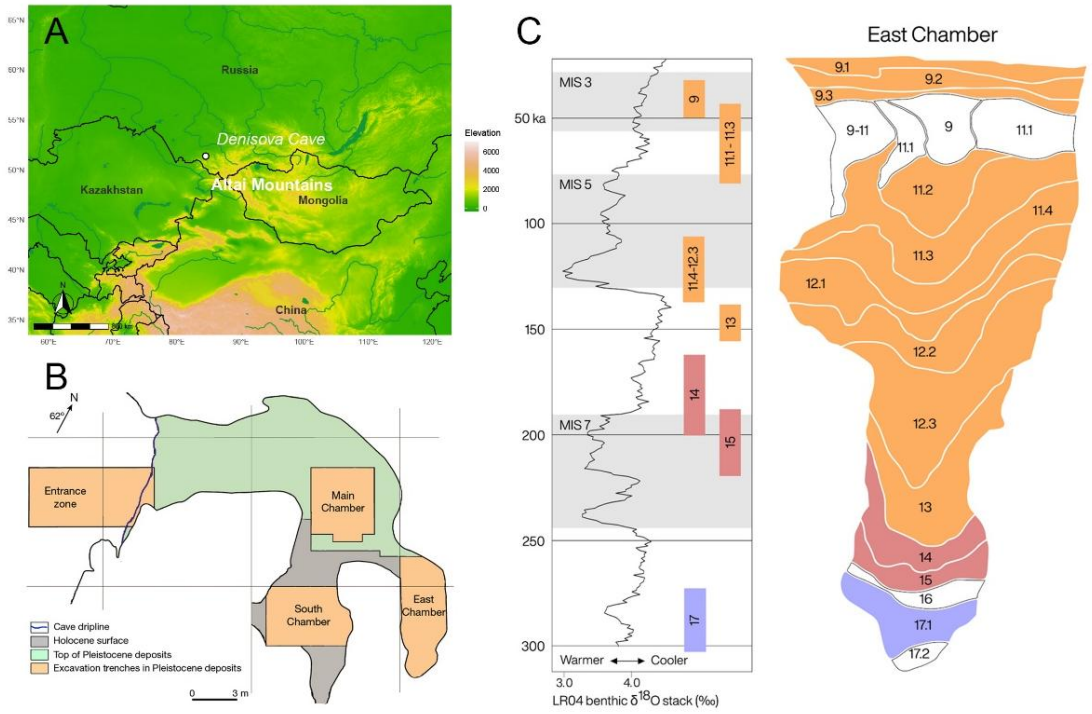



**Figure 1.4. Location and structure of Denisova Cave. (A)** Regional map of Central Asia and site location. The Altai Mountain range is labeled, and the location of Denisova Cave is shown with a white dot. The map was constructed in R and the scale bar may be inaccurate. **(B)** Plan of the cave interior (Jacobs et al., 2019). **(C)** Stratigraphy of the East Chamber (right). The age range of the layers is superimposed on a timescale of Earth's paleoclimate, using marine-oxygen isotope ratio as proxy (left) (Brown et al., 2021). Grey: interglacial intervals; White: glacial intervals.

In Brown et al. (2021), over 8,000 bone fragments from Denisova cave, which prior to this study could not be placed into specific taxa due to being morphologically indistinct, were investigated with Zooarchaeology by Mass Spectrometry (ZooMS). Through the targeted analysis of type I collagen, the main protein present in mineralized bones (Buckley et al., 2009), Brown et al. screened and identified these bones to family/genus level, reconstructing the faunal and hominid diversity at Denisova Cave.

This study also furthered the understanding of taphonomic processes responsible for the extreme state of fragmentation of bones at the site. The differences in the ratios of prey-predator remains between the fragmented assemblage and the more intact specimens suggest a preservation bias. The fragmented assemblage is composed primarily of herbivores (Brown et al., 2021), whereas most of the better-preserved remains belong to carnivores. Furthermore, the fragments show evidence of both hominin processing, as well as gnawing and digestion marks. Together, these findings are interpreted as evidence that bone fragmentation resulted from a combination of hominin hunting activities and carnivore scavenging (Brown et al., 2021).

Among these fragmented bones, ZooMS analysis determined 547 as belonging to the Rhinocerotidae family. However, due to the limitations of ZooMS for large mammals (Buckley et al., 2014), exacerbated in the analysis of deep-time samples, even the most well-resolved results could only determine taxonomy to the family level. Reassessing this taxonomy through a paleogenomic approach could therefore yield a more precise identification to genus and species level. The genomic information recovered from such a vast assemblage is ideal for adding complexity to the existing rhinoceros phylogeny.

### 1.5. Ancient DNA: Challenges and Approaches

The study of paleogenomes can provide relevant insights into the taxonomy and evolutionary history of deep-time samples. However, the state and characteristics under which ancient DNA (aDNA) molecules remain in the samples pose challenges, requiring specific methods suited to its recovery and analysis.

After the death of an organism, DNA suffers degradation and is fragmented to increasingly smaller size, as a result of hydrolysis as well as microbial and endogenous nuclease activity. These shorter fragments are more difficult to isolate through regular methods of DNA recovery than modern DNA, which occurs in much larger fragments and in greater number. Additionally, the chemical mechanisms through which fragmentation occurs induce the accumulation of miscoding lesions that result in information loss. Fragmentation is most often the result of hydrolytic depurination and deamination, though DNA crosslinking and oxidative damage are also involved in this process. Hydrolytic depurination will, over time, result in an overrepresentation of purines

adjacent to strand breaks and interior gaps. Hydrolytic deamination, on the other hand, coverts cytosine bases to uracils, which are sequenced as thymine bases (Dalén et al., 2023; Fulton & Shapiro, 2019), in a process referred to as C→T transitions.

aDNA is therefore highly fragmented (the majority of fragments is often <35 base pairs) (Dalén et al., 2023), chemically damaged, and present in very small quantities. These characteristics demand specific precautions throughout both the laboratory procedures and downstream computational analyses to ensure the recovery of authentic ancient sequences. These precautions extend to the characteristics of the facilities. In the study of aDNA, contamination with modern DNA can signify a loss of nearly all available information, as it overwhelms the small quantities of endogenous DNA available. aDNA laboratories are therefore built with physical separation from other facilities where modern or amplified DNA might be processed, with strict environment control (temperature and air systems regulation), as well as specific user procedures (Fulton & Shapiro, 2019; Orlando et al., 2021). These measures are set to minimize the risk of contamination.

Advances in DNA recovery techniques and the field of next-generation sequencing have enabled researchers to isolate increasingly older DNA molecules, pushing the boundaries of what was previously thought possible (Kjær et al., 2022; van der Valk et al., 2021). aDNA recovery is made possible through extraction protocols optimized for the retention of short DNA fragments using, for example, entrapment in silica spin columns (Dabney & Meyer, 2019). These more sensitive methods can be used to isolate DNA molecules of samples, even when the conditions of preservation are not ideal (Dabney et al., 2013).

Optimization of the library generation protocols for degraded DNA is also required since Illumina sequencing tends to perform poorly with damaged DNA (Gansauge & Meyer, 2013; Kapp et al., 2021). Different methods have been devised, using single or double stranded approaches to library making, each of which have its benefits and drawbacks. Both approaches were implemented in this project.

The protocol developed by Meyer & Kircher (2010) is devised for double-stranded, end-repaired libraries. When implemented in this project, it was preceded by uracil-DNA-glycosylase and endonuclease VIII (USER) treatment, which further repairs damage by removing uracils. The construction of libraries with the Meyer & Kircher protocol has many purification steps, which leads to the loss of many molecules. In this sense, single-stranded approaches such as the Santa Cruz protocol (Kapp et al., 2021) are more sensitive to short or low quantities of DNA. Library construction occurs in a single reaction, followed by a clean-up step, meaning less molecules are lost between intermediate steps. Because the Santa Cruz protocol does not integrate with an initial USER treatment, however, the resulting libraries do retain the damage patterns of ancient DNA.

This is not necessarily a disadvantage, since although C→T transitions can lead to biased sequence analysis, the presence of these and other damage patterns can serve as a method of authentication of genuine ancient molecules. The choice between double-stranded or single-stranded approaches therefore comes down to the goal of the experiment. In this project, the majority of the screening was conducted on single-stranded libraries, whereas to obtain better quality sequences for the deeper sequencing analyses, double-stranded libraries were used.

The computational processing of ancient DNA sequencing data is vital in reducing biases in the analysis and interpretation of the results. Many pipelines have been developed to handle the processing and streamline the mapping of degraded DNA. The pipeline used in this project is DNAharvester, an in-house pipeline developed at the Centre for Palaeogenetics (Sharif et al., in prep.). DNAharvester processes the sequencing data from next generation sequencers (NGS), removing adapter sequences before generating consensus sequences out of the overlapping reads. After mapping to the reference genome, DNAharvester runs MapDamage2 (Jónsson et al., 2013) and AMBER (Dolenz et al., 2024). Both are tools which can be used to assess and quantify post-mortem DNA damage patterns, aiding in authentication.

These and other measures not explicitly mentioned were devised in the field of aDNA for sensitive recovery and processing of genuine ancient fragments, as to offer confidence to the interpretation of the results. Other such methods employed in this project are elaborated on in later sections.

### 1.6. Main Goals

Building on the preliminary taxonomy established by Brown et al. (2021), this project aimed to assess aDNA preservation of ZooMS-identified Rhinocerotidae bone fragments found at Denisova Cave, dating back to the late Middle Pleistocene, between *ca.* 200 - 105 ka. Employing methods for aDNA retrieval and processing, we obtained mitochondrial data which was then used to resolve taxonomy of samples to the species level. Additionally, mitochondrial genomes (mitogenomes) were generated to assess population structure.

The large assemblage recovered from the cave was discovered across several stratigraphical layers of Denisova Cave. The employment of radiocarbon and luminescence dating (Douka et al., 2019; Jacobs et al., 2019) has yielded a chronology of the site, providing context as to the age of the remains. This chronology further enables the study of patterns in the remains' deposition through the layers (Brown et al., 2021). Particularly, using marine-oxygen isotope stages (MIS) as a proxy, it is possible to consider the remains in the context of the Earth's paleoclimate (Figure 1.4 C). Briefly, MIS correspond to units of time relative to a deficit or relative abundance of the $^{18}O$ isotope found at oceanic layers, resulting from the retreat or

formation of glacial ice sheets, respectively. In this sense, even-numbered stages correspond to periods with high levels of $^{18}O$, representing glacial periods, while odd-numbered stages represent interglacial periods, with low levels of $^{18}O$ (Elias, 2013).

Analyzing samples from consecutive stratigraphic layers and across subsequent climatic cycles should enable the comparison of population structure through climate transitions. The contrasting characteristics (both observed and inferred) of the Rhinocerotidae species would suggest that their relative presence was influenced by the climate of the period. As mentioned previously, genomic data (Lord et al., 2020; Yuan et al., 2023) has indicated the Middle Pleistocene as a period with intricate population dynamics, reflected in the genetic diversity of woolly rhinoceroses from the Late Pleistocene.

We have hypothesized that the remains of different species of rhinoceros would be represented in the Denisova Cave assemblage in patterns corresponding to past climate conditions. Specifically, we expected to find the cold-adapted *C. antiquitatis* primarily in the layers corresponding to the glacial period (MIS 6). Conversely, we expected to find a greater number of *Stephanorhinus* remains in the layers corresponding to interglacial intervals (MIS 7 and MIS 5). Thus far, no fossil evidence of *E. sibiricum* has been discovered at Denisova Cave, though some records exist of its presence in the region of the west Siberian plains (Schvyreva, 2015). Denisova Cave is located at the very edge of the geographical range where remains have been found (Kosintsev et al., 2018), making the presence *E. sibiricum* in the assemblage uncertain. If this species were to occur in the assemblage, however, we might expect to find the remains in interglacial-corresponding layers.

This project contributes to a broader effort to resolve the faunal assemblages at Denisova cave, refining the taxonomy and diversity of Rhinocerontidae at the site. It further seeks to resolve rhinoceros population structures from previous analogous studies (Lord et al., 2020; Yuan et al., 2023), adding samples from an earlier age and broadening the geographical range.

# 2. Materials and Methods

## 2.1. Sample information

From the available collection, 135 bone fragments were sampled (Supplementary Table 1), distributed proportionally across layers of the East chamber of Denisova Cave, to produce a well-spread representation of each climatic period (Table 2.1). The attribution of a stage is based on which stage is primarily represented by the temporal range estimated for that layer. As seen in Figure 1.4 C, while layer 14 primarily corresponds to MIS 6, its range also includes the final

stages of MIS 7. Similarly, layers 15 and 11.4 – 12.3 are mainly associated with MIS 7 and 5 respectively but also encompass portions of MIS 6.

**Table 2.1. Distribution of the fragments sampled in this study through layers of the East Chamber and corresponding marine-oxygen isotope stage.** Glacial stage: MIS 6; Interglacial stages: MIS 7 & 5.

| | **MIS 7** | **MIS 6** | | **MIS 5** | |
|---|---|---|---|---|---|
| **Layer** | *Layer 15* | *Layer 14* | *Layer 13* | *Layer 12.1-12.2* | *Layer 11.4-12* |
| **Nº of fragments** | 37 | 16 | 40 | 3 | 39 |
| **Total** | **135** | | | | |

## 2.2. Bone fragment sampling

The laboratory procedures described were conducted at the Centre for Palaeogenetics (Stockholm University, Sweden) in dedicated facilities designed for the handling of aDNA.

Each bone fragment was sampled using clean wire cutting tools, using the pressure to pulverize the bone and expose internal fractures. These sampled fragments were collected (between 22 and 88 mg) and bleached in a < 0.5% sodium hypochlorite solution for about 4 min to reduce exogenous surface contamination. The fragments were then rinsed three times in UltraPure DNase/RNase-Free Distilled Water (Invitrogen), to wash off the bleach.

Some of the samples were instead obtained by drilling. Prior to sampling, the bone fragment was surface-cleaned using a circular drill tip adapted to a Dremel tool. Subsequentially, the fragment was drilled with a sharp-edge drill tip to produce bone powder (between 11 and 28 mg). These powders were not subjected to bleach treatment prior to lysis.

In total, 141 samples were collected, of which nine correspond to samples obtained from the same three fragments (for each of these bone fragments, one sample was obtained through cutting and two with drilling).

## 2.3. Lysis and extraction

The lysis and DNA extraction followed the protocol published by Dabney & Meyer (2019), with modifications for a higher throughput, as described by Gilardet et al. (2025).

The samples were mixed with 1 mL of lysis buffer (900 µL EDTA 0.5 M, pH 8; 75 µL UltraPure DNase/RNase-Free Distilled Water (Invitrogen); 0.5 µL Tween-20; 25 µL Proteinase K 10 µg/µL) and incubated at 37ºC with continuous motion (~12 rpm), to digest as much of the bone as possible. Depending on the size of the samples, incubation lasted between 17 (overnight) and 96 h. If samples were not fully digested after 48 h, an additional 20 µL of Proteinase K was added, and digested for the remaining 48 h. Some samples did not fully digest at the end of the 96 hours, with some bone fragments remaining in the pellet.

Lysates were collected and stored at -20ºC until downstream DNA extraction.

Modifications to the methods of Dabney & Meyer (2019), based on recommendations from the QIAquick 96 PCR Purification Kit (Qiagen), enabled the use of a 96-column plate during extraction and downstream library purification (Gilardet et al., 2025). This adaptation facilitated the efficient processing and screening of the large collection of aDNA samples.

The lysates were thawed to room temperature and centrifuged (6,000 *g* for 3 min) to pellet bone fragments remaining. An aliquot of the lysate was then collected (100 µL) and mixed with 1 mL of binding buffer (5 M guanidine hydrochloride in Ultrapure DNase/RNase-Free Distilled Water (Invitrogen); 40% v/v isopropanol) and 40 µL of sodium acetate (3 M, pH 5.2). The lysate-binding buffer mix was loaded onto the plate columns, and the plate was mounted onto the assembly of a vacuum manifold system QIAvac 96 (Qiagen) coupled to a vacuum pump. The vacuum (-800 to -600 mbar) was used to bind the DNA to the column. The columns were then washed twice with 900 µL of PE buffer (Qiagen), after which the maximum vacuum (~ -300 mbar) was applied for 10 min. The plate was ventilated and dried over absorbent paper, to eliminate PE wash buffer residue. The extracts were eluted with 60 µL of TET buffer (Qiagen; Tris-HCl 10 mM, pH 8.5) into the collection tubes after about 5 min incubation, by applying vacuum (-800 to -600 mbar) for 5 min.

Extracts were stored at -20ºC.

### 2.4. Library preparation

Libraries were generated from extracts RT009 – 088 and RT120 – 125 following the single-stranded Santa Cruz Reaction protocol for low-input DNA (Kapp et al., 2021), with modifications from Nguyen et al. (2023), until the purification step. Library purification was conducted on a 96-column plate following the same procedures as the extraction. Briefly, 125 µL PB buffer (Qiagen) were added to 25 µL aliquots of the libraries and loaded into a 96-column QIAquick plate. Following the vacuum procedure previously described, the columns were washed twice with 900 µL of PE buffer (Qiagen), after which the purified libraries were eluted with 45 µL of TET buffer (10 mM Tris-HCl, 1 mM EDTA, 0.05% Tween-20).

Libraries were generated from extract RT089 – 119 and RT126 – 149 following the protocol for double-stranded Illumina libraries by Meyer & Kircher (2010), with modifications from Kircher et al. (2012), including the step for USER treatment. Library clean-up steps were conducted on 96-column QIAquick plate (Qiagen) or MinElute spin columns (Qiagen). In either case, 200 µL PB buffer (Qiagen) were added to 40 µL aliquots of the libraries and loaded into the column. The vacuum procedure previously described was used on the plate, while a centrifuge was used on the MinElute spin columns (1 min, 13 000 rpm). The columns were then washed

twice with 700 µL of PE buffer (Qiagen), after which the purified libraries were eluted with 22 µL of EB buffer (Qiagen) with 0.05% Tween-20.

Following purification, quantitative PCR (qPCR) was used to assess the quality of the libraries generated, according to recommendations from the corresponding library preparation protocol (Kapp et al., 2021; Meyer & Kircher, 2010). Maxima™ SYBR® Green qPCR Master Mix (2X) and primers 10 µM IS7 and 10 µM IS8 (Gansauge & Meyer, 2013) were used for the reaction, which was run in a real-time PCR detection system (CFX96 Touch Real-Time PCR Detection System, Bio-Rad).

The quantification cycle (Cq) of each library was later used to establish the number of cycles for indexing. In this step, the libraries were indexed through PCR, following the recommendations of the corresponding protocols. The composition of the reaction mixes, thermocycler programs and indexing primers used are detailed in Supplementary Table 2. The PCR was run in a thermocycler (DOPPIO Thermocycler, VWR) with the lid heated to 110ºC.

The results of the indexing PCR were confirmed by electrophoresis in 2% agarose gel in TAE 1X buffer. Based upon the intensity of the band in the gel, the indexed libraries were pooled in amounts that would result in roughly equal concentrations in the pool, which was then bead cleaned and size-selected with Sera-Mag Select (Cytiva) beads, using 0.5X beads to pool ratio to remove fragments larger than the libraries, and 1.8X then 1.5X for smaller fragments, of which adapter dimers. The success of this procedure was verified in the TapeStation (4150 TapeStation system, Agilent) with high sensitivity D1000 DNA screentape, following the manufacturer's protocol.

The indexed libraries were sequenced at the National Genomics Infrastructure (Science for Life Laboratory, Stockholm, Sweden) on the NovaSeq X platform (Illumina) using a paired-end 2 x 150 bp set up.

Throughout the laboratory workflow, blanks and positive controls were added to assess contamination and the efficiency of the workflow. One blank was added for each lysate batch, two for each extraction plate, as well as one blank and one positive control (previously validated extract) for each library plate. Reaction blanks were added to the qPCR and indexing PCR plates as well. Controls with low qPCR Cq values were sent for sequencing in the same pool as the respective libraries to assess potential contamination.

Following sample screening, to improve read depth of samples of interest, full lysate extraction was performed on the selected samples, following the protocol published by Dabney & Meyer (2019). Libraries were then generated from the extracts, following the protocol

published Meyer & Kircher (2010), with Blunt-End repair through USER treatment in MinElute spin columns (Qiagen).

Library quality assessment and indexing were performed as previously described. The libraries were purified with Sera-Mag Select (Cytiva) beads and DNA concentration verified in the TapeStation (4150 TapeStation system, Agilent) with high sensitivity D1000 DNA screentape, following the manufacturer's protocol. Afterwards, the samples were pooled and sent for sequencing.

### 2.5. Sequencing data processing and analysis

The sequencing data was processed using the DNAharvester pipeline (Sharif et al., in prep.). Raw reads were mapped to the reference genome of the closest living relative of the extinct rhinoceroses, the Sumatran rhinoceros (*Dicerorhinus sumatrensis harrissoni*, GCA_014189135.1), in which the mitochondrial region was replaced with the reference mitogenome of woolly rhinoceros (NC_012681.1). Mapping statistics were then determined (Supplementary Table 3). Endogenous content was calculated as the ratio of mapped read to merged reads (that is, reads that, after quality pre-processing, enter the mapping step). Nuclear and mitochondrial depths were calculated using samtools depth (v1.20) (Danecek et al., 2021) with flags -a -q 30 -Q 20 -b, using the BED file for the reference genome or the mitochondrial reference genome, respectively.

The raw reads were also mapped competitively to a panel of reference mitogenomes using BWA aln (v0.7.18) (Li & Durbin, 2009). This was done in two rounds of alignment: first, the samples were aligned to a reference panel containing geographically relevant bait species (Supplementary Table 4), in order to infer the prevalence of ancient rhinoceros signal in the libraries and investigate possible contaminants and by-catches. Duplicates were removed using the same method implemented by the pipeline, with modifications made for implementation into the ancient DNA reads processing pipeline GenErode (Kutschera et al., 2022). Reads aligned to the woolly rhinoceros reference mitogenome (*C. antiquitatis*, NC_012681.1) were isolated and aligned to a second reference panel, containing only the rhinoceros mitogenomes (*C. antiquitatis*, NC_012681.1; *S. kirchbergensis*, OQ730110.1; *E. sibiricum*, MH937513.1).

Cumulative analysis of the mapping statistics and competitive mapping results informed the selection of samples for deeper sequencing. The samples were selected on the basis of endogenous content (> 5%), mitochondrial depth (> 0.1X) and/or competitive mapping profile dominated by ancient rhinoceros hits. Competitive mapping of samples with mitochondrial depth greater than 0.1X is shown in Figure 3.1, competitive mapping of all samples can be found in Supplementary Figure 1.

### 2.6. Mitogenomes and phylogenetic analysis

Iterative mapping was performed on selected samples in order to assemble consensus mitogenomes. Pre-processed fastq files generated by DNAharvester of both screening and deeper sequencing libraries of each sample were merged. These files were used as input for MIA (v1.0) (https://github.com/mpieva/mapping-iterative-assembler.git) assembly, using the parameters given by flags -c -C -U -F -k 14. An ancient DNA substitution matrix was also provided with flag -s to account for the damage patterns of the non-USER-treated libraries. For each sample, the ancient rhinoceros species with the most number of hits in competitive mapping was used as seed: on all samples but one, the *C. antiquitatis* mitogenome (NC_012681.1) was used as a seed; RT082 was assembled from the *S. kirchbergensis* reference mitogenome (OQ730110.1).

For samples with a depth of mitochondrial depth greater than 3X after assembly, the MIA-associated ma program with flag -f 41 was used to parse the output maln file. From this .41 output file, a consensus sequence was generated using create_fasta_from_mia.py (https://github.com/aersoares81/mia-helperscripts), filtering out sites with less than 3X depth and 67% identity. For samples with a depth of mitochondrial coverage between 1 – 3X after MIA, the processed BAM file with duplicates removed given by the DNAharvester pipeline was processed with adna_sslib_damage_removal.py (https://github.com/bilalbioinfo/aDNA_Damage/blob/main/adna_sslib_damage_removal.py) to remove sequence damage. ANGSD (v0.940) (Korneliussen et al., 2014) was then used to make a fasta sequence, filtering sites according to a minimum depth of 0.5X, minimum base quality of 30 and minimum mapping quality 25.

To generate the woolly rhinoceros phylogenetic tree, the mitogenomes generated in this study were aligned with those from Lord et al. (2020) and Yuan et al. (2023). Merck's rhinoceros (*S. kirchbergensis*, OQ730110.1) and/or Sumatran rhinoceros (*D. sumatrensis,* NC_012684.1) were used as outgroup. In a few diagnosis trees, a further two *S. kirchbergensis* mitogenomes (OQ889160.1 and KX646743.1) and/or a partial woolly rhinoceros mitogenome (Seeber et al., 2023) were also introduced into the alignment. The sequences were aligned with muscle (v3.8.31) (Edgar, 2004), then filtered for maximum 20% missingness with fasta_nomissing.py (https://github.com/pontussk/fasta_nomissing.py). Maximum-likelihood phylogenetic trees were generated with IQ-Tree (v2.3.5) (Minh et al., 2020), using Model Finder (Kalyaanamoorthy et al., 2017) to determine the most likely model and ultrafast bootstrapping (Hoang et al., 2018), with 1000 bootstraps. FigTree (v1.4.4) (https://github.com/rambaut/figtree) was used to visualize the trees.

A haplotype network was generated from the woolly rhinoceros mitogenomes with an average depth of coverage > 3X. Haplogroups were determined with DnaSP (v6.12.03) (Rozas et

al., 2017), removing all invariable and missing sites. The median-joining haplotype network was plotted on PopART (v1.7) (Bandelt et al., 1999).

R (v4.3.3) was also used to visualize data and make the figures. The packages used are listed in Supplementary Table 5.

### 2.7. Investigation of by-catch genomic data

One sample from this study (RT108), as well as previously generated libraries from bovid samples (Gilardet & Oppenheimer et al., unpublished data) were investigated for non-endogenous genomic data. These samples stood out as objects of investigation due to strong signal for a non-target species in the first round of competitive mapping, that is a high number of aligned reads. For two samples (AG031 and AG083), *Crocuta crocuta* (spotted hyena) yielded more hits than the target species, which had been bovid. For three samples, including the one processed in this study (AG076, AG216 and RT108), while the first most hits were on the target species (bovid or rhinoceros in either case), there was also a very high number of hits to a secondary species, *C. crocuta*. For another two samples, *Canis lupus* (wolf) and *Ovis aries* (sheep) had the second-most number of hits (AG004 and AG262, respectively).

The raw sequencing data was re-mapped to the non-target species, hereafter referred to as by-catch species, using the methods as described in section 2.5. Iterative mapping was then performed on samples with a depth of coverage at the mitochondria for the by-catch species mitogenome greater than 0.1X, as described in 2.6. The reference genomes used in the mapping and reference mitogenomes used as seed for MIA can be found in Supplementary Table 6, as well as a summary of sequence statistics.

The phylogenetic placement of samples within their respective species was investigated for samples with a mitochondrial depth greater than 1X, AG031 and AG083, both of which hyena by-catches, following the methods described in section 2.6. The hyena by-catch samples were aligned with 43 spotted and cave hyena mitogenomes (*C. crocuta* and *C. crocuta spelaea,* respectively) (Bon et al., 2012; Hu et al., 2021; Krajcarz et al., 2023; Westbury et al., 2020), and the reference mitogenomes of striped hyena (*Hyaena hyaena*, NC_020669.1), brown hyena (*Parahyaena brunnea,* NC_038159.1) and aardwolf (*Proteles cristata*, MH662445.1), which were used as outgroup.

# 3. Results

## 3.1. Report on the screening and taxonomical assignment

Out of 141 samples screened, 19 met one or more of the criteria for deeper sequencing and analysis. The screening libraries of samples RT051, RT053, RT061, RT065 and RT087 had an endogenous DNA content > 5% and mitochondrial depth > 0.1X; RT057 and RT088 had endogenous > 5% (mitochondrial depth: RT057 = 0; RT088 = 0.06X), and RT040, RT041 and RT086 mitochondrial depth > 0.1X (endogenous content: RT040 = 0.64%; RT041 = 4.38%; RT086 = 1.53%). Despite low endogenous (0.143%) and poor mitochondrial depth (0.045X), sample RT082 was selected for deeper sequencing based on the second-round competitive mapping profile, when mapping exclusively to the three rhinoceros' species mitogenomes. Most samples with reads mapping to the *S. kirchbergensis* mitogenome also had a higher number of reads mapping to *C. antiquitatis* in the second round of alignment, when reads were isolated for targeted mapping, resulting from the similarity between the mitogenomes of the rhinoceros species. Unlike these samples, however, sample RT082 was unique in mapping exclusively to *S. kirchbergensis* when the reads mapped to the rhinoceros' species panel (Figure 3.1 C).

A summary of the mapping statistics of deep sequenced samples is presented in Table 3.1.

**Table 3.1. Summary of mapping statistics of deep sequenced samples.** Data shown here corresponds to the mapping of the screening and deep sequencing library for each sample combined. "Endo content" refers to the endogenous content, calculated from the reads mapped/total reads. "Nuclear depth" refers to depth of coverage of the nuclear genome. "Mito depth" and "Mito depth (after MIA)" refer to coverage of the mitochondrial genome, before and after mitogenome reconstruction.

| Sample | Layer | Endo content | Nuclear depth | Mito depth | Mito depth (after MIA) | Species Assignment |
|---|---|---|---|---|---|---|
| RT040 | 13 | 0.361% | 0.0027015 | 1.87947 | 2.268 (4 iterations) | *C. antiquitatis* |
| RT041 | 13 | 3.844% | 0.0517407 | 7.45084 | 8.717 (2 iterations) | *C. antiquitatis* |
| RT051 | 12 | 68.929% | 1.26146 | 40.6648 | 51.083 (5 iterations) | *C. antiquitatis* |
| RT053 | 12 | 53.964% | 0.868678 | 32.3505 | 40.115 (6 iterations) | *C. antiquitatis* |
| RT057 | 12 | 9.580% | 0.0891365 | 0.319664 | 0.793 (6 iterations) | *C. antiquitatis* |
| RT061 | 11.4 | 1.124% | 0.0226222 | 1.33968 | 1.548 (6 iterations) | *C. antiquitatis* |
| RT065 | 11.4 | 64.920% | 0.340086 | 8.51667 | 11.164 (7 iterations) | *C. antiquitatis* |
| RT082 | 15 | 0.059% | 0.000813153 | 0.785897 | 1.567 (6 iterations) | *S. kirchbergensis* |
| RT086 | 12.1-12.2 | 8.840% | 0.0701463 | 28.3076 | 32.602 (2 iterations) | *C. antiquitatis* |
| RT087 | 12.1-12.2 | 15.585% | 0.306414 | 28.3401 | 33.737 (4 iterations) | *C. antiquitatis* |
| RT088 | 12.1-12.2 | 84.216% | 0.693175 | 13.7312 | 18.195 (8 iterations) | *C. antiquitatis* |

Two additional samples, RT113 and RT137, also fulfilled both the endogenous content and mitochondrial depth thresholds. Another 5 (RT108, RT117, RT119, RT129 and RT136) had endogenous content above the threshold. Deeper sequencing was not done on these 7 samples due to time constraints.

For all these samples, the reads recovered were short on average (< 50 bp) and presented patterns of C→T transition concurrent with post-mortem DNA damage. Given its fragmented and damaged state, the genomic data recovered was determined as aDNA. We also investigated the reads recovered from the sequenced blanks and, with the exception of significant endogenous reads present in the extraction blank RTBL5, we found no further evidence of contamination throughout the laboratory procedures. RTBL5 was analyzed, to assess the type of contamination and the effect on the lysis batch. In summary, the low number of reads found at the mitochondrial region did not significantly impair the analysis of samples in that region. For that reason, the samples of the corresponding batch (RT063 – RT087) were incorporated into the analysis in this study. Further context and implication of the contamination of RTBL5 is discussed in detail in Annex I.

Taxonomy was determined for all samples that yielded mitochondrial depth greater than 0.1X (from either screening or deeper sequencing libraries, or the merged libraries), assigned on the basis of the species with the greatest number of hits on competitive mapping (Figure 3.1 C and D). For samples with mitochondrial depth over 1X this taxonomy was then validated by the phylogeny (Supplementary Figure 2 and 3). From this dataset, 12 samples were assigned as woolly rhinoceros (*C. antiquitatis*), and one sample was assigned as Merck's rhinoceros (*S. kirchbergensis*). Competitive mapping (Supplementary Figure 1) further suggested that sample RT109 was bovid, rather than rhinoceros, specifically yak (*Bos mutus*, with close secondary contender *Bison priscus*, the steppe bison). The competitive mapping profile for this sample did not have any secondary genomic signals that would suggest by-catch, unlike the samples reported on section 3.3. We consider this bovid sample was mixed in with the rhinoceros samples by mistake.

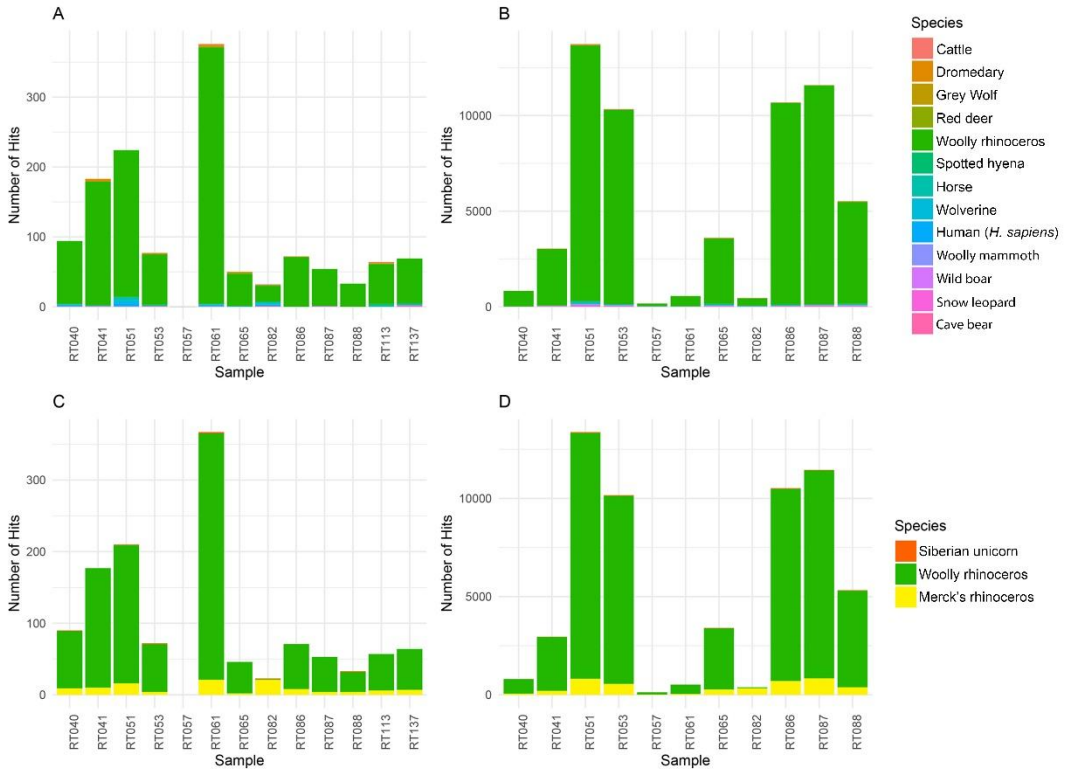



**Figure 3.1. Competitive mapping results for samples with mitochondrial depth >0.1X.** Competitive mapping results are shown as number of hits to each species on the general (first alignment round) and target-specific (second alignment round) panels. Rhinoceros species on the second panel are Siberian unicorn (*E. sibiricum*), woolly rhinoceros (*C. antiquitatis*) and Merck's rhinoceros (*S. kirchbergensis*). Results correspond to the first alignment round on **(A)** screening and **(B)** merged libraries (screening + deep sequencing), and to the second alignment on **(C)** screening and **(D)** merged libraries (screening + deep sequencing) libraries.

## 3.2. Denisova Cave rhinoceros placement in the woolly rhinoceros mitochondrial phylogeny

The seven mitogenomes with depth of coverage greater than 3X (between 8.7X and 51.1X, see Table 3.1) generated in this study were aligned with previously available woolly rhinoceros mitogenomes from Lord et al. (2020) and Yuan et al. (2023) to generate the phylogenetic tree in Figure 3.2. The three major clades A, B and C established in Yuan et al. (2023) are present. Clade D, which was previously represented only by sample CADG744, is now better categorized with the addition of the Denisova Cave samples. However, we infer different phylogenetic relationships between major clades compared to previous work. Specifically, in our tree, clades B and D split from A – C and then from one another; whilst in Yuan et al. (2023), an initial split between D and A – C – B was observed, after which the split between B and A – C occurred. The clustering of clades D and B in Figure 3.2 is poorly supported with 30% bootstrapping value.

It was also observed that, in consensus trees generated from alignments with lower depth of coverage samples (Supplementary Figure 2) and containing a partial European woolly rhinoceros mitogenome (Supplementary Figure 3) we inferred the same topology as the previous studies. Nonetheless, the support toward the clustering of clades A – C – B in these two trees was

also poor, with the grouping of these clades occurring in 57% of bootstraps on the alignment with low-coverage samples (Supplementary Figure 2), and in 61% of bootstraps on the alignment containing the European mitogenome (Supplementary Figure 3).

Along with this topological difference, it was also observed that sample RT087 clustered either with clade D (with 48% bootstrap support) as seen in Figure 3.2, or clade B, as seen in Supplementary Figure 2 and 3 (with 53% and 52% bootstrap support value, respectively). Additionally, in both cases, RT087 occupied a basal position within its assigned clade, which comprises samples that are stratigraphically older than RT087.

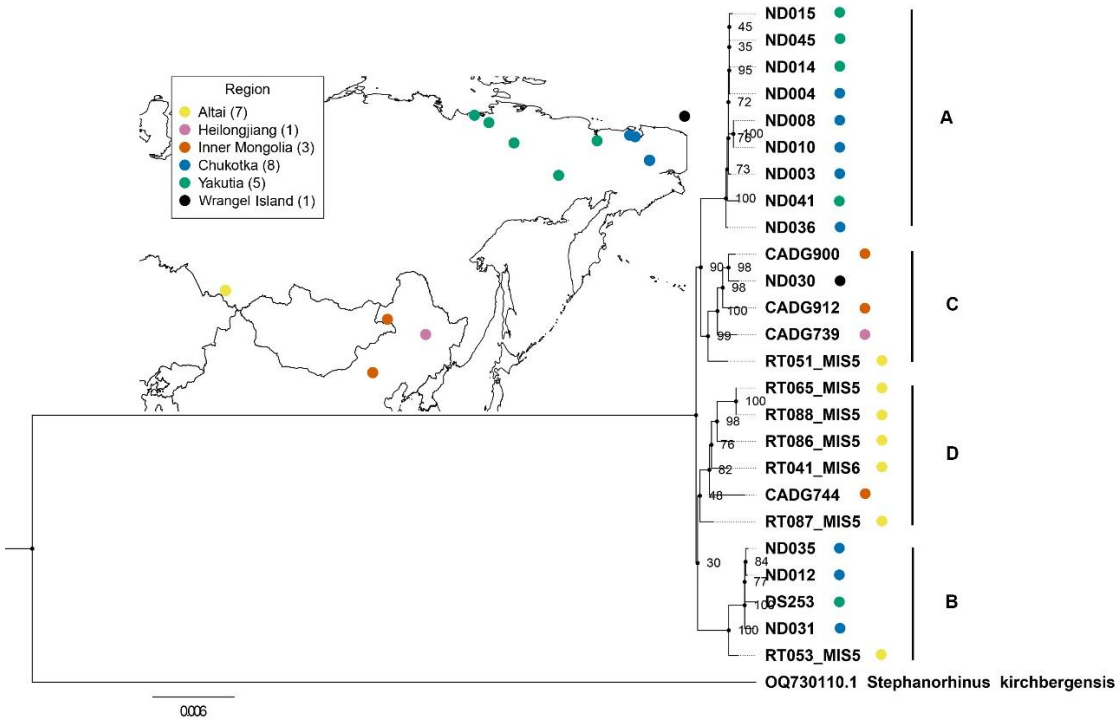
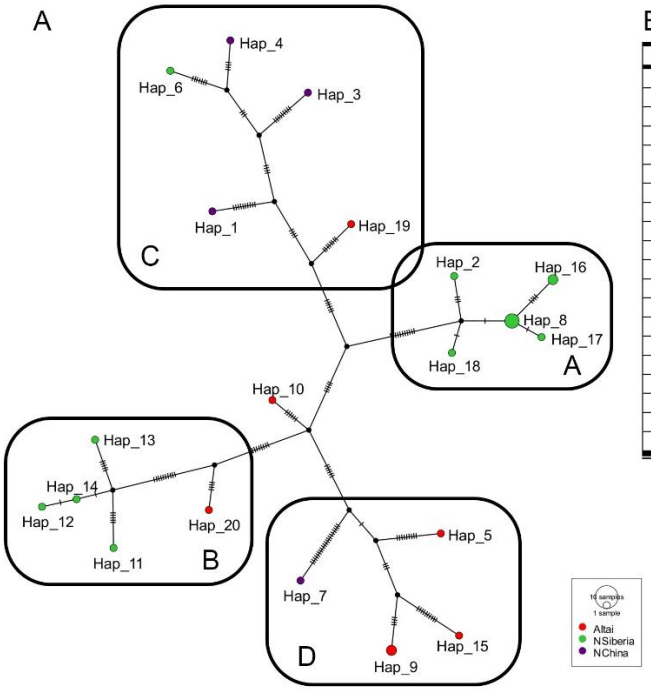



**Figure 3.2. Consensus phylogenetic tree of woolly rhinoceros mitogenomes.** Maximum-likelihood consensus tree of 1000 bootstrap trees. Tree was generated from an alignment of 200 parsimony informative sites using the best-fitting model (TN+F+R3). Bootstrap support values (%) are indicated at each node. Scale represents number of nucleotide replacements per site. *S. kirchbergensis* was used as an outgroup. For the mitogenomes generated in this study, the corresponding marine-isotope stage was added to the taxa label for ease of interpretation. Taxa are color-coded according to their geographical origin, matching the color scheme used in the accompanying map. The map displays the region from which samples used to generate the mitogenomes were sampled, with sample count per region shown in parentheses.

To investigate the relationship between sample RT087 and clades B and D, we generated the median-joining haplotype network in Figure 3.3 A. The haplotype network generated the same pattern as the phylogenetic tree, with four distinct clades. Sample RT087, represented by haplotype 10 in the network, is on a branch that splits from the same ancestral node as clades B and D. Six mutational steps separate the ancestral node from clade D and RT087, and eight from clade B.

A




B

| Haplotype | Samples |
|---|---|
| Hap_1 | CADG739 |
| Hap_2 | ND041 |
| Hap_3 | CADG912 |
| Hap_4 | CADG900 |
| Hap_5 | RT041 |
| Hap_6 | ND030 |
| Hap_7 | CADG744 |
| Hap_8 | ND015, ND004, ND014, ND045 |
| Hap_9 | RT065, RT088 |
| Hap_10 | RT087 |
| Hap_11 | ND031 |
| Hap_12 | ND035 |
| Hap_13 | DS253 |
| Hap_14 | ND012 |
| Hap_15 | RT086 |
| Hap_16 | ND008, ND010 |
| Hap_17 | ND003 |
| Hap_18 | ND036 |
| Hap_19 | RT051 |
| Hap_20 | RT053 |

**Figure 3.3. Median-joining haplotype network of woolly rhinoceros mitogenomes** (A), with rounded squares demarcating the four clades. Haplotype 10 was not placed into either clade B or D. Hatch marks represent the number of mutation sites between two adjacent haplotypes. Haplotypes are color-coded with the region of origin: Altai, namely Denisova Cave; Northern Siberia (NSiberia), namely Chukotka, Yakutia and Wrangle Island; Northern China (NChina), namely Heilongjiang and Inner Mongolia. (B) Samples included in each haplotype. Black dots represent ancient/missing haplotypes.

## 3.3. Mitogenomes generated from by-catch genomic data

Six bovid and one rhinoceros sample were investigated for non-endogenous genomic data. For all these samples, the competitive mapping analysis revealed that the libraries were composed of significant quantities of genomic data from more than one source. For all these at least one of these signals (primary or secondary) matched the results of the ZooMS identification. Furthermore, these libraries yielded endogenous content for bovid or rhinoceros DNA and, for some, it yielded sufficient data to reconstruct the mitogenome (Gilardet & Oppenheimer et al., unpublished data). Within this context, we interpret the strong genetic signal from a non-target species as by-catch alongside the specimen's endogenous DNA.

We generated two partial hyena mitogenomes with depth > 1X from by-catch genomic data of two bovid samples. These sequences covered 70% (AG031) and 40% (AG083) of the hyena mitogenome. The other samples did not yield sufficient data to reconstruct the mitogenomes of the by-catch species. Mapping statistics for all samples investigated for by-catch data are reported in Supplementary Table 6.

The hyena mitogenomes generated were aligned with previously published cave and spotted hyena mitogenomes, ancient and modern. The resulting tree (Supplementary Figure 4) is consistent with the previously established phylogeny (Krajcarz et al., 2023; Westbury et al.,

2020), resolving into four major clades corresponding to haplogroups A through D, and sub-haplogroups A1 and A2. Samples AG083 and AG031 cluster within haplogroup A, which comprises individuals from Eurasia as well as Africa. Within this clade, the two samples are more closely related to the Eurasian hyenas (haplogroup A1), than the sub-haplogroup primarily made up of African hyenas.

The mitogenome generated from sample AG083 covers a substantially shorter portion of the hyena mitogenome compared to other samples in the alignment. As a result of the high amount of gaps and missing data, AG083 is placed in the phylogeny (Supplementary Figure 4) with a long branch.

# 4. Discussion

## 4.1. Temporal patterns of rhinoceros species diversity in Altai

The present work proposed to analyze patterns of rhinoceros species distribution throughout the layers of Denisova Cave and, conversely, through the interglacial and glacial periods of the late Middle Pleistocene. Our results suggest that woolly rhinoceros were continually present in the Altai region through MIS 5 and 6, as evidenced by their remains being found at every layer spanning that period (Figure 4.1). By contrast, no woolly rhinoceros were identified in layers corresponding to MIS 7, and the only sample identified from this stage was Merck's rhinoceros. This is broadly consistent with what has been previously observed in the sedimentary record of the cave, which indicates two major faunal turnover events occurred during the Middle Pleistocene, the earliest of which at the end of MIS 7 (Zavala et al., 2021). Among other changes in the faunal assemblage, this transition is marked by a shift in ursid species from cave bears to brown bears. Similarly, these results could reflect a turnover in the species of rhinoceros that were present in the Denisova Cave region between the interglacial MIS 7 to the glacial MIS 6. However, the evidence presented is not sufficient to substantiate this interpretation.

We speculated that woolly rhinoceros would be found primarily, or in greater number, in layers corresponding to glacial periods, reflecting the adaptative features of the species. Conversely, that Merck's rhinoceros, adapted to moderate environments, would be found in interglacial-corresponding layers. However, the dataset does not allow for an in-depth analysis of such specific interspecies patterns, as the overwhelming majority of samples which yielded sufficient data for taxonomical identification were woolly rhinoceros.

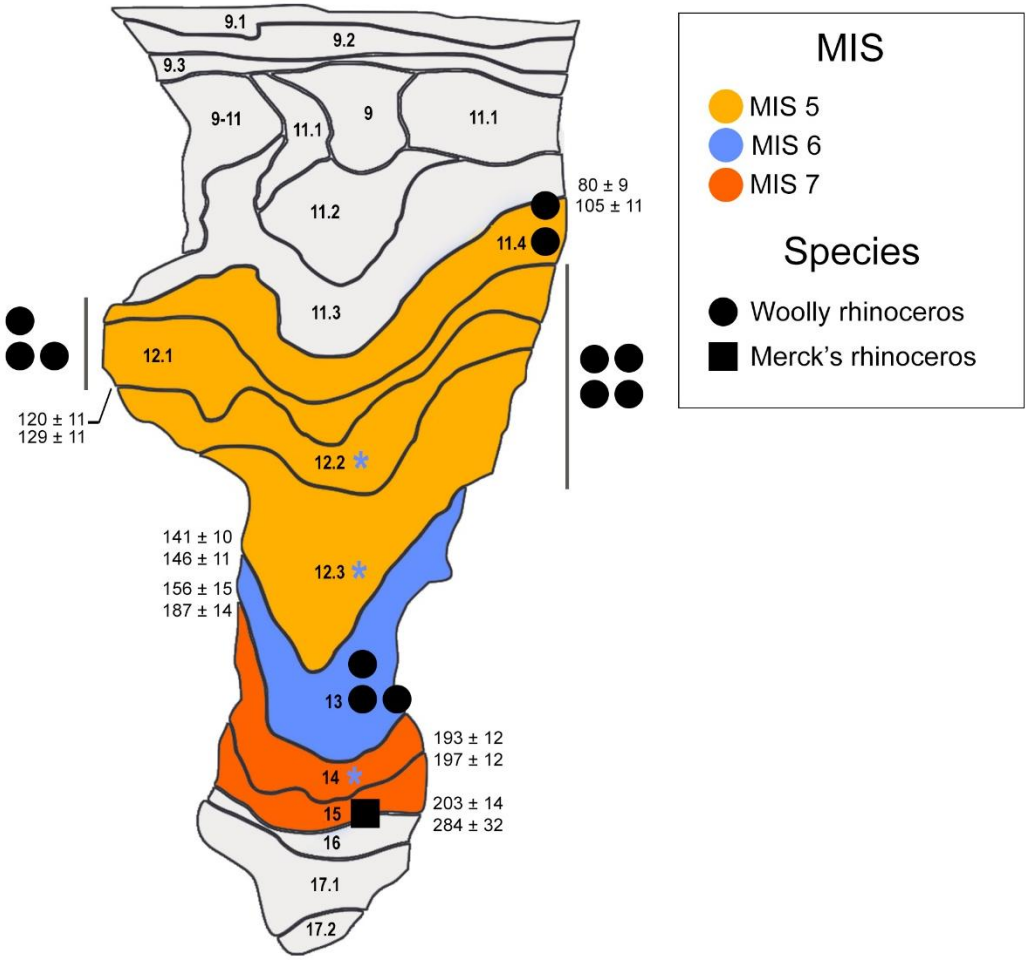



**Figure 4.1. Distribution of taxonomically identified samples through the layers of the East Chamber.** The marine-isotope stages are color-coded for each layer sampled. The blue asterisk next to a layer number indicates the layer partially spans more than one MIS. Dates for layers are shown in ka (Jacobs et al., 2019). The samples are grouped by layer of origin. Samples recorded in layer "12" and layer "12.1 – 11.4" are represented on the side, with a black line marking the full span of the corresponding layer.

Though it is not possible to validate or disprove the proposed hypothesis of rhinoceros species climate-driven turnover in the Altai region from this method alone, certain observed patterns are consistent with it. Firstly, there is the occurrence of the Merck's rhinoceros sample on the layer corresponding to interglacial MIS 7 (Figure 4.1), a warm period. Although MIS 6 is a glacial period, the reconstruction of environment and climate in the Denisova Cave region during this time (Jacobs et al., 2025) indicates that the climate of early MIS 6 (broadly corresponding with the date given to layer 14) was unusually warm in the region, with relation to the global temperature patterns inferred from the marine isotope record. These climatic conditions contrast with the environments the woolly rhinoceros was best suited for, which is to say cold and dry and, correspondingly, we did not find woolly rhinoceros in this layer. The earliest record of woolly rhinoceros in our data corresponds to layer 13 (Figure 4.1) and the late MIS 6, which marked a cold, glacial phase (Jacobs et al., 2025).

MIS 5 represents a climatically unstable period. It is subdivided into five substages, the earliest of which is MIS 5e (*ca.* 116 – 129 ka), during which the last interglacial period of the Pleistocene occurred (Giaccio et al., 2024). MIS 5e was followed by MIS 5d-a, which were much cooler, short glacial/interglacial turnovers. According to the expectations of our hypothesis, fewer woolly rhinoceros samples should be present in layers corresponding to MIS 5e than in those from MIS 6 or MIS 5d-a. However, the stratigraphy of the East Chamber at Denisova Cave during this period does not provide sufficient resolution to distinguish between MIS 5 substages, as layers 12.1 – 11.4 are not individually dated, encompassing MIS 6 and more than one of the MIS 5 substages. In order to test the hypothesis of temporary local extinction or population decrease in the Denisova Cave region during MIS 5e, future studies could compare genetic diversity from rhinoceros samples dated to just before (i.e. late MIS 6) and after MIS 5e (MIS 5d-a). We would expect to find a loss of genetic diversity from MIS 6 to MIS 5d-a, corresponding to a severe bottleneck event during MIS 5e. The current temporal resolution given by the stratigraphical dating would need to be improved/calibrated, in order to distinguish samples from MIS 5e from MIS 5d-a. This could be achieved with molecular dating, further discussed in section 4.3.

### 4.2. Geographical range of woolly rhinoceros clades

In this study, we reconstructed the haplogroup diversity of woolly rhinoceros in the Altai region during MIS 5 and 6, which fits within the greater context of the woolly rhinoceros phylogeny. The inner clustering of the clades from Figure 3.2 is consistent with the previously published phylogenies (Lord et al., 2020; Yuan et al., 2023), with the conflict in major clade grouping address in the results, Section 3.2. Furthermore, the Denisova Cave samples, which according to the stratigraphic context are older than all mitogenomes previously recovered, diverge before the inner branching of the previous, younger specimens. The samples fall into three out of the four established clades. While clade A remains represented only by northern Siberia samples, clade B now encompasses a sample from the Altai region, as well as those from Northern Siberia. One sample generated in this study also falls into clade C, which therefore encompasses northern China, Altai and Wrangel Island. Most of the Altai samples, however, fall within clade D, which is shared with one sample from northern China.

To date, there is a single mitogenome publicly available from a European woolly rhinoceros (Seeber et al., 2023). This partial mitogenome was recovered from a hyena coprolite and covers 60% of the woolly rhinoceros mitogenome after filtering. When aligned with the remaining woolly rhinoceros (Supplementary Figure 3), this sequence clusters with clade D. More European woolly rhinoceros mitogenomes are necessary to clarify the evolutionary history of the species in that region.

Overall, this study identified several mitochondrial lineages co-existing in the Altai region, with ranges spanning large areas. As a result, the woolly rhinoceros phylogeny continues to show no clear geographical structuring. Together, these findings point to an extensive dispersal of the species through their range, with connectivity between the maternal lineagers of woolly rhinoceros across regions. The diversity of lineages identified in this study further indicates to the Denisova Cave region serving as a point of connectivity for populations of different regions, consistent with previous studies, which proposed that Altai functioned as a contact zone for Eurasian mammal populations (Agadjanian & Shunkov, 2018). Alternatively, the basal position of the Altai woolly rhinoceros in the phylogenetic tree (Figure 3.2) may suggest the species expanded out of the Altai to colonize the Northern Siberia and/or China following the Middle Pleistocene. With the current data, it is not possible to discount either interpretation, as the basal position of these samples could equally reflect their older age relative to previously generated genomes.

### 4.3. Resolution of clade divergence

In the phylogeny with the high coverage mitogenomes generated in this study (Figure 3.2), there is low branch support given to the divergence between clades B and D. The structure of the haplotype network of Figure 3.3 A shows that sample RT087 branches off from the node that also corresponds to the divergence between these two clades, forming an effective polytomy. The phylogenetic relationship between this sample and those in clades B and D cannot be fully resolved because the number of mutational steps that separate RT087 from those two clades is very similar. When translated to the phylogenetic tree, this pattern manifests as the short and poorly resolved branch observed in Figure 3.2, which is often associated with rapid divergence events (Whitfield & Lockhart, 2007).

A more extreme case of this is seen with low-coverage sample RT040 (Supplementary Figure 2), which falls within a polytomy along with clade A and C. Sample RT040 is from layer 13 which is dated between 146 ± 11 and 156 ± 15 ka, corresponding to the time interval estimated for the divergence between clades A and C (Lord et al., 2020). Assuming the correct stratigraphical context for the sample, RT040 was sampled from an individual that lived at the time of/soon after the divergence. Evolutionarily speaking, this short window of time to divergence means the accumulation of few mutations, insufficient to resolve a branch in a phylogeny spanning a broad time period.

Stratigraphical information places sample RT087 in layer 12.1 – 12.2, which dates to around 120 – 129 ± 11 ka (as seen in Figure 4.1, there is not a concrete time frame given to this stratigraphical range (Jacobs et al., 2019)). However, its placement in the phylogeny, both with relation to other samples in the same supposed stage, and in relation to clades B and D suggest a

more ancient origin. While tentative, the possibility of a misattribution of the sample's layer of origin cannot be excluded, either as a result of a mix-up in the large-scale process of sample screening, or post-depositional sediment disturbance.

The ages of these fragments are beyond the calibration range of radiocarbon dating (max limit = *ca.* 48 ka) but could potentially be estimated with the use of molecular clock analysis. Following the single-sample dating protocol described by Chacón-Duque et al. (2025) on the BEAST software (Suchard et al., 2018), an estimate age for the samples could be determined in future studies, as well as calibrated estimate divergence times based on the extended dataset. This would allow for more confident inferences into the placement of these samples into the evolutionary history of rhinoceros. Given the results herein discussed, we hypothesize an earlier estimate age for sample RT087 than given by the stratigraphical context, corresponding to sometime after the divergence event between clades B and D.

### 4.4. Implications of by-catch genomic data recovery

The recovery of by-catch DNA highlights the limitations of relying solely on genetic data for taxonomic identification. In contrast, ZooMS is a more reliable alternative because collagen is unlikely to be introduced by predators or other species that may come into contact with the remains after deposition, unlike DNA. We observed that, in the absence of prior contextual information provided by ZooMS analysis, the sole reliance on genomic data for taxonomic identification can lead to conflicting or misleading conclusions. The presence of multiple sources of DNA in the sampled fragments, which cannot be readily attributed to procedural contamination complicates the taxonomical assignment, as well as downstream analysis, including sequence filtering and mitogenome reconstruction.

The latter point is best illustrated by sample AG083. For this sample, MIA ran through an irregularly high number of iterations (Supplementary Table 6) to reach the consensus sequence, generating a very incomplete mitogenome. To improve the reconstruction of this mitogenome, an alternative approach would be to filter the sequencing data, removing endogenous reads (in this case, bovid), retaining only those that map to the by-catch (*viz.* hyena). These reads could then be used as input for MIA, potentially yielding a more complete mitogenome.

The hyena DNA recovered as by-catch from the bovid samples was identified as *C. crocuta* of haplogroup A1 (Supplementary Figure 4). This result agrees with previous research into the hyena population turnover at Denisova cave. The study of sedimentary DNA at the cave conducted by Zavala et al. (2021) indicates the prevalence of haplogroup A hyenas in layers younger than 80 – 120 ka. For AG083 (from layer 11.2, dated to 55 – 63 ± 6 ka), this is consistent. AG031 is reported as coming from layers 12.1 – 12.2, which date around 120 – 129 ± 11 ka,

though the range is not concretely established in the literature (Jacobs et al., 2019). This range coincides with the turnover of hyena population at Denisova Cave, from being dominated by haplogroup D to haplogroup A. Even prior to the turnover, however, though haplogroup D was dominating, A (and B) also occurred (Zavala et al., 2021).

The presence of predator DNA in prey bone fragments also aligns with the assessment made by previous studies that the fragmentation of these remains results, at least partially, from carnivore activity (Brown et al., 2021). This by-catch DNA recovery stands to contribute to the understanding of predator-prey interactions in the Pleistocene, namely ecological dynamics and predation habits. Furthermore, we have shown that it is possible for these samples to yield enough by-catch data to reconstruct mitogenomes, which can contribute to the study of these species' population structure.

# 5. Conclusion

Through the processing and analysis of genomic data, this study yielded species-level taxonomic identification for 13 Rhinocerontidae bone fragments found at Denisova Cave. This screening process allowed for the detection of the continued presence of woolly rhinoceros in the Altai region through late MIS 6 and MIS 5, and the identification of Merck's rhinoceros in the region. We also explored the presence of by-catch DNA in Rhinocerontidae and bovid remains, which yielded two haplogroup A hyena mitogenomes. The presence of a strong genetic signal from a scavenger species provides further taphonomic evidence supporting the hypothesis that carnivore activity was a driving mechanism of herbivore bone fragmentation at Denisova Cave. Together, these findings contribute to the study and interpretation of faunal assemblages from the Altai Mountains, building upon the ever-expanding picture of biodiversity of the region.

We generated 7 high coverage woolly rhinoceros mitogenomes from the late Middle Pleistocene, dated to between 105 ± 11 and 156 ± 15 ka according to the stratigraphic context. These represent the oldest woolly rhinoceros mitogenomes generated to date and expand the temporal reach of the species' phylogeny. Furthermore, this study expands the geographical range of recovered woolly rhinoceros DNA, demonstrating a great dispersal of the species, as well as a degree of connectivity between maternal lineages. Notably, three of four known woolly rhinoceros maternal lineages are present in the Denisova Cave region during MIS 5.

While this study focused on the dynamics of the maternal lineages, the libraries generated in this study also contain nuclear genomic data (Table 3.1). Disregarding the samples in lysis batch corresponding to blank RTBL5, there are three samples for which the generated libraries yielded a depth of coverage of the nuclear genome above 0.1X (0.6 – 1X). To properly interpret

future results within the context of woolly rhinoceros evolutionary history, future studies should first seek to validate the dates provided by the stratigraphical context using BEAST (Suchard et al., 2018) molecular clock dating.

This study brings to light questions regarding woolly rhinoceros population structure and stability across MIS 6 and MIS 5. Future studies of genetic diversity of the woolly rhinoceros during this time period, enabled by the nuclear and mitochondrial genomic data, could help clarify the dynamics of the species in the region, and how it fits into the broader picture of woolly rhinoceros phylogeography. With a more robust chronological framework, such work opens the door to future and more detailed studies into rhinoceros population structure and evolutionary dynamics in the Middle Pleistocene.

# References


Agadjanian, A. K., & Shunkov, M. V. (2018). Paleolithic Man of Denisova Cave and Zoogeography of Pleistocene Mammals of Northwestern Altai. *Paleontological Journal*, *52*(1), 66–89. https://doi.org/10.1134/S0031030118010021

Anderson, D. E., Goudie, A. S., & Parker, A. G. (2013). Pleistocene Climatic Change and Environments of Mid to High Latitudes. In *Global Environments through the Quaternary* (2nd ed., pp. 72–123). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199697267.003.0003

Bai, B., Meng, J., Zhang, C., Gong, Y.-X., & Wang, Y.-Q. (2020). The origin of Rhinocerotoidea and phylogeny of Ceratomorpha (Mammalia, Perissodactyla). *Communications Biology*, *3*(1), 509. https://doi.org/10.1038/s42003-020-01205-8

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Boeskorov, G. G. (2012). Some specific morphological and ecological features of the fossil woolly rhinoceros (*Coelodonta antiquitatis* Blumenbach 1799). *Biology Bulletin*, *39*(8), 692–707. https://doi.org/10.1134/S106235901208002X

Bon, C., Berthonaud, V., Maksud, F., Labadie, K., Poulain, J., Artiguenave, F., Wincker, P., Aury, J.-M., & Elalouf, J.-M. (2012). Coprolites as a source of information on the genome and diet of the cave hyena. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1739), 2825–2830. https://doi.org/10.1098/rspb.2012.0358

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajković, D., Kućan, Ž., Gušić, I., Schmitz, R., Doronichev, V. B., Golovanova, L. V., de la Rasilla, M., Fortea, J., Rosas, A., & Pääbo, S. (2009). Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science*, *325*(5938), 318–321. https://doi.org/10.1126/science.1174462

Brown, S., Wang, N., Oertle, A., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Comeskey, D., Jope-Street, B., Harvey, V. L., Chowdhury, M. P., Buckley, M., Higham, T., & Douka, K. (2021). Zooarchaeology through the lens of collagen fingerprinting at Denisova Cave. *Scientific Reports*, *11*(1), 15457. https://doi.org/10.1038/s41598-021-94731-2

Buckley, M., Collins, M., Thomas-Oates, J., & Wilson, J. C. (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight

mass spectrometry. *Rapid Communications in Mass Spectrometry*, *23*(23), 3843–3854. https://doi.org/10.1002/rcm.4316

Buckley, M., Fraser, S., Herman, J., Melton, N. D., Mulville, J., & Pálsdóttir, A. H. (2014). Species identification of archaeological marine mammals using collagen fingerprinting. *Journal of Archaeological Science*, *41*, 631–641. https://doi.org/10.1016/j.jas.2013.08.021

Cerdeño, E. (1998). Diversity and evolutionary trends of the Family Rhinocerotidae (Perissodactyla). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *141*(1–2), 13–34. https://doi.org/10.1016/S0031-0182(98)00003-0

Chacón-Duque, J. C., Thomas Thorpe, J. A., Li, W., Dehasque, M., Pečnerová, P., Barlow, A., Díez-del-Molino, D., Henneberger, K., Jin, C., Moreland, K. N., Paijmans, J. L. A., van der Valk, T., Westbury, M. V, Wijnands, F., Barnes, I., Germonpré, M., Hall, E., Hewitson, S., Mol, D., … Dalén, L. (2025). A Million Years of Mammoth Mitogenome Evolution. *Molecular Biology and Evolution*, *42*(4). https://doi.org/10.1093/molbev/msaf065

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763. https://doi.org/10.1073/pnas.1314445110

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth. In *Ancient DNA. Methods in Molecular Biology* (Vol. 1963, pp. 25–29). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9176-1_4

Dalén, L., Heintzman, P. D., Kapp, J. D., & Shapiro, B. (2023). Deep-time paleogenomics and the limits of DNA survival. *Science*, *382*(6666), 48–53. https://doi.org/10.1126/science.adh7943

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Dolenz, S., van der Valk, T., Jin, C., Oppenheimer, J., Sharif, M. B., Orlando, L., Shapiro, B., Dalén, L., & Heintzman, P. D. (2024). Unravelling reference bias in ancient DNA datasets. *Bioinformatics*, *40*(7). https://doi.org/10.1093/bioinformatics/btae436

Douka, K., Slon, V., Jacobs, Z., Ramsey, C. B., Shunkov, M. V., Derevianko, A. P., Mafessoni, F., Kozlikin, M. B., Li, B., Grün, R., Comeskey, D., Devièse, T., Brown, S., Viola, B., Kinsley, L., Buckley, M., Meyer, M., Roberts, R. G., Pääbo, S., … Higham, T. (2019). Age

estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. *Nature*, *565*(7741), 640–644. https://doi.org/10.1038/s41586-018-0870-z

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elias, S. A. (2013). Introduction. In S. A. Elias & C. J. Mock (Eds.), *Encyclopedia of Quaternary Science* (2nd ed., pp. xi–xii). Elsevier. https://doi.org/10.1016/B978-0-444-53643-3.09982-9

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. In B. Shapiro, Barlow; A., P. Heintzman, M. Hofreiter, K. Paijmans, & A. Soares (Eds.), *Ancient DNA. Methods in Molecular Biology* (Vol. 1963, pp. 1–13). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9176-1_1

Gansauge, M.-T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, *8*(4), 737–748. https://doi.org/10.1038/nprot.2013.038

Giaccio, B., Bini, M., Isola, I., Hu, H.-M., Rolfo, M. F., Chuan-Chou, S., Ferracci, A., Monaco, L., Pasquetti, F., & Zanchetta, G. (2024). Constraining the end of the Last Interglacial (MIS 5e) relative sea-level highstand in central Mediterranean: New data from Grotta delle Capre, central Italy. *Global and Planetary Change*, *232*, 104321. https://doi.org/10.1016/j.gloplacha.2023.104321

Gilardet, A., Lord, E., García, G. O., Xenikoudakis, G., Douka, K., Wooller, M. J., Rowe, T., Martin, M. D., Le Moullec, M., Anisimov, M., Heintzman, P. D., & Dalén, L. (2025). A High-Throughput Ancient DNA Extraction Method for Large-Scale Sample Screening. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.14077

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hofreiter, M., & Stewart, J. (2009). Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. *Current Biology*, *19*(14), R584–R594. https://doi.org/10.1016/j.cub.2009.06.030

Hu, J., Westbury, M. V., Yuan, J., Zhang, Z., Chen, S., Xiao, B., Hou, X., Ji, H., Lai, X., Hofreiter, M., & Sheng, G. (2021). Ancient mitochondrial genomes from Chinese cave hyenas provide insights into the evolutionary history of the genus *Crocuta*. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1943), 20202934. https://doi.org/10.1098/rspb.2020.2934

Jacobs, Z., Li, B., Shunkov, M. V., Kozlikin, M. B., Bolikhovskaya, N. S., Agadjanian, A. K., Uliyanov, V. A., Vasiliev, S. K., O'Gorman, K., Derevianko, A. P., & Roberts, R. G. (2019). Timing of archaic hominin occupation of Denisova Cave in southern Siberia. *Nature*, *565*(7741), 594–599. https://doi.org/10.1038/s41586-018-0843-2

Jacobs, Z., Zavala, E. I., Li, B., O'Gorman, K., Shunkov, M. V., Kozlikin, M. B., Derevianko, A. P., Uliyanov, V. A., Goldberg, P., Agadjanian, A. K., Vasiliev, S. K., Brink, F., Peyrégne, S., Slon, V., Pääbo, S., Kelso, J., Meyer, M., & Roberts, R. G. (2025). Pleistocene chronology and history of hominins and fauna at Denisova Cave. *Nature Communications*, *16*(1), 4738. https://doi.org/10.1038/s41467-025-60140-6

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *Journal of Heredity*, *112*(3), 241–249. https://doi.org/10.1093/jhered/esab012

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1). https://doi.org/10.1093/nar/gkr771

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., Schmidt, A. M. A., Kjeldsen, K. K., Tesakov, A. S., Snowball, I., Gosse, J. C., Alsos, I. G., Wang, Y., Dockter, C., Rasmussen, M., Jørgensen, M. E., … Willerslev, E. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, *612*(7939), 283–291. https://doi.org/10.1038/s41586-022-05453-y

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kosintsev, P., Mitchell, K. J., Devièse, T., van der Plicht, J., Kuitems, M., Petrova, E., Tikhonov, A., Higham, T., Comeskey, D., Turney, C., Cooper, A., van Kolfschoten, T., Stuart, A. J., & Lister, A. M. (2018). Evolution and extinction of the giant rhinoceros *Elasmotherium*

*sibiricum* sheds light on late Quaternary megafaunal extinctions. *Nature Ecology & Evolution*, *3*(1), 31–38. https://doi.org/10.1038/s41559-018-0722-0

Krajcarz, M. T., Baca, M., Baumann, C., Bocherens, H., Goslar, T., Popović, D., Sudoł-Procyk, M., & Krajcarz, M. (2023). New insights into late pleistocene cave hyena chronology and population history - the case of Perspektywiczna Cave, Poland. *Radiocarbon*, *65*(5), 1038–1056. https://doi.org/10.1017/RDC.2023.89

Kutschera, V. E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D. W. G., Khoonsari, P. E., Nystedt, B., Dalén, L., & Díez-del-Molino, D. (2022). GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. *BMC Bioinformatics*, *23*(1), 228. https://doi.org/10.1186/s12859-022-04757-0

Kuzmin, Y. V., Slavinsky, V. S., Tsybankov, A. A., & Keates, S. G. (2022). Denisovans, Neanderthals, and Early Modern Humans: A Review of the Pleistocene Hominin Fossils from the Altai Mountains (Southern Siberia). *Journal of Archaeological Research*, *30*(3), 321–369. https://doi.org/10.1007/s10814-021-09164-2

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liu, S., Westbury, M. V., Dussex, N., Mitchell, K. J., Sinding, M.-H. S., Heintzman, P. D., Duchêne, D. A., Kapp, J. D., von Seth, J., Heiniger, H., Sánchez-Barreiro, F., Margaryan, A., André-Olsen, R., De Cahsan, B., Meng, G., Yang, C., Chen, L., van der Valk, T., Moodley, Y., … Gilbert, M. T. P. (2021). Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell*, *184*(19), 4874-4885.e16. https://doi.org/10.1016/j.cell.2021.07.032

Lord, E., Dussex, N., Kierczak, M., Díez-del-Molino, D., Ryder, O. A., Stanton, D. W. G., Gilbert, M. T. P., Sánchez-Barreiro, F., Zhang, G., Sinding, M.-H. S., Lorenzen, E. D., Willerslev, E., Protopopov, A., Shidlovskiy, F., Fedorov, S., Bocherens, H., Nathan, S. K. S. S., Goossens, B., van der Plicht, J., … Dalén, L. (2020). Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros. *Current Biology*, *30*(19), 3871-3879.e7. https://doi.org/10.1016/j.cub.2020.07.046

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli,

K.-P., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364. https://doi.org/10.1038/nature10574

Massicotte, P., & South, A. (2017). rnaturalearth: World Map Data from Natural Earth. In *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.package.rnaturalearth

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6). https://doi.org/10.1101/pdb.prot5448

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Nguyen, R., Kapp, J. D., Sacco, S., Myers, S. P., & Green, R. E. (2023). A computational approach for positive genetic identification and relatedness detection from low-coverage shotgun sequencing data. *Journal of Heredity*, *114*(5), 504–512. https://doi.org/10.1093/jhered/esad041

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*(1), 14. https://doi.org/10.1038/s43586-020-00011-0

Pebesma, E. (2016). sf: Simple Features for R. In *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.package.sf

Prothero, D. R. (1993). Fifty million years of rhinoceros evolution. In *Scientific American book of dinosaurs* (pp. 82–91).

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Schvyreva, A. K. (2015). On the importance of the representatives of the genus *Elasmotherium* (Rhinocerotidae, Mammalia) in the biochronology of the Pleistocene of Eastern Europe. *Quaternary International*, *379*, 128–134. https://doi.org/10.1016/j.quaint.2015.03.052

Seeber, P. A., Palmer, Z., Schmidt, A., Chagas, A., Kitagawa, K., Marinova-Wolff, E., Tafelmaier, Y., & Epp, L. S. (2023). The first European woolly rhinoceros mitogenomes, retrieved from cave hyena coprolites, suggest long-term phylogeographic differentiation. *Biology Letters*, *19*(11). https://doi.org/10.1098/rsbl.2023.0343

South, A. (2011). rworldmap : a new R package for mapping global data. *The R Journal*, *3*(1), 35. https://doi.org/10.32614/RJ-2011-006

South, A. (2012). rworldxtra: Country boundaries at high resolution. In *CRAN: Contributed Packages*. https://doi.org/10.32614/CRAN.package.rworldxtra

South, A., Michael, S., & Massicotte, P. (2017). rnaturalearthdata: World Vector Map Data from Natural Earth Used in "rnaturalearth." In *CRAN: Contributed Packages*. R package version 1.0.0.9000. https://doi.org/10.32614/CRAN.package.rnaturalearthdata

Stefaniak, K., Stachowicz-Rybka, R., Borówka, R. K., Hrynowiecka, A., Sobczyk, A., Moskal-del Hoyo, M., Kotowski, A., Nowakowski, D., Krajcarz, M. T., Billia, E. M. E., Persico, D., Burkanova, E. M., Leshchinskiy, S. V., van Asperen, E., Ratajczak, U., Shpansky, A. V., Lempart, M., Wach, B., Niska, M., … Kovalchuk, O. (2021). Browsers, grazers or mix-feeders? Study of the diet of extinct Pleistocene Eurasian forest rhinoceros *Stephanorhinus kirchbergensis* (Jäger, 1839) and woolly rhinoceros *Coelodonta antiquitatis* (Blumenbach, 1799). *Quaternary International*, *605–606*, 192–212. https://doi.org/10.1016/j.quaint.2020.08.039

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/ve/vey016

van Asperen, E. N., & Kahlke, R.-D. (2015). Dietary variation and overlap in Central and Northwest European *Stephanorhinus kirchbergensis* and *S. hemitoechus* (Rhinocerotidae, Mammalia) influenced by habitat diversity. *Quaternary Science Reviews*, *107*, 47–61. https://doi.org/10.1016/j.quascirev.2014.10.001

van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter,

M., … Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*, *591*(7849), 265–269. https://doi.org/10.1038/s41586-021-03224-9

Westbury, M. V., Hartmann, S., Barlow, A., Preick, M., Ridush, B., Nagel, D., Rathgeber, T., Ziegler, R., Baryshnikov, G., Sheng, G., Ludwig, A., Wiesel, I., Dalen, L., Bibi, F., Werdelin, L., Heller, R., & Hofreiter, M. (2020). Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Science Advances*, *6*(11). https://doi.org/10.1126/sciadv.aay0456

Whitfield, J. B., & Lockhart, P. J. (2007). Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, *22*(5), 258–265. https://doi.org/10.1016/j.tree.2007.01.012

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Vaughan Davis, & Girlich Maximilian. (2025). *tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org

Yuan, J., Sun, G., Xiao, B., Hu, J., Wang, L., Taogetongqimuge, Bao, L., Hou, Y., Song, S., Jiang, S., Wu, Y., Pan, D., Liu, Y., Westbury, M. V., Lai, X., & Sheng, G. (2023). Ancient mitogenomes reveal a high maternal genetic diversity of Pleistocene woolly rhinoceros in Northern China. *BMC Ecology and Evolution*, *23*(1), 56. https://doi.org/10.1186/s12862-023-02168-0

Zavala, E. I., Jacobs, Z., Vernot, B., Shunkov, M. V., Kozlikin, M. B., Derevianko, A. P., Essel, E., de Fillipo, C., Nagel, S., Richter, J., Romagné, F., Schmidt, A., Li, B., O'Gorman, K., Slon, V., Kelso, J., Pääbo, S., Roberts, R. G., & Meyer, M. (2021). Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature*, *595*(7867), 399–403. https://doi.org/10.1038/s41586-021-03675-0

# Supplementary Material

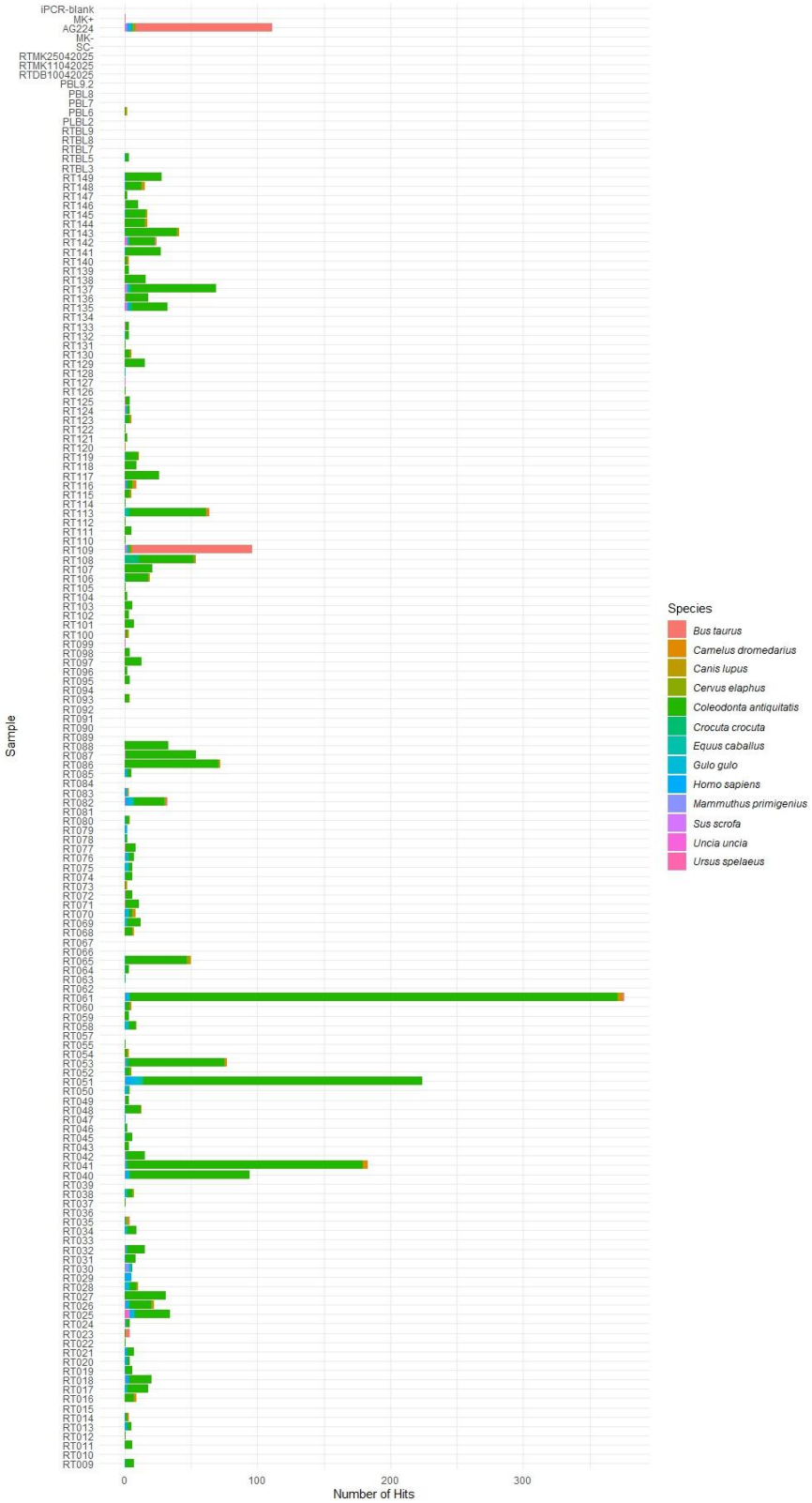
## Annex I. Remarks on RTBL5 contamination

A total of 141 samples and 18 controls (lysis, extraction and library blanks, and library positive controls) were sequenced, and the resulting libraries were screened for endogenous rhinoceros DNA. Contamination of lysis blank RTBL5 was detected and analyzed. Considering that none of the other downstream blanks associated with this screening batch are contaminated in this manner, it suggests the contamination occurred at the lysis step, affecting samples from the same lysis batch as RTBL5 (RT063 – RT087).

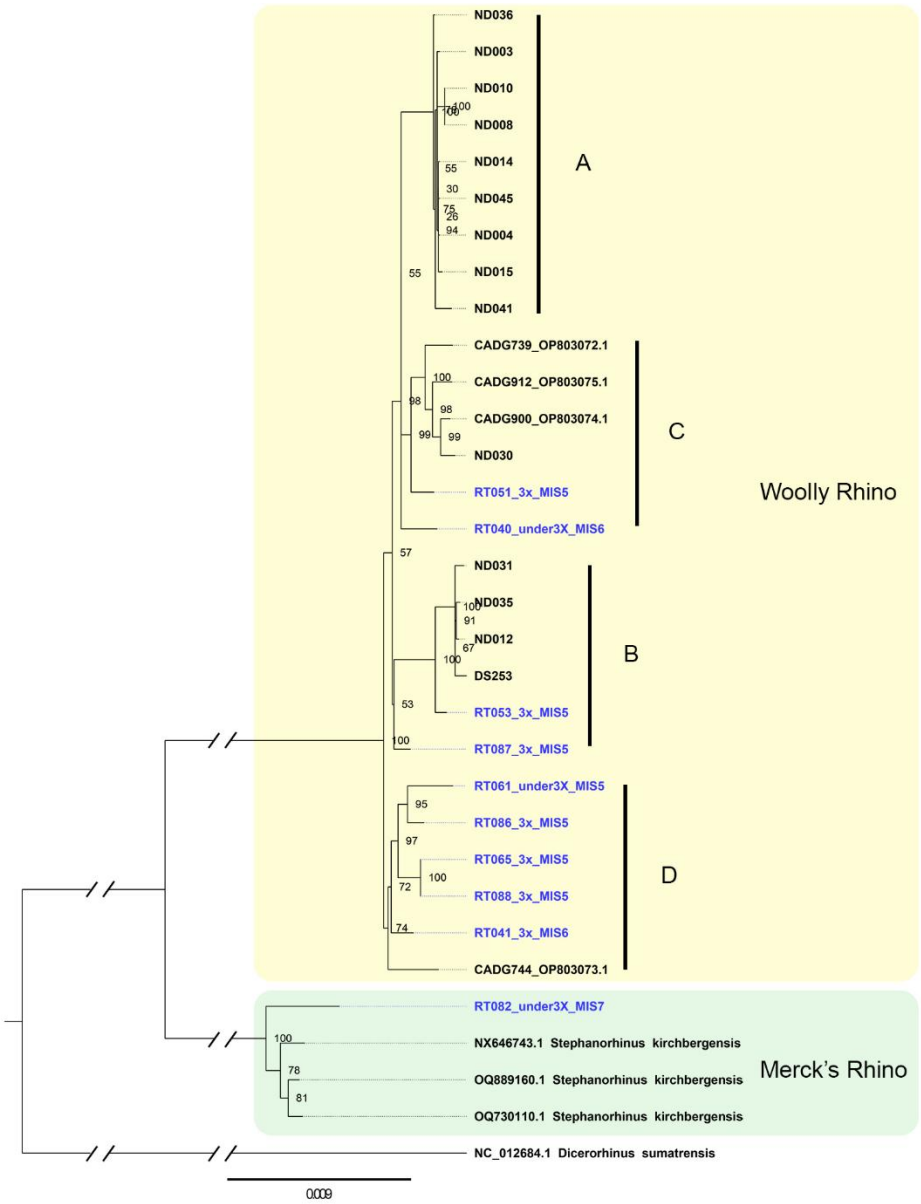
Out of 630k reads after fastp preprocessing, 1.065% mapped to the reference genome. The AMBER plot (Dolenz et al., 2024) (Supplementary Figure 5) show patterns of read length (C) and sequence damage consistent with aDNA damage patterns (B). The distribution of hits across the reference scaffolded was investigated, and it was seen that the reads aligned to various scaffolds across the assembly. When mapping to woolly rhinoceros alone, no reads were found in the mitochondrial region. During competitive mapping, two reads mapped to the *C. antiquitatis* mitochondrial genome, and one to *Homo sapiens*. Although it is most likely the blank was contaminated with genomic data from another sample within the same batch, we cannot confidently say this is the case. This number of reads in the mitochondrial region is negligible and, for that reason, we consider that the analysis of the samples in the same lysis batch as RTBL5 would not be affected by this contamination.

With regard to future studies, samples RT063 – RT087 would need to be re-sampled and new libraries generated, to allow analysis at the nuclear level.

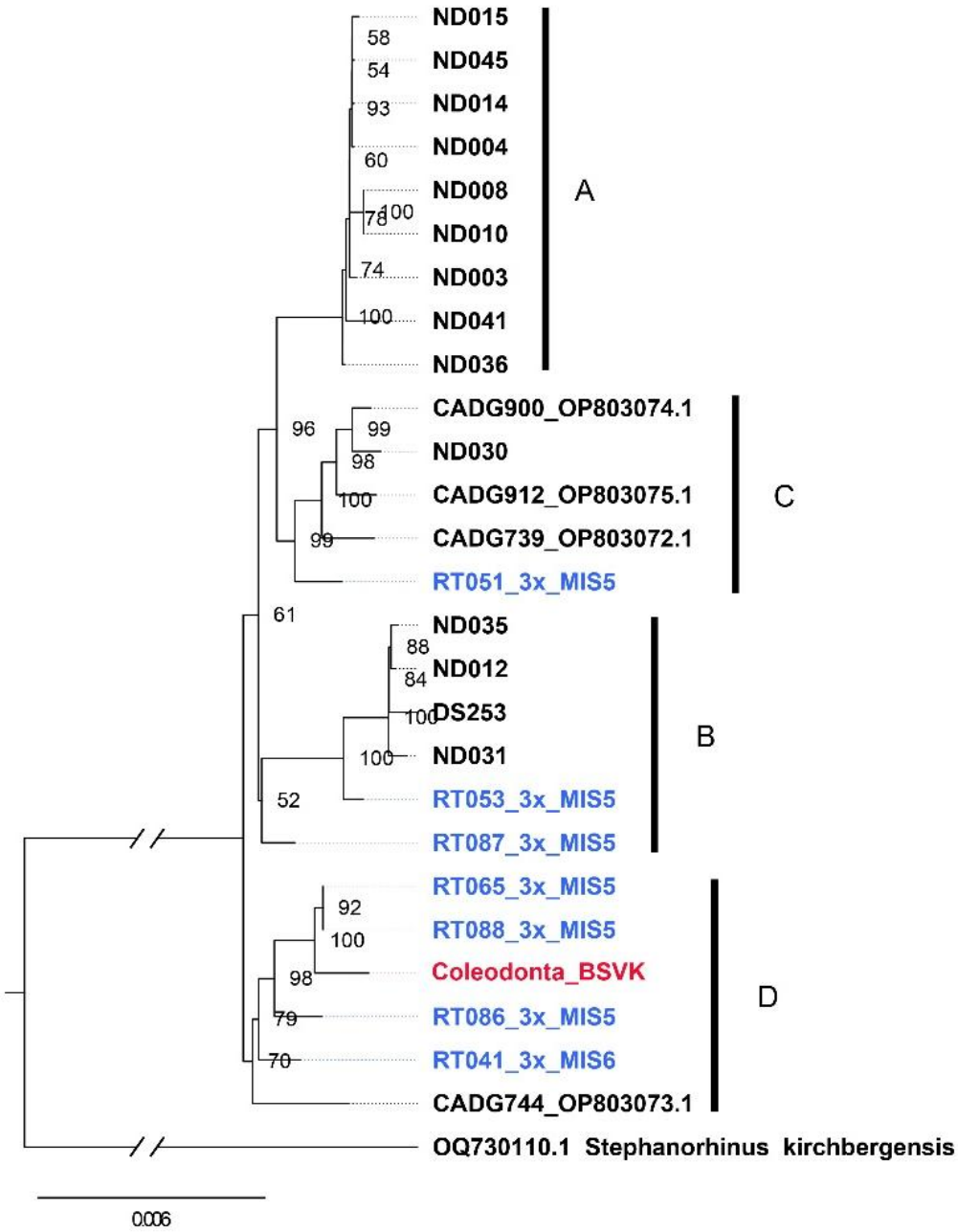
## Annex II. Supplementary Figures



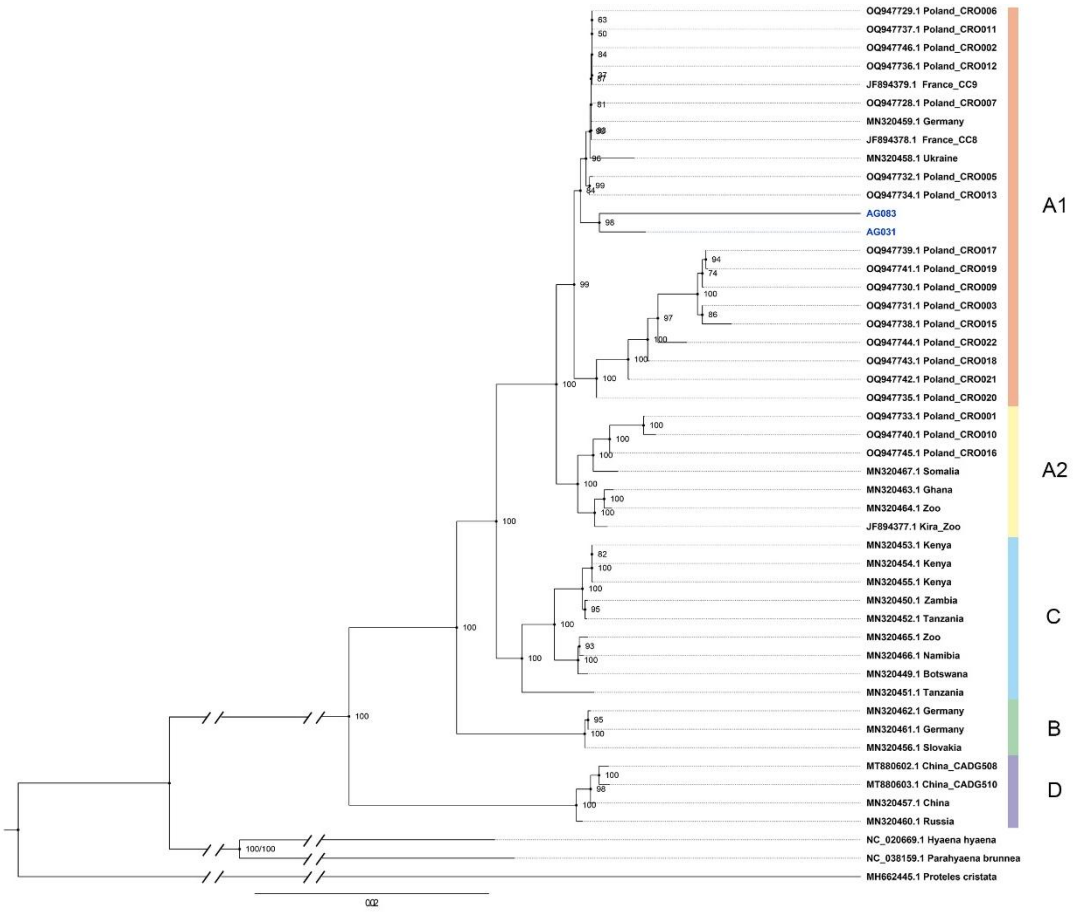

**Supplementary Figure 1. Competitive mapping results of the full screening data.** Competitive mapping results are shown as number of hits to each species to the general (first alignment round) species panel. Lysis blanks: RTBL3-9; Extraction blanks (on plate): PBL2-9; Extraction blanks (on column): RTDB21042025 & RTDB10052025; Library blanks (on plate): SC-, MK-; Library blanks (on column): RTMK26032025, RTMK11042025 & RTMK25042025. Positive library controls: AG224 & MK+ (AVM_002a).



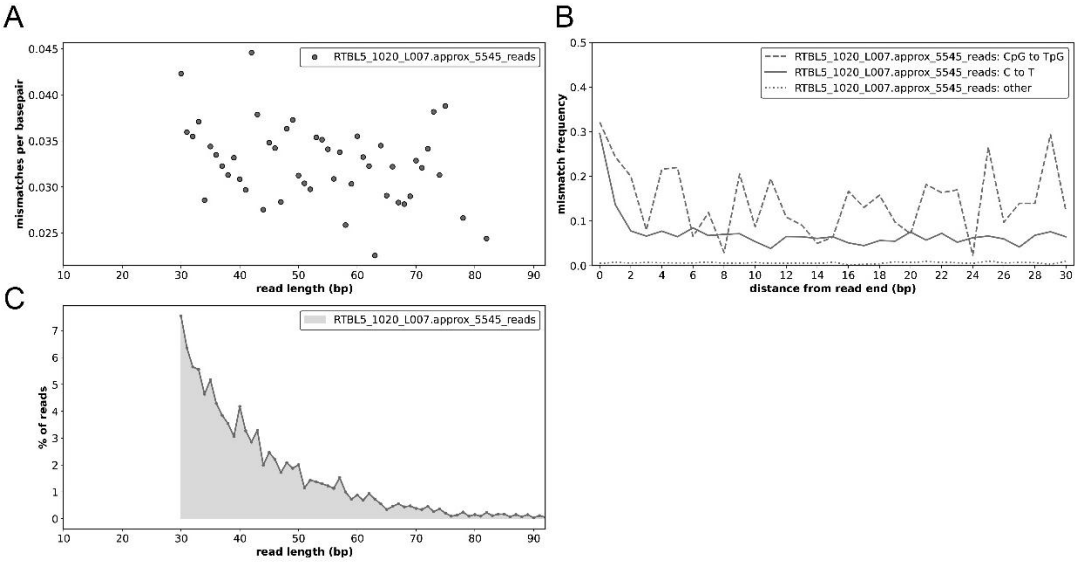

**Supplementary Figure 2. Consensus phylogenetic tree of low and high depth of coverage mitogenomes.** Maximum-likelihood consensus tree of 1000 bootstrap trees. Tree was generated from an alignment of 793 parsimony informative sites using the best-fitting model (TN+F+R3). Bootstrap support values (%) are indicated at each node. Scale represents number of nucleotide replacements per site. Mitogenomes generated in this study are highlighted blue and include in the label the corresponding marine-isotope stage. Low-coverage samples are labeled as "under3X", high-coverage samples are labeled as "3X", irrespective of the actual mitochondrial depth of coverage value. *D. sumatrensis* was used as an outgroup.




**Supplementary Figure 3. Consensus phylogenetic tree generated from the alignment of high depth of coverage mitogenomes and European woolly rhinoceros from hyena coprolite.** Maximum-likelihood consensus tree of 1000 bootstrap trees. Tree was generated from an alignment of 798 parsimony informative sites using the best-fitting model (TN+F+R3). Bootstrap support values (%) are indicated at each node. Scale represents number of nucleotide replacements per site. Mitogenomes generated in this study are highlighted blue and include in the label the corresponding marine-isotope stage. European woolly rhinoceros (Seeber et al., 2023) is highlighted in red. Low-coverage samples are labeled as "under3X", high-coverage samples are labeled as "3X", irrespective of the actual mitochondrial depth of coverage value. *S. kirchbergensis* was used as an outgroup.




**Supplementary Figure 4. Hyena phylogenetic tree.** Maximum-likelihood consensus consensus tree of 1000 bootstrap trees. Tree was generated from an alignment of 1701 parsimony informative sites using the best-fitting model (HKY+F+I+R3). Bootstrap support values (%) are indicated at each node. Scale represents number of nucleotide replacements per site. Clade labels correspond to haplogroups as in Westbury et al. (2020) and Krajcarz et al. (2023). Mitogenomes generated in this study are colored blue.




**Supplementary Figure 5. AMBER plots of extraction blank RTBL5.** (A) Base mismatches between reads and the reference per length of the read. (B) Rate of C→T transitions across sequence reads. (C) Fragment length distribution.

## Annex III. Supplementary Tables

**Supplementary Table 1. Record of samples used in this study.** Sample ID is the name given to the sample after sampling. Fragment ID is the name given in the collection.

| Sample ID | Fragment ID | Bone Powder (mg) | Layer | Sampling method |
|---|---|---|---|---|
| RT009 | DC.8931 | 32 | 14 | cut |
| RT010 | DC.8983 | 44 | 14 | cut |
| RT011 | DC.8885 | 50 | 14 | cut |
| RT012 | DC.8919 | 47 | 14 | cut |
| RT013 | DC.8912 | 32 | 14 | cut |
| RT014 | DC.8926 | 35 | 14 | cut |
| RT015 | DC.8910 | 38 | 14 | cut |
| RT016 | DC.8906 | 61 | 14 | cut |
| RT017 | DC.8960 | 44 | 14 | cut |
| RT018 | DC.8986 | 37 | 14 | cut |
| RT019 | DC.5802 | 32 | 14 | cut |
| RT020 | DC.8440 | 45 | 14 | cut |
| RT021 | DC.8909 | 35 | 14 | cut |
| RT022 | DC.8293 | 38 | 14 | cut |
| RT023 | DC.15822 | 36 | 13 | cut |
| RT024 | DC.11578 | 60 | 14 | cut |
| RT025 | DC.11572 | 31 | 14 | cut |
| RT026 | DC.5900 | 55 | 13 | cut |
| RT027 | DC.5888 | 49 | 13 | cut |
| RT028 | DC.5912 | 44 | 13 | cut |
| RT029 | DC.15824 | 39 | 13 | cut |
| RT030 | DC.15692 | 25 | 13 | cut |
| RT031 | DC.15654 | 47 | 13 | cut |
| RT032 | DC.15652 | 35 | 13 | cut |
| RT033 | DC.15641 | 35 | 13 | cut |
| RT034 | DC.15689 | 22 | 13 | cut |
| RT035 | DC.15688 | 41 | 13 | cut |
| RT036 | DC.15644 | 42 | 13 | cut |
| RT037 | DC.15747 | 44 | 13 | cut |
| RT038 | DC.15716 | 37 | 13 | cut |
| RT039 | DC.15453 | 75 | 13 | cut |
| RT040 | DC.15431 | 88 | 13 | cut |
| RT041 | DC.15454 | 84 | 13 | cut |
| RT042 | DC.15474 | 83 | 13 | cut |
| RT043 | DC.15814 | 66 | 13 | cut |
| RT044 | DC.15451 | 43 | 13 | cut |
| RT045 | DC.15471 | 28 | 13 | cut |

| | | | | |
|---|---|---|---|---|
| RT046 | DC.15775 | 37 | 13 | cut |
| RT047 | DC.15467 | 52 | 13 | cut |
| RT048 | DC.15462 | 29 | 13 | cut |
| RT049 | DC.8048 | 51 | 11.4 | cut |
| RT050 | DC.8183 | 50 | 12 | cut |
| RT051 | DC.8225 | 30 | 12 | cut |
| RT052 | DC.8171 | 46 | 12 | cut |
| RT053 | DC.8170 | 47 | 12 | cut |
| RT054 | DC.8036 | 34 | 11.4 | cut |
| RT055 | DC.8029 | 51 | 11.4 | cut |
| RT056 | DC.8087 | 49 | 11.4 | cut |
| RT057 | DC.8153 | 54 | 12 | cut |
| RT058 | DC.8086 | 55 | 11.4 | cut |
| RT059 | DC.8128 | 42 | 12 | cut |
| RT060 | DC.8111 | 31 | 12 | cut |
| RT061 | DC.6810 | 63 | 11.4 | cut |
| RT062 | DC.6771 | 67 | 11.4 | cut |
| RT063 | DC.6899 | 64 | 11.4 | cut |
| RT064 | DC.6923 | 52 | 11.4 | cut |
| RT065 | DC.6831 | 57 | 11.4 | cut |
| RT066 | DC.8862 | 50 | 15 | cut |
| RT067 | DC.8666 | 60 | 15 | cut |
| RT068 | DC.8831 | 56 | 15 | cut |
| RT069 | DC.8792 | 48 | 15 | cut |
| RT070 | DC.8634 | 62 | 15 | cut |
| RT071 | DC.8577 | 42 | 15 | cut |
| RT072 | DC.8512 | 62 | 15 | cut |
| RT073 | DC.8624 | 47 | 15 | cut |
| RT074 | DC.8658 | 60 | 15 | cut |
| RT075 | DC.8615 | 47 | 15 | cut |
| RT076 | DC.7660 | 44 | 15 | cut |
| RT077 | DC.7638 | 53 | 15 | cut |
| RT078 | DC.7699 | 56 | 15 | cut |
| RT079 | DC.8567 | 39 | 15 | cut |
| RT080 | DC.7778 | 53 | 15 | cut |
| RT081 | DC.7652 | 45 | 15 | cut |
| RT082 | DC.7782 | 63 | 15 | cut |
| RT083 | DC.8452 | 60 | 15 | cut |
| RT084 | DC.7698 | 48 | 15 | cut |
| RT085 | DC.7757 | 53 | 15 | cut |
| RT086 | DC.6102 | 57 | 12.1-12.2 | cut |

| | | | | |
|---|---|---|---|---|
| RT087 | DC.6090 | 44 | 12.1-12.2 | cut |
| RT088 | DC.6071 | 40 | 12.1-12.2 | cut |
| RT089 | DC.5962 | 51 | 15 | cut |
| RT090 | DC.7661 | 35 | 15 | cut |
| RT091 | DC.7621 | 31 | 15 | cut |
| RT092 | DC.7735 | 31 | 15 | cut |
| RT093 | DC.7719 | 32 | 15 | cut |
| RT094 | DC. 5951 | 63 | 15 | cut |
| RT095 | DC.7444 | 62 | 15 | cut |
| RT096 | DC.7531 | 65 | 15 | cut |
| RT097 | DC.7501 | 52 | 15 | cut |
| RT098 | DC.7449 | 55 | 15 | cut |
| RT099 | DC.7552 | 48 | 15 | cut |
| RT100 | DC.7299 | 54 | 15 | cut |
| RT101 | DC.7597 | 37 | 15 | cut |
| RT102 | DC.7662 | 49 | 15 | cut |
| RT103 | DC.7384 | 56 | 15 | cut |
| RT104 | DC.7388 | 51 | 15 | cut |
| RT105 | DC.7415 | 54 | 15 | cut |
| RT106 | DC.6979 | 52 | 11.4 | cut |
| RT107 | DC.7227 | 37 | 12 | cut |
| RT108 | DC.7220 | 47 | 12 | cut |
| RT109 | DC.6963 | 50 | 11.4 | cut |
| RT110 | DC.3909 | 41 | 12 | cut |
| RT111 | DC.6933 | 64 | 11.4 | cut |
| RT112 | DC.6854 | 44 | 11.4 | cut |
| RT113 | DC.3916 | 66 | 12 | cut |
| RT114 | DC.3919 | 84 | 12 | cut |
| RT115 | DC.3903 | 31 | 12 | cut |
| RT116 | DC.3884 | 46 | 12 | cut |
| RT117 | DC.4001 | 58 | 12 | cut |
| RT118 | DC.3879 | 41 | 12 | cut |
| RT119 | DC.3955 | 62 | 12 | cut |
| RT120 | DC.3879 | 11 | 12 | drilling |
| RT121 | DC.3879 | 13 | 12 | drilling |
| RT122 | DC.4001 | 17 | 12 | drilling |
| RT123 | DC.4001 | 28 | 12 | drilling |
| RT124 | DC.3884 | 12 | 12 | drilling |
| RT125 | DC.3884 | 13 | 12 | drilling |
| RT126 | DC.3895 | 54 | 12 | cut |
| RT127 | DC.3940 | 53 | 12 | cut |

| | | | | |
|---|---|---|---|---|
| RT128 | DC.3878 | 53 | 12 | cut |
| RT129 | DC.3887 | 39 | 12 | cut |
| RT130 | DC.3886 | 69 | 12 | cut |
| RT131 | DC.3974 | 63 | 12 | cut |
| RT132 | DC.15455 | 44 | 13 | cut |
| RT133 | DC.15460 | 31 | 13 | cut |
| RT134 | DC.15788 | 30 | 13 | cut |
| RT135 | DC.15828 | 48 | 13 | cut |
| RT136 | DC.15622 | 48 | 13 | cut |
| RT137 | DC.15484 | 30 | 13 | cut |
| RT138 | DC.15564 | 44 | 13 | cut |
| RT139 | DC.15658 | 35 | 13 | cut |
| RT140 | DC.15697 | 44 | 13 | cut |
| RT141 | DC.15574 | 67 | 13 | cut |
| RT142 | DC.15515 | 50 | 13 | cut |
| RT143 | DC.15566 | 47 | 13 | cut |
| RT144 | DC.15615 | 59 | 13 | cut |
| RT145 | DC.15554 | 59 | 13 | cut |
| RT146 | DC.15486 | 56 | 13 | cut |
| RT147 | DC.15675 | 62 | 13 | cut |
| RT148 | DC.6929 | 50 | 11.4 | cut |
| RT149 | DC.6841 | 49 | 11.4 | cut |

**Supplementary Table 2. Indexing PCR reaction and cycle programs for libraries prepared with the Santa Cruz Reaction (SC) and Meyer & Kircher (MK) protocols.**

| Library preparation protocol | Master Mix Reagents | Volume library | Final volume | Thermocycler program |
|---|---|---|---|---|
| SC | 5 U of AmpliTaq Gold DNA Polymerase, 1X TaqGold Buffer, 2.5 mM MgCl2 (Applied Biosystems), 0.25 mM dNTPs, 0.4 µM (each) P7 and P5 indexing primers | 5 µL | 50 µL | Initial denaturation:<br>94ºC for 12 min<br>Cq + 1 cycles of:<br>- Denaturation: 94ºC for 30 s<br>- Annealing: 60ºC for 30 s<br>- Elongation: 72ºC for 45 s<br>Final extension:<br>72ºC for 10 min |
| MK | 1X AccuPrime reaction mix (10X), 7 U AccuPrime Pfx, 0.3 µM (each) P7 and P5 indexing primers (Meyer & Kircher, 2010) | 3 µL | 25 µL | Initial denaturation:<br>95ºC for 2 min<br>Cq + 1 cycles of:<br>- Denaturation: 95ºC for 15 s<br>- Annealing: 60ºC for 30 s<br>- Elongation: 68ºC for 60 s |

**Supplementary Table 3. Screening mapping statistics.** "Merged reads" refers to number of reads after fastp processing. "Unique reads" refers to reads after quality and duplicate filtering. "Endo content" refers to endogenous content, calculated from the reads mapped/total read. "Nuclear depth" refers to depth of coverage of the nuclear genome. "Mito depth" refers to depth of coverage of the mitochondrial genome. Lysis blanks: RTBL3-9; Extraction blanks (on plate): PBL2-9; Extraction blanks (on column): RTDB21042025 & RTDB10052025; Library blanks (on plate): SC-, MK-; Library blanks (on column): RTMK26032025, RTMK11042025 & RTMK25042025.

| Sample | Raw reads | Merged reads | Mapped reads | Unique reads | Endo content | Nuclear depth | Mito depth | Read length (mean) |
|---|---|---|---|---|---|---|---|---|
| RT009 | 7 692 631 | 3 460 154 | 2 046 | 402 | 0.059% | 6.15E-06 | 0.014 | 34 |
| RT010 | 12 628 843 | 3 810 149 | 27 647 | 1 556 | 0.726% | 2.81E-05 | - | 43 |
| RT011 | 13 683 818 | 4 945 595 | 34 181 | 5 702 | 0.691% | 1.01E-04 | 0.012 | 41 |
| RT012 | 5 346 306 | 1 903 209 | 1 726 | 391 | 0.091% | 5.75E-06 | 0.002 | 33 |
| RT013 | 9 709 100 | 3 994 585 | 8 354 | 4 617 | 0.209% | 6.89E-05 | 0.011 | 35 |
| RT014 | 8 519 119 | 3 085 018 | 3 337 | 1 262 | 0.108% | 1.75E-05 | 0.002 | 33 |
| RT015 | 6 016 786 | 2 529 306 | 2 463 | 540 | 0.097% | 8.86E-06 | - | 37 |
| RT016 | 16 266 178 | 7 008 378 | 93 591 | 9 939 | 1.335% | 1.71E-04 | 0.016 | 40 |
| RT017 | 12 463 196 | 6 435 296 | 53 768 | 8 086 | 0.836% | 1.87E-04 | 0.044 | 55 |
| RT018 | 13 702 327 | 7 459 668 | 171 577 | 47 637 | 2.300% | 9.16E-04 | 0.048 | 46 |
| RT019 | 14 487 951 | 6 654 803 | 17 099 | 4 711 | 0.257% | 8.90E-05 | 0.011 | 44 |
| RT020 | 15 111 292 | 3 548 233 | 13 082 | 4 349 | 0.369% | 7.57E-05 | 0.004 | 42 |
| RT021 | 22 555 617 | 8 434 865 | 17 911 | 3 383 | 0.212% | 5.93E-05 | 0.015 | 41 |
| RT022 | 7 231 156 | 1 396 166 | 6 073 | 3 238 | 0.435% | 4.78E-05 | 0.002 | 36 |
| RT023 | 15 227 470 | 5 647 079 | 6 715 | 1 499 | 0.119% | 2.05E-05 | 0.004 | 32 |
| RT024 | 10 827 838 | 3 614 013 | 20 155 | 10 205 | 0.558% | 1.54E-04 | 0.004 | 36 |
| RT025 | 8 980 194 | 3 284 046 | 40 252 | 23 487 | 1.226% | 3.72E-04 | 0.052 | 37 |
| RT026 | 9 354 106 | 5 395 194 | 25 046 | 12 466 | 0.464% | 2.25E-04 | 0.035 | 42 |
| RT027 | 5 802 364 | 1 737 383 | 7 021 | 2 269 | 0.404% | 3.59E-05 | 0.072 | 37 |
| RT028 | 21 434 871 | 7 211 134 | 40 464 | 4 288 | 0.561% | 7.04E-05 | 0.010 | 39 |
| RT029 | 7 045 901 | 3 932 772 | 1 966 | 261 | 0.050% | 4.36E-06 | - | 35 |
| RT030 | 10 949 676 | 4 679 842 | 4 303 | 532 | 0.092% | 7.27E-06 | 0.002 | 32 |
| RT031 | 7 684 721 | 3 929 752 | 9 681 | 4 329 | 0.246% | 7.34E-05 | 0.013 | 39 |
| RT032 | 14 906 681 | 6 476 664 | 227 477 | 49 517 | 3.512% | 8.85E-04 | 0.025 | 41 |
| RT033 | 8 955 582 | 2 614 400 | 6 040 | 345 | 0.231% | 5.86E-06 | - | 39 |
| RT034 | 11 623 773 | 4 976 440 | 24 678 | 7 113 | 0.496% | 1.27E-04 | 0.014 | 42 |
| RT035 | 12 994 596 | 4 856 406 | 9 504 | 3 454 | 0.196% | 4.69E-05 | - | 32 |
| RT036 | 16 172 445 | 3 741 049 | 2 534 | 96 | 0.068% | 1.61E-06 | - | 33 |
| RT037 | 5 190 130 | 2 571 421 | 18 251 | 9 351 | 0.710% | 1.56E-04 | 0.002 | 39 |
| RT038 | 14 186 422 | 6 831 471 | 22 087 | 7 072 | 0.323% | 1.19E-04 | 0.007 | 40 |
| RT039 | 32 130 570 | 8 248 246 | 7 268 | 291 | 0.088% | 4.38E-06 | - | 34 |
| RT040 | 14 911 050 | 7 661 816 | 48 871 | 20 346 | 0.638% | 3.27E-04 | 0.177 | 37 |
| RT041 | 22 923 417 | 13 418 356 | 588 270 | 120 055 | 4.384% | 2.50E-03 | 0.453 | 50 |
| RT042 | 7 872 774 | 3 745 033 | 20 794 | 7 372 | 0.555% | 1.31E-04 | 0.034 | 41 |
| RT043 | 25 711 919 | 9 162 209 | 88 961 | 2 560 | 0.971% | 4.88E-05 | 0.006 | 45 |
| RT045 | 23 573 282 | 9 182 629 | 43 267 | 3 406 | 0.471% | 5.34E-05 | 0.011 | 36 |
| RT046 | 9 336 850 | 3 968 112 | 5 046 | 627 | 0.127% | 9.65E-06 | 0.002 | 35 |
| RT047 | 11 809 142 | 3 892 847 | 5 668 | 775 | 0.146% | 1.27E-05 | - | 38 |
| RT048 | 9 495 499 | 3 094 254 | 122 327 | 10 924 | 3.953% | 1.98E-04 | 0.019 | 43 |
| RT049 | 9 500 308 | 4 036 012 | 7 474 | 2 331 | 0.185% | 3.40E-05 | 0.004 | 35 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RT050 | 12 712 975 | 5 607 903 | 11 499 | 3 567 | 0.205% | 5.96E-05 | - | 39 |
| RT051 | 14 638 662 | 8 765 800 | 2 023 662 | 702 772 | 23.086% | 1.38E-02 | 0.525 | 46 |
| RT052 | 12 741 848 | 3 011 513 | 8 781 | 2 068 | 0.292% | 2.98E-05 | 0.006 | 34 |
| RT053 | 12 610 796 | 6 603 878 | 376 191 | 163 155 | 5.697% | 3.69E-03 | 0.204 | 54 |
| RT054 | 4 514 693 | 1 579 125 | 8 644 | 3 814 | 0.547% | 6.90E-05 | 0.002 | 42 |
| RT055 | 16 702 333 | 3 461 403 | 11 827 | 493 | 0.342% | 8.18E-06 | - | 39 |
| RT057 | 17 485 746 | 4 738 379 | 305 460 | 9 942 | 6.447% | 1.57E-04 | - | 37 |
| RT058 | 10 528 560 | 2 267 223 | 8 463 | 3 951 | 0.373% | 5.83E-05 | 0.010 | 35 |
| RT059 | 11 607 642 | 4 562 579 | 43 620 | 6 018 | 0.956% | 1.20E-04 | 0.010 | 47 |
| RT060 | 9 317 083 | 3 724 071 | 9 359 | 1 726 | 0.251% | 2.77E-05 | 0.006 | 39 |
| RT061 | 7 095 507 | 3 618 020 | 1 542 800 | 939 425 | 42.642% | 1.88E-02 | 1.002 | 47 |
| RT062 | 11 995 763 | 2 919 795 | 3 079 | 429 | 0.105% | 6.65E-06 | - | 33 |
| RT063 | 16 010 708 | 8 229 813 | 14 345 | 1 620 | 0.174% | 2.68E-05 | - | 38 |
| RT064 | 6 932 459 | 3 179 624 | 10 172 | 5 661 | 0.320% | 9.14E-05 | 0.002 | 38 |
| RT065 | 5 452 240 | 3 113 823 | 396 617 | 171 265 | 12.737% | 3.15E-03 | 0.120 | 43 |
| RT066 | 5 929 921 | 2 349 805 | 3 932 | 576 | 0.167% | 1.02E-05 | - | 39 |
| RT067 | 7 165 165 | 2 836 914 | 3 188 | 371 | 0.112% | 6.17E-06 | - | 39 |
| RT068 | 14 523 078 | 4 893 097 | 16 197 | 1 595 | 0.331% | 3.03E-05 | 0.019 | 43 |
| RT069 | 17 560 911 | 5 117 329 | 160 555 | 30 003 | 3.137% | 5.45E-04 | 0.017 | 42 |
| RT070 | 15 399 864 | 6 558 946 | 11 865 | 5 582 | 0.181% | 9.29E-05 | 0.006 | 38 |
| RT071 | 7 190 979 | 2 150 735 | 17 355 | 9 358 | 0.807% | 1.73E-04 | 0.023 | 44 |
| RT072 | 17 883 507 | 5 133 958 | 56 778 | 5 284 | 1.106% | 9.56E-05 | 0.011 | 42 |
| RT073 | 6 417 822 | 3 247 137 | 5 377 | 2 522 | 0.166% | 4.52E-05 | 0.002 | 41 |
| RT074 | 19 557 134 | 6 768 582 | 4 891 | 833 | 0.072% | 1.10E-05 | 0.010 | 32 |
| RT075 | 18 464 201 | 6 125 259 | 7 028 | 3 114 | 0.115% | 4.20E-05 | 0.007 | 33 |
| RT076 | 9 458 616 | 3 143 585 | 15 616 | 8 304 | 0.497% | 1.36E-04 | 0.009 | 39 |
| RT077 | 4 560 258 | 1 223 946 | 7 909 | 5 097 | 0.646% | 8.77E-05 | 0.016 | 41 |
| RT078 | 5 789 961 | 1 624 127 | 6 909 | 2 939 | 0.425% | 5.43E-05 | 0.002 | 44 |
| RT079 | 9 682 621 | 2 441 989 | 4 550 | 946 | 0.186% | 1.26E-05 | - | 33 |
| RT080 | 10 494 332 | 2 397 387 | 6 632 | 3 317 | 0.277% | 4.70E-05 | 0.005 | 34 |
| RT081 | 14 834 351 | 2 910 460 | 7 940 | 276 | 0.273% | 4.61E-06 | - | 40 |
| RT082 | 20 787 517 | 7 279 072 | 10 440 | 5 906 | 0.143% | 8.15E-05 | 0.045 | 33 |
| RT083 | 8 832 703 | 3 796 559 | 2 712 | 314 | 0.071% | 4.54E-06 | - | 33 |
| RT084 | 6 531 038 | 1 547 109 | 1 656 | 174 | 0.107% | 2.98E-06 | - | 34 |
| RT085 | 5 541 392 | 3 008 041 | 5 083 | 2 217 | 0.169% | 4.32E-05 | 0.007 | 46 |
| RT086 | 9 999 448 | 5 330 795 | 81 618 | 22 729 | 1.531% | 4.38E-04 | 0.154 | 45 |
| RT087 | 11 764 405 | 5 296 963 | 342 431 | 53 959 | 6.465% | 1.06E-03 | 0.123 | 46 |
| RT088 | 4 761 642 | 1 777 134 | 508 190 | 98 295 | 28.596% | 1.92E-03 | 0.063 | 46 |
| RT089 | 5 589 284 | 2 644 973 | 3 375 | 902 | 0.128% | 1.46E-05 | - | 34 |
| RT090 | 2 484 480 | 1 103 624 | 4 097 | 740 | 0.371% | 1.25E-05 | - | 38 |
| RT091 | 3 234 718 | 910 108 | 2 621 | 265 | 0.288% | 3.90E-06 | - | 32 |
| RT092 | 3 612 329 | 769 040 | 2 241 | 351 | 0.291% | 5.13E-06 | - | 34 |
| RT093 | 7 076 135 | 2 092 132 | 7 725 | 1 997 | 0.369% | 3.29E-05 | 0.006 | 35 |
| RT094 | 10 034 158 | 4 337 996 | 7 607 | 2 080 | 0.175% | 3.47E-05 | - | 36 |
| RT095 | 10 499 326 | 5 990 521 | 9 166 | 3 036 | 0.153% | 5.23E-05 | 0.009 | 37 |
| RT096 | 14 603 283 | 10 904 075 | 10 097 | 5 476 | 0.093% | 9.70E-05 | 0.006 | 40 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RT097 | 18 235 375 | 8 253 357 | 14 924 | 6 219 | 0.181% | 1.02E-04 | 0.029 | 35 |
| RT098 | 18 927 855 | 5 327 589 | 7 289 | 2 073 | 0.137% | 3.21E-05 | 0.006 | 32 |
| RT099 | 12 388 361 | 4 404 425 | 5 568 | 1 642 | 0.126% | 2.67E-05 | - | 34 |
| RT100 | 9 214 605 | 2 503 939 | 3 491 | 1 166 | 0.139% | 1.74E-05 | 0.002 | 33 |
| RT101 | 16 927 121 | 5 310 263 | 14 904 | 3 742 | 0.281% | 5.93E-05 | 0.011 | 34 |
| RT102 | 8 279 447 | 3 344 086 | 28 836 | 9 065 | 0.862% | 1.56E-04 | 0.009 | 38 |
| RT103 | 7 023 210 | 2 444 365 | 4 012 | 1 011 | 0.164% | 1.56E-05 | 0.016 | 33 |
| RT104 | 7 797 376 | 2 650 116 | 2 297 | 563 | 0.087% | 9.14E-06 | 0.006 | 31 |
| RT105 | 10 062 504 | 3 058 596 | 5 200 | 1 408 | 0.170% | 2.15E-05 | 0.002 | 34 |
| RT106 | 6 343 691 | 3 137 368 | 5 971 | 1 946 | 0.190% | 2.95E-05 | 0.034 | 34 |
| RT107 | 6 558 761 | 2 490 676 | 106 214 | 24 816 | 4.264% | 4.20E-04 | 0.045 | 38 |
| RT108 | 6 111 783 | 3 587 445 | 1 036 688 | 262 766 | 28.898% | 4.35E-03 | 0.076 | 37 |
| RT109 | 8 438 583 | 4 374 744 | 5 813 | 2 025 | 0.133% | 3.27E-05 | 0.028 | 34 |
| RT110 | 1 381 289 | 375 015 | 1 370 | 207 | 0.365% | 3.24E-06 | 0.002 | 34 |
| RT111 | 765 924 | 225 697 | 11 278 | 7 433 | 4.997% | 1.19E-04 | 0.009 | 36 |
| RT112 | 11 847 975 | 2 205 021 | 4 961 | 1 930 | 0.225% | 2.94E-05 | 0.002 | 34 |
| RT113 | 7 142 891 | 2 514 104 | 886 897 | 318 347 | 35.277% | 4.93E-03 | 0.102 | 35 |
| RT114 | 1 952 706 | 784 169 | 1 369 | 222 | 0.175% | 3.58E-06 | 0.002 | 34 |
| RT115 | 5 763 714 | 2 167 528 | 75 896 | 18 702 | 3.502% | 2.92E-04 | 0.009 | 35 |
| RT116 | 15 999 822 | 6 624 693 | 15 459 | 7 229 | 0.233% | 1.13E-04 | 0.009 | 34 |
| RT117 | 4 696 163 | 2 241 871 | 283 023 | 85 963 | 12.624% | 1.37E-03 | 0.049 | 36 |
| RT118 | 5 761 183 | 2 224 352 | 42 119 | 16 344 | 1.894% | 2.58E-04 | 0.014 | 35 |
| RT119 | 1 208 343 | 481 952 | 49 688 | 18 674 | 10.310% | 2.97E-04 | 0.018 | 36 |
| RT120 | 4 822 485 | 1 864 166 | 4 355 | 526 | 0.234% | 9.00E-06 | - | 38 |
| RT121 | 10 310 043 | 5 409 716 | 13 297 | 1 890 | 0.246% | 3.35E-05 | 0.005 | 40 |
| RT122 | 8 564 860 | 3 373 758 | 13 913 | 1 637 | 0.412% | 3.50E-05 | 0.002 | 51 |
| RT123 | 20 235 843 | 7 753 952 | 86 327 | 8 159 | 1.113% | 1.59E-04 | 0.007 | 45 |
| RT124 | 6 872 597 | 2 402 096 | 4 361 | 1 901 | 0.182% | 2.87E-05 | 0.004 | 35 |
| RT125 | 9 054 090 | 2 191 849 | 6 379 | 1 751 | 0.291% | 2.65E-05 | 0.006 | 34 |
| RT126 | 14 007 934 | 6 071 007 | 7 017 | 2 732 | 0.116% | 4.41E-05 | 0.002 | 34 |
| RT127 | 3 005 490 | 1 072 959 | 9 420 | 3 266 | 0.878% | 5.22E-05 | - | 35 |
| RT128 | 16 059 031 | 5 435 559 | 7 341 | 2 349 | 0.135% | 3.73E-05 | - | 34 |
| RT129 | 3 532 874 | 1 788 628 | 124 610 | 32 774 | 6.967% | 5.71E-04 | 0.039 | 39 |
| RT130 | 10 261 052 | 5 436 553 | 20 725 | 11 014 | 0.381% | 1.75E-04 | 0.008 | 36 |
| RT131 | 9 045 346 | 3 676 453 | 5 711 | 2 022 | 0.155% | 3.28E-05 | 0.002 | 33 |
| RT132 | 5 951 792 | 1 860 060 | 8 734 | 2 752 | 0.470% | 4.30E-05 | 0.004 | 34 |
| RT133 | 9 305 050 | 3 328 122 | 13 307 | 4 970 | 0.400% | 7.64E-05 | 0.004 | 34 |
| RT134 | 13 138 259 | 5 499 035 | 6 164 | 2 844 | 0.112% | 4.17E-05 | - | 31 |
| RT135 | 10 865 539 | 5 506 893 | 62 213 | 25 550 | 1.130% | 4.34E-04 | 0.052 | 38 |
| RT136 | 8 071 643 | 3 804 628 | 384 373 | 97 437 | 10.103% | 1.55E-03 | 0.036 | 36 |
| RT137 | 6 893 311 | 3 262 464 | 511 402 | 219 151 | 15.675% | 3.47E-03 | 0.123 | 35 |
| RT138 | 4 136 614 | 2 110 522 | 21 300 | 8 982 | 1.009% | 1.54E-04 | 0.028 | 39 |
| RT139 | 2 327 849 | 941 081 | 7 550 | 2 934 | 0.802% | 4.74E-05 | 0.005 | 36 |
| RT140 | 10 869 229 | 4 521 684 | 9 606 | 3 702 | 0.212% | 5.92E-05 | 0.004 | 34 |
| RT141 | 6 022 474 | 4 954 356 | 50 444 | 35 716 | 1.018% | 7.50E-04 | 0.080 | 48 |
| RT142 | 9 176 251 | 7 064 361 | 7 277 | 5 114 | 0.103% | 1.04E-04 | 0.054 | 44 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RT143 | 7 098 804 | 5 628 831 | 35 020 | 24 835 | 0.622% | 4.72E-04 | 0.094 | 43 |
| RT144 | 13 562 081 | 10 397 484 | 16 336 | 11 523 | 0.157% | 2.22E-04 | 0.032 | 43 |
| RT145 | 11 500 912 | 8 562 780 | 13 719 | 8 910 | 0.160% | 1.67E-04 | 0.033 | 42 |
| RT146 | 6 559 272 | 4 765 820 | 55 810 | 39 757 | 1.171% | 8.21E-04 | 0.024 | 47 |
| RT147 | 2 630 887 | 2 035 061 | 1 926 | 1 408 | 0.095% | 2.49E-05 | 0.005 | 41 |
| RT148 | 26 796 217 | 21 654 575 | 15 322 | 8 172 | 0.071% | 1.50E-04 | 0.027 | 40 |
| RT149 | 15 390 274 | 12 755 290 | 33 472 | 19 749 | 0.262% | 3.76E-04 | 0.064 | 43 |
| **Positive control** | | | | | | | | |
| AG224 | 3 913 871 | 1 464 964 | 36 553 | 18 418 | 2.495% | 2.71E-04 | 0.014 | 35 |
| MK+ | 3 026 481 | 2 187 753 | 1 289 | 534 | 0.059% | 7.71E-06 | - | 32 |
| **Negative control** | | | | | | | | |
| RTBL3 | 1 227 616 | 721 918 | 445 | 140 | 0.062% | 2.30E-06 | - | 33 |
| RTBL5 | 1 474 136 | 629 327 | 6 705 | 4 386 | 1.065% | 8.23E-05 | 0.002 | 44 |
| RTBL7 | 713 522 | 327 298 | 572 | 78 | 0.175% | 1.52E-06 | - | 35 |
| RTBL8 | 532 158 | 293 268 | 407 | 76 | 0.139% | 1.80E-06 | - | 33 |
| RTBL9 | 1 166 636 | 586 513 | 2 711 | 169 | 0.462% | 2.42E-06 | - | 33 |
| RTDB10042025 | 41 | 15 | 1 | 0 | 6.667% | 0.00E+00 | - | - |
| PBL2 | 1 641 656 | 755 424 | 582 | 302 | 0.077% | 3.99E-06 | - | 31 |
| PBL6 | 4 661 683 | 3 777 059 | 1 935 | 570 | 0.051% | 9.68E-06 | 0.002 | 33 |
| PBL7 | 2 867 975 | 2 312 729 | 1 086 | 354 | 0.047% | 6.00E-06 | - | 33 |
| PBL8 | 491 573 | 279 527 | 422 | 74 | 0.151% | 1.77E-06 | - | 33 |
| PBL9.2 | 2 431 163 | 1 998 398 | 496 | 203 | 0.025% | 3.60E-06 | - | 34 |
| SC- | 1 207 417 | 772 803 | 76 | 17 | 0.010% | 6.54E-07 | - | 36 |
| MK- | 26 535 | 2 920 | 44 | 5 | 1.507% | 1.68E-06 | - | 36 |
| RTMK11042025 | 62 129 | 13 585 | 35 | 8 | 0.258% | 9.75E-07 | - | 36 |
| RTMK25042025 | 162 797 | 36 876 | 89 | 18 | 0.241% | 1.47E-06 | - | 41 |
| iPCR_blank | 28 456 | 521 | 8 | 5 | 1.536% | 1.04653E-06 | - | 36 |

**Supplementary Table 4. Bait species on the panel for the first round of competitive mapping.**

| Accession number | Species (binomial nomenclature) | Species (common name) |
|---|---|---|
| NC_006853.1 | *Bos taurus* | Cattle |
| NC_009849.1 | *Camelus dromedarius* | Dromedary |
| NC_008092.1 | *Canis lupus* | Grey wolf |
| NC_007704.1 | *Cervus elaphus* | Red deer |
| NC_007596.1 | *Mammuthus primigenius* | Woolly mammoth |
| NC_001640.1 | *Equus caballus* | Horse |
| NC_010638.1 | *Uncia uncia* | Snow leopard |
| NC_012920.1 | *Homo sapiens* | Human |
| NC_020670.1 | *Crocuta crocuta* | Spotted hyena |
| NC_009685.1 | *Gulo gulo* | Wolverine |
| NC_012681.1 | *Coelodonta antiquitatis* | Woolly rhinoceros |
| NC_000845.1 | *Sus scrofa* | Wild boar |
| NC_011112.1 | *Ursus spelaeus* | Cave bear |

**Supplementary Table 5. R packages used in data visualization.**

| Package | Version | Application | Reference |
|---|---|---|---|
| dplyr | v1.1.4 | Data manipulation | https://dplyr.tidyverse.org |
| elevatr | v0.99.0 | Geographical mapping (Fig 1.4) | https://cran.r-project.org/package=elevatr |
| ggplot2 | v3.5.0 | Plotting (Fig 3.1) | (Wickham, 2016) |
| ggspatial | v1.1.9 | Plotting (Fig 3.1) | https://cran.r-project.org/package=ggspatial |
| raster | v3.6-32 | Geographical mapping (Fig 1.4) | https://cran.r-project.org/package=raster |
| readr | v2.1.5 | Data manipulation | https://readr.tidyverse.org/ |
| rnaturalearth | v1.0.1 | Geographical mapping (Fig 1.4) | (Massicotte & South, 2017) |
| rnaturalearthdata | v1.0.0 | Geographical mapping (Fig 1.4) | (South et al., 2017) |
| rworldmap | v1.3-8 | Geographical mapping (Fig 1.4) | (South, 2011) |
| rworldxtra | v1.01 | Geographical mapping (Fig 1.4) | (South, 2012) |
| sf | v1.0-20 | Geographical mapping (Fig 1.4) | (Pebesma, 2016) |
| tidyr | v1.3.1 | Data manipulation | (Wickham et al., 2025) |

**Supplementary Table 6. Mapping statistics of the libraries investigated for by-catch mitogenome content.**
**"**Endo content" refers to the endogenous content, calculated from the reads mapped/total reads. "Mito depth" and "Mito depth (after MIA)" refer to coverage of the mitochondrial genome.

| Sample | Layer | Species | By-catch species | By-catch mapping statistics | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Mapping reference | Mitogenome reference | Endo content | Total depth | Mito depth | Mito depth (after MIA) |
| AG004 | 15 | Bos/Bison* | Wolf (*C. lupus*) | [1] | NC_008092.1 | 2.57% | 0.0056 | 0.8886 | 0.914 (30 iterations) |
| AG031 | 12.1-12.2 | Aurochs (*Bos primigenius*) | Hyena (*C. crocuta*) | GCA_008692635.1 | NC_020670.1 | 4.99% | 0.0085 | 1.2248 | 2.272 (8 iterations) |
| AG076 | 12 | European bison (*Bison bonasus*) | Hyena (*C. crocuta*) | GCA_008692635.1 | NC_020670.1 | 11.15% | 0.0029 | 0.2169 | 0.559 (4 iterations) |
| AG083 | 11.2 | Bos/Bison* | Hyena (*C. crocuta*) | GCA_008692635.1 | NC_020670.1 | 8.71% | 0.0112 | 1.7691 | 3.189 (30 iterations) |
| AG216 | 11.2 | Aurochs (*Bos primigenius*) | Hyena (*C. crocuta*) | GCA_008692635.1 | NC_020670.1 | 2.83% | 0.0014 | 0.1873 | 0.312 (4 iterations) |
| RT108 | 12 | Rhinocerontidae* | Hyena (*C. crocuta*) | GCA_008692635.1 | NC_020670.1 | 4.96% | 0.0005 | 0.0145 | *NA* |
| AG262 | 13 | Bos/Bison* | Sheep (*O. aries*) | GCF_016772045.2 | NC_001941.1 | 5.07% | 0.0017 | 0.2458 | 0.291 (3 iterations) |

[1] https://sid.erda.dk/wsgi-bin/ls.py?share_id=f1ppDgUPQG
*ZooMS data; species was not genetically determined.