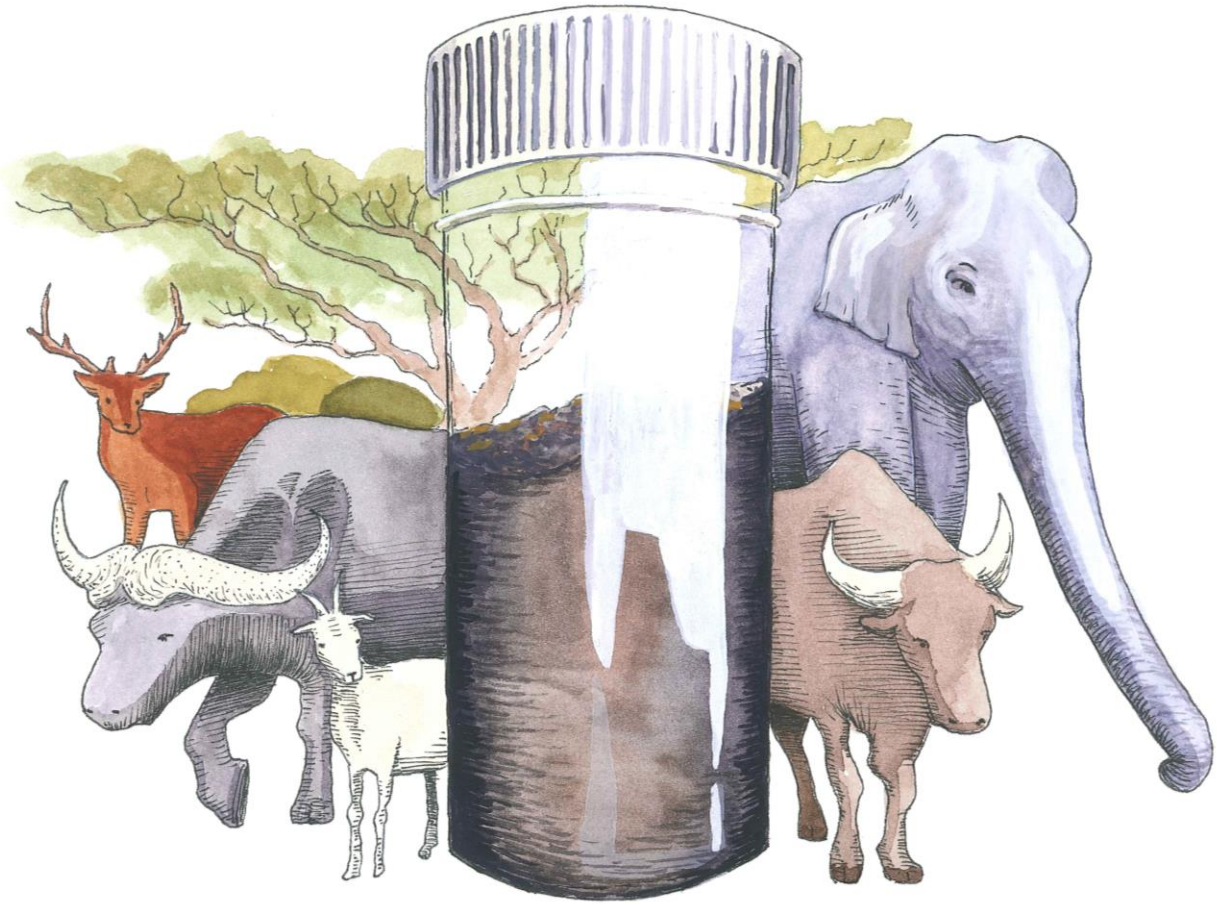
Dung, Dirt & DNA

Unearthing past and present biodiversity using environmental DNA metabarcoding

Anneke T.M. ter Schure

Dissertation presented for the degree of
Philosophiae Doctor (PhD)
2021



Centre for Ecological and Evolutionary Synthesis
Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2462

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Contents

Preface page iii
Acknowledgements iv
Summary v

Introduction 1
The multidimensionality of biodiversity
What causes/-d biodiversity change?
From dirt and faeces to species
Aims and outline

1 Ancient DNA from sediments 17
Molecular Identification of Plants: From Sequence to Species

2 eDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary 41
Environmental DNA 3, 681-696, 2021

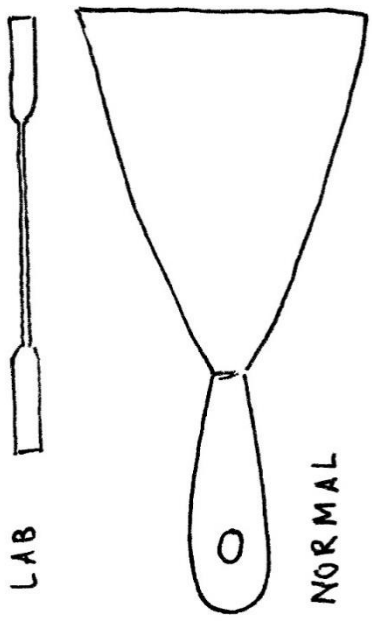
3 Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna 79
Quaternary Science Reviews 267, 107084, 2021

4 Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene 117
Quaternary Science Reviews 270, 107175, 2021

5 *Sed*aDNA metabarcoding as a tool for assessing prehistoric plant use at Upper Palaeolithic cave site Aghitu-3, Armenia 173
Manuscript

Discussion 209
Dung: reconstructing diets
Dirt: reconstructing human-environment interactions
DNA: a window to things past

SPATULAS



Preface

I was wearing a synthetic lab coat over my clothes, two sets of gloves (long nitrile ones covering the ends of my sleeves and topped by a set of latex gloves), a hairnet and a facemask that was fogging up my glasses. It was a few days into my PhD and I found myself in an industrial area of Oslo, armed with lab spatulas not much bigger than an average pen, against a massive PVC tube of several meters in length, full of dense mud. We, the ancient DNA team consisting this time of Agata, Giada, Sanne and me, made a slight miscalculation in what a 'spatula' means for us and what it means for the rest of the world. We needed something to neatly split the sediment core lengthwise into two seperate sections and fortunately, the rest of the team was better prepared: they brought proper scrapers at least ten times the width of our tiny lab spatulas.

The mud we were up against was from a lake, and this particular sediment core was taken earlier that summer by a group of archaeologists interested in the human-mediated introduction of fish to freshwater lakes, an early form of fish farming. Perhaps the mishap with the spatulas was an omen, because despite our best efforts that day and the coming months in the ancient DNA lab, we ended up without any usable DNA. As far as ancient DNA projects go, I learned that a flexible mindset goes a long way. On the plus side, our elaborate outfits protecting the taken samples from contamination with unwanted DNA seemed to work wonderfully.

It was some time later during *Norskkurs* - a course to learn Norwegian organised for international researchers - that I met Manon. Manon was just starting her postdoc at the geology department and was at the start of a project where she wanted to include DNA analysis of lake sediments. What serendipity! This time the sampling team consisted of Manon, Eirik, Sanne and me, and with the proper spatulas and the aid of two circular saws attached to a custom-made cutting bench, opening up this lake sediment core went much smoother than last time. More than that, this time we did recover usable DNA and the resulting collaborative project has turned into one of the major chapters of this book.

For someone with mostly a background in ecology, very limited practice in DNA lab work or the realm of bioinformatics, starting a PhD on DNA metabarcoding of ancient sediments and faeces was, to say the least, slightly daunting. From first learning about ancient DNA sampling and lab practices, making reference libraries from public databases, metabarcoding data processing and analyses, diving into the different topics and working in diverse interdisciplinary teams, each chapter in this book is a testament to my progress as well as a bunch of really wonderful collaborations.

Acknowledgements

My PhD journey has led me from trying to push a 6 meter long PVC tube into lake sediments on a rainy day in Norway, to sunny cave excavations in Armenia, many days in a full-body suit in the ancient DNA lab and especially many days in front of my computer at the office or at home. All in all, it was an unforgettable experience, with much laughter, social things, a few minor setbacks, but mostly one big adventure. My utmost thanks to my incredible supervisor team, and particularly my main supervisor Sanne Boessenkool. Thank you so much for the many things you have taught me, your keen sense of what I needed and perhaps more importantly, our easy conversations and including me into the Friday-night pancake tradition with the whole family when I first arrived in Oslo. Thank you Hugo de Boer and Anne Brysting for your constructive supervision approaches, positive attitudes, very helpful advice and comments and keen eye for detail. I would also like to thank my co-authors for their contributions to the papers in this thesis.

This adventure would definitely not be the same without the wonderful atmosphere within the Archaeogenomics group. Thank you Bastiaan for your scientific insights, baking the pancakes, and generally providing cheer. Thank you Agata and Giada, for handling the dirt with me and teaching me how to process everything in the ancient DNA lab. Thanks Oliver, Albína, Lulú, Lane and our newest members, for our lunches and coffees and generally making the working days at the office as well as the home office, a little brighter. Working at CEES over the last few years has been a pleasure and I am grateful for my friends at work, including the Grad-forum as a whole, and the other board members Ada, Lisa, Melanie and Synne in particular, for sharing coffee breaks, lunches, stories, beers, pizza and pie.

To my parents and family, who have encouraged me to pursue this adventure and even helped me move several times, and to my friends. Thank you for our many zoom dates, sending countless care packages with chocolate sprinkles, stroopwafels and other treats that I missed, for visiting me, sending cards, messages of encouragement, forgiving me when I forgot to respond, and for simply always picking up where we left off.

Finally, thank you Roland, for motivating me to start this PhD, for joining me on this journey, filling my days with happiness and cheer, supporting me when I needed support, and for laughing with me when things do not turn out as planned.

Summary

The rapid loss of biodiversity highlights the urgency of large-scale biodiversity monitoring for conservation, as well as the importance of understanding biodiversity in all its dimensions, especially in light of environmental change. In order to understand the causes of these changes in biodiversity, we need to track them over time, either through continued biomonitoring, or through reconstructions of past changes. Trace DNA in environmental samples such as sediments and faeces, allows for the detection of species and biodiversity monitoring without the need to sight or sample the actual organisms. This is not only useful for conservation purposes, but also for the detection of species in palaeo records, such as those presented by lakes, caves, and permafrost. Particularly eDNA metabarcoding, a tool for the simultaneous identification of many organisms in an environmental sample, is a useful tool that can provide a window to the biological past.

With initial technological issues involved in these methods largely worked out, and many guides are being published on how to apply them (including chapter 1), the field of eDNA metabarcoding can now move towards enhancing the interpretation of the resulting data. This is especially complex in interdisciplinary contexts and testing of its application to a range of topics and contexts is therefore needed.

By applying these methods to faecal and sediment samples, this thesis provides insight into its applications through the biomonitoring of herbivores in a wildlife reserve in India (chapter 2), reconstructing the diets and habitats of extinct and extant megafauna (chapter 3), the untangling of human-environment interactions (chapter 4) and the assessment of DNA metabarcoding for reconstructing prehistoric human plant use (chapter 5). In interdisciplinary collaborations such as many of these chapters, good communication is very important. Environmental DNA metabarcoding, combined with statistical approaches as well as clear visualisation of the obtained results, such as those employed here, provide a way forward, untangling the complexity of biodiversity in all its dimensions. Application of these methods to modern (chapter 2) and palaeo (chapters 3-5) records can provide unprecedented insight into past and present biodiversity, including biological interactions, effects of climate change and human activities.

Introduction

Introduction

The air we breathe, the water we drink and the food we eat all ultimately rely on biodiversity. Less obvious are the interactions between species, evolved over millions of years, together producing intricate and resilient systems, which become vulnerable to collapse when too many species disappear. Biodiversity is currently in crisis as we are facing catastrophic species loss. Understanding what causes these changes in biodiversity is important for its conservation. Trace DNA in environmental samples such as water, soil or faeces, collectively referred to as eDNA, allows for the detection of species and biodiversity monitoring without the need to sight or sample the actual organisms. Not only is this useful for monitoring current endangered or invasive species, when applied to palaeo records these technologies provide a window to the biological past, including biological interactions, past effects of climate change and human activities. Particularly eDNA metabarcoding, a tool for the simultaneous identification of many organisms in an environmental sample, can provide unprecedented insight into past and present biodiversity and human-environment interactions, as presented in this thesis.

The multidimensionality of biodiversity and the importance of interactions

The 6th mass extinction, the Age of Extinction, or the biodiversity crisis all refer to the current high rate of species loss (Barnosky et al., 2011; Díaz et al., 2019; Koh et al., 2004; Naeem et al., 2012; Novacek & Cleland, 2001; Pievani, 2014). Together with the impacts of anthropogenic climate change on the biosphere, the biodiversity crisis highlights the urgency of large-scale biodiversity monitoring for conservation, as well as the need to understand biological interactions and processes in light of environmental change. But, even without taking into account these environmental changes, biodiversity is a complex concept to monitor and study.

Biodiversity – a contraction of "biological diversity" – has many definitions but is generally described as the diversity of life, or as per the UN Convention of Biological Diversity: "the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part: this includes diversity within species, between species and of ecosystems." Different dimensions of biodiversity can be distinguished, including genetic, phylogenetic, taxonomic, interaction (also known as network), functional, spatial, temporal, and landscape diversity, and many of these are interrelated (Naeem et al., 2012). For example, the presence of different evolutionary lineages (phylogenetic diversity) and the richness and abundance of species (taxonomic diversity) are directly correlated. And both of these forms of diversity can also be correlated to functional traits, such as the diversity of growth

forms and resource use strategies (functional diversity). Despite the term 'DNA' in the title of this thesis, it is not genetic diversity that is under study here, but more so taxonomic, temporal, landscape and interaction diversity. DNA is in this case merely a tool to assess the taxonomic diversity in the samples under study. We can infer other dimensions of biodiversity from the taxonomic composition either through formal analysis or in a more descriptive fashion. Temporal diversity refers to the change in taxonomic diversity over time. Landscape diversity to the different habitat types within a landscape. Interaction diversity can be defined as the characteristics of the linkages in a biotic network, or the number and abundance of biotic interactions, such as competition and predation, in a community (Dyer et al., 2010). Trophic networks and food webs can be seen as subsets of such networks (Naeem et al., 2012).

While biodiversity is often measured as the number of biological items (such as genes, species or ecosystems) and their relative abundance, the term can also refer to biological assemblages, activities and interactions, and it is these connections that form the structure, or interaction network of functional biological communities and ecosystems (Sadava et al., 2014; Swingland, 2001). Ecosystems do not only include biological communities, but also abiotic components and processes. Ecosystems with higher biological diversity tend to have a higher resilience, i.e. the ability to regain fundamental structures, processes and functioning after perturbations, or in other words, the capacity to recover after environmental change or disturbance (Scheffer et al., 2012; Seidl et al., 2016). With all dimensions of biodiversity playing an essential role in ecosystems, local and global species losses could threaten ecosystem resilience and diversity, especially when a disturbance or change in the environment is severe enough to push the ecosystem into a zone where it is no longer resilient, leading to permanent alteration or loss of the ecosystem. Continuous biodiversity monitoring is therefore needed, as well as identification of the factors causing biodiversity change.

What causes/-d biodiversity change?

The current loss of biodiversity is mainly caused by human activities and this has previously been linked to the industrial revolution (Crutzen, 2002; Lewis & Maslin, 2015). Anthropogenic drivers of biodiversity loss include land transformation, habitat fragmentation, exploitation (such as hunting and fishing), pollution, species introductions and anthropogenic climate change. Most records used to assess these changes are based on monitoring data from only the last few decades, while our modern ecosystems are the result of a longer history of climate and anthropogenic influences (Boivin et al., 2016; Boivin & Crowther, 2021; Ellis, 2015; Ruddiman, 2003, 2013; Vitousek et al., 1997; Williams et al., 2015; Willis & Birks, 2006). To understand the shaping of these ecosystems, a long-term perspective is needed.

Recent archaeological assessments on past land-use indicate that we, humans, became a global environmental force at least 3000 years ago (Stephens et al., 2019) and have shaped most terrestrial ecosystems for 12,000 years (Ellis et al., 2021), long before the industrial revolution. The transformation of land for human use alters the structure and functioning of ecosystems as well as the interactions of these ecosystems with each other, with the atmosphere, and with surrounding land (Vitousek et al., 1997) and these effects can be long-lasting (Boivin et al., 2016). For example, a study in France identified reduced plant species richness in forested areas that were once Roman farms (Dambrine et al., 2007), indicating long-term impact of agricultural practices on forest biodiversity. Even changes to individual species have been shown to have potentially long-term cascading effects transforming the landscape. A famous example is the relatively recent re-introduction of the wolf in Yellowstone park, where elk initially responded to the wolves by eating somewhere else (Fortin et al., 2005). Subsequent decreasing elk populations and the resulting increase in available woody plants and herbaceous forage further allowed the number of bison, as well as beaver - well known for their effects on the landscape as 'ecosystem engineers' - to increase (Ripple & Beschta, 2012).

Increased recognition for the importance of understanding environmental history has resulted in a growing number of palaeoecological studies related to conservation, for example by using past data to: identify invasive species, quantify temporal biodiversity, reconstruct past responses to environmental change, including climate and anthropogenic stressors, and establish modern baselines for conservation (for reviews see Boivin & Crowther, 2021; Dietl et al., 2015). In addition, technological advances in the recovery of biological information from palaeoecological records, as well as archaeological contexts, have greatly facilitated our understanding of relationships within past ecosystems and interactions between past human societies and their environments.

Particularly lake sediments have proven useful for obtaining a continuous sequence of environmental change that can be associated with human activities. Examples include increased erosion related to deforestation and pastoralism in the French Alps (Giguet-Covex et al., 2014), and pastoral farming, on-site slaughtering, and storage of crops revealed by a detailed reconstruction of human activities at an early Medieval Celtic settlement in Ireland (Brown et al., 2021). Both studies used DNA from lake sediments for the identification of plants and mammals.

From dirt and faeces to species

Biodiversity monitoring and reconstruction of past biodiversity can provide a basis for evaluating the integrity of an ecosystem, assess biodiversity change, as well as the effect of disturbances, environmental change or management actions. Such monitoring and reconstruction has traditionally relied on morphological

identification and counting of individuals, which can be complicated due to difficulties associated with correct identifications (e.g. juvenile stages and cryptic species), a continuous decline in taxonomic expertise, and the invasiveness of some survey techniques (Hebert et al., 2003; Thomsen & Willerslev, 2015). Advances in non-invasive genetics, especially in environmental DNA (eDNA) techniques, has made it possible to efficiently detect organisms and monitor biodiversity with unprecedented precision and depth (Barnes & Turner, 2016; Pawlowski et al., 2021).

The term eDNA refers to genetic material obtained directly from environmental samples (such as soil, water, and air), including extracellular DNA, cells, tissues and possibly whole organisms (Barnes & Turner, 2016; Taberlet, Coissac, Hajibabaei, et al., 2012; Thomsen & Willerslev, 2015). Technologies for eDNA analysis have been widely applied to soil (e.g. Edwards et al., 2018; Yoccoz, 2012), sediment (e.g. Hofreiter et al., 2003; Willerslev, 2003; Willerslev et al., 2014; see Capo et al., 2021, for an extensive review of lake sediment studies), water (e.g. Ficetola et al., 2008; Thomsen et al., 2012), and faeces (for reviews see Ando et al., 2020; de Sousa et al., 2019). More recently eDNA has also been successfully used for the detection of insects from surfaces of bark and leaves (Valentin et al., 2020), and even air samples have been used for the detection of DNA from birds, reptiles, amphibians and mammals (Clare et al., 2021; Lynggaard et al., 2021). In short, eDNA represents an efficient, non-invasive and easy-to-standardize sampling approach which is applicable to range of different environments, ancient and modern, terrestrial and aquatic (for reviews see e.g. Barnes & Turner, 2016; Capo et al., 2021; Thomsen & Willerslev, 2015; Veilleux et al., 2021). In this thesis, I focus specifically on eDNA from "dirt" and faeces; "dirt" DNA being the colloquial term for DNA from sediments, more formally it is referred to as sedDNA, or in case of ancient contexts, sedaDNA (Epp et al., 2019).

How can we get from DNA in environmental samples, such as dirt and faeces, to species? The answer to this question can be divided into two parts, as the recovery of biodiversity data from environmental samples requires not only the identification of taxa based on their DNA, but doing this for many taxa at the same time. Two main methodological advances have facilitated this. Firstly, the analysis of sequence variation in short, standardized gene regions (i.e. DNA barcodes) has allowed the identification of individual species and has proven an invaluable technology (Hebert et al., 2003). Secondly, high-throughput sequencing technologies allow the simultaneous acquisition of millions of these DNA barcodes, facilitating the emergence of DNA metabarcoding (Taberlet, Coissac, Pompanon, et al., 2012). DNA metabarcoding is an approach using DNA barcode regions for the automated identification of multiple species from either bulk samples containing entire organisms or environmental samples. When applied to samples that contain damaged and fragmented DNA, such as most environmental samples, DNA metabarcoding requires short enough DNA barcoding regions to still allow amplification of the degraded DNA (preferably <100-150 bp; Taberlet, Coissac,

Pompanon, et al., 2012). Primers for these short regions have been developed for a wide range of taxa, including fungi (Bellemain et al., 2010; Buée et al., 2009; Tedersoo et al., 2015, 2018), plants (Taberlet et al., 2007; Willerslev et al., 2014), insects (Elbrecht et al., 2016; Epp et al., 2012), fish (Evans et al., 2016; Valentini et al., 2016), birds (Epp et al., 2012), and mammals (Giguet-Covex et al., 2014; Taylor, 1996).

Early studies using these techniques exposed potential sources of bias, causing false detection of organisms (false positives) as well as false non-detection (false negatives). Sources of bias include: PCR artefacts, tag jumps and sequencing errors. When working with samples with degraded DNA present in low concentrations, contamination is an especially common issue. The development of rigorous standards for ancient DNA research (Cooper & Poinar, 2000) has set an example for tackling issues of contamination and data reproducibility, particularly relevant for eDNA metabarcoding of ancient samples, as well as other samples containing degraded and low concentration DNA. In recent years there has been an increase in the number of guides for best practices concerning the eDNA metabarcoding workflow, from sampling to bioinformatic analyses (e.g. Alberdi et al., 2018; Creer et al., 2016; Ficetola et al., 2016; Pollock et al., 2018; Taberlet et al., 2018; van der Loos & Nijland, 2021; Zinger et al., 2019). With initial technological issues largely worked out, the field of eDNA metabarcoding can now move towards enhancing the interpretation of the resulting data, which is especially complex in interdisciplinary contexts. Further testing of its application to a range of topics and contexts is therefore needed.

Aims and outline

The current biodiversity crisis and the impacts of anthropogenic climate change on the biosphere highlight the urgency of large-scale biodiversity monitoring for conservation, as well as the need to understand both current and past biological interactions and processes in light of environmental change. The aim of this thesis revolves around the untangling of these complex interactions and processes. To make this feasible for a PhD project spanning only 3 years, I focused on several studies (described in chapters 2-5) and made two main delineations regarding the approach and the topics covered.

As came forward in the previous section, the analysis of DNA from environmental samples is a non-invasive approach for species detection, and DNA metabarcoding allows the simultaneous detection of many organisms from one sample. Improvements in the eDNA metabarcoding approach in the last few decades have moved it from its developmental phase to a phase of application and interpretation. The first delineation is therefore to use eDNA metabarcoding with the aim to test the applicability of this approach in a range of cases, while exploring data analytical and visualisation approaches to enhance ecological interpretation of the obtained data.

The second delineation concerns the topics covered in these case studies. With food webs and human-environment interactions as major themes in ecosystem functioning, these also form the main themes here, concentrating on the diversity of macro-organisms (particularly plants and herbivores) in terrestrial ecosystems. Plants fulfill important roles in ecosystems, protecting soils against erosion, transporting water from soils to the atmosphere, and as photoautotrophs they use light energy and inorganic carbon to produce organic materials, forming the basis of any food chain. Traces of these plants can end up in the faeces of herbivores after being eaten but not fully digested, as well as in sediments through various taphonomic processes (the transfer, deposition, and preservation of remains), motivating the use of faecal or sediment samples in the different chapters of this thesis.

Analysis of herbivore faeces allows not only the reconstruction of which-herbivore-eats-which plant, but also inferences about the habitat of those herbivores and their potential for dietary competition. These topics come forward specifically in chapters 2 and 3, studying competition between wild and domestic herbivores while presenting a starting point for biomonitoring in a wildlife reserve in India (chapter 2), and reconstructing the diets and habitats of extinct and extant megafauna (chapter 3). The study of human-environment interactions further requires the integration of environmental records with evidence of human activities. Traces of past human-activities can be recovered from ancient sediments associated with human occupation through, for example, archaeological evidence. The studies presented in chapters 4 and 5 accordingly include a substantive archaeological component, applying plant sedaDNA metabarcoding to lake sediments as basis for untangling the climatic and human factors impacting vegetation change in southeastern Norway during the Holocene (chapter 4), and to Palaeolithic cave sediments as a means to learn more about the prehistoric environment in Armenia (chapter 5).

Thesis outline

The first chapter is written as part of an educational book entitled “Molecular Identification of Plants: From Sequence to Species” and is meant to guide students on what to consider when making choices concerning the research strategy and design of a sedimentary ancient DNA study. Despite the difference in sample types, the eDNA present in both ancient sediment samples and in faecal samples is degraded and fragmented, motivating similar choices in study design for the four studies described in the remaining chapters. Chapters 2 and 3 are published herbivore dietary reconstructions using modern and ancient faecal samples. Chapter 4 is a published paper on the environmental drivers of vegetation change using lake sediments. Finally, chapter 5 is a manuscript assessing the potential to use cave sediments to reconstruct prehistoric human plant use.

As can be seen in table 1, there are several ways in which the different chapters could be categorized, based on their research themes, the biodiversity

dimensions covered, their statistical approaches or their potential applications. I will provide a short description of the individual chapters here, starting with the first chapter.

As part of an educational book, Chapter 1 guides students on some of the basics of doing research on sedimentary ancient DNA. The book itself, entitled Molecular Identification of Plants: From Sequence to Species, is focused particularly on the study of plants, and is a product of the Plant.ID collaborative network. From the planning stages to extracting the DNA, this chapter is meant to help students to develop an appropriate research design, which questions to ask, how to sample, and what to keep in mind when choosing an extraction protocol. Downstream methods such as metabarcoding, shotgun sequencing, targeted enrichment and bioinformatic analysis are topics of other chapters in Molecular Identification of Plants: From Sequence to Species.

Table 1. Overview of the chapters in this thesis including information on the approach, with details on the sample type and target organisms for eDNA metabarcoding, relevant biodiversity dimensions and relevant field of application per chapter.

Chapter		Approach sample type target DNA	Relevant biodiversity dimensions	Relevant field of application
1	aDNA from sediments	*Literature review; methodology guide for sedaDNA studies*		(Plant) palaeoecology
2	eDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary	Herbivore faeces Plant (*trn*L) and herbivore (species specific primers)	Taxonomic diversity Interaction diversity *diet, competition, human-environment*	Biomonitoring, modern ecology
3	Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna	Herbivore faeces Plant (*trn*L, ITS)	Taxonomic diversity Interaction diversity *diet* Landscape diversity *habitat*	Palaeoecology
4	Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene	Lake sediments Plant (*trn*L) and mammal (16S)	Taxonomic diversity Temporal diversity *vegetation change* Interaction diversity *human-environment*	Palaeoecology, Environmental archaeology
5	*Sed*aDNA metabarcoding as a tool for assessing prehistoric plant use at Upper Palaeolithic cave site Aghitu-3, Armenia	Cave sediments Plant (*trn*L)	Taxonomic diversity Interaction diversity *human-environment*	Plant palaeoecology, Archaeobotany

Chapter 2 applies DNA metabarcoding to faecal samples from domestic and wild herbivores in a wildlife sanctuary in southern India (ter Schure, Pillai, et al., 2021). The Malai Mahadeshwara Wildlife Sanctuary supports a range of wildlife (including several species of deer, bonnet macaque, and Asian elephant) as well as a considerable number of domestic herbivores (cattle, goats and water buffalo), owned by the long-standing ethnic and tribal groups that live in the sanctuary. The main aim of this chapter is to test the application of the eDNA metabarcoding approach for reconstructing the diets of the herbivores that live in the Malai Mahadeshwara Wildlife Sanctuary and assess potential dietary competition, thereby providing a starting point for biomonitoring.

Having set up a bioinformatic pipeline for dietary reconstructions of herbivores for Chapter 2, I was able to adapt these methods for Chapter 3, of which I'm second author. This chapter also deals with faeces for the reconstruction of diet, in this case of herbivores that have long been extinct, roaming the earth during the Pleistocene and Holocene, and also one still surviving species to validate our findings. Obtained from permafrost and ice-preserved woolly mammoth, horse, steppe bison and caribou, these samples are analysed using a multiproxy approach, combining visual analysis of the plant remains, i.e. macrofossil and pollen analysis, with DNA metabarcoding using several sets of primers (Polling et al., 2021). The aim: to provide a more complete picture of what these animals ate than previously possible with only visual techniques, and potentially make inferences about their habitat based on the ecology of the identified plant taxa.

In Chapter 4, we try to distinguish different environmental factors impacting vegetation change in southeastern Norway during the Holocene (ter Schure, Bajard, et al., 2021). We reconstruct abiotic and biotic environments based on the sediments of Lake Ljøgottjern, spanning the last 10,000 years, by using plant and mammal sedaDNA metabarcoding, pollen analysis and geochemical analysis. These reconstructions are integrated with climate data, archaeological evidence of local human settlement and regional human population dynamics in a formal as well as a descriptive analysis, aiming to untangle effects of human land-use and environmental changes on vegetation dynamics.

With the numerous successful studies analysing plant sedaDNA from lake sediments (including chapter 4) and the first publications analysing plant sedaDNA from cave sediments already in 2003 (Willerslev et al., 2003), I was surprised to find that the application to cave sediments had so far been very limited. Sure, there were studies targeting and retrieving hominin and mammal DNA from cave sediments in Eurasia, but DNA from plants had, at the start of this project, only been analysed for cave sediments in New Zealand and Australia (Haile et al., 2007; Haouchar et al., 2014). With the sheltered conditions in dry caves, and many caves having been used by humans, the analysis of sedaDNA from cave sediments could provide a good opportunity to assess prehistoric plant use. Chapter 5 therefore applies plant sedaDNA metabarcoding to the sediments of

Aghitu-3 cave in Armenia, with the aim to assess the potential of this method to reconstruct prehistoric human plant use.

The discussion at the end of this thesis summarizes the results from each chapter and discusses these in light of the topics raised in this introduction. This includes the application and challenges of eDNA methods for biodiversity research and their implications for the inferences we can make when using these methods.

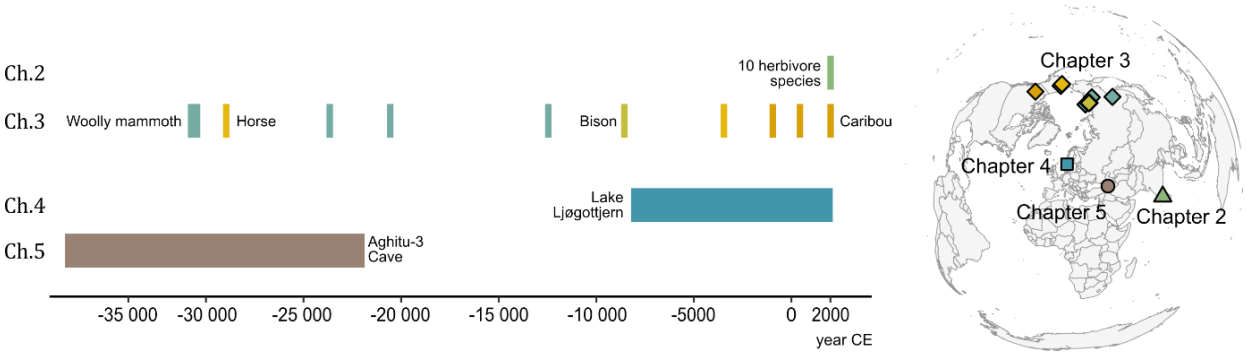


Figure 1. Overview of the locations and ages of the eDNA samples analysed in chapters 2-5.

Note on the supplementary information

Supplementary information consisting of primary data and raw analyses are not included in this book due to the size of tables. Access to these supplemental materials is provided through an online GitHub repository: https://github.com/terschure/dung-dirt-dna

References

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Ando, H., Mukai, H., Komura, T., Dewi, T., Ando, M., & Isagi, Y. (2020). Methodological trends and perspectives of animal dietary studies by noninvasive fecal DNA metabarcoding. Environmental DNA, 2(4), 391–406.

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics, 17(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., & Maguire, K. C. (2011). Has the Earth's sixth mass extinction already arrived? Nature, 471(7336), 51–57.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. BMC Microbiology, 10(1), 1–9.

Boivin, N. L., & Crowther, A. (2021). Mobilizing the past to shape a better Anthropocene. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-020-01361-4

Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., Denham, T., & Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. Proceedings of

the National Academy of Sciences, 113(23), 6388–6396. https://doi.org/10.1073/pnas.1525200113

Brown, A. G., Van Hardenbroek, M., Fonville, T., Davies, K., Mackay, H., Murray, E., Head, K., Barratt, P., McCormick, F., Ficetola, G. F., Gielly, L., Henderson, A. C. G., Crone, A., Cavers, G., Langdon, P. G., Whitehouse, N. J., Pirrie, D., & Alsos, I. G. (2021). Ancient DNA, lipid biomarkers and palaeoecological evidence reveals construction and life on early medieval lake settlements. Scientific Reports, 11(1), 11807. https://doi.org/10.1038/s41598-021-91057-x

Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R., Uroz, S., & Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytologist, 184(2), 449–456.

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A. G., Clarke, C. L., Crump, S. E., Debroas, D., Englund, G., Ficetola, G. F., Garner, R. E., … Parducci, L. (2021). Lake Sedimentary DNA Research on Past Terrestrial and Aquatic Biodiversity: Overview and Recommendations. Quaternary, 4(1), 6. https://doi.org/10.3390/quat4010006

Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F. J., Drinkwater, R., & Littlefair, J. E. (2021). eDNAir: Proof of concept that animal DNA can be collected from air sampling. PeerJ, 9, e11030.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: do it right or not at all. Science, 289(5482), 1139. https://doi.org/10.1126/science.289.5482.1139b

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution, 7(9), 1008–1018. https://doi.org/10.1111/2041-210X.12574

Crutzen, P. J. (2002). Geology of mankind. Nature, 415(6867), 23–23. https://doi.org/10.1038/415023a

Dambrine, É., Dupouey, J.-L., Laüt, L., Humbert, L., Thinon, M., Beaufils, T., & Richard, H. (2007). Present forest biodiversity patterns in France related to former Roman agriculture. Ecology, 88(6), 1430–1439.

de Sousa, L. L., Silva, S. M., & Xavier, R. (2019). DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. Environmental DNA, 1(3), 199–214.

Díaz, S. M., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., & Butchart, S. (2019). The global assessment report on biodiversity and ecosystem services: Summary for policy makers.

Dietl, G. P., Kidwell, S. M., Brenner, M., Burney, D. A., Flessa, K. W., Jackson, S. T., & Koch, P. L. (2015). Conservation Paleobiology: Leveraging Knowledge of the Past to Inform Conservation and Restoration. Annual Review of Earth and Planetary Sciences, 43(1), 79–103. https://doi.org/10.1146/annurev-earth-040610-133349

Dyer, L. A., Walla, T. R., Greeney, H. F., Stireman III, J. O., & Hazen, R. F. (2010). Diversity of interactions: A metric for studies of biodiversity. Biotropica, 42(3), 281–289.

Edwards, M. E., Alsos, I. G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., Haile, J., Langdon, C. T., Tribsch, A., Binney, H. A., & others. (2018). Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. The Holocene, 28(12), 2006–2016.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.-N., Coissac, E., Boyer, F., & Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ, 4, e1966.

Ellis, E. C. (2015). Ecology in an anthropogenic biosphere. Ecological Monographs, 85(3), 287–331. https://doi.org/10.1890/14-2274.1

Ellis, E. C., Gauthier, N., Goldewijk, K. K., Bird, R. B., Boivin, N. L., Díaz, S., Fuller, D. Q., Gill, J. L., Kaplan, J. O., & Kingston, N. (2021). People have shaped most of terrestrial nature for at least 12,000 years. Proceedings of the National Academy of Sciences, 118(17).

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., Erséus, C., Gusarov, V. I., Edwards, M. E., Johnsen, A., Stenøien, H. K., Hassel, K., Kauserud, H., Yoccoz, N. G., Bråthen, K. A., Willerslev, E., Taberlet, P., Coissac, E., & Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. Molecular Ecology, 21(8), 1821–1833.

https://doi.org/10.1111/j.1365-294X.2012.05537.x

Epp, L. S., Zimmermann, H. H., & Stoof-Leichsenring, K. R. (2019). Sampling and extraction of ancient DNA from sediments. In Ancient DNA (pp. 31–44). Springer.

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R., Pfrender, M. E., Lamberti, G. A., & Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources, 16(1), 29–41.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biology Letters, 4(4), 423–425.

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? Molecular Ecology Resources, 16(3), 604–607. https://doi.org/10.1111/1755-0998.12508

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. Ecology, 86(5), 1320–1330.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J., & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications, 5(1). https://doi.org/10.1038/ncomms4211

Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M. T. P., Nielsen, R., Munch, K., Ho, S. Y. W., Shapiro, B., & Willerslev, E. (2007). Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a Factor? Molecular Biology and Evolution, 24(4), 982–989. https://doi.org/10.1093/molbev/msm016

Haouchar, D., Haile, J., McDowell, M. C., Murray, D. C., White, N. E., Allcock, R. J. N., Phillips, M. J., Prideaux, G. J., & Bunce, M. (2014). Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–Holocene cave deposit on Kangaroo Island, South Australia. Quaternary Science Reviews, 84, 56–64. https://doi.org/10.1016/j.quascirev.2013.11.007

Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313–321.

Hofreiter, M., Mead, J. I., Martin, P., & Poinar, H. N. (2003). Molecular caving. Current Biology, 13(18), R693–R695. https://doi.org/10.1016/j.cub.2003.08.039

Koh, L. P., Dunn, R. R., Sodhi, N. S., Colwell, R. K., Proctor, H. C., & Smith, V. S. (2004). Species Coextinctions and the Biodiversity Crisis. Science, 305(5690), 1632–1634. https://doi.org/10.1126/science.1101101

Lewis, S. L., & Maslin, M. A. (2015). Defining the Anthropocene. Nature, 519(7542), 171–180. https://doi.org/10.1038/nature14258

Lynggaard, C., Bertelsen, M. F., Jensen, C. V., Johnson, M. S., Froslev, T. G., Olsen, M. T., & Bohmann, K. (2021). Airborne environmental DNA for terrestrial vertebrate community monitoring. BioRxiv.

Naeem, S., Duffy, J. E., & Zavaleta, E. (2012). The functions of biological diversity in an age of extinction. Science, 336(6087), 1401–1406.

Novacek, M. J., & Cleland, E. E. (2001). The current biodiversity extinction event: Scenarios for mitigation and recovery. Proceedings of the National Academy of Sciences, 98(10), 5466–5470.

Pawlowski, J., Bonin, A., Boyer, F., Cordier, T., & Taberlet, P. (2021). Environmental DNA for biomonitoring. Authorea Preprints.

Pievani, T. (2014). The sixth mass extinction: Anthropocene and the human impact on biodiversity. Rendiconti Lincei, 25(1), 85–93. https://doi.org/10.1007/s12210-013-0258-9

Polling, M., ter Schure, A. T., van Geel, B., van Bokhoven, T., Boessenkool, S., MacKay, G., Langeveld, B. W., Ariza, M., van der Plicht, H., Protopopov, A. V., & others. (2021). Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna. Quaternary Science Reviews, 267, 107084.

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. Applied and Environmental Microbiology, 84(7), e02627-17.

Ripple, W. J., & Beschta, R. L. (2012). Trophic cascades in Yellowstone: The first 15 years

after wolf reintroduction. Biological Conservation, 145(1), 205–213.

Ruddiman, W. F. (2003). The anthropogenic greenhouse era began thousands of years ago. Climatic Change, 61(3), 261–293.

Ruddiman, W. F. (2013). The Anthropocene. Annual Review of Earth and Planetary Sciences, 41(1), 45–68. https://doi.org/10.1146/annurev-earth-050212-123944

Sadava, D. E., Hillis, D., Heller, H. C., & Berenbaum, M. R. (2014). Biodiversity and Conservation Biology. In Life: The Science of Biology (10th ed.). Macmillan.

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., Koppel, J. van de, Leemput, I. A. van de, Levin, S. A., Nes, E. H. van, Pascual, M., & Vandermeer, J. (2012). Anticipating Critical Transitions. Science, 338(6105), 344–348. https://doi.org/10.1126/science.1225244

Seidl, R., Spies, T. A., Peterson, D. L., Stephens, S. L., & Hicke, J. A. (2016). REVIEW: Searching for resilience: addressing the impacts of changing disturbance regimes on forest ecosystem services. Journal of Applied Ecology, 53(1), 120–129. https://doi.org/10.1111/1365-2664.12511

Stephens, L., Fuller, D., Boivin, N. L., Rick, T., Gauthier, N., Kay, A., Marwick, B., Armstrong, C. G., Barton, C. M., Denham, T., Douglass, K., Driver, J., Janz, L., Roberts, P., Rogers, J. D., Thakar, H., Altaweel, M., Johnson, A. L., Sampietro Vattuone, M. M., … Ellis, E. (2019). Archaeological assessment reveals Earth's early transformation through land use. Science, 365(6456), 897–902. https://doi.org/10.1126/science.aax1192

Swingland, I. R. (2001). Biodiversity, definition of. Encyclopedia of Biodiversity, 1, 377–391.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (Eds.). (2018). Environmental DNA: For biodiversity research and monitoring. https://books.google.no/books?hl=nl&lr=&id={1e9IDwAAQBAJ}&oi=fnd&pg={PP1}&dq=taberlet+environmental+dna&ots={UY}-{Vr6wbnO}&sig=9ry5-{qcq2GdL9E6d}-{SdT5Ul6mWM}&redir_esc=y#v=onepage&q=taberlet%20environmental%20dna&f=false

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. Molecular Ecology, 21, 1789–1793.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14–e14. https://doi.org/10.1093/nar/gkl938

Taylor, P. G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Molecular Biology and Evolution, 13(1), 283–285. https://doi.org/10.1093/oxfordjournals.molbev.a025566

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., & others. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys, 10, 1.

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: Errors, biases and perspectives. New Phytologist, 217(3), 1370–1385.

ter Schure, A., Bajard, M., Loftsgarden, K., Høeg, H., Ballo, E., Bakke, J., Støren, E., Iversen, F., Kool, A., Brysting, A., & others. (2021). Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene. Quaternary Science Reviews, 270, 107175.

ter Schure, A., Pillai, A. A., Thorbek, L., Bhavani Shankar, M., Puri, R., Ravikanth, G., de Boer, H. J., & Boessenkool, S. (2021). EDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary. Environmental DNA, 3(3), 681–696.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology, 21(11), 2565–2573.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation, 183, 4–18.

https://doi.org/10.1016/j.biocon.2014.11.019

Valentin, R. E., Fonseca, D. M., Gable, S., Kyle, K. E., Hamilton, G. C., Nielsen, A. L., & Lockwood, J. L. (2020). Moving eDNA surveys onto land: Strategies for active eDNA aggregation to detect invasive forest insects. Molecular Ecology Resources, 20(3), 746–755.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology, 25(4), 929–942. https://doi.org/10.1111/mec.13428

van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology, 30(13), 3270–3288.

Veilleux, H. D., Misutka, M. D., & Glover, C. N. (2021). Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. Science of The Total Environment, 782, 146891. https://doi.org/10.1016/j.scitotenv.2021.146891

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of earth's ecosystems. Science, 277(5325), 494–499.

Willerslev, E. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47–51. https://doi.org/10.1038/nature12921

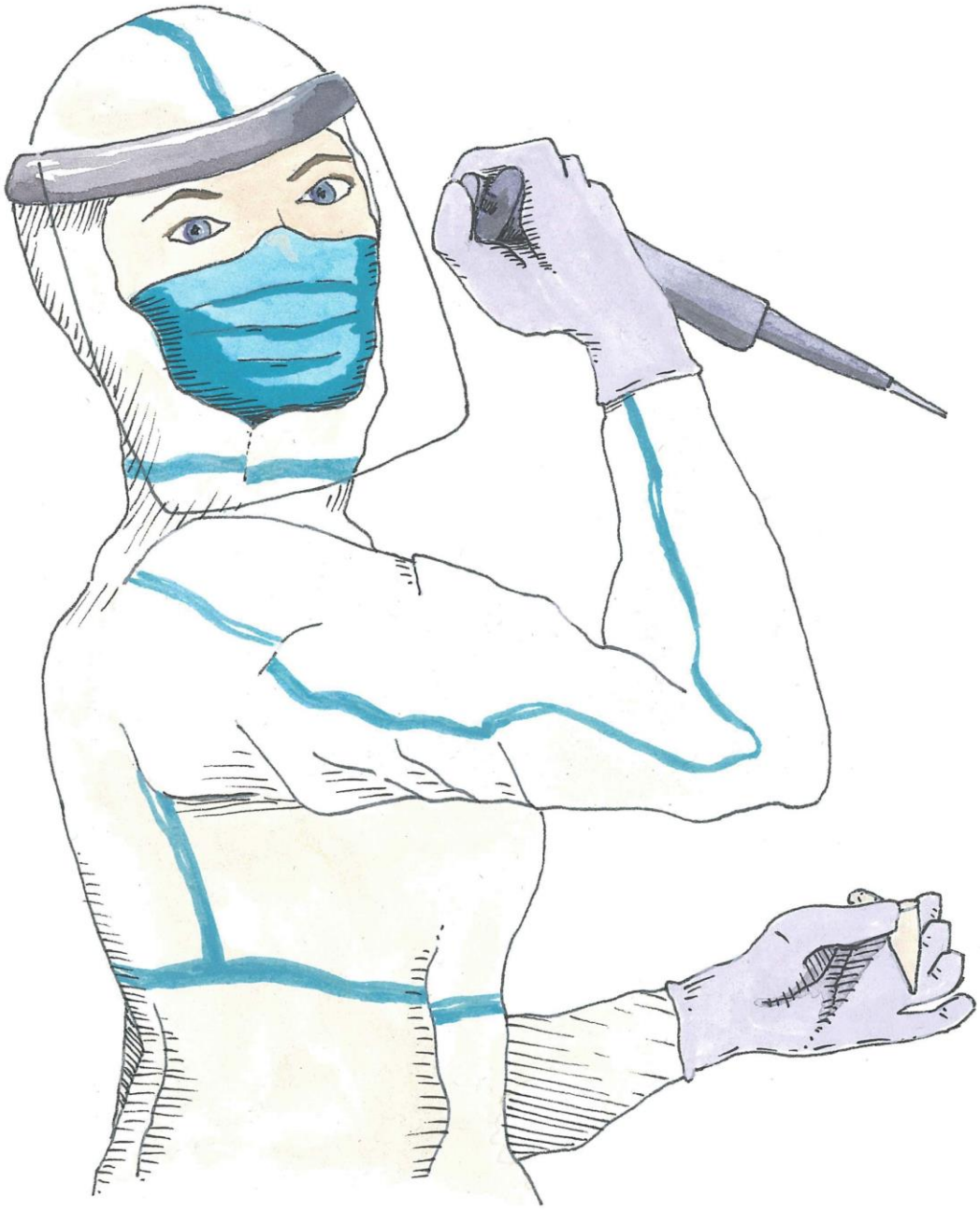
Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Williams, M., Zalasiewicz, J., Haff, P., Schwägerl, C., Barnosky, A. D., & Ellis, E. C. (2015). The Anthropocene biosphere. The Anthropocene Review, 24.

Willis, K. J., & Birks, H. J. B. (2006). What Is Natural? The Need for a Long-Term Perspective in Biodiversity Conservation. Science, 314(5803), 1261–1265. https://doi.org/10.1126/science.1122667

Yoccoz, N. G. (2012). The future of environmental DNA in ecology: NEWS AND VIEWS: OPINION. Molecular Ecology, 21(8), 2031–2038. https://doi.org/10.1111/j.1365-294X.2012.05505.x

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology, 28(8), 1857–1862. https://doi.org/10.1111/mec.15060



Chapter 1

aDNA from sediments

aDNA from sediments

Anneke TM ter Schure[1], Yi Wang[2], Anna LJ Chagas[2], and Laura S Epp[2]

Book Chapter 8 in: Molecular Identification of Plants: From Sequence to Species
Editors: Hugo de Boer[3], Brecht Verstraete[3], Marcella Rydmark[3], and Barbara Gravendeel[4]

[1]Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway
[2]University of Konstanz, Germany
[3]Natural History Museum, University of Oslo, Norway
[4]Naturalis Biodiversity Center & Natural History Museum, Leiden, Netherlands

Sedimentary ancient DNA studies aim to reconstruct the biology and ecology of past environments using the DNA present in the sediment record. Compared to modern soil and sedimentary DNA (see Chapter 4 DNA from soil and sediments), these analyses can be more challenging due to the prolonged exposure of the DNA to degradation processes. This has major implications for the scope of the study and the appropriate study design, which will be discussed in this chapter.

Background

What is sedimentary ancient DNA?

In order to use sedimentary ancient DNA for paleoecological studies (*sed*aDNA; Haile et al., 2009) it is important to understand some aspects of its physical nature and the local environment's role in transforming modern DNA into *sed*aDNA. We will start by breaking down the term into its components.

Ancient DNA is the hereditary genetic content of cells from organisms that died a long time ago. There is no consensus on how old DNA should be in order to be called ancient, as the age is generally less important than the exposure to degradation processes that make it more degraded than modern DNA. *Sed*aDNA degradation processes are primarily related to environmental and sedimentary properties, such as temperature, pH, water content, oxygen levels, and minerals present in the sediment (Giguet-Covex et al., 2019; Torti et al., 2015), whereas time plays a secondary role: providing opportunity for these processes to take place. Permafrost in general provides excellent conditions for preserving DNA, due to its neutral pH, anaerobic conditions, and near-constant subzero temperatures that ensure it remains constantly frozen for 2 years or longer. Optimal conditions in ice cores from Greenland have allowed the detection of plant DNA as old as 450 to 800 thousand years (Willerslev et al., 2007). To date, the oldest amplifiable DNA from sediments is from ca. 400 thousand years old permafrost (Willerslev et al., 2004, 2003).

How does DNA end up in the sediment? Sediment is a result of erosion, weathering and biological processes and consists of organic and inorganic

particles (e.g., sand and silt) that are transported by wind, water, or people (Masselink et al., 2014). These transportation processes also explain the main distinctive quality between sediments and soils: soils develop precisely because of the absence of horizontal transport, allowing biological, physical, and chemical weathering of the local substrate, thereby forming soil horizons rich in organic matter (see Chapter 4 DNA from soil and sediments). Deposition of sediment happens when the sediments stop being transported and stay in place. The incorporation of organismal remains into the sediment are similarly a result of transportation by wind, water, or people, or it can originate from the organisms that are living at that location (Alsos et al., 2018; Parducci et al., 2018). The processes involved in the transfer, deposition, and preservation of organismic remains are called taphonomic processes. Bacterial and fungal DNA make up a very large part of sedimentary DNA, since they are natural inhabitants of sediments and outrate macroorganisms in terms of total biomass. Animal DNA that is found in sediment typically comes from skin flakes, faeces, urine, saliva, hair, feathers, and other animal tissues, while plant DNA typically originates from plant debris, leaves, seeds, fruits, and other plant tissues. Living cells can actively secrete DNA into sediment (e.g., plant root tips; Wen et al., 2017), while dead tissues can degrade, releasing the intracellular DNA (iDNA), along with the rest of the cell contents, when cell lysis occurs. Both active secretion of DNA as well as cell lysis result in iDNA becoming extracellular DNA (exDNA).

Once exposed to the sedimentary environment, exDNA can undergo different post-depositional taphonomic processes that determine the quality of the DNA on longer timescales. ExDNA can be internalized by microbial cells (Overballe-Petersen and Willerslev, 2014), degraded by extracellular microbial nucleases that break it up into smaller fragments, damaged by abiotic processes such as hydrolysis and oxidation, or preserved by adsorption onto particles such as humic acids, sand and clay minerals (Torti et al., 2015; Willerslev and Cooper, 2005). An overview of DNA degradation processes is provided in Figure 8.1. Chemical alkylation can lead to cross-links within (intra) and between (inter) DNA molecules making it impossible to PCR amplify the DNA (Fulton and Shapiro, 2019). Low pH, high temperatures, high oxygen and water content can also lead to strand breaks, deamination of nitrogen bases, and base modifications (Dabney et al., 2013; Willerslev and Cooper, 2005). These processes can result in a decrease in the amount of detectable DNA, shorter DNA fragments, and changes in chemical properties as damage accumulates over time. DNA is better preserved in sediments with a high mineral content and at low temperatures. Minerals can inactivate nucleases as well as bind to and protect DNA, while low temperatures thermally stabilize DNA against chemical degradation (Torti et al., 2015). Desiccated dry and anoxic sediments will putatively also strongly decrease the effects of hydrolysis and oxidation, respectively. The preserved exDNA together with the iDNA preserved in dead cells make up the total DNA that can be recovered using *sed*aDNA methods.

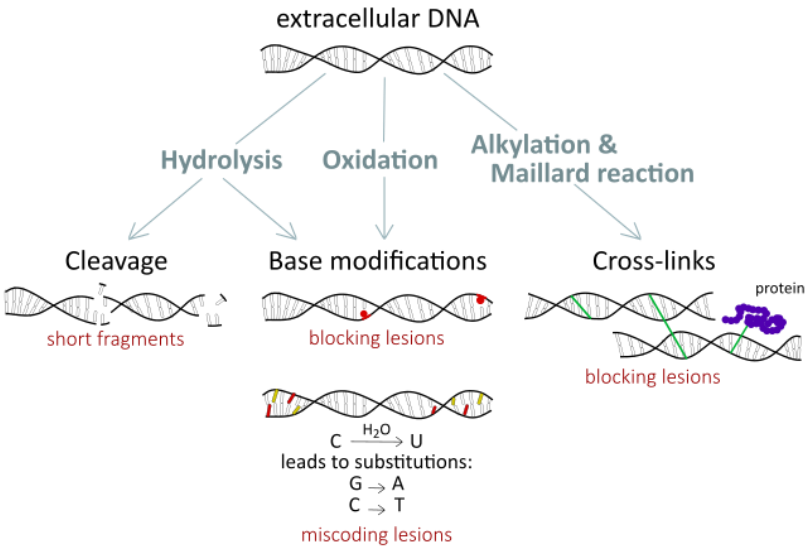


Figure 8.2. Schematic overview of DNA degradation processes (hydrolysis, oxidation, alkylation and Maillard reaction) that can cause DNA damage in the form of cleavage, base modifications or cross-links. The major mechanism leading to miscoding lesions in aDNA is the hydrolysis of cytosine to uracil, which leads to G to A and C to T substitutions by DNA polymerases, whereas blocking lesions can obstruct the movement of DNA polymerases during PCR (Dabney et al., 2013).

Advantages and limitations of *sed*aDNA as palaeoecological proxy

By analysing the ancient DNA present in the sediment (Haile et al., 2009; Slon et al., 2017) it is possible to identify the source species of archaeological artefacts and deposits, and even detect organisms in the absence of any visible remains. For plants, the detection of taxa that do not leave traces in the fossil record (e.g., Alsos et al., 2016; Bremond et al., 2017; Brown et al., 2021; Pedersen et al., 2013) opens up new ways of studying past vegetation complementary to more traditional palaeoecological proxies such as pollen and macrofossils.

Macrofossils and plant *sed*aDNA originate close to the sample location and give a similar local signal (Alsos et al., 2018; Jørgensen et al., 2012; Niemeyer et al., 2017), while the pollen record generally includes taxa that originated from further away from the sample location (Parducci et al., 2018) as pollen, especially of wind-pollinated species, may originate from a wide area as they are distributed regionally through the air (Birks and Bjune, 2010). Pollen does not contribute much to the total pool of *sed*aDNA (Clarke et al., 2020; Sjögren et al., 2017). This can be partially explained by the low DNA content of pollen grains and the robustness of the pollen grain wall, hindering the retrieval of the DNA. At the source, DNA can be considered more consistent than pollen, as all plant tissues contain DNA, but not all plants produce pollen, and insect-pollinated plants produce fewer pollen than wind-pollinated plants.

In general, palaeovegetation data are the result of the attributes of the original vegetation, combined with depositional factors and preservation, as well as the experimental procedures to produce the data. For *sed*aDNA analyses, this includes every step of the data generation itself: sampling, transport, storage, processing of the DNA in the laboratory, and finally, the bioinformatic pipelines used. In terms of the data generation, pollen analyses and macrofossil analyses rely on taxonomic identification by microscopy, which is labour-intensive and requires

a high level of taxonomic knowledge. Although some training is needed to work in an ancient DNA laboratory, in principle, taxonomic identification by DNA can be carried out without prior taxonomic knowledge. However, familiarity with plant taxonomy, phylogenetic placement, and biology of different groups is invaluable in the interpretation of the automated identifications. For example, it is important to check if the automated DNA identifications make sense for the sample location, because contamination, DNA degradation, and the quality of the reference library can cause false DNA identifications (see Chapter 18 Sequence to species for details).

A combination of *sed*aDNA, macrofossils, and pollen proxies gives the most complete overview of plant diversity and community composition through time. The choice for these proxies is dependent on the aims of the study. Table 8.1 summarises the main differences.

Table 8.2. Comparison of pollen, plant macrofossils, and *sed*aDNA as proxies for palaeoecological reconstructions on the levels of: source and sediment, data generation and data interpretation. Sources: Ahmed et al., 2018; Birks and Bjune, 2010; Parducci et al., 2018, 2017.

Category	Pollen	Plant macrofossils	*Sed*aDNA
Source and sediment			
- Scale	Regional	Local	Local
- Taxonomic groups	Pollen-producers	All plants	All organisms
Potential sources of bias	*High pollen-producing plants; vegetation cover close to sampling area; differential preservation*	*Differential preservation of tissue-types and species*	*Differential DNA degradation and decay*
Data generation			
- Labour-intensive	Yes	Yes	No
- Need for taxonomic knowledge	Yes	Yes	No
- Taxonomic resolution	Limited to identifiable pollen types, generally to genus level	Generally to species-level	Depends on the marker, possible to species-level
Potential sources of bias	*Identifiability of the remains*	*Identifiability of the remains; random occurrence*	*DNA contamination; choice of lab techniques; completeness of reference library*
Data interpretation			
- Qualitative	Yes	Yes	Yes
- Quantitative	Partial	Limited	Debated

*Sed*aDNA research applications

The first study using *sed*aDNA of macroorganisms was published in 2003, demonstrating the possibility to detect plant and animal DNA in both permafrost sediments and temperate cave sediments (Willerslev et al., 2003). Since then, the number of *sed*aDNA studies and applications has increased as enhanced understanding of ancient DNA and methodological developments allowed better reconstructions, as also illustrated by a recent comprehensive synthesis of current analytical procedures (Capo et al., 2021). *Sed*aDNA methods are relevant for a range of research fields across biology, conservation, and archaeology and have been applied for roughly two main purposes: understanding natural environmental processes and reconstructing past human-environmental interactions.

Environmental reconstructions can range from polar, to temperate and tropical regions, although they are limited to sampling sites that allow preservation of *sed*aDNA, such as permafrost, lake sediments, and dry cave sediments. Permafrost sediment can be used to assess vegetational development in polar regions under climate change (e.g., Willerslev et al., 2014; Zimmermann et al., 2017). *Sed*aDNA from archaeological sites can reveal human past activities such as plant and animal cultivation, migration and settlement history (e.g., Hebsgaard et al., 2009; Smith et al., 2015), and Neanderthal and Denisovan DNA have been recovered from cave sediments (Slon et al., 2017; Vernot et al., 2021). Lake sediments can be reliable archives of the palaeoenvironment, integrating environmental information across the lake catchment area and displaying a very clear temporal stratification. Many *sed*aDNA studies use lake sediments to focus on past vegetation dynamics, which can be used to establish natural baselines for conservation (e.g., Boessenkool et al., 2014; Wilmshurst et al., 2014), reconstruct the effects of past climate change on the environment (e.g., Alsos et al., 2016, 2020; Clarke et al., 2020; Jørgensen et al., 2012), show long-lasting effects of biological invasions (e.g., Ficetola et al., 2018), or track past human impacts (e.g., Giguet-Covex et al., 2014; Pansu et al., 2015). This list illustrates the wide range of potential applications, but for further discussion, please see Section 3 of this book, especially Chapter 21 Palaeobotany and Chapter 25 Environment and biodiversity assessments can be relevant for *sed*aDNA.

Experimental design

*Sed*aDNA research strategy

Due to its low concentration, retrieving ancient DNA from sediment samples requires strict protocols to avoid contamination by modern DNA or further degradation (Cooper and Poinar, 2000; Capo et al., 2021). However, once these protocols are followed *sed*aDNA can be a powerful tool providing novel insights to palaeoecology reconstructions that are not possible through traditional methods.

The previous section described some *sed*aDNA studies focusing on palaeoecological and archaeological questions. In both cases, choices of location and methods are very much steered by the research focus and what is already known about the area, such as past changes in climate, geology, ecology, or human impacts. Although details in the study design can differ, all *sed*aDNA studies follow the same steps: site selection, collection of samples and metadata, DNA extraction, further processing of the DNA in the lab, sequencing, and finally, bioinformatic sequence quality filtering and data analyses (Figure 8.2).

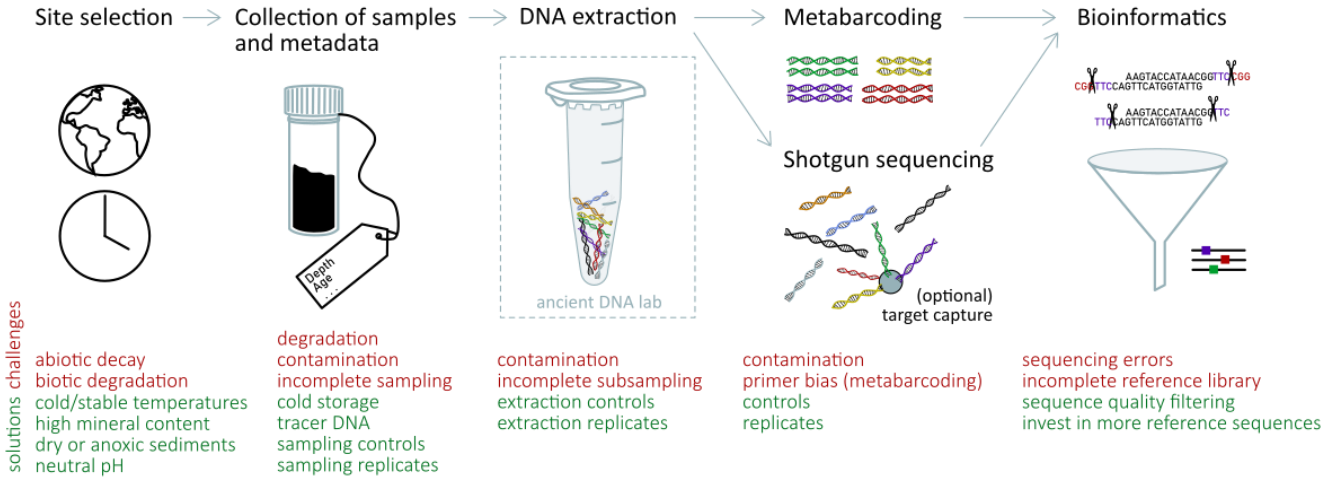


Figure 8.3. Simplified overview of the *sed*aDNA research process, including some of the major challenges and potential solutions indicated at each step.

Choices for the different options at each step depend on the aims of the study. For example, when performing a reconstruction of overall plant community dynamics with universal plant metabarcoding primers, the most common taxa and major trends in community change will be reliably retrieved in the first PCR performed (Alsos et al., 2016), with no specific sampling strategy. However, the detection of rare plant species will require a number of repeats (Alsos et al., 2016), and possibly sampling at several locations (Capo et al., 2021). The following questions can help to develop a *sed*aDNA research strategy and these topics will be discussed throughout this chapter:

1. What is my study aim?
2. What spatial and temporal scale do I need to cover?
3. What contextual information and metadata do I need?
4. What taxa should I target and at what taxonomic resolution?
5. What laboratory and analytical methods should I use?
6. How will I minimize / control for contamination, biases, and false positives?

Site selection

The aims of the study define the temporal and spatial scale needed to achieve them, thereby steering the selection of relevant sampling sites. Lake sediments provide a record of the plants that occurred in the lake catchment, being the area of land from which water and surface runoff drains into the lake (Giguet-Covex et al., 2019). A lake sediment record can only go as far back as the formation of the lake itself. Other terrestrial sediments may primarily contain the DNA that is deposited by plants growing at that particular location, or by humans, animals, or abiotic factors such as wind and water. For example, DNA in cave sediments will come primarily from organisms that have lived or died in the cave, or from remains that are transported into the cave (Hofreiter et al., 2003). The likelihood of finding *sed*aDNA should also be considered. However, more often than not the sampling location is opportunity driven, especially when it comes to archaeological sites, and *sed*aDNA retrieval can prove difficult.

General conditions under which *sed*aDNA preserves well are: cold and stable temperatures, neutral pH, dry or anoxic sediments with a high mineral content. Sediments from rockshelters, dry caves, and lake sediments are generally preferred as they are protected and provide stable conditions: rockshelter and dry cave sediments are sheltered from rain and have stable temperatures and there is some evidence that calcite has a high adsorption capacity for DNA (Capo et al., 2021; Freeman et al., 2020). Lake sediments on the other hand are often anoxic and generally undisturbed, especially when they are below the wave disturbance depth and subsurface slopes are gentle.

Dating of sediments

Dating is important in any study that involves ancient samples. Only with accurate dating can the timing of events be compared and their rates of change estimated. Commonly applied sediment dating methods are radioisotopic dating (in particular 210Pb, 14C, and luminescence dating) and dating based on chemostratigraphy or marker minerals (in particular tephrochronology), and the choice for a method depends on the type and age of the sediments (see Table 8.2 for an overview). Many sources describe these methods in detail (e.g., Bradley, 1999) and we provide a brief introduction here.

Radioisotopic dating is based on the principle of radioactive decay. When a nucleus breaks down, it emits energy and forms a daughter product. The time this takes is expressed as the half-life, i.e., the time that it takes for 50% of a parent element to transmute into the daughter product. The relative quantity of a radioactive parent element in a sample can be used to infer its age. Relatively young aquatic sediments, with ages up to 150 years are commonly dated with 210Pb (half-life: 22.27 years; Barsanti et al., 2020). 210Pb occurs naturally in the atmosphere and settles in sediments through dry fallout or precipitation. The supply of this 210Pb is not constant but the decline of this excess 210Pb along a sediment sequence is a proxy for the sedimentation rate. Additionally, if the age at

a point of the sequence is known, a chronology can be determined. Radiocarbon (14C, half-life: 5730 years) is a radioactive isotope of carbon that naturally occurs in the atmosphere. Plants fix atmospheric carbon during photosynthesis, so the level of 14C in plants and animals upon death approximately equals the level of 14C in the atmosphere at that time. After death, it decreases as 14C decays to 14N at a rate of 50% per 5730 years, allowing the date of death to be estimated. Limited by its half-life, radiocarbon dating is only possible for samples younger than 50,000 years. As the concentration of atmospheric 14C is not constant over time, radiocarbon dates are calibrated against a global calibration curve obtained from tree rings and varved lake sediments (Reimer et al., 2020). This produces calendrical dates, which are expressed as calibrated years before present (cal years BP) with present being 1950 (before large-scale testing of nuclear weapons). The most reliable age-depth models for both marine and lake sediments use accelerator mass-spectrometry (AMS) dating of macroscopic plant or animal fragments (as little as 0.1 mg) as this can avoid the problems of both mixed material and also the so-called hard-water error associated with carbonate waters.

Luminescence dating is based on the phenomenon that mineral crystals absorb electrons from the ionizing radiation of surrounding sediments over time, and when stimulated in a laboratory by heat or light, they release the accumulated radiation as luminescence. The intensity of measured luminescence indicates the length of time between this in-lab stimulation and the last natural event of similar stimulation. Heat stimulated or thermoluminescence (TL) dating is used to date baked pottery from archeological sites or sediments once in contact with molten lava; optically stimulated luminescence (OSL) dating is used to date sediments once exposed to sunlight. The time range for luminescence dating can be from a few decades to over 1 Ma, depending on the ability of a mineral to absorb radiation over time. For studies concerning relatively young samples, OSL dating of quartz grains are generally used, covering from a few decades to ~150 ka.

Tephrochronology uses the chemical signature of tephra (volcanic ash) to pinpoint the age of that specific layer in a sediment sequence by reference to known or unknown dated volcanic eruptions. Terrestrial sediments (Froese et al., 2006), marine deposits (Larsen et al., 2002), and ice cores (Davies et al., 2008) from areas once under the influence of dated volcanic eruption events can be dated with this method. With accurate geochemical fingerprinting, tephrochronology can be used to corroborate or even extend the dating limits of other techniques.

Table 8.3. Summary of sediment dating methods, their applicability and limitations. Sources: Barsanti et al., 2020; Bradley, 1999; Fattahi and Stokes, 2003.

Dating method	Suitable sample types	Age limit	Sources of error and uncertainty
^{210}Pb dating	Materials from aquatic environments such as lacustrine and marine deposits	~100 to 150 years	Complex sedimentation processes that break the dating model assumptions, such as compaction, local mixing, erosion etc.
^{14}C (radiocarbon) dating	Organic remains (charcoal, wood, animal tissue), carbonates (corals, sediments, stalagmites and stalactites), water, air and organic matter from various sediments, soil, paleosol and peat deposits	Up to 50,000 years	Atmospheric ^{14}C content fluctuation due to changes in cosmogenic production rate and exchange between the atmosphere and ocean
Luminescence dating: - Thermoluminescence (TL)	TL: materials containing crystalline minerals, such as sediments, lava, clay, and ceramics	TL: A few years to over 1,000,000 years	Variations in environmental radiation dose; saturation of electron traps in sample minerals
- Optical stimulated luminescence (OSL)	OSL: materials containing quartz or potassium feldspar sand-sized grains, or fine-grained mineral deposits	OSL: A few decades to ~150,000 years for quartz.	
Tephrochronology	Terrestrial and lake sediments, marine deposits and ice cores that contain tephra	Up to 35,000 years, extendable under good conditions	Can only obtain indirect dates within the ^{14}C age range

Prepare to work cleanly

DNA is everywhere - including in the air - and contamination can come from many different sources. When collecting and working with *sed*aDNA samples, it is important to keep in mind that the DNA you are interested in will probably be present in very low concentrations. Contamination with modern DNA can easily overpower the *sed*aDNA signal in which you are interested. Therefore it is important to absolutely minimize the amount of modern DNA coming into your samples and limit further degradation of the *sed*aDNA.

The precautions you can take include: work cleanly, use equipment that is free of DNA and nucleases, and try to keep the samples in a stable and cold environment. In practice this is not so easy, which is why dedicated ancient DNA facilities are set-up to avoid any form of contamination. These facilities should be physically isolated - ideally in a separate building - from any location where PCRs are performed (Fulton and Shapiro, 2019) and strict cleaning regimes and clean lab practices should be upheld. How to set-up and work in an ancient DNA lab is

described in detail by e.g., Cooper and Poinar (2000) and Fulton and Shapiro (2019). Here we summarize general clean lab practices. We note that working cleanly and consistently will require practice and adequate training.

You should assume that everything that you bring into the lab is contaminated with DNA. Therefore, before entering the lab, you should have showered and changed into clean clothes and everything you bring into the lab should be decontaminated. Inside the lab, you should wear a hairnet, face mask, full body suit with hood, shoe covers, and gloves at all times. Wearing two layers of gloves will allow you to change the outer gloves while still covering your hands, and you should change your outer gloves regularly while working. All tools and equipment should be decontaminated before use, and regular cleaning of the aDNA workspace is needed. Decontamination can be achieved by using a DNA decontamination product (e.g., 3-10% bleach or DNA-ExitusPlusTM) for surfaces, ideally supplemented with UV irradiation of the workspace. To prevent cross-contamination, tools should be cleaned between working with each sample or sample-extract. Tools should be left in a DNA decontamination product for at least 10 minutes, rinsed with UV irradiated milliQ water, and ideally also UV irradiated using a UV crosslinker with irradiation at the shortest distance possible to the UV source (Champlot et al., 2010).

Collection, transport, and storage of ancient sediment samples

Choices for sampling and personal protective equipment will depend on the setting, as the sampling of sediments at an archaeological site can be very different from the sub-sampling of a lake sediment core in a lab facility. It is important to try to limit the amount of potential contamination, but practical considerations and the target DNA can also be leading. For example, a study aiming to recover human aDNA will require stricter use of personal protective equipment than a study focussing on plant aDNA. Sampling of sediments can be done directly in the field or by subsampling of sediment cores in a clean, sheltered environment. When collecting sediment cores for *sed*aDNA, closed-chamber piston-type corers are preferred (Parducci et al., 2017) as they enclose the sediment in a plastic tube that can be opened in the laboratory. As frozen sediments should be kept at freezing temperatures, subsampling of these types of cores requires a climate chamber (Epp et al., 2019).

A general *sed*aDNA sampling kit contains personal protective equipment, sampling equipment, and cleaning products, including: full bodysuits, face masks, hairnets, nitrile gloves, sterile scalpels, sample tubes, clean ziplock bags, DNA decontamination products, distilled water, 70% ethanol, trays or beakers for cleaning the tools, paper towels, trash bags and pens for labelling. To limit potential contamination, much of the preparation for the sampling kit takes place in the ancient DNA lab facility: making sure the sampling tools and collection tubes are prepared and DNA-free. Aluminum foil can be helpful for covering your workspace and provides a clean surface for all of the sampling materials at a

sampling site. Sterile syringes with the tip cut off can be useful mini-corers, speeding up the sample-taking (Epp et al., 2019). If you are taking sub-samples in a lab facility, make sure it is isolated from any PCR machine as the high number of DNA copies produced with PCR can become airborne and may enter your samples through the building air supply (Fulton and Shapiro, 2019; Willerslev and Cooper, 2005). Tracing of contamination during sampling can be done by placing several open sample tubes with DNA-free water in your work area (Parducci et al., 2017), or using tracer DNA during coring or on the outside of the sediment core (Epp et al., 2019; Pedersen et al., 2016).

The sampling itself follows aDNA lab procedures where possible, even if it takes place elsewhere: clean the workspace, use personal protective equipment, do not hover over the sediment you are sampling and change outer gloves and tools between each individual sample. In order to avoid contamination, sampling should start at the oldest part of the sediment, working your way up to the youngest parts and subsamples from sediment cores should be taken from inside the undisturbed centre (Parducci et al., 2017). Sampling procedures for both non-frozen and frozen sediment cores are described in detail by Epp et al. (2019). Collected samples should be kept in a stable and low-temperature environment (i.e. freeze at -20 for longer term storage), as degradation slows down with lower temperatures and temperature fluctuations can be additionally damaging to the DNA. An ice-box with ice packs can be used for temporary storage and transport of the taken samples. Further processing of the *sed*aDNA samples should be done in a laboratory dedicated to working with ancient DNA.

Sedimentary ancient DNA extraction

The choice for a specific DNA extraction protocol depends on a range of factors, including the aim of your study, sample characteristics, available laboratory facilities and equipment, and costs of the reagents or extraction kits. The latter can be a consideration of investing either time or finances as it can be cheaper to make the buffers needed for extraction yourself, but this also increases the preparation time and could introduce additional contamination to your samples. There are several protocols that can be used for *sed*aDNA extraction (see Capo et al., 2021; for a detailed review) and general steps are: sample homogenization, lysis, binding, washing, and elution of the DNA. Here we discuss some of the most commonly used extraction protocols and we summarize their main advantages and limitations in Table 8.3.

All extraction protocols include similar steps for the isolation of sedimentary DNA (Figure 8.3), but due to the differences in chemical composition of the buffers, input volume, use of equipment, and targeted DNA (total DNA, iDNA, or exDNA), results of these protocols can vary. You can decide to extract only

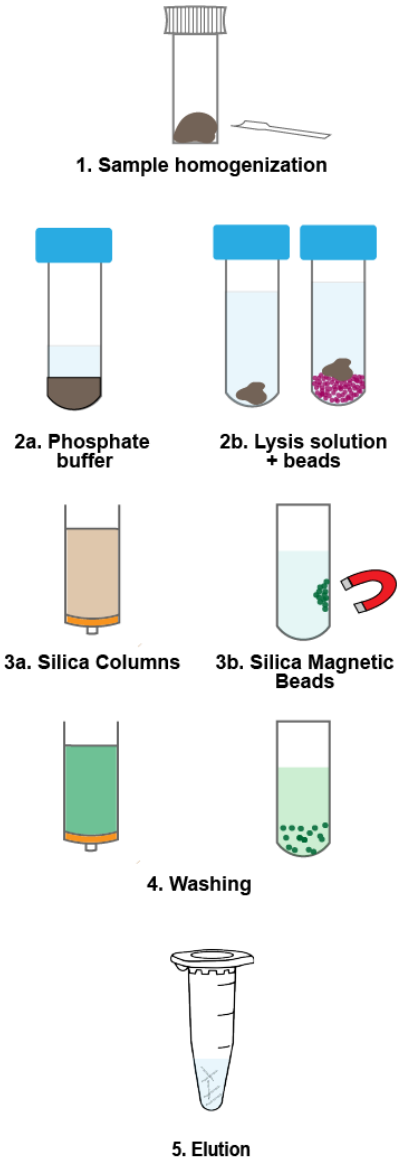


Figure 8.4. Common DNA extraction steps: (1) samples are first homogenized using a sterile scalpel and later on go through a step, in which either (2a) extracellular DNA is washed off the sedimentary matrix (Taberlet et al., 2012) and/or (2b) intracellular DNA is freed through lysis, which can include beating with garnet beads. The free DNA suspended in a high salt buffer can now bind to either (3a) a silica column or (3b) silica magnetic beads, (4) samples are washed with an ethanol based buffer to remove impurities, and finally (5) DNA is eluted in an elution buffer. Figure based on Rohland et al. (2018).

exDNA using the "Taberlet protocol", where samples are first incubated in a saturated phosphate buffer and later on purified with an extraction kit, skipping the lysis step (Taberlet et al., 2012). An advantage is that a large sample volume can be processed, minimizing the possible effects of heterogeneous distribution of DNA in the sediment. However, DNA yield and purity can be lower in comparison to the DNeasy PowerMax Kit (Qiagen), formerly known as the PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc.; Zinger et al., 2016) and probably also to other protocols targeting total DNA (e.g., the Rohland protocol; Rohland et al., 2018).

*Sed*aDNA studies employing protocols developed for the extraction of modern environmental DNA from soils and sediments generally add additional steps to increase the yield of DNA from low concentration ancient sediment samples. A lysis step can be added to extract iDNA from intact cells present in the samples through chemical lysis, and/or mechanical shearing of cell membranes using beads. Adding certain chemicals to the lysis buffer can also increase yield: N-phenacylthiazolium bromide (PTB) breaks down cross-links between DNA and proteins (Vasan et al., 1996; Poinar et al., 1998), and adding proteinase K and dithiothreitol (DTT) during the lysis step of the PowerMax and PowerSoil kits allows better recovery of DNA (Epp et al., 2019). It has also been suggested to concentrate the DNA before further processing (Taberlet et al., 2018), as *sed*aDNA concentrations are likely to be low (Zimmermann et al., 2020). The Rohland protocol is specifically designed to target degraded DNA from ancient samples (Rohland et al., 2018) and should yield a higher concentration of short fragments compared to the other extraction protocols, especially when silica magnetic beads are used for DNA binding.

Be aware that the presence of certain substances may inhibit further amplification or sequencing steps. These can be derived from humic substances (important components of humus), which are commonly present in sediments and might inhibit downstream analysis. Moreover, the amount of humic substances is site-specific, and it might be necessary to repurify the samples or use inhibitor removal columns. During DNA extraction, contamination may be introduced from the laboratory facilities, tools, reagents and other consumables. It is essential to track this contamination by including a negative control. It is suggested to add one such extraction control for each batch of 11 samples, and include it in all subsequent steps (e.g., metabarcoding, library preparation, sequencing; Rohland et al., 2018). It is common for the extraction of modern DNA to add a positive control with a known DNA content, but due to the contamination risk this is not recommended for *sed*aDNA (Willerslev and Cooper, 2005).

Table 8.4. Overview of the advantages and limitations of several commonly used extraction protocols and some example publications using these protocols.

Extraction protocol	Sample size	Advantages	Limitations	Used by
DNeasy PowerMax kit (Qiagen)	≤ 10 g	- Large initial sample volume - Few inhibitors in the resulting extract	- Expensive - DNA can be lost with inhibitor removal solution	Epp et al., 2018; Zimmermann et al., 2017
DNeasy PowerSoil kit (Qiagen)	≤ 250 mg	- Few amplification and sequencing inhibitors in the resulting extract - Easy processing of large sets of samples	- DNA can be lost with inhibitor removal solution - Smaller initial sample volume compared to the PowerMax kit	Lejzerowicz et al., 2013; Monchamp et al., 2016; Dommain et al., 2020;
Rohland protocol (Rohland et al., 2018)	≤ 50 mg	- Developed to recover small DNA fragments - Easy processing of large sets of samples	- Small starting amount of sediment - Potential coextraction of inhibitors - Homemade buffers can increase contamination risk	Zavala et al., 2021; Vernot et al., 2021;
Phosphate buffer + NucleoSpin® Soil kit (Taberlet et al., 2012)	≤ 15 g	- Large initial sample volume	- Extracts only extracellular DNA - Processes a 2 mL subsample of the phosphate buffer and sample mixture	Giguet-Covex et al., 2014; Pansu et al., 2015;
Murchie protocol (Murchie et al., 2020)	≤ 250 mg	- High DNA yields - Uses a high volume binding buffer to improve the recovery of small DNA fragments	- Optimized for permafrost samples and may not perform as well in lake sediment	Murchie et al., 2020;

Molecular methods for *sed*aDNA

After extracting the DNA, the *sed*aDNA needs to be further processed before sequencing and several approaches are continuously being improved and new ones developed.

Most *sed*aDNA studies apply a DNA metabarcoding approach, using PCR amplification primers to target short DNA sequences (< 300 bp, preferentially around or below 100 bp) from taxonomic marker genes to identify specific taxonomic groups (see Chapter 11 Amplicon metabarcoding). It is relatively low cost and some of the metabarcoding primers give high taxonomic resolution. However, this method can introduce amplification bias (Bellemain et al., 2010) and is susceptible to errors introduced in the PCR. More recently, shotgun sequencing became another option for these types of samples (Pedersen et al., 2016). This approach converts the DNA extracts directly to a library for sequencing, allowing the analyses of the entire diversity of taxonomic groups in the samples including microorganisms (Ahmed et al., 2018), plants (Parducci et al., 2019; Pedersen et al., 2016), animals (Graham et al., 2016; Pedersen et al., 2016), and humans (Slon et al., 2017; Vernot et al., 2021). Shotgun sequencing requires a high sequencing depth and can be costly as most sequences will be from non-target organisms. Target capture has recently been applied to *sed*aDNA samples to enrich the concentration of taxa of interest in a shotgun approach by using DNA (Schulte et al., 2020) or RNA (Murchie et al., 2020; Seeber et al., 2019) baits. These methods are described in detail in Chapter 11 Amplicon metabarcoding, Chapter 12 Metagenomics, and Chapter 14 Target capture, and are followed by library preparation and sequencing (see Chapter 9 Sequencing platforms and data types).

Sequencing data can be processed using bioinformatic tools, where strict quality filtering of the sequence data is followed by taxonomic assignment. Further filtering allows removal of sequences with low identity scores, contaminants (i.e., sequences present in the controls), and false-positives (see Chapter 18 Sequence to species for details). False identifications can be caused by the quality of the reference library, but also by technical errors, contamination, or errors in the DNA sequences, especially as *sed*aDNA is generally highly degraded and of low concentration. It is therefore important to check if the identifications make sense for the sampling location and age before further analyses of the *sed*aDNA data.

Glossary

Alkylation - Addition or substitution of an alkyl group (C_nH_{2n+1}) to an organic molecule.

Accelerator Mass-Spectrometry (AMS) dating - A dating method that determines the age of an organic material (i.e., macroscopic remains of plants or animals) by measuring their radiocarbon concentration.

Cell lysis - The process whereby the membrane(s) of a cell breaks down, thereby releasing the cell contents.

exDNA - Extracellular DNA; all DNA located outside cell membranes.

Geochemical fingerprinting - A method using chemical signals to infer the origin, the formation and/or the environment of a geological sample.

Half-life - The time necessary for half of a radioactive atom's nucleus to decay by emission of matter and energy to form a new daughter product. The half-life is specific to a radioactive element, and can be used for dating purposes.

iDNA - Intracellular DNA; all DNA present within cell membranes.

Lake catchment - Area of land from which water and surface runoff drains into a lake.

Luminescence dating - A group of methods to determine how long ago mineral grains were last exposed to sunlight or sufficient heating by measuring the luminescence emitted by the mineral grain upon stimulation.

Metabarcoding - Method for the simultaneous identification of many taxa within the same complex DNA extract. This is achieved by high throughput sequencing (HTS) of amplicons from taxonomic marker genes (barcodes).

Next Generation Sequencing (NGS) - Massively parallel sequencing technology allowing high throughput of DNA.

Nucleases - Diverse group of enzymes able to hydrolyze the phosphodiester bonds of DNA and RNA thereby cleaving them into smaller fragments.

Thermoluminescence (TL) dating - Dating method that determines the age of a sample by measuring the luminescence it emits in response to heat.

Optically stimulated luminescence (OSL) dating - Dating method that determines the age of a sample by measuring the luminescence it emits in response to visible or infrared light.

Taphonomic processes - The processes involved in the transfer, deposition and preservation or organismal remains, including DNA.

Tephrochronology - A geochronological technique that uses layers of tephra (volcanic ash from a single volcanic eruption) to create a chronological framework for the sedimentary record.

Target capture - A technique that allows the capture of the DNA of interest by hybridization to target-specific probes (baits).

Palaeoecology - The study of the relationship between past organisms and their ancient environments.

Permafrost - Soil, sediment, or rock that is continuously exposed to temperatures of < 0°C for at least two consecutive years.

Radioactive isotope - An atom with excess nuclear energy and prone to undergo radioactive decay.

Reference library - A database of known DNA sequences with their taxonomic identifications, used in bioinformatics as a reference to identify the DNA sequences obtained in a *sed*aDNA study.

***Sed*aDNA** - Sedimentary ancient DNA; this is the aged and degraded DNA from dead organisms now incorporated in the sediment record, either as iDNA in dead tissues, or as exDNA free in the sediment matrix or adsorbed to sediment particles.

Shotgun sequencing - A method for the random sequencing of all of the DNA within a DNA extract.

Total DNA - The intracellular and extracellular DNA combined.

Tree-ring dating - Also called dendrochronology; a method of dating tree rings to the exact year they were formed.

Questions

1. Name and explain two main advantages of using *sed*aDNA as a proxy for past plant presence compared to pollen. Motivate your answer.
2. Imagine you have a long lake sediment core that is thought to be between 50 000 and 10 000 years old. What dating methods could be used to date this core and why?
3. What are the main sources of bias when working with *sed*aDNA (name at least 3) and how can you limit the resulting false positives?

Answers

1. Possible advantages of *sed*aDNA compared to pollen as a proxy for past plant presence are: the possibility of detecting past plant presence even in the absence of visible remains; less labour-intensive as taxonomic identification is automated; in principle, no prior taxonomic knowledge is needed for the data generation with *sed*aDNA (although it is highly called for in the interpretation of the data); and it is possible to obtain a higher taxonomic resolution depending on the choice of marker.
2. For mineral-rich sediments, luminescence dating can be used as this method can be applied to sediments from a few decades old to over a million years old, and is based on the phenomenon that mineral crystals absorb electrons from ionizing radiation of surrounding sediments over time. For sediment rich in organic materials, AMS radiocarbon dating of identified macroscopic remains (with calibration) is a good option. Radiocarbon

dating is based on the concentration of C^{14} in organismic remains. The half-life of C^{14} (5730 years) makes it an appropriate method for samples under 50,000 years old. To increase confidence in the dating results, multiple dating techniques could be used for creating an age model for the core.

3. Biases when working with *sed*aDNA can come from: taphonomic processes including differential DNA degradation and preservation, choice of metabarcoding primers, completeness of reference library, and contamination during sampling, DNA extraction and other lab processes. False positives can be limited by inclusion of multiple replicates and controls and prevention of contamination at every step of the experimental design, preparation of an appropriate reference database, and checking if the identifications fit with what is known for the age and location of the sample by a taxonomic expert.

References

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J.E., Han, L., Muschitiello, F., Pedersen, M.W., Smittenberg, R.H., Yamoah, K.A., Slotte, T., Wohlfarth, B., 2018. Archaeal community changes in Lateglacial lake sediments: Evidence from ancient DNA. Quat. Sci. Rev. 181, 19–29. doi:10.1016/j.quascirev.2017.11.037

Alsos, I.G., Lammers, Y., Yoccoz, N.G., Jørgensen, T., Sjögren, P., Gielly, L., Edwards, M.E., 2018. Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PLoS ONE 13, e0195403. doi:10.1371/journal.pone.0195403

Alsos, I.G., Sjögren, P., Brown, A.G., Gielly, L., Merkel, M.K.F., Paus, A., Lammers, Y., Edwards, M.E., Alm, T., Leng, M., Goslar, T., Langdon, C.T., Bakke, J., van der Bilt, W.G.M., 2020. Last Glacial Maximum environmental conditions at Andøya, northern Norway; evidence for a northern ice-edge ecological "hotspot." Quat. Sci. Rev. 239, 106364. doi:10.1016/j.quascirev.2020.106364

Alsos, I.G., Sjögren, P., Edwards, M.E., Landvik, J.Y., Gielly, L., Forwick, M., Coissac, E., Brown, A.G., Jakobsen, L.V., Foreid, M.K., Pedersen, M.W., 2016. Sedimentary ancient DNA from Lake Skartjorna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene 26, 627–642. doi:10.1177/0959683615612563

Barsanti, M., Garcia-Tenorio, R., Schirone, A., Rozmaric, M., Ruiz-Fernández, A.C., Sanchez-Cabeza, J.A., Delbono, I., Conte, F., De Oliveira Godoy, J.M., Heijnis, H., Eriksson, M., Hatje, V., Laissaoui, A., Nguyen, H.Q., Okuku, E., Al-Rousan, S.A., Uddin, S., Yii, M.W., Osvath, I., 2020. Challenges and limitations of the 210Pb sediment dating method: Results from an IAEA modelling interlaboratory comparison exercise. Quat. Geochronol. 59, 101093. doi:10.1016/j.quageo.2020.101093

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., Kauserud, H., 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 10, 189. doi:10.1186/1471-2180-10-189

Birks, H.H., Bjune, A.E., 2010. Can we detect a west Norwegian tree line from modern samples of plant remains and pollen? Results from the DOORMAT project. Veg. Hist. Archaeobot. 19, 325–340. doi:10.1007/s00334-010-0256-0

Boessenkool, S., McGlynn, G., Epp, L.S., Taylor, D., Pimentel, M., Gizaw, A., Nemomissa, S., Brochmann, C., Popp, M., 2014. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. Conserv. Biol. 28, 446–455. doi:10.1111/cobi.12195

Bradley, R.S., 1999. Paleoclimatology: reconstructing climates of the Quaternary. Elsevier.

Bremond, L., Favier, C., Ficetola, G.F., Tossou, M.G., Akouégninou, A., Gielly, L., Giguet-Covex, C., Oslisly, R., Salzmann, U., 2017. Five thousand

years of tropical lake sediment DNA records from Benin. Quat. Sci. Rev. 170, 203–211. doi:10.1016/j.quascirev.2017.06.025

Brown, A.G., Van Hardenbroek, M., Fonville, T., Davies, K., Mackay, H., Murray, E., Head, K., Barratt, P., McCormick, F., Ficetola, G.F., Gielly, L., Henderson, A.C.G., Crone, A., Cavers, G., Langdon, P.G., Whitehouse, N.J., Pirrie, D., Alsos, I.G., 2021. Ancient DNA, lipid biomarkers and palaeoecological evidence reveals construction and life on early medieval lake settlements. Sci. Rep. 11, 11807. doi:10.1038/s41598-021-91057-x

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P.D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A.G., Clarke, C.L., Crump, S.E., Debroas, D., Englund, G., Ficetola, G.F., Garner, R.E., Gauthier, J., Parducci, L., 2021. Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: overview and recommendations. Quaternary 4, 6. doi:10.3390/quat4010006

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E.A., Grange, T., Geigl, E.-M., 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS ONE 5. doi:10.1371/journal.pone.0013042

Clarke, C.L., Alsos, I.G., Edwards, M.E., Paus, A., Gielly, L., Haflidason, H., Mangerud, J., Regnéll, C., Hughes, P.D.M., Svendsen, J.I., Bjune, A.E., 2020. A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. Quat. Sci. Rev. 247, 106564. doi:10.1016/j.quascirev.2020.106564

Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. Science 289, 1139. doi:10.1126/science.289.5482.1139b

Dabney, J., Meyer, M., Pääbo, S., 2013. Ancient DNA damage. Cold Spring Harb. Perspect. Biol. 5. doi:10.1101/cshperspect.a012567

Davies, S.M., Wastegård, S., Rasmussen, T.L., Svensson, A., Johnsen, S.J., Steffensen, J.P., Andersen, K.K., 2008. Identification of the Fugloyarbanki tephra in the NGRIP ice core: a key tie-point for marine and ice-core sequences during the last glacial period. J. Quaternary Sci. 23, 409–414. doi:10.1002/jqs.1182

Dommain, R., Andama, M., McDonough, M.M., Prado, N.A., Goldhammer, T., Potts, R., Maldonado, J.E., Nkurunungi, J.B., Campana, M.G., 2020. The Challenges of Reconstructing Tropical Biodiversity With Sedimentary Ancient DNA: A 2200-Year-Long Metagenomic Record From Bwindi Impenetrable Forest, Uganda. Front. Ecol. Evol. 8. doi:10.3389/fevo.2020.00218

Epp, L.S., Kruse, S., Kath, N.J., Stoof-Leichsenring, K.R., Tiedemann, R., Pestryakova, L.A., Herzschuh, U., 2018. Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. Sci. Rep. 8, 17436. doi:10.1038/s41598-018-35550-w

Epp, L.S., Zimmermann, H.H., Stoof-Leichsenring, K.R., 2019. Sampling and Extraction of Ancient DNA from Sediments. Methods Mol. Biol. 1963, 31–44. doi:10.1007/978-1-4939-9176-1_5

Fattahi, M., Stokes, S., 2003. Dating volcanic and related sediments by luminescence methods: a review. Earth-Science Reviews 62, 229–264. doi:10.1016/S0012-8252(02)00159-9

Ficetola, G.F., Poulenard, J., Sabatier, P., Messager, E., Gielly, L., Leloup, A., Etienne, D., Bakke, J., Malet, E., Fanget, B., Støren, E., Reyss, J.-L., Taberlet, P., Arnaud, F., 2018. DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Sci. Adv. 4, eaar4292. doi:10.1126/sciadv.aar4292

Freeman, Dieudonné, Collins, Sand, 2020. Survival of environmental DNA in natural environments: Surface charge and topography of minerals as driver for DNA storage. BioRxiv. doi:10.1101/2020.01.28.922997

Froese, D.G., Zazula, G.D., Reyes, A.V., 2006. Seasonality of the late Pleistocene Dawson tephra and exceptional preservation of a buried riparian surface in central Yukon Territory, Canada. Quat. Sci. Rev. 25, 1542–1551. doi:10.1016/j.quascirev.2006.01.028

Fulton, T.L., Shapiro, B., 2019. Setting up an ancient DNA laboratory. Methods Mol. Biol. 1963, 1–13. doi:10.1007/978-1-4939-9176-1_1

Giguet-Covex, C., Ficetola, G.F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A.L., David, F., Taberlet, P., Brisset, E., Guiter, F., Sinet, R., Arnaud, F., 2019. New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. Sci. Rep. 9, 14676. doi:10.1038/s41598-019-50339-1

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J., Taberlet, P.,

2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat. Commun. 5, 3211. doi:10.1038/ncomms4211

Graham, R.W., Belmecheri, S., Choy, K., Culleton, B.J., Davies, L.J., Froese, D., Heintzman, P.D., Hritz, C., Kapp, J.D., Newsom, L.A., Rawcliffe, R., Saulnier-Talbot, É., Shapiro, B., Wang, Y., Williams, J.W., Wooller, M.J., 2016. Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. Proc Natl Acad Sci USA 113, 9310–9314. doi:10.1073/pnas.1604903113

Haile, J., Froese, D.G., Macphee, R.D.E., Roberts, R.G., Arnold, L.J., Reyes, A.V., Rasmussen, M., Nielsen, R., Brook, B.W., Robinson, S., Demuro, M., Gilbert, M.T.P., Munch, K., Austin, J.J., Cooper, A., Barnes, I., Möller, P., Willerslev, E., 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. Proc Natl Acad Sci USA 106, 22352–22357. doi:10.1073/pnas.0912510106

Hebsgaard, M.B., Gilbert, M.T.P., Arneborg, J., Heyn, P., Allentoft, M.E., Bunce, M., Munch, K., Schweger, C., Willerslev, E., 2009. 'The Farm Beneath the Sand' – an archaeological case study on ancient 'dirt' DNA. Antiquity 83, 430–444. doi:10.1017/S0003598X00098537

Hofreiter, M., Mead, J.I., Martin, P., Poinar, H.N., 2003. Molecular caving. Curr. Biol. 13, R693-5.

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N.H., Andersen, K., Orlando, L., Gilbert, M.T.P., Willerslev, E., 2012. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Mol. Ecol. 21, 1989–2003. doi:10.1111/j.1365-294x.2011.05287.x

Kromer, B., 2009. Radiocarbon and dendrochronology. Dendrochronologia 27, 15–19. doi:10.1016/j.dendro.2009.03.001

Larsen, G., Eiríksson, J., Knudsen, K.L., Heinemeier, J., 2002. Correlation of late Holocene terrestrial and marine tephra markers, north Iceland: implications for reservoir age changes. Polar Res. 21, 283–290. doi:10.1111/j.1751-8369.2002.tb00082.x

Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., Arbizu, P.M., Pawlowski, J., 2013. Ancient DNA complements microfossil record in deep-sea subsurface sediments. Biol. Lett. 9, 20130283. doi:10.1098/rsbl.2013.0283

Masselink, G., Hughes, M., Knight, J., 2014. Introduction to Coastal Processes and Geomorphology.

Monchamp, M.-E., Walser, J.-C., Pomati, F., Spaak, P., 2016. Sedimentary DNA Reveals Cyanobacterial Community Diversity over 200 Years in Two Perialpine Lakes. Appl. Environ. Microbiol. 82, 6472–6482. doi:10.1128/AEM.02174-16

Murchie, T.J., Kuch, M., Duggan, A.T., Ledger, M.L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., MacPhee, R., Froese, D., Poinar, H., 2020. Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quaternary Research 1–24. doi:10.1017/qua.2020.59

Niemeyer, B., Epp, L.S., Stoof-Leichsenring, K.R., Pestryakova, L.A., Herzschuh, U., 2017. A comparison of sedimentary DNA and pollen from lake sediments in recording vegetation composition at the Siberian treeline. Mol. Ecol. Resour. 17, e46–e62. doi:10.1111/1755-0998.12689

Overballe-Petersen, S., Willerslev, E., 2014. Horizontal transfer of short and degraded DNA has evolutionary implications for microbes and eukaryotic sexual reproduction. Bioessays 36, 1005–1010. doi:10.1002/bies.201400035

Pansu, J., Giguet-Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., Choler, P., 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Mol. Ecol. 24, 1485–1498. doi:10.1111/mec.13136

Parducci, L., Alsos, I.G., Unneberg, P., Pedersen, M.W., Han, L., Lammers, Y., Salonen, J.S., Väliranta, M.M., Slotte, T., Wohlfarth, B., 2019. Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from southern sweden. Front. Ecol. Evol. 7. doi:10.3389/fevo.2019.00189

Parducci, L., Bennett, K.D., Ficetola, G.F., Alsos, I.G., Suyama, Y., Wood, J.R., Pedersen, M.W., 2017. Ancient plant DNA in lake sediments. New Phytol. 214, 924–942. doi:10.1111/nph.14470

Parducci, L., Nota, K., Wood, J., 2018. Reconstructing Past Vegetation Communities Using Ancient DNA from Lake Sediments, in: Lindqvist, C., Rajora, O.P. (Eds.), Paleogenomics: Genome-

Scale Analysis of Ancient DNA, Population Genomics. Springer International Publishing, Cham, pp. 163–187. doi:10.1007/13836_2018_38

Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., Willerslev, E., Kjær, K.H., 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quat. Sci. Rev. 75, 161–168. doi:10.1016/j.quascirev.2013.06.006

Pedersen, M.W., Ruter, A., Schweger, C., Friebe, H., Staff, R.A., Kjeldsen, K.K., Mendoza, M.L.Z., Beaudoin, A.B., Zutter, C., Larsen, N.K., Potter, B.A., Nielsen, R., Rainville, R.A., Orlando, L., Meltzer, D.J., Kjær, K.H., Willerslev, E., 2016. Postglacial viability and colonization in North America's ice-free corridor. Nature 537, 45–49. doi:10.1038/nature19085

Poinar, H.N., Hofreiter, M., Spaulding, W.G., Martin, P.S., Stankiewicz, B.A., Bland, H., Evershed, R.P., Possnert, G., Pääbo, S., 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. Science 281, 402–406. doi:10.1126/science.281.5375.402

Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kromer, B., Manning, S.W., Muscheler, R., Palmer, J.G., Talamo, S., 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon 1–33. doi:10.1017/RDC.2020.41

Rohland, N., Glocke, I., Aximu-Petri, A., Meyer, M., 2018. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. 13, 2447–2461. doi:10.1038/s41596-018-0050-5

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H.H., Pestryakova, L.A., Epp, L.S., Herzschuh, U., 2020. Hybridization capture of larch (Larix Mill) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. Mol. Ecol. Resour. doi:10.1111/1755-0998.13311

Seeber, P.A., McEwen, G.K., Löber, U., Förster, D.W., East, M.L., Melzheimer, J., Greenwood, A.D., 2019. Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. Mol. Ecol. Resour. 19, 1486–1496. doi:10.1111/1755-0998.13069

Sjögren, P., Edwards, M.E., Gielly, L., Langdon, C.T., Croudace, I.W., Merkel, M.K.F., Fonville, T., Alsos, I.G., 2017. Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytol. 213, 929–941. doi:10.1111/nph.14199

Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., Stewart, J.R., Derevianko, A.P., Jacobs, Z., Li, B., Roberts, R.G., Shunkov, M.V., de Lumley, H., Perrenoud, C., Gušić, I., Kućan, Ž., Meyer, M., 2017. Neandertal and Denisovan DNA from Pleistocene sediments. Science 356, 605–608. doi:10.1126/science.aam9695

Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V., Allaby, R.G., 2015. Archaeology. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. Science 347, 998–1001. doi:10.1126/science.1261278

Taberlet, P., Bonin, A., Zinger, L., Coissac, E. (Eds.), 2018. Environmental DNA: For Biodiversity Research and Monitoring.

Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., Coissac, E., 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Mol. Ecol. 21, 1816–1820. doi:10.1111/j.1365-294X.2011.05317.x

Torti, A., Lever, M.A., Jørgensen, B.B., 2015. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Mar. Genomics 24 Pt 3, 185–196. doi:10.1016/j.margen.2015.08.007

Vasan, S., Zhang, X., Zhang, X., Kapurniotu, A., Bernhagen, J., Teichberg, S., Basgen, J., Wagle, D., Shih, D., Terlecky, I., Bucala, R., Cerami, A., Egan, J., Ulrich, P., 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. Nature 382, 275–278. doi:10.1038/382275a0

Vernot, B., Zavala, E.I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., Pablos, A., Aranburu, A., de Castro, J.M.B., Carbonell, E., Li, B., Krajcarz, M.T., Krivoshapkin, A.I., Kolobova, K.A., Kozlikin, M.B., Shunkov, M.V., Meyer, M., 2021. Unearthing Neanderthal population

history using nuclear and mitochondrial DNA from cave sediments. Science 372. doi:10.1126/science.abf1667

Wen, F., Curlango-Rivera, G., Huskey, D.A., Xiong, Z., Hawes, M.C., 2017. Visualization of extracellular DNA released during border cell separation from the root cap. Am. J. Bot. 104, 970–978. doi:10.3732/ajb.1700142

Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M.B., Brand, T.B., Hofreiter, M., Bunce, M., Poinar, H.N., Dahl-Jensen, D., Johnsen, S., Steffensen, J.P., Bennike, O., Schwenninger, J.-L., Nathan, R., Armitage, S., de Hoog, C.-J., Alfimov, V., Christl, M., Beer, J., Collins, M.J., 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science 317, 111–114. doi:10.1126/science.1141758

Willerslev, E., Cooper, A., 2005. Ancient DNA. Proc. Biol. Sci. 272, 3–16. doi:10.1098/rspb.2004.2813

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Bråthen, K.A., Yoccoz, N., Binney, H., Taberlet, P., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature 506, 47–51. doi:10.1038/nature12921

Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A., 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300, 791–795. doi:10.1126/science.1084114

Willerslev, E., Hansen, A.J., Rønn, R., Brand, T.B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D., Cooper, A., 2004. Long-term persistence of bacterial DNA. Curr. Biol. 14, R9–R10. doi:10.1016/j.cub.2003.12.012

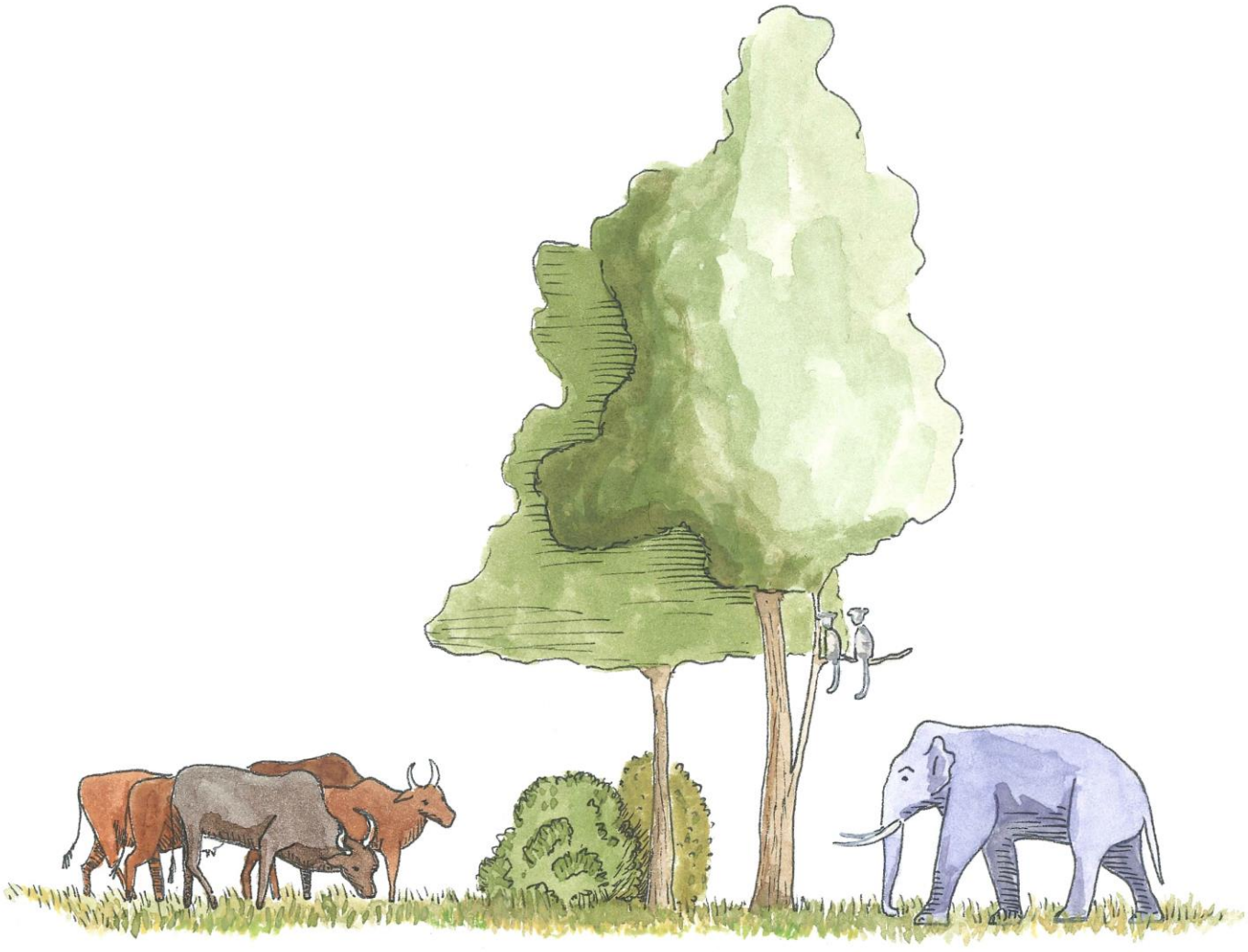
Wilmshurst, J.M., Moar, N.T., Wood, J.R., Bellingham, P.J., Findlater, A.M., Robinson, J.J., Stone, C., 2014. Use of pollen and ancient DNA as conservation baselines for offshore islands in New Zealand. Conserv. Biol. 28, 202–212. doi:10.1111/cobi.12150

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. doi:10.1038/s41586-021-03675-0

Zimmermann, H.H., Raschke, E., Epp, L.S., Stoof-Leichsenring, K.R., Schirrmeister, L., Schwamborn, G., Herzschuh, U., 2017. The History of Tree and Shrub Taxa on Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial Uncovered by Sedimentary Ancient DNA and Pollen Data. Genes (Basel) 8. doi:10.3390/genes8100273

Zimmermann, H.H., Stoof-Leichsenring, K.R., Kruse, S., Müller, J., Stein, R., Tiedemann, R., Herzschuh, U., 2020. Changes in the composition of marine and sea-ice diatoms derived from sedimentary ancient DNA of the eastern Fram Strait over the past 30 000 years. Ocean Sci. 16, 1017–1032. doi:10.5194/os-16-1017-2020

Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., Schilling, V., Schimann, H., Sommeria-Klein, G., Taberlet, P., 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biology and Biochemistry 96, 16–19. doi:10.1016/j.soilbio.2016.01.008



Chapter 2

eDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary

eDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary

Anneke T.M. ter Schure, Anusree A.S. Pillai, Lisbeth Thorbek, Maradani Bhavani Shankar, Rajindra Puri, Gudasalamani Ravikanth, Hugo J. de Boer, Sanne Boessenkool

Published in Environmental DNA 3, 681-696, 2021

Abstract

As many ecosystems are under increasing pressure from invasive species, habitat degradation, overgrazing and overharvesting, pollution, and climate change, dietary niche monitoring is gaining importance. The Malai Mahadeshwara Wildlife Sanctuary (MMH) in southern India is home to several long-standing ethnic and tribal groups and supports a considerable number of domestic herbivores (cattle, goats and water buffalo) as well as a range of wildlife (including several species of deer, bonnet macaque, and Asian elephant). We reconstructed dietary niche partitioning of the herbivores occurring in MMH using eDNA metabarcoding to quantify diet richness, composition, and overlap. In total, we distinguish 134 diet items (molecular operational taxonomic units), covering 31 plant families. Overall, our results indicate 35% overlap in domestic and wild herbivore diet items. The greatest overlap is found for the dietary niches of cattle and sambar deer (Pianka's niche overlap index: 0.68), and the dietary niche of cattle also overlaps considerably with those of Indian hare (0.65) and Asian elephant (0.46). This suggests that these herbivores may compete for these food plants in the case of limited availability, which could lead to exclusion of some herbivore species. Particular concern should go to bonnet macaque and Asian elephant as their below average dietary richness could make them vulnerable to changes in their environment. With increasing pressures on local wildlife from a range of different factors, DNA metabarcoding of fecal samples is a non-invasive method for monitoring changes in animal diets, providing valuable information for the management of biodiversity in mosaic natural and anthropogenic landscapes.

1. Introduction

In areas with species of similar ecology, the partitioning of ecological niches can reduce competition for resources, thus aiding species coexistence and biodiversity (Hutchinson, 1959; MacArthur & Levins, 1967; Pianka, 2011a). Given that diet represents a fundamental aspect of a species' niche (Simberloff & Dayan, 1991), it is unsurprising that dietary niche analysis has been recognized as important for understanding the mechanistic processes behind community ecology (Pompanon et al., 2012) and diversification (Cantalapiedra et al., 2014). More specifically, the dietary niche width of a species can provide information on the extent of dietary specialization (e.g. Sato et al. 2018), on the potential for competition between coexisting species (e.g. Lopes et al. 2015) as well as adaptive responses to environmental changes (Devictor et al., 2010; Pianka, 2011a). Species with narrow niches are deemed more vulnerable (Carscadden et al., 2020; Clavel et al., 2011; Devictor et al., 2010) and thus should be monitored closely in light of climate change, invasive species and other anthropogenic pressures.

Many natural ecosystems are currently under pressure from invasive species, hunting, habitat degradation and destruction, threatening 26% of all mammal species (IUCN, 2020) and more than half of all large wild herbivore species with extinction (Ripple et al., 2015). Moreover, wild herbivores can be threatened by growing livestock populations (Food and Agriculture Organization of the United Nations, 2017). This is worrying as domesticated herbivores generally have competitive advantage over local wild herbivores owing to support they get from humans, for example through supplemental feeding during periods of scarcity. This encroachment by domestic herbivores can potentially lead to competitive exclusion of wild herbivores if they occupy similar niches (Hardin, 1960; Pianka, 2011b). Similarly, establishing nature reserves and wildlife sanctuaries in traditionally managed lands can disempower local communities as wild herbivores encroach on cattle grazing lands and raid croplands (Lamarque et al., 2009; Anand and Radhakrishna 2017).

Traditional methods for studying dietary niche partitioning have provided insight into herbivore niche overlap, but can be time-consuming and dependent on the presence of undigested and identifiable plant remains, as well as direct observation of foraging behaviour. More recently, advances in eDNA metabarcoding have enabled broad application of this method in biodiversity monitoring (for reviews see e.g. Bohmann et al., 2014; Cristescu & Hebert, 2018), and the application to faecal samples provides a valuable alternative approach for dietary reconstruction. This value is evidenced both by direct comparison of different approaches (e.g. Newmaster et al., 2013) and by the rapidly increasing number of studies using this method (for recent reviews see e.g. Ando et al., 2020; de Sousa, Silva & Xavier et al., 2019). So far, faecal DNA metabarcoding has been successfully applied to reconstruct the diets of a range of different herbivores, including birds, insects, molluscs (e.g. Valentini et al., 2009) and a wide range of

mammalian herbivores such as small rodents (e.g. Lopes et al., 2015; Sato et al., 2018; Soininen et al., 2014, 2015), a number of deer species (e.g Bison et al., 2015; Czernik et al., 2013; Fløjgaard, De Barba, Taberlet, & Ejrnæs, 2017; Rayé et al., 2011), tapirs (e.g. Hibert et al., 2013), several primate species (e.g Bradley et al., 2007), the European bison (e.g. Kowalczyk et al., 2011, 2019), and large herbivore assemblages in Kenya (Kartzinel et al., 2015; Kartzinel & Pringle, 2020) and Mozambique (Pansu et al., 2019). From a different perspective, dietary niche analysis should also be able to triangulate traditional ecological knowledge on grazing of domestic and wild herbivores. All of the above studies analysing herbivore diet used the P6 loop of the chloroplast trnL (UAA) intron (Taberlet et al., 2007), a universal plant marker specifically suited to environmental samples with degraded DNA. Most studies applying the trnL approach have focussed on the dietary reconstruction of a few species, providing valuable insight into the trophic ecology of these particular species. However, the approach has also been applied to analyse dietary niche partitioning of more complete herbivore assemblages (e.g. in African large herbivores; Kartzinel et al., 2015; Pansu et al., 2019).

Much of what is currently known about large mammalian herbivore diet comes from research in North America, Europe and Africa, while data from Asia is particularly scarce (Schieltz & Rubenstein, 2016; Öllerer et al., 2019). Furthermore, despite globally growing livestock numbers, there are relatively few studies specifically investigating impacts of domesticated herbivores on wild herbivores (see Schieltz & Rubenstein, 2016 for a review). Extrapolation from these other regions to Asia is typically not straightforward as effects of livestock on wildlife are highly context dependent and species assemblages and biogeography differ greatly (Ahrestani & Sankaran, 2016). The potential competition from livestock is, however, of particular concern in Asia, and specifically in India. Here, the world's second largest livestock population is found (Food and Agriculture Organization of the United Nations, 2017) and many wildlife reserves are being grazed by livestock. Previous studies from India have shown that excessive livestock grazing seriously threatens elephant habitat contiguity (Silori & Mishra, 2001), and suggest livestock-mediated resource limitation as declining livestock numbers resulted in recovery of wild large herbivore densities (Madhusan, 2004). Further insight into seasonal variation in diet and niche-overlap among some of the most common large mammalian herbivores in India comes primarily from microhistological analyses (e.g. Ahrestani, Heitkönig, & Prins, 2012), and deer in particular are suggested to be impacted by livestock grazing (Bagchi, Goyal, & Sankar, 2003). Overall, the limited current knowledge from this area is based on traditional methods that can be time-consuming and dependent on the presence of undigested and identifiable plant remains, as well as direct observation of foraging behaviour that is extra challenging as most of the species-rich large herbivore assemblages are found in densely forested areas (Ahrestani et al., 2012; Ahrestani & Sankaran 2016).

In the present study, we use eDNA metabarcoding of faecal samples to test for dietary niche partitioning by livestock and wild mammalian herbivores in the Malai Mahadeshwara Hills Wildlife Sanctuary (MMH) in southern India (Figure 1). The MMH is home to people from long-standing ethnic and tribal groups with their domestic animals (Harisha & Padmavathy, 2013; Kent & Dorward, 2015) as well as a wide range of wildlife. Although livestock rearing (forest grazing) has traditionally been part of the livelihoods of the local communities (Kent & Dorward, 2015), there is currently an effort to regulate forest access and livestock grazing in MMH (Thornton, Puri, Bhagwat & Howard, 2019). At the same time, resource impacts from tens of thousands of pilgrims annually (Soumya et al., 2019a), invasive plant species and modernization, including developmental activities and tourism, are reported to reduce biodiversity in the area. This is evident from, among others, interviews with local communities (Harisha et al., 2015) and research on the impacts of the invasive plant *Lantana camara* L. on vegetation (Varghese et al., 2015; Soumya et al., 2019b), bird assemblages (Aravind et al., 2010) and human adaptive responses (Kent & Dorward, 2015; Thornton et al., 2019). Our results provide a starting point for tracking the effects of the environmental changes in the area, and urge the need to understand and monitor dietary niches of both local wildlife and livestock, especially in regions where their potential overlap is high and/or under-examined. The application of eDNA technology for such monitoring proves to be an efficient tool to address this need, allowing non-invasive analyses of faecal samples that provide valuable data for improving biodiversity management.

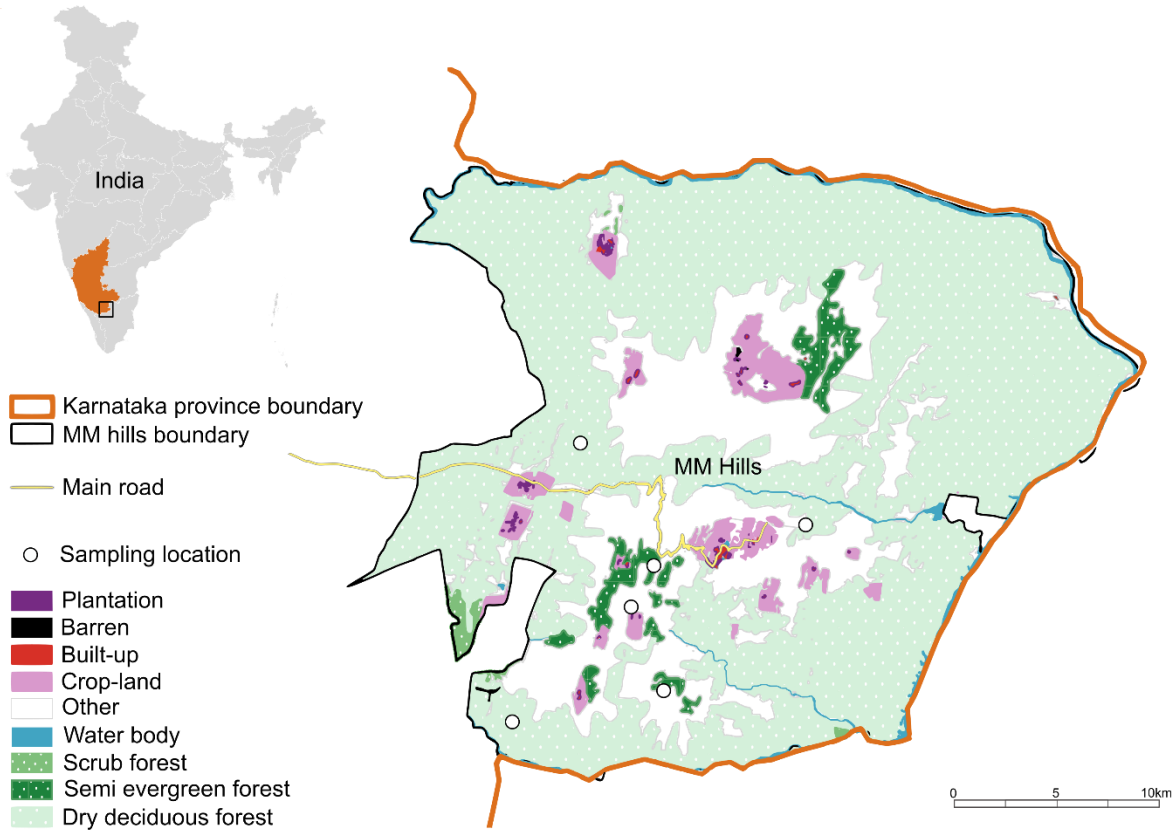


Figure 1. Map of Malai Mahadeshwara Hills Wildlife Sanctuary land use and sampling locations

2. Materials and Methods

2.1. Study site and sampling

Malai Mahadeshwara Hills Wildlife Sanctuary (MMH; 12.1.60N, 77.35.21E, 906 km2) is a protected area of the Kollegala Range Forest in the state of Karnataka, southern India. The area is part of tiger habitat and acts as an important elephant corridor between two adjacent wildlife sanctuaries (i.e. Cauvery and Biligiriranganatha Swamy Temple Wildlife Sanctuary (BRT); Bawa, Joseph, & Setty, 2007; Gubbi et al., 2017). Most of the area is dry deciduous forest (64.3%) with scrub woodland (20.5%) and patches of moist deciduous and riparian forest (2.5%, Harisha & Padmavathy, 2013). MMH is home to approximately 12,000 people from several ethnic and tribal groups, but throughout the year the population is heavily elevated due to tens of thousands of religious pilgrims who visit the main temple and other shrines (Harisha & Padmavathy, 2013; Kent & Dorward, 2015; Soumya et al., 2019a). Despite its history of human interactions and anthropogenic character, the MMH forests host a wide range of wildlife.

We collected 116 faecal samples from 16 different mammal species in the winter and summer pre-monsoon seasons in 2015-2016 (December to April), 2017 (March and April) and 2018 (April). Seventy-seven samples were identified

as from herbivorous animals, and 62 provided usable results after quality control filtering of the herbivore and plant DNA sequence data. These samples represent 10 herbivore species that can be subdivided into domestic herbivores, goat (*Capra hircus*), cattle (*Bos taurus*), water buffalo (*Bubalis bubalus*), and wild herbivores, sambar (*Rusa unicolor*) and barking deer (*Muntiacus muntjak*), Indian hare (*Lepus nigricollis*), Indian crested porcupine (*Hystrix indica*), bonnet macaque (*Macaca radiata*), wild boar (*Sus scrofa*), and Asian elephant (*Elephas maximus*). Samples were dried or placed in ethanol and stored at –20 °C.

2.2. DNA analyses

We followed standard procedures for working with low copy DNA, such as the regular cleaning of work surfaces with bleach and changing gloves between the handling of each sample. Subsamples of faecal material were obtained by spreading the faecal sample on a Petri dish and randomly collecting 200 mg from the dish. Excess ethanol from storing the faecal samples was evaporated by briefly heating the sample to 50 °C. We extracted the DNA using the PSP Spin Stool DNA Kit (Stractec Biomedical, Berlin, Germany) following manufacturer's instructions, using 100 µL elution buffer supplied with the kit and omitting the heating step (10 min at 95 °C) to prevent further DNA degradation. Extraction blanks (6) were included in each extraction round, and these were pooled per two during the PCR step resulting in sequences from 3 sets of extraction blanks in the final dataset. DNA extraction, PCR preparation and post-PCR work took place in separate dedicated rooms.

2.3. Herbivore DNA amplification and sequencing

The identity of the herbivore faecal samples was confirmed using specifically designed primers for each target species (Table S1). Herbivore DNA amplifications were carried out in a final volume of 12.5 µL using 1 µL of DNA extract and 0.24 µM of each primer. The amplification mixture contained 0.5 U of AmpliTaq Gold DNA Polymerase with buffer II (Applied Biosystems, Foster City, CA), 1x Buffer II, 2.5 mM MgCl2, 0.48 mM of each dNTP, and 0.048 µg/µL of bovine serum albumin (BSA, Roche Diagnostic, Basel, Switzerland). The mixture was denatured at 95 °C for 10 min, followed by 35 cycles of 30 s at 95 °C, 30 s at 48 – 55 °C depending on the primer pair used (Table S1), 45 s at 72 °C and a 3 min final elongation at 72 °C.

The PCR products were visualised with agarose gel electrophoresis and cleaned for sequencing by adding 2 µL 1:10 dilution of Illustra ExoProStar (GE Healthcare, USA) to the PCR products and incubating them at 37 °C for 45 min before enzyme inactivation at 80 °C for 15 min. The cleaned PCR products were bidirectionally sequenced on an ABI 3730xl at Macrogen Europe BV (Amsterdam, the Netherlands).

2.4. Plant DNA amplification and sequencing

Plant DNA metabarcoding was done using the *trn*L *g* and *h* primers (Taberlet et al., 2007). Both primers were tagged with a unique 8 or 9 bp barcode at the 5' end to allow for multiplexing as described by Voldstad et al. (2020). We conducted three PCR replicates per sample and both the extraction negative controls and PCR negative controls were included in the PCRs.

Plant DNA amplifications were carried out in a final volume of 25 µL containing 2 µL of DNA extract and 0.24 µM of each primer. The amplification mixture further contained 1 U of AmpliTaq Gold DNA Polymerase with Buffer II (Applied Biosystems), 1 x Buffer II, 2.5 mM MgCl2, 0.48 mM of each dNTP, and 0.048 µg/µL of bovine serum albumin (BSA, Roche Diagnostic). The mixture was denatured at 95 °C for 10 min, followed by 35 cycles of 30 s at 95 °C, and 30 s at 55 °C, 45 s at 72 °C and a 2 min final elongation at 72 °C.

Amplicons were quantified using a Bio-Rad Gel doc XR+ and the Image Lab v.6.0.0 software (Bio-Rad Laboratory, Inc.), and subsequently cleaned as described above. A Biomek 4000 liquid handling robot (Beckman Coulter) was used to pool amplicons equimolarly into three pools whereby each pool contained one of the three replicate PCRs. For amplicons with concentrations lower than 1 ng/µL the maximum amount of 15 µL was added to the pool. The resulting three pools were cleaned two times with Ampure XP (Beckman Coulter) according to the manufacturer's protocol using first a 1.4:1 and then a 2:1 ratio between Ampure XP beads and pool. Concentrations of the pools were measured on a Qubit 2.0 with the Qubit dsDNA HS kit (ThermoFisher) and pools were visualised on a Fragment Analyzer using the DNF-488 kit (Advanced Analytical Technologies Inc.). Libraries were built from the pools with plant PCRs using the KAPA HyperPrep DNA kit (Roche) and pools were sequenced on the Illumina HiSeq 4000 at the Norwegian Sequencing Centre.

2.5. Data processing and analyses

2.5.1. Herbivore DNA identification

Sequence reads from the herbivore DNA were aligned and trimmed manually using Geneious Prime 2019.1.3 (https://www.geneious.com). The resulting consensus sequences were then checked against the NCBI nucleotide collection using megaBLAST (Morgulis et al., 2008). Sequences resulting in percentage ID < 95%, or poor quality reads (HQ% < 35) were excluded from further analyses.

2.5.2. Plant DNA sequence analyses and filtering

Initial analyses and filtering of the plant DNA sequence data were performed using the OBITools package (http://metabarcoding.org/obitools/doc/index.html; Boyer et al., 2016). Assembling of the forward and corresponding reverse reads was done using *illuminapairedend*, followed by sample assignment with *ngsfilter*. We removed reads with a quality score <40, <100% tag match, >3 mismatches

with the primers, shorter lengths then expected (<8 bp), singletons and those containing ambiguous nucleotides. Amplification and sequencing errors were identified using *obiclean*, with a threshold ratio of 5% for reclassification of sequences identified as 'internal' to their corresponding 'head' sequence. Finally, sequences were compared to two taxonomic reference libraries using *ecotag*. These two reference libraries were prepared by performing an in-silico PCR with the *ecoPCR* software (Ficetola et al., 2010) and the NCBI Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy). The first, local reference library contained 555 sequences of 134 plant taxa known to occur in MMH and the surrounding area from monitoring data (information provided by the Ashoka Trust for Research in Ecology and the Environment, ATREE) and published species lists (Appendix 1 of Harisha, Padmavathy & Nagaraja, 2015). The majority of the sequences for this library were obtained from the EMBL database (release 137). An additional 35 species of locally occurring Poaceae were sequenced for the database (at ATREE, Bengaluru, see Appendix S2). As the plant taxa in the local reference library occur in MMH and neighbouring areas, we prioritized matches against this library. To mitigate erroneous or missing taxonomic assignment due to lacking references in this library, we used a second reference library based on the global EMBL database (release 137), containing 111,146 sequences of 18,101 plant taxa.

In order to minimise any misidentifications, we filtered the identified sequences in R (version 3.5.2) to remove: (1) sequences that were identified only as 'internal' in the *obiclean* step, (2) sequences with higher occurrence (i.e. more reads) in negative controls than in samples, (3) sequences with a percentage identity <95%, (4) 0.001% of each sequence read count per sample to correct for potential leakage, (5) unreliable PCR replicates, and (6) sequences that make up <1% of the sample (as advised during a workshop; see Appendix S3 for details). We identified unreliable PCR replicates by estimating Euclidian distances between all PCR replicates and their centroid based on square rooted rarefied read counts (similar to Kowalczyk et al., 2019). We estimated kernel densities for non-replicates and replicates and compared them to identify the distance where the kernel density was higher for non-replicates compared to replicates. Replicates were discarded when distances among replicates were the same or larger than this threshold distance. Remaining replicates were merged while averaging the read counts per molecular operational taxonomic unit (MOTU). In order to check and where possible narrow down some of the taxonomic identifications, the identified plant taxa were checked by a taxonomist with extensive knowledge of the locally occurring plants. Remaining unique plant sequences were designated as MOTUs. An overview of these steps and the remaining reads can be found in Appendix S3, and the resulting processed data in tables S4 – S7.

2.6. Dietary niche analysis

All further data processing and statistical analyses were done in R (version 3.5.2). In order to quantify the diet composition, the obtained plant MOTU-by-samples matrix of the read counts was transformed using two distinct approaches: 1) the presence/absence of each plant MOTU in each faecal sample, and 2) the relative read abundance (RRA), i.e., the proportional representation of each plant MOTU in each faecal sample. All further analyses were performed on both of the resulting transformed datasets. RRA data has been used in numerous other dietary metabarcoding studies (e.g. Kartzinel et al., 2015; Kartizinel & Pringle 2020; Mychek-Londer, Chaganti, & Heath, 2020; Pansu et al., 2019), as results based on RRA have been shown to be less sensitive to rare MOTUs (i.e. low-abundant reads that may for example result from PCR or sequencing errors or contamination; Deagle et al., 2019). We only present figures and analyses performed on the RRA data in the main text. Analyses on presence/absence data can be found in the Appendix (S8 – S12).

We assessed dietary niche width by calculating average MOTU richness and the Shannon diversity index for each sample using the *spaa* package (Zhang, 2016). We subsequently computed and visualized intersections of herbivore species' diets and calculated intersection sizes in number of shared MOTUs using *UpSetR* (Conway, Lex, & Gehlenborg, 2017). Similar to Pansu et al. (2019), we used two complementary metrics to describe the (dis-)similarities between the different dietary niches: Bray-Curtis dissimilarity index and Pianka's niche overlap index (Pianka, 1974). Bray-Curtis dissimilarity was calculated between each pair of faecal samples in order to quantify dietary dissimilarity. We subsequently ordinated these values in two dimensions using non-metric multidimensional scaling (NMDS) in the *vegan* package (Oksanen et al., 2019) to allow the visualization of the patterns of dietary dissimilarity among samples, and species (groups of samples). A stress level for the NMDS of < 0.2 is considered acceptable (Clarke, 1993). We tested for dietary difference among species by doing a permutational multivariate analysis of variance (perMANOVA) with 999 permutations, using the *adonis*-function in *vegan*. We tested the perMANOVA assumption of homogeneity of multivariate dispersions using the *permutest*-function in *vegan* with 999 permutations. Pairwise calculations of Pianka's niche overlap index (Pianka, 1974) were performed with the *spaa* package (Zhang, 2016). We evaluated these results with reference to 1000 permutations of a null model that retains the dietary niche width of each species while randomizing the values for the diet items using *EcoSimR* (Gotelli, Hart, & Ellison, 2015). See Appendix S12 for further details.

3. Results

3.1. Description of the raw dataset

The 77 analysed samples yielded 62 retained samples. Five samples were discarded due to poor quality sequences from the herbivore PCR, six samples were discarded due to percentage herbivore identity of <95 and four samples were discarded due to poor yield of plant DNA. After filtering and merging the replicates, the dataset contains a total of 12.5 M reads distributed over 134 plant MOTUs from 62 faecal samples, representing the diets of 10 herbivore species (Table S6). Read depth per sample ranged from 16 929 to 820 667 (average: 201 718 ± 16 615). 31% of the MOTUs are annotated with a plant species name, 13% with a genus name, and 40% with a plant family name. The remaining 16% are annotated to higher taxonomic ranks (Table S5). A total of 31 different plant families are distinguished. The most abundant plant families in the dataset are Fabaceae and Poaceae, both in number of MOTUs (24 and 22) and in percentage read counts (30% and 35% respectively; Table S4). These are followed by Malvaceae in terms of MOTUs (7), and by Anacardiaceae in terms of percentage of read counts (6%).

3.2. Dietary niche width and composition

Average dietary niche width over all studied individuals is 8.23 ± 0.55 in MOTU richness and 1.24 ± 0.09 in Shannon diversity. MOTU richness is greatest for domestic goat, water buffalo and sambar deer (Table 1), with averages of 11.75 ± 1.53, 13.33 ± 2.73 and 12.40 ± 0.93, respectively. The Shannon diversity index also indicates the greatest dietary richness for these herbivore species, with index values: 1.78 ± 0.18, 1.88 ± 0.30 and 2.00 ± 0.18, respectively. The narrowest dietary niches are found for bonnet macaque and cattle samples (Table 1). Bonnet macaques have a dietary niche width of 4.83 ± 0.91 in MOTU richness and 0.78 ± 0.19 in Shannon diversity. The bulk of the bonnet macaque diet consists of Fabaceae (63%) of which 85% represents the Senegalia genus, followed by Malvaceae (13%) and Rhamnaceae (10%, Figure 2 and Table S7). Contrarily, the diet of cattle consists primarily of grass, as indicated by the RRA for Poaceae of 84%. Although the remaining 16% of the cattle diet is composed of 15 other plant families and at least 24 genera, the MOTU richness and Shannon diversity are still below average, with 6.22 ± 0.69 and 0.87 ± 0.13, respectively.

Table 1. Overview of the sample size, a priori feeding guild assignment and niche width described by average MOTU count and Shannon diversity index per herbivore species with standard errors (SE) for both measures.

	Herbivore species	Scientific name	N	Feeding guild	MOTUs	(± SE)	Shannon	(± SE)
domestic	Cattle	*Bos taurus indicus*	23	folivore (grazer)	6.22	(± 0.69)	0.87	(± 0.13)
	Domestic goat	*Capra hircus*	8	folivore (mixed feeder)	11.75	(± 1.53)	1.78	(± 0.18)
	Water buffalo	*Bubalus bubalis*	3	folivore (mixed feeder)	13.33	(± 2.73)	1.88	(± 0.30)
wild	Asian elephant	*Elephas maximus*	5	folivore (mixed feeder)	8.40	(± 2.84)	1.09	(± 0.46)
	Barking deer	*Muntiacus muntjak*	2	folivore (browser)	7.50	(± 2.50)	1.22	(± 0.59)
	Bonnet macaque	*Macaca radiata*	6	frugivore	4.83	(± 0.91)	0.78	(± 0.19)
	Indian hare	*Lepus nigricollis*	3	folivore (grazer)	7.67	(± 1.76)	1.40	(± 0.19)
	Indian porcupine	*Hystrix indica*	4	frugivore	8.75	(± 1.97)	1.43	(± 0.43)
	Sambar deer	*Rusa unicolor*	5	folivore (mixed feeder)	12.40	(± 0.93)	2.00	(± 0.12)
	Wild boar	*Sus scrofa*	3	omnivore (browser)	9.33	(± 1.20)	1.51	(± 0.18)

Note: Feeding guild assignments are based on Nowak and Walker (1999), Ahrestani et al. (2012), Ahrestani and Sankaran (2016), and IUCN (2020).

Comparing the dietary composition for the 10 studied herbivores, five species primarily consume Poaceae, whereas four species primarily consume Fabaceae. Members of the Poaceae family make up more than 50% of the diet of cattle, Asian elephant, wild boar, Indian hare, and more than 30% of the sambar deer diet. By contrast, the diets of domestic goat, bonnet macaque, Indian porcupine and barking deer consist primarily of Fabaceae (43%–63%). The diet of water buffalo forms the exception with 31% of reads from the Anacardiaceae family (which consists for 84% of Mangifera), followed by 22% from the Moraceae family (which consists for 94% of Ficus) and 20% Poaceae reads (Figure 2 and Table S7).

3.3. Dietary niche overlap

From the 134 plant MOTUs in the total dataset, 48 MOTUs are exclusive for domestic herbivore samples and 39 MOTUs are only found in wild herbivore samples (Figure 3 and Table S7). These sets make up 17% and 7% of the RRA dataset respectively. The remaining 47 MOTUs (35% of all MOTUs) are shared between domestic and wild herbivores and represent the remaining 76% of RRAs in the total dataset. These 47 MOTUs are spread over 38 intersections each representing the number of plant MOTUs shared by a particular combination of herbivore species (Figure 3). The total number of shared plant MOTUs for a specific combination of herbivore species can be found by taking the sum of

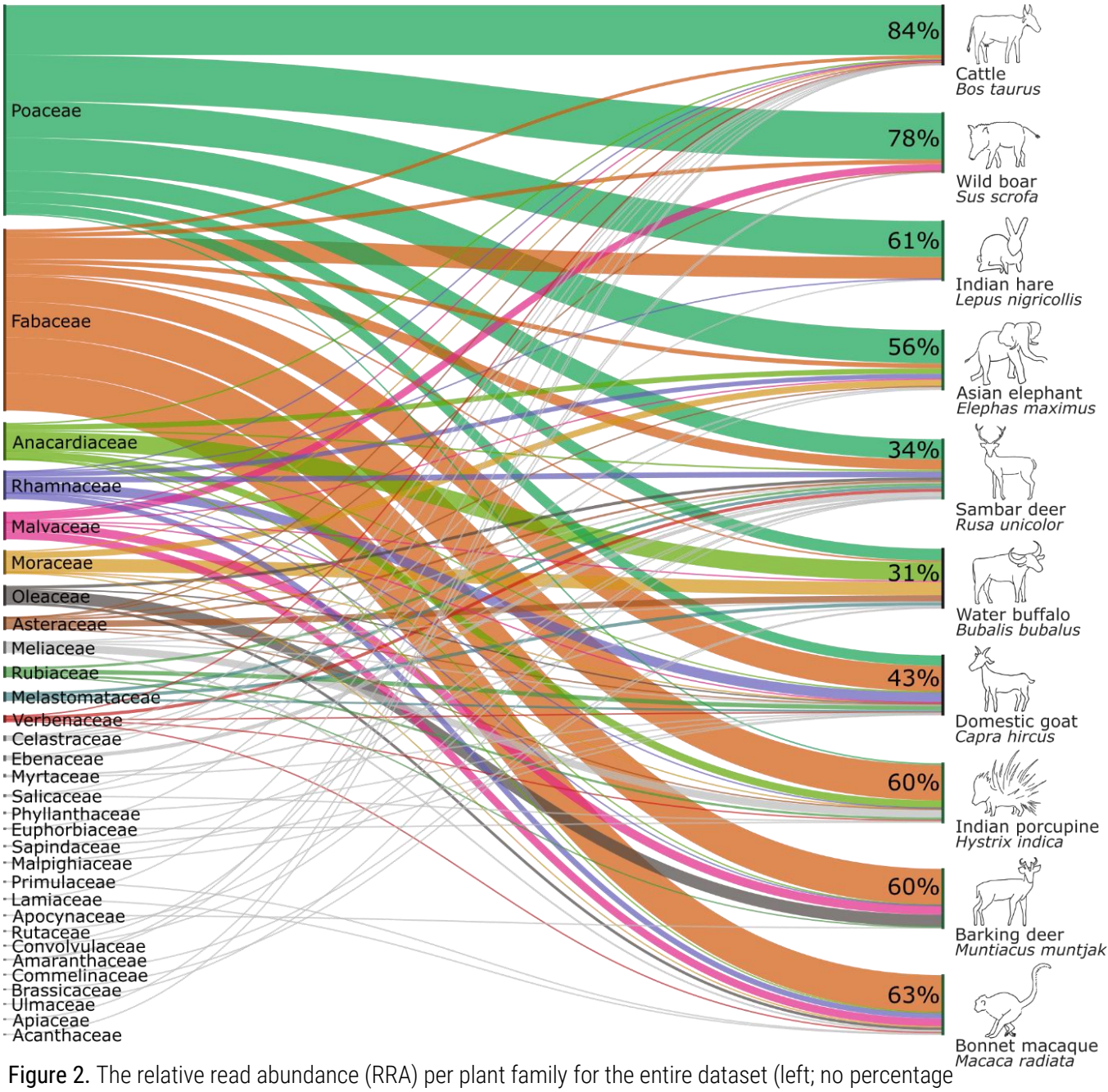


Figure 2. The relative read abundance (RRA) per plant family for the entire dataset (left; no percentage indication means RRA < 1%) and for each herbivore species (right), where percentages indicate the RRA of the most abundant plant family in the diet.

intersection sizes for all combinations that include the pair or group of herbivore species of interest. The largest number of shared MOTUs between two species is 17 and is found for two pairs of species: cattle and domestic goat, and domestic goat and sambar deer. This is followed by 14 shared MOTUs for cattle and water buffalo, and 13 shared MOTUs for domestic goat and Asian elephant.

In order to quantify the degree of overlap in dietary niches between the different herbivores, we calculated Bray-Curtis dissimilarity (BC; 0: similar; 1: dissimilar) and complementary Pianka niche overlap indices (O; 0: no overlap, 1: full overlap) based on MOTU relative read abundances. Resulting Bray-Curtis dissimilarity index values are in the range of 0.61 to 1.00 with an average of 0.88

± 0.02, whereas Pianka index values range from 0.00 to 0.68 with an average of 0.15 ± 0.04 (Table 2). 35 MOTUs, representing 79% of total RRAs are shared between wild herbivore species. Seven of these 35 MOTUs (13% of total RRAs) do not occur in the diets of the domestic herbivore species. The highest degree of dietary niche overlap for wild herbivores was observed between barking deer and porcupine (BC: 0.72, O: 0.63; Table 2) and between sambar deer and Indian hare (BC: 0.67, O: 0.60). Within the group of domestic herbivores, 30 MOTUs representing 53% of total RRAs, are shared among species. Of these 30 MOTUs, nine (5% of total RRAs) do not occur in the diets of the wild herbivore species. The dietary niche of cattle overlaps with those of the other domestic herbivores (goat, BC: 0.75, O: 0.28; water buffalo, BC: 0.82, O: 0.23; Table 2), but comparison between goat and water buffalo reveals little overlap of their dietary niches (BC: 0.92, O: 0.05).

A perMANOVA on Bray-Curtis dissimilarities derived from the RRA data indicates significant dietary differences among species ($F_{9,52} = 3.79$, $r^2 = 0.40$, $p \leq 0.001$). The assumption of homogeneity of dispersion among species was supported by a non-significant permutest result ($p = 0.789$). In order to uncover which particular herbivore species drive these results, we performed a post-hoc pairwise perMANOVA (using a Bonferroni correction of the p-values). Four of the 45 comparisons were statistically significant ($p < 0.05$): cattle vs domestic goat, cattle vs bonnet macaque, domestic goat vs bonnet macaque, and cattle vs wild boar (see Table S11). A perMANOVA on the presence/absence data also identifies these pairs of herbivores to differ significantly in their dietary niches, but further indicates different niches for cattle compared to Asian elephant, Indian porcupine, and sambar deer (see Table S11).

Herbivore pairs with significantly different dietary niches according to the pairwise perMANOVA on RRA based Bray-Curtis dissimilarities also score below average on Pianka's overlap index (O: 0.00 - 0.06), except for cattle vs domestic goat (O: 0.28). In contrast, the highest overlap in dietary niches of domestic and wild herbivore species is observed for cattle and sambar deer (BC: 0.68, O: 0.68), and cattle and Indian hare (BC: 0.61, O: 0.65). These are followed by domestic goat and Asian elephant (BC: 0.84, O: 0.52), cattle and Asian elephant (BC: 0.70, O: 0.49), and domestic goat and sambar deer (BC: 0.65, O: 0.46).

Overall, 20 out of 45 comparisons showed statistically significant niche overlap based on comparison with 1000 null models (Table 2). Three out of seven wild herbivores have significant dietary overlap with cattle. In order of Pianka's niche overlap values from highest to lowest, these are: sambar deer, Indian hare and Asian elephant (O: 0.68, 0.65 and 0.49 respectively). Four wild herbivores have significant dietary overlap with domestic goat: Asian elephant, sambar deer, Indian porcupine and barking deer (O: 0.52, 0.46, 0.11 and 0.08). In the case of water buffalo, significant dietary overlap is also found for four wild herbivore diets: sambar deer, Indian hare, wild boar and Asian elephant (O: 0.24, 0.20, 0.12 and 0.12).

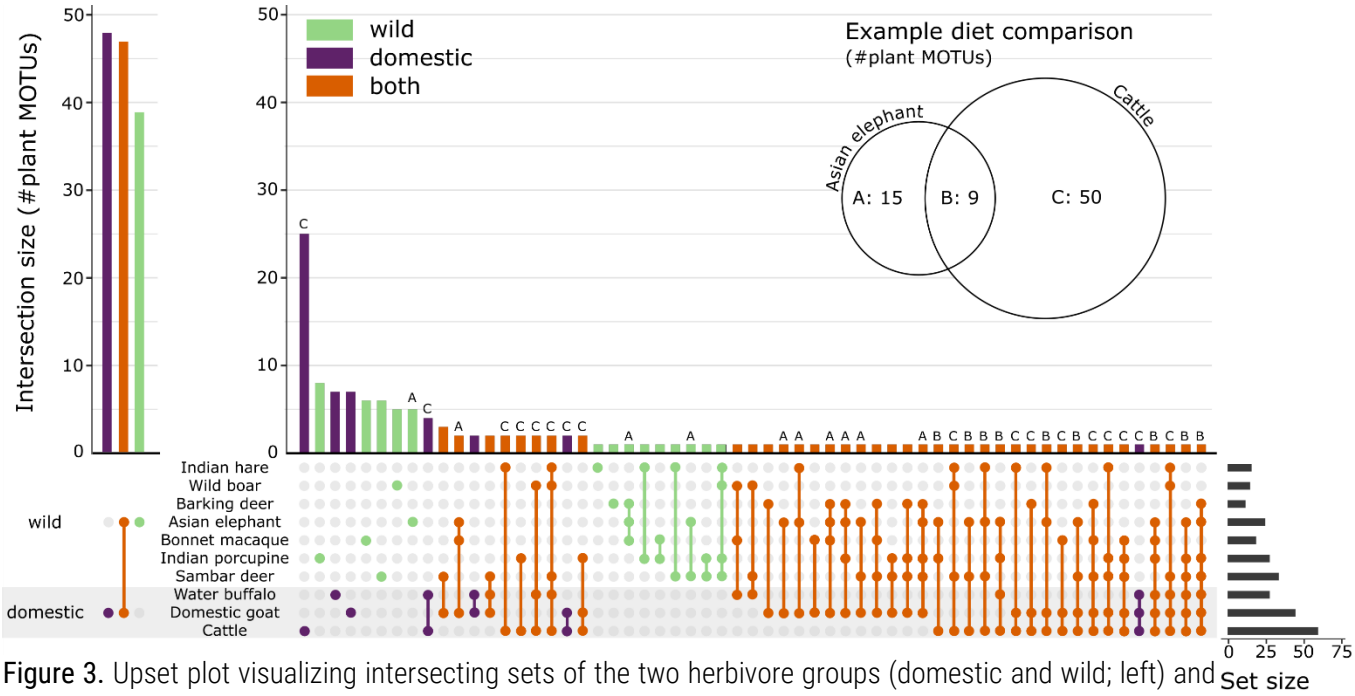


Figure 3. Upset plot visualizing intersecting sets of the two herbivore groups (domestic and wild; left) and herbivore species (right). Intersection size is presented in the number of shared plant MOTUs observed only for that particular combination of herbivore species indicated by the dots. Herbivore combinations and their respective intersection size bars are colored to indicate their composition of exclusively wild herbivore species (green), exclusively domestic herbivore species (purple), or both wild and domestic herbivore species (orange). Set size indicates the total number of MOTUs found per herbivore species. Scientific names for the herbivore species from top to bottom are as follows: *Lepus nigricollis, Sus scrofa, Muntiacus muntjak, Elephas maximus, Macaca radiata, Hystrix indica, Rusa unicolor, Bubalis bubalus, Capra hircus,* and *Bos taurus indicus.* The total number of shared plant MOTUs for a specific combination of herbivore species can be found by taking the sum of intersection sizes for all combinations that include the pair or group of herbivore species of interest. An example is shown for Asian elephant and cattle, where the total number of shared plant MOTUs is found by taking the sum of the bars indicated with "B." The bars specifying the number of MOTUs found for the Asian elephant but not for cattle are indicated with "A," and those found for cattle but not for Asian elephant are indicated with "C."

Table 2. Bray–Curtis dissimilarity (bottom left; 0: similar, 1: dissimilar) and Pianka's overlap index (top right; 0: no overlap, 1: full overlap) based on RRA data.

	Cattle	Domestic goat	Water buffalo	Asian elephant	Barking deer	Bonnet macaque	Indian hare	Indian porcupine	Sambar deer	Wild boar
Cattle		0.28*	0.23*	0.49*	0.00	0.00	0.65*	0.01	0.68*	0.03
Domestic goat	0.75		0.05	0.52*	0.08*	0.06	0.04	0.11*	0.46*	0.00
Water buffalo	0.82	0.92		0.12*	0.00	0.02	0.20*	0.04	0.24*	0.12*
Asian elephant	0.70	0.76	0.84		0.02	0.05	0.00	0.03	0.10*	0.00
Barking deer	0.98	0.92	0.99	0.93		0.19*	0.00	0.63*	0.03	0.00
Bonnet macaque	0.99	0.87	0.97	0.94	0.79		0.00	0.02	0.04	0.00
Indian hare	0.61	0.93	0.85	0.97	1.00	1.00		0.10*	0.60*	0.24*
Indian porcupine	0.94	0.87	0.94	0.90	0.72	0.95	0.90		0.06	0.01
Sambar deer	0.68	0.65	0.77	0.87	0.96	0.89	0.67	0.90		0.08*
Wild boar	0.96	0.99	0.88	1.00	1.00	0.98	0.80	0.99	0.93	

4. Discussion

Environmental DNA metabarcoding of faecal samples has enabled us to reconstruct the dietary niche partitioning of 10 mammalian herbivore species present in the MMH in southern India. We specifically focused on dietary overlap that may arise by shared use of the forest and village habitats by domesticated animals and wildlife and we argue for monitoring of potential effects of environmental change as restrictions on grazing are enforced and impacts of invasive species change.

4.1. Dietary niche reconstruction

The reconstructed diets represent a continuum of grazers, through mixed feeders and browsers to frugivorous mammals based on the RRAs of Poaceae compared to other plant families (Figure 2). Despite limited sample sizes for some herbivore species, these assignments are in agreement with the priori feeding guild assignments of the herbivores under study (Table 1) as based on Nowak and Walker (1999), Ahrestani et al. (2012), Ahrestani and Sankaran (2016), and IUCN (2020). One exception is observed: wild boar (Sus scrofa) is considered an omnivorous browser, but with 78% Poaceae in our study its diet is categorized together with the grazers (Figure 2). Wild boar are some of the most persistent crop raiders in the area, especially when local food staples finger millet (Eleusine coracana L.), sorghum (Sorghum bicolor L.) and maize (Zea mays L.) ripen and are harvested (November to January). This interpretation also fits with other studies that show a variable diet for wild boar across geographic regions and habitats (Ickes, 2001; Gray et al., 2016; Robeson et al., 2018).

Considering the ability of mixed feeders to switch between grazing and browsing (Ahrestani & Sankaran, 2016), we expected these species to have a generalist diet and therefore a relatively large dietary niche width compared to more specialised feeders. In accordance with these expectations, the narrowest dietary niches were found for two specialised feeders: bonnet macaque and cattle with Shannon diversities of 0.78 ± 0.19 and 0.87 ± 0.13 respectively (Table 1). Bonnet macaques are conventionally described as frugivores and we primarily found diet items originating from the Fabaceae family (63% RRA, Figure 2) in their diet. Contrarily, cattle are grazers and primarily eat grass, as indicated by the high RRA for Poaceae of 84%. Although the remaining 16% of the cattle diet is composed of 15 other plant families, the average number of MOTUs and Shannon diversity are low (6.22 ± 0.69 and 0.87 ± 0.13, respectively), which include wild grasses that grow in the villages and in the forest, as well as bamboos and cereal and vegetable crops. The hare is also conventionally described as primarily grazing (Nowak & Walker, 1999), but the analysed samples of the Indian hare contained a high proportion of Fabaceae (36%) as well as the expected Poaceae (61%). The high proportion of Fabaceae might be explained by crop raiding on various species of beans grown by farmers, and Cajanus sp. (pigeon pea) was

indeed detected in the hare diet (Table S7). Species assigned to the feeding guild of mixed feeders scored above average in dietary richness in both average number of MOTUs and Shannon diversity (Table 1 and Table S8.1). The Asian elephant is the exception to this pattern, scoring below average on Shannon diversity for both datasets. We found the diet of elephants to consist mainly of grasses (56%), which is consistent with other reports from southern India, although their diets are suggested to shift towards less woody plants and more graminoids (Poaceae, Juncaceae and Cyperaceae) in the wet season (Ahrestani et al., 2012; Baskaran, Balasubramanian, Swaminathan & Desai, 2010; Sukumar, 2006). The classification of mixed feeder is therefore only appropriate if one takes into account the complete diet, while it seems that within seasons they should be considered as either grazer or browser.

In the present study, samples were collected in the winter and summer pre-monsoon seasons in three subsequent years. Considering the seasonal availability of plants and the evidence for differences in herbivore feeding patterns over wet and dry seasons approximately 100 km from the study area (Ahrestani et al., 2012), it is likely that the dietary niches of the herbivores in MMH would shift over the seasons. Such shifts may also result in temporal variation in the dietary niche overlap between species pairs, though Ahrestani et al. (2012) found that the overlap in dietary niches of sambar deer and elephants remained constant across dry and wet seasons.

As species with narrow niches are deemed more vulnerable to environmental changes (Clavel et al., 2011; Devictor et al., 2010), it is especially important to closely monitor the dietary niches of bonnet macaque and Asian elephant. To obtain a more complete view of diet and dietary niche overlap, samples should be analysed across different seasons, and long-term monitoring of diet and its overlap should include temporally spaced sample collection ensuring both the wet and the dry season are covered.

4.2. Dietary niche partitioning

Due to the limited number of samples for several herbivore species, an amount of uncertainty in their dietary reconstructions needs to be acknowledged, and dietary niche dissimilarity or overlap may be under- or overestimated as a result. Nonetheless, the dietary compositions of the samples within herbivore species are similar and clear clusters can be distinguished (Figure 4, Supplemental Information S9) indicating dietary niche partitioning between the studied species. Herbivore species pairs of diverse feeding guilds show high niche dissimilarity and low overlap, which is especially the case for cattle vs bonnet macaque, cattle vs wild boar, and domestic goat vs bonnet macaque (Table 2). Corresponding herbivore pairs also clearly segregate in the nMDS plot (Figure 4, Supplemental Information S9).

As expected based on previous indications of livestock-mediated resource limitation in India (Bagchi et al., 2003; Madhusudan, 2004), we found significant

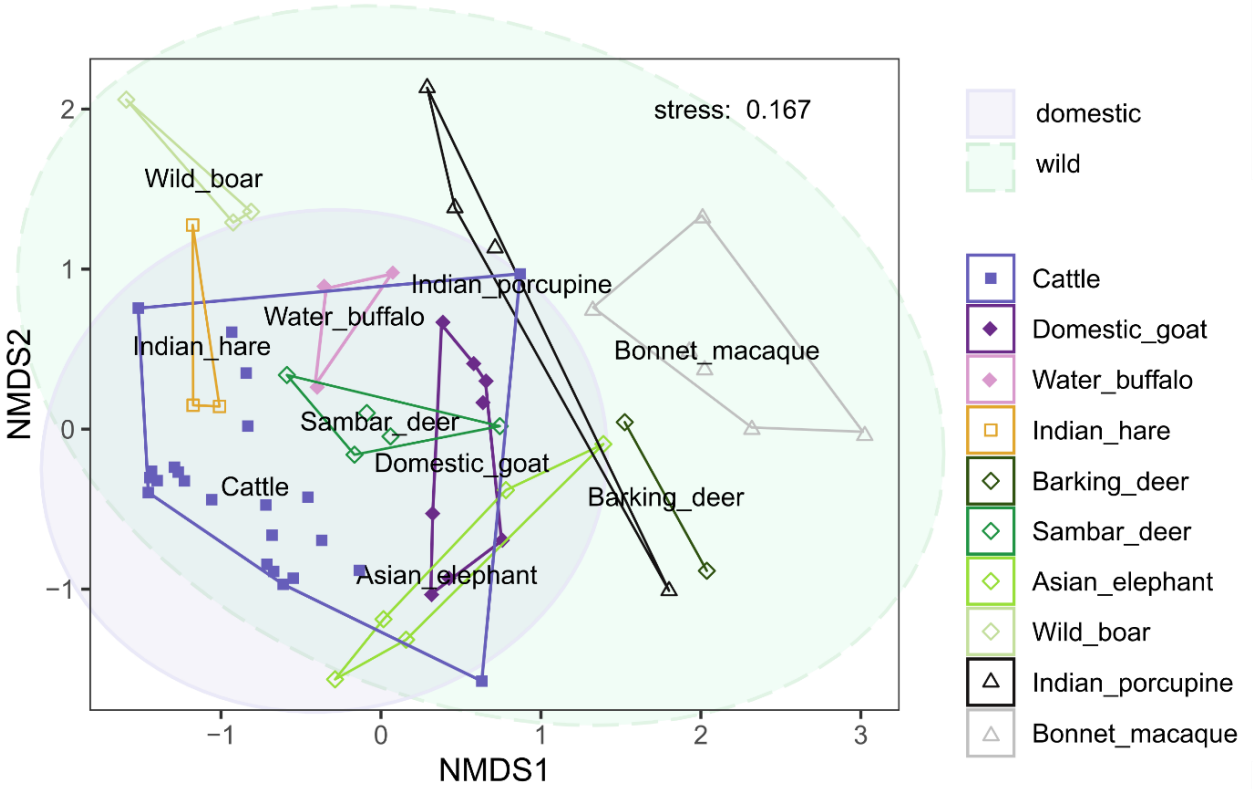


Figure 4. Dietary niche partitioning within and among domestic and wild herbivore species by nMDS of RRA-based Bray–Curtis dissimilarity of samples (adonis $F_{9,52} = 3.79$, $R^2 = 0.40$, $p \leq .001$). The positioning of the species label indicates the mean for that species. Depicted are as follows: cattle (*Bos taurus indicus*), domestic goat (*Capra hircus*), water buffalo (*Bubalis bubalus*), Indian hare (*Lepus nigricollis*), barking deer (*Muntiacus muntjak*), sambar deer (*Rusa unicolor*), Asian elephant (*Elephas maximus*), wild boar (*Sus scrofa*), Indian porcupine (*Hystrix indica*), and bonnet macaque (*Macaca radiata*). Samples from domestic herbivore species are indicated with filled symbols and the shaded ellipses indicate the standard deviation from the mean of domestic and wild herbivore groups. The shapes of the symbols refer to the different feeding guilds: grazer (square), mixed feeder (diamond), and frugivore (triangle). The stress level of 0.167 is under the cut-off value of 0.2 as posed by Clarke (1993) to indicate an interpretable ordination

overlap in consumed MOTUs (35%) between domestic and wild herbivores in MMH representing 76% of RRAs in the dataset. The greatest overlap was found for the dietary niches of cattle and sambar deer (O: 0.68). The cattle diet also overlaps with the diets of Indian hare and Asian elephant and the diet of domestic goat overlaps with the diets of sambar deer and Asian elephant (Table 2). Bray-Curtis dissimilarities and perMANOVA results support these findings with below average dissimilarities and no significant differences in dietary niches for these herbivore pairs. Similar overlap between cattle diet and the diets of several wild herbivore species across dietary guilds such as elephant (mixed feeder), impala (mixed feeder) and Dik-Dik (browser), have previously been reported in Kenya (Kartzinel et al. 2015). Likewise, a study in the Greater Himalayas indicates high trophic niche overlap between livestock (sheep and goats) and wild ungulates, including sambar deer (Bhattacharya et al., 2012) further supporting our findings of dietary overlap between domesticates and wildlife.

Compared to RRA data, presence/absence data transformations result in a larger differentiation in some herbivore pairs (cattle paired with Asian elephant, Indian porcupine, and sambar deer) and a smaller differentiation in seven other pairs (see Table S10). This reflects the difference in the amount of rare, low-abundant MOTUs in the diet of the various animals, since results based on RRAs are less sensitive to the presence of rare MOTUs compared to results based on presence/absence alone (Deagle et al. 2019). Rare MOTUs also play a role in the seemingly contradictory results of dietary niche similarity indices of two domestic herbivores. Domestic goat and cattle have significantly different dietary niches according to the RRA-based perMANOVA test, but score significantly high on Pianka's niche overlap index. Comparison of quantitative metrics of niche overlap from species distributions by Rödder and Engler (2011) suggests that Bray-Curtis values more accurately reflect niche overlap than most other tested methods. Especially for species distributions made up of many grid cells with low occurrence, Pianka's niche overlap was shown to be prone to both under- as well as overestimation (Rödder & Engler, 2011), suggesting a potential bias when herbivore diet is made up of many MOTUs with low RRAs. As this is the case for the comparison of domestic goat and cattle in this study, we should conclude that despite the suggested overlap according to Pianka's niche overlap index, their dietary niches are different.

Although eDNA metabarcoding provides a relatively cost-effective and time-efficient alternative to microhistological analyses (Pompanon et al., 2012), a limitation of the use of eDNA metabarcoding for dietary niche partitioning studies is the lack of differentiation between plant tissues. In cases where some herbivores prefer to eat the fruits, while others eat the leaf material or the roots of a plant, the dietary niche overlap can be overestimated as partitioning does not take place on a taxonomic level, but is instead based on the consumption of different parts of the plants. For example, this could be the case for bonnet macaque and barking deer based on their feeding guild assignments (Table 2), although both are reported to prefer young, fresh leaves and fruits (Ahrestani & Sankaran, 2016; Krishnamani, 1994) and thus partitioning would more likely take place based on which parts of the plants the animals can reach. Otherwise, we found relatively low taxonomic overlap in diets between herbivore species that are known to feed on different plant tissues.

4.3. Land use and invasive species

Apart from identifying dietary niches and quantifying dietary niche overlap, eDNA metabarcoding data from herbivore faecal samples also provides an opportunity to monitor the available plant taxa in the foraging areas of the herbivores under study. MMH is home to several ethnic groups that largely depend on the forest for their livelihoods (Harisha & Padmavathy, 2013; Kent & Dorward, 2015), and the area consists of forests, interspersed with anthropogenic lands for crops,

plantations or buildings (Figure 1). Human influences on dietary composition are therefore likely.

The traditional ecological knowledge in the MMH villages together with the locally occurring plant species and their use have been mapped in a previous study (Harisha et al. 2015). The local communities depend to a large degree on agriculture for their livelihood and grow both subsistence crops (e.g. the cereals and beans) as well as cash crops (e.g. maize and sunflower; Harisha et al. 2015). Not all known cultivated species could be identified to species level in our study; nevertheless, we detected several known crop species in the diets of the studied herbivores. For example, in the diet of the water buffalo, we observed *Brassica rapa* L. which is cultivated in the area for food and medicinal purposes (Harisha et al., 2015). In the cattle diet, *Achyranthes aspera* L. is present –a species used for food and cultural reasons (Harisha et al., 2015)– and *Sorghum bicolor* (L.) Moench –commonly cultivated as cash crop for use as fodder as well as human consumption (Harisha et al., 2015)– was found in the diet of both cattle and water buffalo. These domestic herbivores are likely to obtain these food sources by way of supplemental feeding.

We also found evidence for the consumption of agricultural crops by wild herbivores, e.g. in the presence of *Amaranthus* spp. and Poaceae spp. in the diet of wild boar, and *Cajanus* spp. in the diet of Indian hare, all of which are cultivated for food (Harisha et al., 2015). Many other members from the Poaceae plant family are commonly used as fodder (e.g. *Apluda mutica* L., *Cynodon dactylon* (L.) Pers., Melinis repens (Willd.) Zizka, *Heteropogon contortus* (L.) P.Beauv. ex Roem. & Schult. and *Themeda triandra* Forssk.), but a large proportion of Poaceae found could not be identified to genus or species level, thereby limiting inferences about the proportion of cultivated fodder versus wild Poaceae species. However, using the diet items that do have genus or species level identifications, some instances can be identified where typical fodder species are also eaten by wild herbivores. One example is *Cynodon dactylon*, which is present in the diets of water buffalo (4%) and cattle (< 1%), but also in the diets of sambar deer (1%), Indian hare (7%) and wild boar (37%). *Vachellia farnesiana* (L.) Wight & Arn. (Fabaceae) is also a known fodder species and is present in the diet of domestic goat and cattle at percentages below 1%, but at 17% represents a much larger proportion of the Indian porcupine diet. Wild boar are some of the most persistent crop raiders in the area, especially as local food staples ripen (finger millet, sorghum, and maize in November to January), and farmers stay up all night with smoky fires to scare away the boar, as well as deer, hare, and elephants. Reports from farmers in MMH indicate an increase in crop raiding with the increase of invasive *Lantana camara* L. (Verbenaceae) in the forest understory, while Mundoli et al. (2016) documented similar increases of crop raiding by boar in neighbouring BRT Tiger Reserve over a 7-year period, leading to many farmers giving up growing food crops altogether in favour of commercial coffee.

Many plant species occur both in the wild as well as on agricultural lands, or are collected from the wild for use as food, medicine or cultural purposes. Harisha et al. (2015) identified 96 wild plant species that are used for food, 118 for medicine, 26 for cultural and 14 for economic purposes in the area. An example is the tamarind (Tamarindus indica L.) of which local communities use the fruits as food and as a source of income (Shaanker et al., 2004). We found sequence reads from *Tamarindus indica* in the diets of bonnet macaque, water buffalo and wild boar. Another example is *Semecarpus anacardium* L.f.; its fruits are used for food as well as medicine (Harisha et al., 2015) and sequence reads for this species were identified in the diets of Asian elephant, barking deer, cattle, domestic goat, Indian porcupine, sambar deer and water buffalo. These herbivores are probably eating the leaves as the use of fruits is limited to the months of May to October (Harisha et al., 2015) and faecal samples were obtained between December and April. In addition to niche overlap between wild and domestic herbivores, there is a potential overlap between herbivores and humans in utilized plant species.

Finally, the introduction and spread of invasive species may influence the diet of herbivores in MMH. *L. camara* is a very abundant invasive plant species in the area, and has been given academic attention (e.g. Aravind et al., 2010; Varghese et al., 2015) as well as in conservation management: local communities are encouraged to use it for the production of non-timber forest products (Kannan et al., 2016). The species makes up a small part of our dataset (0.78%) and is found in 13 of the 62 samples. Cattle and domestic goats eat it in small quantities, and also some of the wild herbivores, i.e. porcupines and macaques that reportedly mostly eat the fruits. However, it seems to make up a more substantive part of the diet of particularly sambar deer (5%). Large herbivores are reported to avoid L. camara as its leaves and fruits contain toxins that cause cholestasis and hepatoxicity, which could ultimately lead to death (Sharma et al., 2007). Furthermore, the spread of *L. camara* in MMH is likely to reduce the availability of more suitable diet items, as its presence is associated to a decline in tree sapling densities and grass volume (Prasad, 2012; Varghese et al., 2015) and reduces access to the forests for wildlife, domesticates and humans (Thornton et al., 2019). Widespread expansion of this invasive plant may therefore restrict resource availability and consequently change the foraging ecology of herbivores in invaded areas (Wilson et al., 2013). Continued monitoring of the presence of L. camara is therefore recommended.

4.4. Wildlife management in MMH

MMH is known as an important elephant corridor and forms a large tiger habitat together with the adjacent BRT wildlife sanctuary (Bawa, Joseph, & Setty, 2007; Gubbi et al., 2017). Of the wild herbivores under study, the sambar deer and Asian elephant are respectively considered vulnerable and endangered (IUCN, 2020), while the other herbivores are considered of least concern under IUCN 3.1. A study of the food habits of tigers in northern India indicated that the sambar deer,

together with wild boar and chital, constitutes a major part of the tiger's diet (Biswas & Sankar, 2002), which further indicates the importance of studying the wildlife in MMH.

The dietary niche overlap we identified between wild and domestic herbivores, combined with previous indications of livestock mediated resource limitation in India (Madhusudan, 2004), suggests potential for competition between domestic and wild herbivores in the MMH area, especially in case of limited resource availability. Niche overlap does not necessarily equate to competition (Pianka, 2011a), and assessment of the resource availability is needed to establish if there is direct competition. Previous authors have suggested that competitive exclusion of wild herbivores occupying similar niches may eventually occur if domestic herbivores are given an artificial competitive advantage (e.g. through extra feed in periods of scarcity) (Hardin, 1960; Pianka, 2011b). For instance, an experimental study in Kenya showed a reduction of land-use by wildlife (including zebra, oryx, buffalo, steenbok, gazelle, eland and elephant) with the presence of cattle (Kimuyu et al., 2016). The effect of livestock presence on wild herbivores will vary per species and geographic area, depending on the overlap in dietary niches, social intolerance, required forage quantity and quality, and several other factors (see Schieltz & Rubenstein, 2016 for a review). For example, Madhusudan (2004) described a muted effect for sambar deer, but a sharp decline in elephant population densities with increased livestock presence, followed by a clear increase after reduced livestock numbers in Bandipur national park, southern India. Furthermore, domestic herbivores can act as carrier of disease, such as foot-and-mouth disease, potentially spreading to wild herbivores as suggested for two wildlife sanctuaries close to MMH (Chandranaik, Shivashankar, Giridhar & Nagaraju, 2016; Silori & Mishra, 2001).

Since becoming a designated wildlife sanctuary in 2013, forest access has become more regulated and only daily livestock grazing is permitted as cowsheds have been forbidden in MMH (Thornton et al, 2019). Such measures are likely to reduce resource competition and interaction between wild and domestic species. Likewise, the encouraged use of the invasive plant L. camara by local communities may limit the negative impacts of this plant on the habitat and resource availability of herbivores in MMH. Continued monitoring could show if these particular conservation management strategies prove to be effective.

Based on our findings, particular concern should go to bonnet macaque and Asian elephant as their narrow dietary niches could make them vulnerable to changes in their environment (Clavel et al., 2011; Devictor et al., 2010), such as climate change, invasive species and other anthropogenic pressures. Indeed, the range extension of rhesus macaque has already been suggested to threaten the declining bonnet macaque populations in southern India (Kumar, Radhakrishna & Sinha, 2011) and Asian elephant habitats in India are under continuous threats of forest fragmentation and loss (Padalia, et al., 2019). Moreover, our observations of dietary niche overlap suggest that especially the overlap of cattle and domestic

goat with sambar deer and Asian elephants should be monitored closely, as these latter species are already considered vulnerable and endangered (IUCN, 2020).

Environmental DNA metabarcoding of faecal samples has provided a starting point for tracking the effects of the environmental changes in MMH on local wildlife. As environmental change continues to threaten biodiversity in the area, the need to continue monitoring both the wildlife species themselves and the interaction between wildlife and domestic livestock becomes more urgent. This is not only true for the wildlife sanctuary under study, but for many ecosystems across the world as they are under increasing pressure from globally increasing livestock population sizes, invasive species, habitat degradation and other anthropogenic factors. DNA metabarcoding of faecal samples is an ideal, non-invasive method for such monitoring, providing a wide variety of valuable information for biodiversity management.

Author contributions

GR, RP, HdB and SB designed the study. GR, RP and BS collected the samples. AP and LT performed the laboratory work. BS performed the laboratory work for the reference library. AtS did the bioinformatics and statistical analyses. AtS made the figures and wrote the paper with input from RP, HdB and SB. All authors edited and made suggestions on the final text.

Acknowledgements

We thank Dr. Harisha and Dr. Siddappa Setty for their botanical expertise of the MMH area and help with the taxonomical identifications and Dr. Ravichandran for his identification of fodder grasses. We gratefully acknowledge the help from Mr. Madesha and Mr. Narayanan during sample collections. We are further grateful to the Ecoinformatics lab at ATREE for use of their map and The Norwegian Sequencing Centre for their advice and plant DNA sequencing. The analyses were performed on the Abel computer cluster, owned by the University of Oslo (UiO) and Uninett/Sigma2, and operated by the Department for Research Computing at USIT (UiO). GR acknowledges the support from Norwegian embassy for the travel grant to the UiO. Our work was conducted under Malai Mahadeshwara Hills Wildlife Sanctuary Permit -PCCF/[WF]/E2/CR/22/2014-15.

Data availability statement

All read data are available at the European Nucleotide Archive (ENA) under study accession number PRJEB41139. All processed data are available in the supporting information of this article.

References

Ahrestani, F. S., & Sankaran, M. (Eds.). (2016). The Ecology of Large Herbivores in South and Southeast Asia. https://doi.org/10.1007/978-94-017-7570-0

Ahrestani, F. S., Heitkönig, I. M. A., & Prins, H. H. T. (2012). Diet and habitat-niche relationships within an assemblage of large herbivores in a seasonal tropical forest. Journal of Tropical Ecology, 28(4), 385–394. https://doi.org/10.1017/S0266467412000302

Anand, S., & Radhakrishna, S. (2017). Investigating trends in human-wildlife conflict: is conflict escalation real or imagined?. Journal of Asia-Pacific Biodiversity, 10(2), 154-161.

Ando, H., Mukai, H., Komura, T., Dewi, T., Ando, M., & Isagi, Y. Methodological trends and perspectives of animal dietary studies by noninvasive fecal DNA metabarcoding. Environmental DNA. https://doi.org/10.1002/edn3.117

Aravind, N. A., Rao, D., Ganeshaiah, K. N., Shaanker, R. U., & Poulsen, J. G. (2010). Impact of the invasive plant, Lantana camara, on bird assemblages at Malé Mahadeshwara Reserve Forest, South India. Tropical Ecology, 51(2), 325-338. https://doi.org/10.1111/aje.12698

Bison, M., Ibanez, S., Redjadj, C., Boyer, F., Coissac, E., Miquel, C., … Loison, A. (2015). Upscaling the niche variation hypothesis from the intra- to the inter-specific level. Oecologia, 179(3), 835–842. https://doi.org/10.1007/s00442-015-3390-7

Biswas, S., & Sankar, K. (2002). Prey abundance and food habit of tigers (Panthera tigris tigris) in Pench National Park, Madhya Pradesh, India. Journal of Zoology, 256(3), 411-420. https://doi.org/10.1017/s0952836902000456

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution, 29(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. (2016) obitools: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16, 176–182. https://doi.org/10.1111/1755-0998.12428

Bradley, B. J., Stiller, M., Doran-Sheehy, D. M., Harris, T., Chapman, C. A., Vigilant, L., & Poinar, H. (2007). Plant DNA sequences from feces: Potential means for assessing diets of wild

Bagchi, S., Goyal, S. P., & Sankar, K. (2003). Niche Relationships in an Ungulate Assemblage in a Dry Tropical Forest. Journal of Mammalogy, 84(3), 981–988. https://doi.org/10.1644/BBa-024

Bawa, K. S., Joseph, G., & Setty, S. (2007). Poverty, biodiversity and institutions in forest-agriculture ecotones in the Western Ghats and Eastern Himalaya ranges of India. Agriculture, Ecosystems & Environment, 121(3), 287–295. https://doi.org/10.1016/j.agee.2006.12.023

Baskaran, N., Balasubramanian, M., Swaminathan, S., & Desai, A. A. (2010). Feeding ecology of the Asian elephant Elephas maximus Linnaeus in the Nilgiri Biosphere Reserve, southern India. Journal of the Bombay Natural History Society, 107(1), 3.

Bhattacharya, T., Kittur, S., Sathyakumar, S., & Rawat, G. S. (2012). Diet overlap between wild ungulates and domestic livestock in the greater Himalaya: implications for management of grazing practices. Proceedings of the Zoological Society 65(1), 11-21. https://doi.org/10.1007/s12595-012-0025-4 primates. American Journal of Primatology, 69(6), 699–705. https://doi.org/10.1002/ajp.20384

Carscadden, K. A., Emery, N. C., Arnillas, C. A., Cadotte, M. W., Afkhami, M. E., Gravel, D., ... & Wiens, J. J. (2020). Niche Breadth: Causes and Consequences for Ecology, Evolution, and Conservation. The Quarterly Review of Biology, 95(3), 179-214. https://doi.org/10.1086/710388

Cantalapiedra, J. L., FitzJohn, R. G., Kuhn, T. S., Fernández, M. H., DeMiguel, D., Azanza, B., … Mooers, A. Ø. (2014). Dietary innovations spurred the diversification of ruminants during the Caenozoic. Proceedings of the Royal Society B: Biological Sciences, 281(1776), 20132746. https://doi.org/10.1098/rspb.2013.2746

Chandranaik, B. M., Shivashankar, B. P., Giridhar, P., & Nagaraju, D. N. (2016). Molecular characterisation and serotyping of Pasteurella multocida isolates from Asiatic elephants (Elephas maximus). European Journal of Wildlife Research, 62(6), 681-685.

Clavel, J., Julliard, R., & Devictor, V. (2011).

Worldwide decline of specialist species: toward a global functional homogenization?. Frontiers in Ecology and the Environment, 9(4), 222-228.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics, 33(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. Annual Review of Ecology, Evolution, and Systematics, 49(1), 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

Czernik, M., Taberlet, P., Świsłocka, M., Czajkowska, M., Duda, N., & Ratkiewicz, M. (2013). Fast and efficient DNA-based method for winter diet analysis from stools of three cervids: Moose, red deer, and roe deer. Acta Theriologica, 58(4), 379–386. https://doi.org/10.1007/s13364-013-0146-9

de Sousa, L. L., Silva, S. M., & Xavier, R. (2019). DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. Environmental DNA, 1(3), 199-214. https://doi.org/10.1002/edn3.27

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., … Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? Molecular Ecology, 28(2), 391–406. https://doi.org/10.1111/mec.14734

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., … Mouquet, N. (2010). Defining and measuring ecological specialization. Journal of Applied Ecology, 47(1), 15–25. https://doi.org/10.1111/j.1365-2664.2009.01744.x

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 434. https://doi.org/10.1186/1471-2164-11-434

Fløjgaard, C., De Barba, M., Taberlet, P., & Ejrnæs, R. (2017). Body condition, diet and ecosystem function of red deer (Cervus elaphus) in a fenced nature reserve. Global Ecology and Conservation, 11, 312–323. https://doi.org/10.1016/j.gecco.2017.07.003

Food and Agriculture Organization of the United Nations. (2017). FAOSTAT statistical database [http://www.fao.org/faostat/en/#data].

Gotelli, N., Hart, E., & Ellison, A. (2015). EcoSimR: Null model analysis for ecological data. R Package Version 0.1. 0. Available at Github. Com/Gotellilab/EcoSimR. Accessed August, 29, 2017. https://doi.org/10.1016/0304-3800(94)00117-z

Gray, T. N., Prum, S., & Phan, C. (2016). Density and activity patterns of the globally significant large herbivore populations of Cambodia's Eastern Plains landscape. In: The Ecology of Large Herbivores in South and Southeast Asia (pp. 207-222). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-7570-0_9

Gubbi, S., Harish, N., Kolekar, A., Poornesha, H., Reddy, V., Mumtaz, J., & Madhusudan, M. (2017). From intent to action: A case study for the expansion of tiger conservation from southern India. Global Ecology and Conservation, 9, 11–20. https://doi.org/10.1016/j.gecco.2016.11.001

Hardin, G. (1960). The competitive exclusion principle. Science, 131(3409), 1292–1297.

Harisha, R. P., Padmavathy, S., & Nagaraja, B. C. (2015). Traditional Ecological Knowledge (TEK) and its Importance in South India: Perspecive from Local Communities. Applied Ecology and Environmental Research, 14(1), 311–326. https://doi.org/10.15666/aeer/1401_311326

Harisha, R., & Padmavathy, S. (2013). Knowledge and Use of Wild Edible Plants in Two Communities in Malai Madeshwara Hills, Southern India. International Journal of Botany, 9, 64–72. https://doi.org/10.3923/ijb.2013

Hibert, F., Taberlet, P., Chave, J., Scotti-Saintagne, C., Sabatier, D., & Richard-Hansen, C. (2013). Unveiling the Diet of Elusive Rainforest Herbivores in Next Generation Sequencing Era? The Tapir as a Case Study. PLoS ONE, 8(4), e60799. https://doi.org/10.1371/journal.pone.0060799

Hutchinson, G. E. (1959). Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? The American Naturalist, 93(870), 145–159. https://doi.org/10.1086/282070

Ickes, K. (2001). Hyper-abundance of Native Wild Pigs (Sus scrofa) in a Lowland Dipterocarp Rain Forest of Peninsular Malaysia 1. Biotropica, 33(4), 682-690. https://doi.org/10.1646/0006-3606(2001)033[0682:haonwp]2.0.co;2

IUCN 2020. The IUCN Red List of Threatened Species. Version 2019-3. Available at https://www.iucnredlist.org

Kannan, R., Shackleton, C. M., Krishnan, S., & Shaanker, R. U. (2016). Can local use assist in

controlling invasive alien species in tropical forests? The case of Lantana camara in southern India. Forest Ecology and Management, 376, 166-173. https://doi.org/10.1016/j.foreco.2016.06.016

Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., … Pringle, R. M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences, 112(26), 8019–8024. https://doi.org/10.1073/pnas.1503283112

Kartzinel, T. R., & Pringle, R. M. (2020). Multiple dimensions of dietary diversity in large mammalian herbivores. Journal of Animal Ecology. https://doi.org/10.1111/1365-2656.13206

Kent, R., & Dorward, A. (2015). Livelihood responses to Lantana camara invasion and biodiversity change in southern India: Application of an asset function framework. Regional Environmental Change, 15(2), 353–364. https://doi.org/10.1007/s10113-014-0654-4

Kimuyu, D. M., Veblen, K. E., Riginos, C., Chira, R. M., Githaiga, J. M., & Young, T. P. (2017). Influence of cattle on browsing and grazing wildlife varies with rainfall and presence of megaherbivores. Ecological Applications, 27(3), 786-798. https://doi.org/10.1002/eap.1482

Kowalczyk, R., Taberlet, P., Coissac, E., Valentini, A., Miquel, C., Kamiński, T., & Wójcik, J. M. (2011). Influence of management practices on large herbivore dietCase of European bison in Białowieża Primeval Forest (Poland). Forest Ecology and Management, 261(4), 821–828. https://doi.org/10.1016/j.foreco.2010.11.026

Kowalczyk, R., Wójcik, J. M., Taberlet, P., Kamiński, T., Miquel, C., Valentini, A., … Coissac, E. (2019). Foraging plasticity allows a large herbivore to persist in a sheltering forest habitat: DNA metabarcoding diet analysis of the European bison. Forest Ecology and Management, 449, 117474. https://doi.org/10.1016/j.foreco.2019.117474

Krishnamani, R. (1994). Diet composition of the bonnet macaque (Macaca radiata) in a tropical dry evergreen forest of southern India. Tropical Biodiversity, 2(2), 285-302.

Kumar, R., Radhakrishna, S., & Sinha, A. (2011). Of least concern? Range extension by rhesus macaques (Macaca mulatta) threatens long-term survival of bonnet macaques (M. radiata) in peninsular India. International Journal of Primatology, 32(4), 945-959.

Lamarque, F., Anderson, J., Fergusson, R., Lagrange, M., Osei-Owusu, Y., Bakker, L. (2009).

Human-wildlife conflict in Africa: causes, consequences and management strategies (No. 157). Food and Agriculture Organization of the United Nations (FAO).

Lopes, C. M., De Barba, M., Boyer, F., Mercier, C., da Silva Filho, P. J. S., Heidtmann, L. M., … Taberlet, P. (2015). DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: A case study on subterranean rodents. Heredity, 114(5), 525–536. https://doi.org/10.1038/hdy.2014.109

MacArthur, R., & Levins, R. (1967). The Limiting Similarity, Convergence, and Divergence of Coexisting Species. The American Naturalist, 101(921), 377–385. https://doi.org/10.1086/282505

Madhusudan, M. D. (2004). Recovery of wild large herbivores following livestock decline in a tropical Indian wildlife reserve: Livestock decline and wild herbivore recovery. Journal of Applied Ecology, 41(5), 858–869. https://doi.org/10.1111/j.0021-8901.2004.00950.x

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. Bioinformatics, 24(16), 1757-1764. https://doi.org/10.1093/bioinformatics/btn554

Mundoli, S., Joseph, G. and Setty, S. (2016). "Shifting agriculture": The changing dynamics of

Adivasi farming in the forest-fringes of a tiger reserve in South India, Journal of Agroecology and Sustainable Food Systems, 40(8), 759-782. https://doi.org/10.1080/21683565.2016.1189475

Mychek-Londer, J. G., Chaganti, S. R., & Heath, D. D. (2020). Metabarcoding of native and invasive species in stomach contents of Great Lakes fishes. Plos one, 15(8), e0236077. https://doi.org/10.1371/journal.pone.0236077

Newmaster, S. G., Thompson, I. D., Steeves, R. A., Rodgers, A. R., Fazekas, A. J., Maloles, J. R., … Fryxell, J. M. (2013). Examination of two new technologies to assess the diet of woodland caribou: Video recorders attached to collars and DNA barcoding. Canadian Journal of Forest Research, 43(10), 897–900. https://doi.org/10.1139/cjfr-2013-0108

Nowak, R. M., & Walker, E. P. (1999). Walker's Mammals of the World (Vol. 1). JHU press.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2019). Vegan: Community ecology package.

Öllerer, K., Varga, A., Kirby, K., Demeter, L., Biró, M., Bölöni, J., & Molnár, Z. (2019). Beyond the obvious impact of domestic livestock grazing on temperate forest vegetation–A global review. Biological Conservation, 237, 209-219.

Padalia, H., Ghosh, S., Reddy, C. S., Nandy, S., Singh, S., & Kumar, A. S. (2019). Assessment of historical forest cover loss and fragmentation in Asian elephant ranges in India. Environmental Monitoring and Assessment, 191(3), 802.

Pansu, J., Guyton, J. A., Potter, A. B., Atkins, J. L., Daskin, J. H., Wursten, B., … Pringle, R. M. (2019). Trophic ecology of large herbivores in a reassembling African ecosystem. Journal of Ecology, 107(3), 1355–1376. https://doi.org/10.1111/1365-2745.13113

Pianka, E. R. (1974). Niche overlap and diffuse competition. Proceedings of the National Academy of Sciences, 71(5), 2141–2145. https://doi.org/10.1073/pnas.71.5.2141

Pianka, E. R. (2011a). The Ecological Niche. In Evolutionary Ecology (7th ed., pp. 267–289). Eric R. Pianka.

Pianka, E. R. (2011b). Competition. In Evolutionary Ecology (7th ed., pp. 240–264). Eric R. Pianka.

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing: NGS DIET ANALYSIS. Molecular Ecology, 21(8), 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Prasad, A. E. (2012). Landscape-scale relationships between the exotic invasive shrub Lantana camara and native plants in a tropical deciduous forest in southern India. Journal of Tropical Ecology, 28(1), 55-64.

Rayé, G., Miquel, C., Coissac, E., Redjadj, C., Loison, A., & Taberlet, P. (2011). New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: Chamois diet in autumn as a case study. Ecological Research, 26(2), 265–276. https://doi.org/10.1007/s11284-010-0780-5

Ripple, W. J., Newsome, T. M., Wolf, C., Dirzo, R., Everatt, K. T., Galetti, M., … Van Valkenburgh, B. (2015). Collapse of the world's largest herbivores. Science Advances, 1(4), e1400103. https://doi.org/10.1126/sciadv.1400103

Robeson, M. S., Khanipov, K., Golovko, G., Wisely, S. M., White, M. D., Bodenchuck, M., ... & Piaggio, A. J. (2018). Assessing the utility of metabarcoding for diet analyses of the omnivorous wild pig (Sus scrofa). Ecology and Evolution, 8(1), 185-196. https://doi.org/10.1002/ece3.3638

Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. Global Ecology and Biogeography, 20(6), 915-927. https://doi.org/10.1111/j.1466-8238.2011.00659.x

Sato, J. J., Shimada, T., Kyogoku, D., Komura, T., Uemura, S., Saitoh, T., & Isagi, Y. (2018). Dietary niche partitioning between sympatric wood mouse species (Muridae: Apodemus) revealed by DNA meta-barcoding analysis. Journal of Mammalogy, 99(4), 952–964. https://doi.org/10.1093/jmammal/gyy063

Schieltz, J. M., & Rubenstein, D. I. (2016). Evidence based review: positive versus negative effects of livestock grazing on wildlife. What do we really know? Environmental Research Letters, 11(11), 113003. https://doi.org/10.1088/1748-9326/11/11/113003

Shaanker, R. U., Ganeshaiah, K. N., Krishnan, S., Ramya, R., Meera, C., Aravind, N. A., ... & Gauthier, R. (2004). Livelihood gains and ecological costs of non-timber forest product dependence: assessing the roles of dependence, ecological knowledge and market structure in three contrasting human and ecological settings in south India. Environmental Conservation, 31(3), 242-253.https://doi.org/10.1017/s0376892904001596

Sharma, O. P., Sharma, S., Pattabhi, V., Mahato, S. B., & Sharma, P. D. (2007). A review of the hepatotoxic plant Lantana camara. Critical reviews in toxicology, 37(4), 313-352. https://doi.org/10.1080/10408440601177863

Silori, C. S., & Mishra, B. K. (2001). Assessment of livestock grazing pressure in and around the elephant corridors in Mudumalai Wildlife Sanctuary, south India. Biodiversity & Conservation, 10(12), 2181-2195.

Simberloff, D., & Dayan, T. (1991). The Guild Concept and the Structure of Ecological Communities. 31.

Soininen, E. M., Ehrich, D., Lecomte, N., Yoccoz, N. G., Tarroux, A., Berteaux, D., … Ims, R. A. (2014). Sources of variation in small rodent trophic

niche: New insights from DNA metabarcoding and stable isotope analysis. Isotopes in Environmental and Health Studies, 50(3), 361–381. https://doi.org/10.1080/10256016.2014.915824

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., … Yoccoz, N. G. (2015). Highly Overlapping Winter Diet in Two Sympatric Lemming Species Revealed by DNA Metabarcoding. PLOS ONE, 10(1), e0115335. https://doi.org/10.1371/journal.pone.0115335

Soumya, K.V., Shackleton, C.M., & Setty, S. R. (2019a). Harvesting and Local Knowledge of a Cultural Non-Timber Forest Product (NTFP): Gum-Resin from Boswellia serrata Roxb. in Three Protected Areas of the Western Ghats, India. Forests, 10(10), 907. https://doi.org/10.3390/f10100907.

Soumya, K.V., Shackleton, C.M., & Setty, S. R. (2019b). Impacts of gum-resin harvest and Lantana camara invasion on the population structure and dynamics of Boswellia serrata in the Western Ghats, India. Forest Ecology and Management, 453, 117618

Sukumar, R. (2006). A brief review of the status, distribution and biology of wild Asian elephants Elephas maximus. International Zoo Yearbook, 40(1), 1-8.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14–e14. https://doi.org/10.1093/nar/gkl938

[dataset] ter Schure, A.T.M., Pillai, A.A.S, Thorbek, L., Bhavani Shankar, M., Puri, R., Ravikanth, G., de Boer, H.J., & Boessenkool, S.; European Nucleotide Archive; study accession number: PRJEB41139

Thornton, T. F., Puri, R. K., Bhagwat, S., & Howard, P. (2019). Human adaptation to biodiversity change: An adaptation process approach applied to a case study from southern India. Ambio, 48(12), 1431-1446.

Varghese, A., Ticktin, T., Mandle, L., Nath, S. (2015). Assessing the Effects of Multiple Stressors on the Recruitment of Fruit Harvested Trees in a Tropical Dry Forest, Western Ghats, India. PLoS ONE, 10(3).

Voldstad, L. H., Alsos, I.G., Farnsworth, W. R, Heintzmann, P. D., Håkansson, L., Kjellman, S. E., Rouillard, A., Schomacker, A., & Eidesen, P. B. (2020). A complete Holocene lake sediment ancient DNA record reveals long standing high-arctic plant diversity hotspot in northern Svalbard. Quaternary Science Reviews. https://doi.org/10.1016/j.quascirev.2020.106207

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., … Taberlet, P. (2009). New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The Trn L approach. Molecular Ecology Resources, 9(1), 51–60. https://doi.org/10.1111/j.1755-0998.2008.02352.x

Wilson, G., Desai, A. A., Sim, D. A., & Linklater, W. L. (2013). The influence of the invasive weed Lantana camara on elephant habitat use in Mudumalai Tiger Reserve, southern India. Journal of Tropical Ecology, 29(3), 199-207. https://doi.org/10.1017/s0266467413000205

Zhang, J. (2016). Spaa: SPecies association analysis.

Supporting information

Additional supporting information can be found online in the Supporting Information section. https://doi.org/10.1002/edn3.168

S1. Herbivore primers

Table S1. The herbivore specific primers designed for this study. GenBank identities refer to the sequences used for the primer design.

Target species	Target gene	Primer name	Primer sequence	Primer size(bp)	Product size(bp)	Tm (°C)
Macaca radiata	12S	macacF	TGGAAGGTGCGCTTGGATAA	20	150	55
GenBank: AY224262.1		macacR	ATGGTTTAGTGTGCGTTGGC	20		
Elephas maximus	Cytb	elmaF	TCTAGCTTTCCTACCAATTGCAG	23	150	53
GenBank: AY589512.2		elmaR	TTGATAGCGAGGTAAGTGGACC	22		
Hystrix indica	16S	hyinF	TGCCCAGTGACAAACCAGTT	20	150	53
GenBank: JN714145.1		hyinR	CACGGGAAGGTCAATTTCACTG	22		
Bos bubalus, Bos taurus indicus, Bos gaurus	CO1	bosbubF	GTAACCGCACACGCATTTGT	20	149	52
GenBank: KF808255.1, MF667930.1, JQ735454.1		bosbubR	GGAGGGAGRAGTCAGAAGCT	20		
Rusa unicolor, Axis axis, Muntiacus muntjak	CO1	rusaxismF	YGTAACCGCACATGCATTCG	20	150	52
GenBank: KT372090.1, KT372098.1, KT372093.1		rusaxismR	GGWGGRAGRAGYCAAAAGCTT	21		
Tetracerus quadricornis, Capra hircus	CO1	tqcaF	AACAGAYCGAAACCTAAACACAACC	25	150	51
GenBank: EF536355.1, JN245994.1		tqcaR	TAGGTTACGATGTGAGARATTATTCC	26		
Melursus ursinus	Cytb	slothbF	ATCAGACACAACCACAGCCT	20	150	54
GenBank: EF196662.1		slothbR	GAGCCATAGTACAGACCCCG	20		
Lepus spp.	Cytb	lepusF	TGGCTCCAATAACCCATCAGG	21	180	48
GenBank: JQ219662.1, NC_025316.1, KU250092.1, HQ596461.1		lepusR	TTGAGGGGATTGGCAGGGG	19		
Sus scrofa	Cytb	susF	CGCTACCTACATGCAAACGGA	21	180	50
GenBank: JN242241.1, AB015083.1		susR	TGATATTTGTCCTCAGGGCAGG	22		

S2. Plant DNA Reference libraries

Two plant DNA reference libraries were prepared for the taxonomic identification of the plant DNA sequences: a global reference library and a local reference library. Both reference libraries were created by performing an in-silico PCR with the *ecoPCR* software (Ficetola et al., 2010) using taxonomic information from the NCBI Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy). We ran ecoPCR with the *trnL g* and *h* primers, setting the product size to 10 - 150 bp and allowing a maximum of 3 errors per primer, but no errors on the last base of the 3' end.

The resulting global reference library is a primer-specific reference database based on the global EMBL database (release 137) and contains 111146 sequences of 18101 plant taxa. The local reference library is more specific and only contains DNA sequences of plant taxa known to occur in MM Hills and the surrounding area based on monitoring data (ATREE) and published species lists (appendix 1 from Harisha, 2015). The local reference library contains a total 555 sequences of 134 plant taxa and is constructed of sequences obtained from the EMBL database (release 137) plus an additional 34 species of locally occurring *poaceae* that were sequenced for this study (table S2).

Fresh leaves were collected from Male Mahadeshwara Hills Wildlife Sanctuary (MM Hills) and stored at -20 °C until DNA extraction. Total genomic DNA was extracted from leaf material using a cetyltrimethylammonium bromide (CTAB) method (Sambrook et al., 1989). DNA of the plant samples with high contents of various pigments were purified by Qiagen-DNeasy Plant Mini Kit following the manufacturer's protocols. Plant DNA metabarcoding was done using the trnL *c* and *d* primers from Taberlet et al. (1991). Plant DNA amplifications were carried out in a final volume of 50 µL containing 50–100 ng of template DNA and 0.2 µM of each primer. The amplification mixture further contained 1 U of Taq DNA Polymerase with 1X Buffer (Bangalore Genei, Bangalore, India) and 0.25 mM of each dNTP. The mixture was denatured at 95 °C for 5 min, followed by 35 cycles of 30 s at 95 °C, and 30 s at 54 °C, 30 s at 72 °C and a 10 min final elongation at 72 °C. PCR products were cleaned with a MinElute PCR Purification Kit (Qiagen, CA). Cycle sequencing reactions were performed in 10 µl reactions using 1 µl of BigDyeH Terminator cycle sequencing chemistry (v3.1; ABI; Warrington, Cheshire, UK) and run on ABI 3730XL sequencer (Applied Biosystems, USA). The qualities of generated sequences were estimated by assembling the original forward and reverse sequences using CodonCode Aligner 3.0 (CodonCode Co., USA).

Supporting references

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 434. https://doi.org/10.1186/1471-2164-11-434

Harisha, R. P., Padmavathy, S., & Nagaraja, B. C. (2015). Traditional Ecological Knowledge (TEK) and its Importance in South India: Perspecive from Local Communities. Applied Ecology and Environmental Research, 14(1), 311–326. https://doi.org/10.15666/aeer/1401_311326

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Taberlet,P., Gielly,L., Pautou,G. and Bouvet,J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol., 17, 1105–1109.

Table S2. The poaceae species sequenced for this study and their GenBank accession numbers.

Genus	Species name	*trn*L c,d sequence length	GenBank accession
Apluda	*Apluda mutica* L.	714	MT263035
Aristida	*Aristida funiculata* Trin. & Rupr.	801	MT263049
	Aristida setacea Retz.	811	MT263050
	Aristida adscensionis L.	808	MT263053
Bothriochloa	*Bothriochloa pertusa* (L.) A.Camus	707	MT263046
	Bothriochloa insculpta (Hochst. ex A.Rich.) A.Camus	742	MT263047
Cymbopogon	*Cymbopogon flexuosus* (Nees ex Steud.) W.Watson	670	MT263037
	Cymbopogon citratus (DC.) Stapf	645	MT263060
	Cymbopogon nardus (L.) Rendle	668	MT263061
	Cymbopogon commutatus (Steud.) Stapf	383	MT263062
	Cymbopogon goeringii (Steud.) A.Camus	654	MT263063
Cynodon	*Cynodon dactylon* (L.) Pers.	736	MT263039
	Cynodon nlemfuensis Vanderyst	446	MT263064
Dactyloctenium	*Dactyloctenium aegyptium* (L.) Willd.	760	MT263036
	Dactyloctenium radulans (R.Br.) P.Beauv.	605	MT263059
Dichanthium	*Dichanthium annulatum* (Forssk.) Stapf	712	MT263038
Digitaria	*Digitaria bicornis* (Lam.) Roem. & Schult.	793	MT263041
	Digitaria sanguinalis (L.) Scop.	788	MT263051
	Digitaria ischaemum (Schreb.) Muhl.	769	MT263052
Eleusine	*Eleusine indica* (L.) Gaertn.	816	MT263040
	Eleusine coracana (L.) Gaertn.	444	MT263065
Hackelochloa	*Hackelochloa granularis* (L.) Kuntze	771	MT263054
Heteropogon	*Heteropogon contortus* (L.) P.Beauv. ex Roem. & Schult.	700	MT263034
Megathyrsus	*Megathyrsus maximus* (Jacq.) B.K.Simon & S.W.L.Jacobs	818	MT263057
Melinis	*Melinis repens* (Willd.) Zizka	704	MT263048
Panicum	*Panicum virgatum* L.	737	MT263043
	Panicum maximum Jacq.	583	MT263056
	Panicum bulbosum Kunth	539	MT263058
Setaria	*Setaria pumila* (Poir.) Roem. & Schult.	754	MT263042
Sorghum	*Sorghum bicolor* (L.) Moench	768	MT263055
Sporobolus	*Sporobolus diandrus* (Retz.) P.Beauv.	512	MT263066
	Sporobolus blakei B.K.Simon	513	MT263067
Themeda	*Themeda triandra* Forssk.	716	MT263044
	Themeda arundinacea (Roxb.) A.Camus	701	MT263045

S3. Data processing steps in OBITools and R

Raw read data are available at the European Nucleotide Archive (ENA) under study accession number PRJEB41139: http://www.ebi.ac.uk/ena/data/view/PRJEB41139

A public GitHub repository with code and supporting files used to process the raw data is available here: https://github.com/terschure/dataprocessing_MMH

Table S3. Number of sequence reads and unique sequences remaining after each filtering step.

Filtering steps	Program/command	Total reads	Unique sequences
Raw reads		168 357 706	
Pairwise alignment	*illuminapairedend*	84 178 853	
Assignment to samples	*ngsfilter*	69 623 377	
Merged identical reads	*obiuniq* & *obiannotate*		968 157
Removal of reads with count =1 & < 8 bp length	*obigrep*	61 537 259	130 021
Identification & removal of PCR/sequencing errors	*obiclean* & R	55 868 956	22 164
Reduction of read counts of each sequence per sample (each cell in the MOTU table) by 0.001%	R	55 778 708	16 856
Removal of sequences with maximum abundance in negative controls & sequences with ≤ 95% match	R	52 375 610	3 801
Removal of unreliable PCR replicates (Euclidian distances between replicates ≥ non-replicates) & samples represented by < 2 replicates	R	48 428 300	3 725
Removal of samples only present once (Asian palm civet, sloth bear) & merging replicates (mean)	R	13 702 072	3 725
Removal of sequences that make up ≤ 1 % of the sample & taxa that are not in the region	R	12 509 596	134

S8. Niche width (Shannon) based on presence/absence data

Table S8.1. Overview of niche width described by average MOTU count and Shannon diversity index per herbivore species with standard errors for both measures. Feeding guild assignments are based on Nowak and Walker (1999), Ahrestani et al. (2012), Ahrestani and Sankaran (2016), and IUCN (2020).

Herbivore species	Latin name	MOTUs	(± SE)	Shannon	(± SE)
Cattle	*Bos taurus indicus*	6.22	(± 0.69)	1.67	(± 0.12)
Domestic goat	*Capra hircus*	11.75	(± 1.53)	2.41	(± 0.12)
Water buffalo	*Bubalus bubalis*	13.33	(± 2.73)	2.54	(± 0.23)
Asian elephant	*Elephas maximus*	8.40	(± 2.84)	1.83	(± 0.41)
Barking deer	*Muntiacus muntjak*	7.50	(± 2.50)	1.96	(± 0.35)
Bonnet macaque	*Macaca radiata*	4.83	(± 0.91)	1.49	(± 0.19)
Indian hare	*Lepus nigricollis*	7.67	(± 1.76)	1.98	(± 0.23)
Indian porcupine	*Hystrix indica*	8.75	(± 1.97)	2.05	(± 0.32)
Sambar deer	*Rusa unicolor*	12.40	(± 0.93)	2.51	(± 0.07)
Wild boar	*Sus scrofa*	9.33	(± 1.20)	2.22	(± 0.14)

Table S8.2. Kruskal-Wallis rank sum test of differences in Shannon diversity for the presence/absence data (chi-squared = 23.044, df = 9, p-value = 0.006098). No significant differences are found when looking at pairwise comparisons using the Holm adjustment of p-values. The upper number for each comparison is Dunn's pairwise z test statistic, the bottom number is the adjusted p-value. alpha = 0.05 Reject H_0 if p <= alpha/2.

	Asian_el	Barking_	Bonnet_m	Cattle	Domestic	Indian_h	Indian_p	Sambar_d	Water_bu
Barking_	0.163074								
	1.0000								
Bonnet_m	1.236957	0.750250							
	1.0000	1.0000							
Cattle	0.856750	0.388374	-0.711720						
	1.0000	1.0000	1.0000						
Domestic	-1.39689	-1.17989	-2.861461	-2.97007					
	1.0000	1.0000	0.0886	0.0640					
Indian_h	0.053378	-0.10675	-1.00413	-0.62518	1.233868				
	0.4787	1.0000	1.0000	1.0000	1.0000				
Indian_p	-0.35696	-0.43405	-1.531342	-1.22238	0.909395	-0.3645			
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000			
Sambar_d	-1.78745	-1.51426	-3.103889	-3.14779	-0.58610	-1.6013	-1.3282		
	1.0000	1.0000	0.0420	0.0370	1.0000	1.0000	1.0000		
Water_bu	-1.48443	-1.33700	-2.592382	-2.45471	-0.42499	-1.3754	-1.1058	0.063545	
	1.0000	1.0000	0.1954	0.2820	1.0000	1.0000	1.0000	0.9493	
Wild_boa	-0.51853	-0.56428	-1.594807	-1.30558	0.616934	-0.5115	-0.1822	1.029443	0.863925
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

S9. NMDS of Bray-Curtis dissimilarities based on presence/absence data

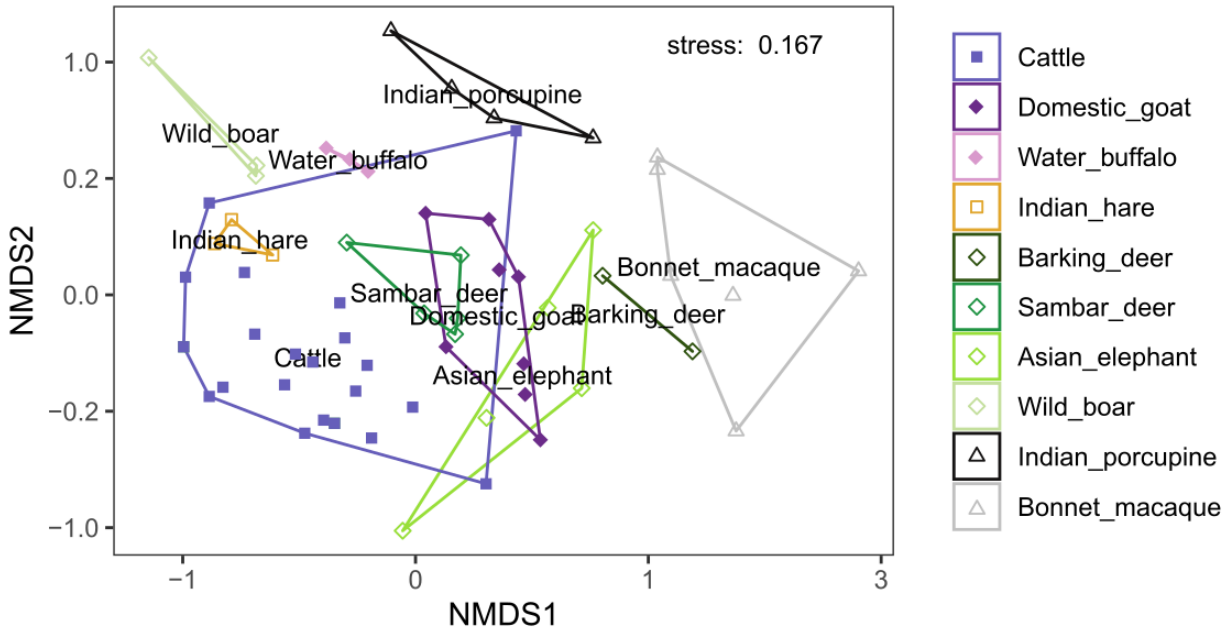


Figure S9: NMDS plot based on the presence/absence data per faecal sample using Bray-Curtis dissimilarity (adonis $F_{9,52} = 4.31$, $R2 = 0.42$, $P \leq 0.001$). The positioning of the species label indicates the mean for that species. Samples from domestic herbivore species are indicated with filled symbols and the shapes of the symbols refer to the different feeding guilds: grazer (square), mixed feeder (diamond), and frugivore (triangle).

S10. Bray-Curtis dissimilarity and Pianka's niche overlap values based on presence/absence data

Table S10. Bray-Curtis dissimilarity (bottom left; 0:similar 1:dissimilar) and Pianka's overlap index (top right; 0:no overlap 1:full overlap) based on presence/absence data. Asterisks indicate statistically significant niche overlap (i.e., greater than expected by chance based on comparison with 1000 null models, α = 0.05).

	Cattle	Domestic goat	Water buffalo	Asian elephant	Barking deer	Bonnet macaque	Indian hare	Indian porcupine	Sambar deer	Wild boar
Cattle		0.34*	0.42*	0.27*	0.09	0.06	0.53*	0.12	0.52*	0.28*
Domestic goat	0.78		0.23*	0.47*	0.30*	0.28*	0.23*	0.30*	0.64*	0.03
Water buffalo	0.78	0.82		0.16	0.02	0.15	0.27*	0.13	0.26*	0.41*
Asian elephant	0.81	0.67	0.90		0.32*	0.27*	0.09	0.24*	0.31*	0.00
Barking deer	0.94	0.75	0.97	0.76		0.41*	0.00	0.20	0.23*	0.00
Bonnet macaque	0.93	0.78	0.93	0.79	0.73		0.00	0.12	0.26*	0.06
Indian hare	0.70	0.84	0.81	0.89	1.00	1.00		0.15	0.36*	0.48*
Indian porcupine	0.87	0.74	0.90	0.79	0.78	0.90	0.91		0.25*	0.05
Sambar deer	0.72	0.54	0.79	0.73	0.80	0.88	0.73	0.76		0.22*
Wild boar	0.85	0.98	0.65	1.00	1.00	0.95	0.65	0.97	0.85	

S11. Dietary difference among species (adonis)

Table S11. Results of pairwise adonis tests based on Bray-Curtis dissimilarities for the RRA and the presence/absence data with and without Bonferroni adjusted p-values.

	Relative read abundance data				Presence/absence data			
pairs	F Model	R2	p-value	p-adj	F Model	R2	p-value	p-adj
Cattle vs Indian_hare	2.12	0.0812	0.058	1.000	2.50	0.0945	0.005	0.225
Cattle vs Sambar_deer	3.02	0.1040	0.007	0.315	3.97	0.1326	0.001	0.045*
Cattle vs Bonnet_macaque	7.78	0.2238	0.001	0.045*	7.59	0.2195	0.001	0.045*
Cattle vs Domestic_goat	6.82	0.1904	0.001	0.045*	6.73	0.1884	0.001	0.045*
Cattle vs Barking_deer	3.37	0.1277	0.010	0.450	4.20	0.1545	0.003	0.135
Cattle vs Water_buffalo	3.24	0.1188	0.003	0.135	3.11	0.1146	0.006	0.270
Cattle vs Indian_porcupine	4.20	0.1439	0.002	0.090	4.78	0.1605	0.001	0.045*
Cattle vs Asian_elephant	3.77	0.1267	0.005	0.225	5.15	0.1654	0.001	0.045*
Cattle vs Wild_boar	5.54	0.1877	0.001	0.045*	4.70	0.1638	0.001	0.045*
Indian_hare vs Sambar_deer	2.24	0.2718	0.041	1.000	4.60	0.4341	0.015	0.675
Indian_hare vs Bonnet_macaque	4.28	0.3797	0.012	0.540	4.79	0.4062	0.012	0.540
Indian_hare vs Domestic_goat	4.04	0.3098	0.007	0.315	4.36	0.3262	0.008	0.360
Indian_hare vs Barking_deer	2.91	0.4924	0.100	1.000	7.14	0.7042	0.100	1.000
Indian_hare vs Water_buffalo	2.69	0.4020	0.100	1.000	4.01	0.5007	0.100	1.000
Indian_hare vs Indian_porcupine	2.04	0.2893	0.035	1.000	3.10	0.3828	0.034	1.000
Indian_hare vs Asian_elephant	4.13	0.4078	0.051	1.000	5.42	0.4747	0.018	0.810
Indian_hare vs Wild_boar	3.54	0.4698	0.100	1.000	3.45	0.4632	0.100	1.000
Sambar_deer vs Bonnet_macaque	4.40	0.3284	0.006	0.270	4.64	0.3404	0.003	0.135
Sambar_deer vs Domestic_goat	2.36	0.1769	0.012	0.540	1.94	0.1500	0.033	1.000
Sambar_deer vs Barking_deer	2.72	0.3525	0.046	1.000	5.38	0.5183	0.047	1.000
Sambar_deer vs Water_buffalo	2.47	0.2914	0.018	0.810	4.72	0.4403	0.015	0.675
Sambar_deer vs Indian_porcupine	2.45	0.2590	0.006	0.270	3.46	0.3307	0.006	0.270
Sambar_deer vs Asian_elephant	3.98	0.3321	0.004	0.180	4.90	0.3799	0.012	0.540
Sambar_deer vs Wild_boar	4.86	0.4474	0.019	0.855	7.25	0.5471	0.026	1.000
Bonnet_macaque vs Domestic_goat	5.06	0.2965	0.001	0.045*	3.94	0.2473	0.002	0.090
Bonnet_macaque vs Barking_deer	2.05	0.2548	0.068	1.000	1.64	0.2149	0.214	1.000
Bonnet_macaque vs Water_buffalo	3.44	0.3298	0.011	0.495	3.63	0.3416	0.007	0.315
Bonnet_macaque vs Indian_porcupine	2.60	0.2451	0.021	0.945	2.85	0.2630	0.005	0.225
Bonnet_macaque vs Asian_elephant	4.32	0.3243	0.011	0.495	3.61	0.2862	0.005	0.225
Bonnet_macaque vs Wild_boar	5.00	0.4165	0.009	0.405	5.02	0.4176	0.011	0.495
Domestic_goat vs Barking_deer	2.54	0.2410	0.022	0.990	3.13	0.2814	0.030	1.000
Domestic_goat vs Water_buffalo	3.40	0.2742	0.005	0.225	3.90	0.3022	0.007	0.315
Domestic_goat vs Indian_porcupine	2.62	0.2078	0.002	0.090	2.68	0.2116	0.002	0.090
Domestic_goat vs Asian_elephant	2.95	0.2115	0.016	0.720	2.86	0.2063	0.004	0.180
Domestic_goat vs Wild_boar	5.24	0.3680	0.007	0.315	6.42	0.4164	0.007	0.315
Barking_deer vs Water_buffalo	2.28	0.4318	0.100	1.000	5.56	0.6497	0.100	1.000
Barking_deer vs Indian_porcupine	1.05	0.2077	0.467	1.000	2.08	0.3422	0.067	1.000
Barking_deer vs Asian_elephant	2.35	0.3195	0.051	1.000	3.06	0.3796	0.047	1.000
Barking_deer vs Wild_boar	3.93	0.5670	0.100	1.000	10.35	0.7753	0.100	1.000
Water_buffalo vs Indian_porcupine	1.88	0.2728	0.074	1.000	2.72	0.3519	0.067	1.000
Water_buffalo vs Asian_elephant	2.92	0.3271	0.033	1.000	4.49	0.4280	0.015	0.675
Water_buffalo vs Wild_boar	3.47	0.4643	0.100	1.000	3.80	0.4871	0.100	1.000
Indian_porcupine vs Asian_elephant	2.38	0.2533	0.063	1.000	2.67	0.2762	0.020	0.900
Indian_porcupine vs Wild_boar	2.92	0.3690	0.028	1.000	4.09	0.4499	0.037	1.000
Asian_elephant vs Wild_boar	5.17	0.4627	0.014	0.630	7.05	0.5402	0.020	0.900

S12. Pianka's niche overlap models

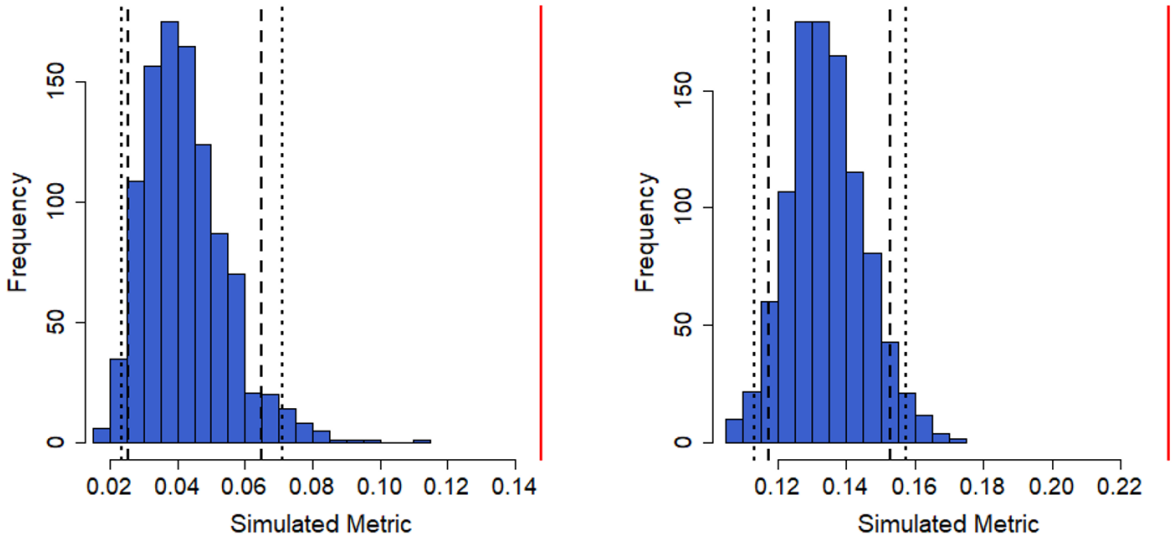


Figure S12.1: Histogram of simulated Pianka values based on all RRA data (left; 95% CI [0.024, 0.073]) and all presence/absence data (right; 95% CI [0.115, 0.157]). The blue bars are the results from the null model created with the ra3 algorithm and 1000 replications (EcoSimR) while the red line indicates the observed Pianka value for all herbivores species combined.

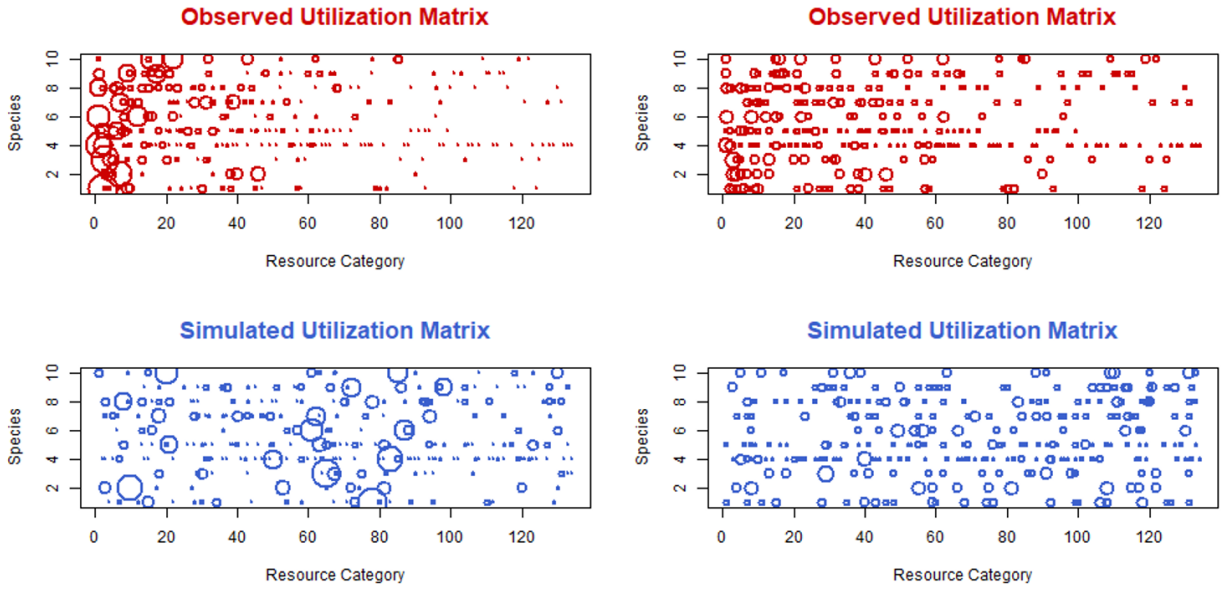
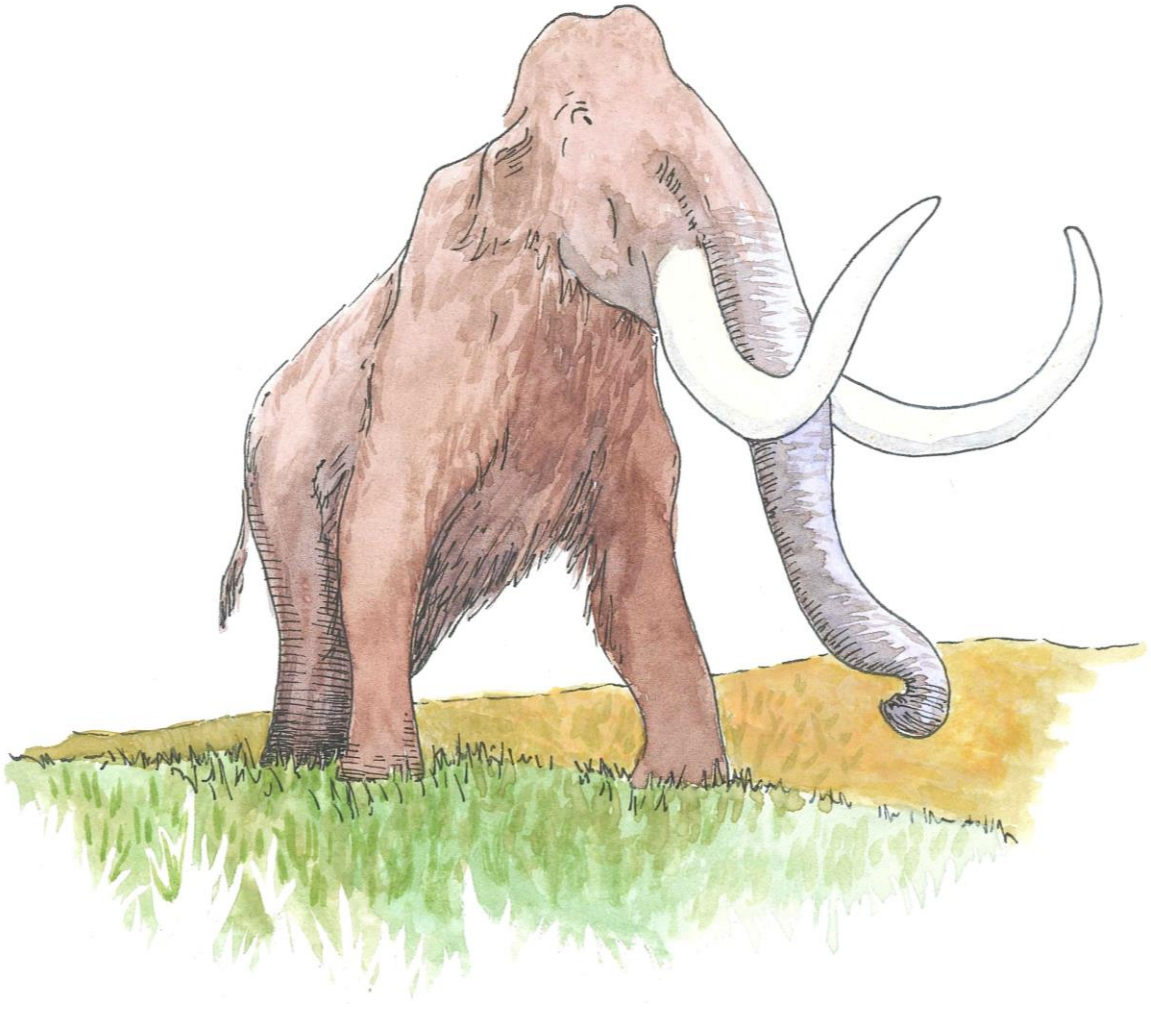


Figure S12.2: Utilization matrices based on RRA data (left) and the occurrence data (right) of observed (red) and simulated (blue) values from a null model created with the ra3 algorithm and 1000 replications (EcoSimR). The area of each circle depicted is proportional to the utilization of a resource category by a species. If no circle is shown, the utilization was 0.



Chapter 3

Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna

Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna

Marcel Polling, Anneke T.M. ter Schure, Bas van Geel, Tom van Bokhoven, Sanne Boessenkool, Glen MacKaye, Bram W. Langeveld, María Arizab, Hans van der Plicht, Albert V. Protopopov, Alexei Tikhonov, Hugo de Boer, Barbara Gravendeel

Published in Quaternary Science Reviews 267, 107084, 2021

Highlights

- The first integrated analysis of DNA, pollen and macrofossils from permafrost faeces.
- Successful amplification of up to 28,6 kyr old DNA using long, plant-specific nrITS.
- High taxonomic resolutions allow detailed insights in extinct megafaunal habitat.
- Macrofossils and DNA show diverse woolly mammoth diet and use of ‘mammoth steppe'.

Abstract

The study of faecal samples to reconstruct the diets and habitats of extinct megafauna has traditionally relied on pollen and macrofossil analysis. DNA metabarcoding has emerged as a valuable tool to complement and refine these proxies. While published studies have compared the results of these three proxies for sediments, this comparison is currently lacking for permafrost preserved mammal faeces. Moreover, most metabarcoding studies have focused on a single plant-specific DNA marker region. In this study, we target both the commonly used chloroplast *trn*L P6 loop as well as nuclear ribosomal ITS (nrITS). The latter can increase taxonomic resolution of plant identifications but requires DNA to be relatively well preserved because of the target length (~300–500 bp). We compare DNA results to pollen and macrofossil analyses from permafrost and ice-preserved faeces of Pleistocene and Holocene megafauna. Samples include woolly mammoth, horse, steppe bison as well as Holocene and extant caribou. Most plant identifications were found using DNA, likely because the studied faeces contained many vegetative remains that could not be identified using macrofossils or pollen. Several taxa were, however, identified to lower taxonomic levels uniquely with macrofossil and pollen analysis. The nrITS marker provides species level taxonomic resolution for commonly encountered plant families that are hard to distinguish using the other proxies (e.g. Asteraceae, Cyperaceae and Poaceae).

Integrating the results from all proxies, we are able to accurately reconstruct known diets and habitats of the extant caribou. Applying this approach to the extinct mammals, we find that the Holocene horse and steppe bison were not strict grazers but mixed feeders living in a marshy wetland environment. The mammoths showed highly varying diets from different non-analogous habitats. This confirms the presence of a mosaic of habitats in the Pleistocene 'mammoth steppe' that mammoths could fully exploit due to their flexibility in food choice.

1. Introduction

During much of the Late Pleistocene epoch, Siberia, Alaska and northern Canada were connected, forming a dry and largely treeless landmass known as Beringia (Hopkins, 1959; Hopkins et al., 1982). The landscape was dominated by emblematic megafauna such as woolly mammoths and steppe bison, and in terms of biomass some authors have compared this period to the current African savannah (Zimov et al., 2012). Mammals had a major role in shaping vegetation community and structure by reducing vegetation density, enhancing nutrient turnover, dispersing seeds and reducing fire potential (Guthrie, 2001; Hester et al., 2006; Johnson, 2009). Reconstructing the species composition of this former plant community without a modern analogue, as well as the corresponding diets of the mammals that roamed it has been challenging.

According to Guthrie (1990) there were mainly open landscapes with highly productive graminoids and Artemisia sp. in a steppe-tundra biome that is often designated the 'mammoth steppe'. Recent studies have changed the view of the mammoth steppe vegetation into a more heterogeneous mosaic of different habitats. This mosaic consisted of areas rich in shrubs combined with permanent moist areas and productive grasslands (Chytrý et al., 2019; Lozhkin et al., 2019; Zazula et al., 2006). Willerslev et al. (2014) further showed that forbs (non-graminoid herbaceous vascular plants) were more abundant in the environment than previously thought, and featured in megafaunal diets to provide important proteins. Relatively little is known, however, about the specific plant species in megafaunal diets.

The shift in appreciation of the Beringian megafaunal habitats has been catalysed by a growing body of research that uses a multidisciplinary approach, combining pollen and plant macrofossils with DNA metabarcoding (Boast et al., 2018; Gravendeel et al., 2014; Haarsma et al., 2016; Hofreiter et al., 2000; Sønstebø et al., 2010; van Geel et al., 2008, 2011a, 2011b; Van Geel et al., 2014; Willerslev et al., 2014). By improving taxonomic resolution and finding complementary taxa, DNA metabarcoding can help to resolve vegetation classifications where species resolution is required (e.g. steppe and tundra, partly defined on distinct species of grass; Swanson, 2006). Several studies on lake sediments have shown that instead of replacing traditional methods, DNA metabarcoding acts as a complementary proxy by revealing both additional taxa and providing increased taxonomic

resolution (see e.g. Boessenkool et al., 2014; Parducci et al., 2019; Pedersen et al., 2013; Rawlence et al., 2014). While pollen grains mostly show a regional signal due to dominant wind-dispersed pollen (grasses and Artemisia sp.), DNA may represent a more local signal that is more similar to the spectrum of macrofossil taxa (Alsos et al., 2018; Boessenkool et al., 2014; Jorgensen et al., 2012).

While the studies cited above provide a good overview of the advantages and drawbacks of the different proxies used, all of these studies focussed on lake sediments. So far, there are few studies comparing these proxies in megafaunal faecal samples (e.g. Gravendeel et al., 2014; Hofreiter et al., 2003; van Geel et al., 2008). Strictly speaking, the faecal samples of extinct megafauna are not coprolites since they are not fossilized but perfectly preserved in permafrost. However, the plant macrofossils in these samples are drastically affected by masticatory and digestive processes, which may result in differential preservation of taxa and fragments becoming unidentifiable (van Geel et al., 2008). For pollen recovered from faeces an additional complicating factor is that the faecal samples are often dominated by wind-transported pollen or pollen deriving from ingestion of inflorescences from plants that were flowering at the time of consumption (Van Geel et al., 2014). The advantage of DNA as a proxy for dietary reconstruction is that it does not depend on flowering time or time of fruit setting, as vegetative plant remains are included in the DNA record (Willerslev et al., 2014). However, as in ancient sediments, not all taxa are recorded using DNA metabarcoding due to incomplete reference libraries, PCR bias, primer mismatches and DNA degradation (Jorgensen et al., 2012).

Most studies of ancient DNA from sediments have relied either on the P6 loop of the chloroplast *trn*L (UAA) intron or the *rbc*L gene, and both give good taxonomic resolution for some plant taxa but limited for others (Sønstebø et al., 2010; Taberlet et al., 2006). While in the animal kingdom the mitochondrial marker COI can be used as a universal barcode for identifying species (Hebert et al., 2003) no such universal barcode has been identified for plants. For this reason a combination of markers has been advised for plants, including both a nuclear marker and a plastid marker (CBOL Plant Working Group et al., 2011). Since permafrost acts as an excellent natural freezer, even long DNA fragments (up to 510 bp) have been recovered from sediments as old as 400 kyr (Lydolph et al., 2005; Willerslev et al., 2014). Yet in the study of ancient megafaunal faeces, the relatively long nuclear ribosomal ITS (nrITS) has rarely been used, and only to amplify relatively short amplicons (e.g. 240 bp in the Cape Blossom mammoth; van Geel et al., 2011b). Due to its length, nrITS has the advantage of being able to provide a higher taxonomic resolution, which in turn can give better insight into the palaeoenvironmental conditions represented by the taxa in a sample.

In this study, we aim to 1) investigate the potential of using the nrITS marker on megafaunal faeces, 2) compare the nrITS results to *trn*L, pollen and macrofossil records and 3) integrate results of all proxies to obtain a detailed reconstruction of ancient megafaunal diets and habitats. To this end, we applied

DNA metabarcoding, pollen and macrofossil analysis on a variety of permafrost and ice-preserved faecal samples from extinct and extant megafauna, specifically woolly mammoth, steppe bison, horse and caribou. In addition to the *trn*L P6 loop, we target the nrITS regions nrITS1 and nrITS2. The wide temporal range of the samples (28,000 to modern) further allows us to capture potential taphonomic effects on the recovery of the different marker regions and read counts, while inclusion of faecal samples from extant caribou with known diets and habitats enables validation of the diet and habitat reconstructions of the extinct megafauna.

2. Materials and methods

2.1. Material

Eleven faecal samples from four mammal species were included (Table 1; for detailed information about location and dating see Table S1). Several of the samples we used here have been studied previously and DNA from the original material - which was stored at −80 °C - was re-extracted and analysed here, except for the Oyogas Yar horse and Yakutian bison of which DNA extracts from previous studies were used (CTAB DNA extraction; Doyle and Doyle, 1987). All samples are derived from Russia, Canada and USA (Fig. 1) and are briefly discussed below.

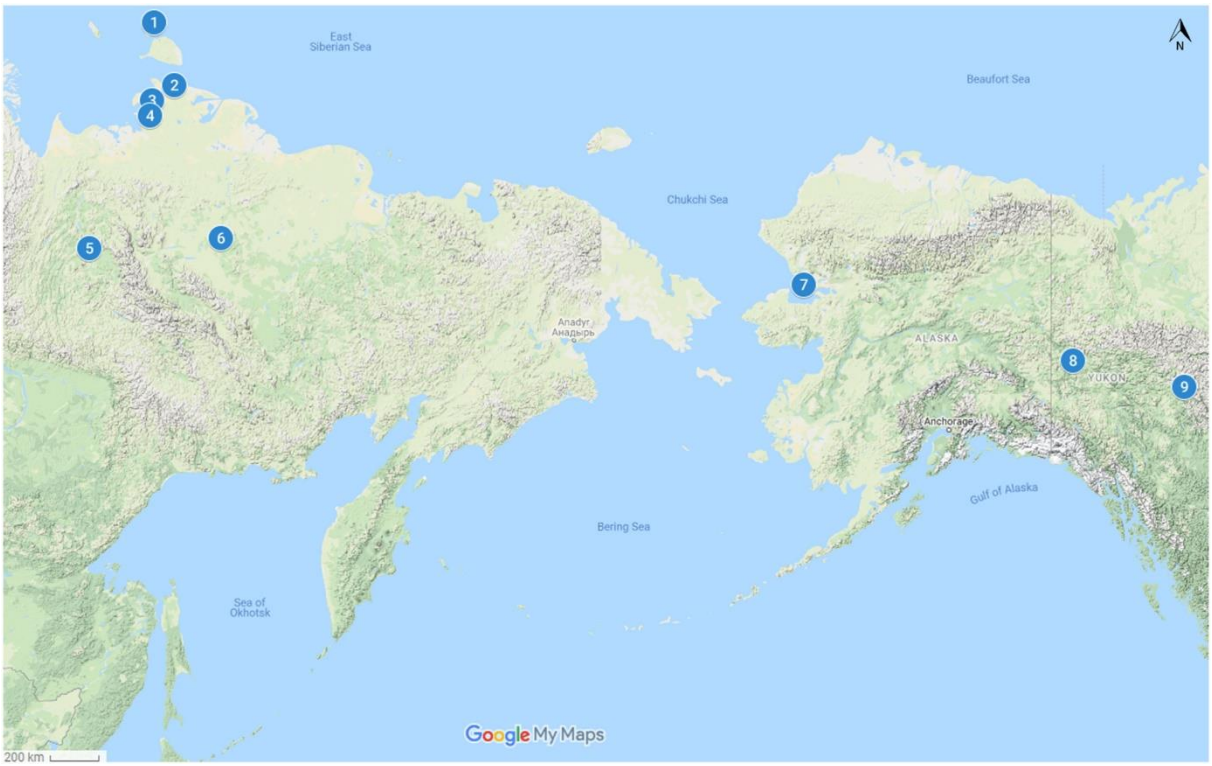


Figure 1. Sample localities. (1) Maly Lyakhovsky mammoth, (2) Oyogas Yar horse, (3) Yakutian bison, (4) Yukagir mammoth, (5) Adycha mammoth, (6) Abyland mammoth, (7) Cape Blossom mammoth (8) Yukon horse and (9) Selwyn caribou A, B and C.

2.1.1. Holocene and modern mountain caribou

Three northern mountain caribou (Rangifer tarandus caribou (Gmelin, 1788)) faecal samples were collected from cores in ice patch deposits in the Selwyn Mountains, Northwest Territories, Canada. Caribou visit these ice patches during the summer months to escape summer heat and insect harassment and their faeces are subsequently buried by snow creating stratigraphically discrete faecal bands that are very well preserved. The samples include faeces from modern caribou collected from the surface near the ice patch (Selwyn A), and two samples of late Holocene age collected from the ice core, Selwyn B and Selwyn C. From Selwyn A, DNA was retrieved by Galloway et al. (2012) confirming that caribou was indeed the producer of the faeces. For the other samples, the faecal material was identified as being deposited by caribou based on the general shape, size and texture of the pellets, without additional DNA confirmation.

2.1.2. Holocene bison and horse

A colon sample of a horse (Oyogas Yar or Yukagir horse; Equus cf. lenensis Russanov, 1968) of middle Holocene age and a rumen sample of a Yakutian steppe bison (Bison priscus (Bojanus, 1825)) of early Holocene age were taken directly from permafrost preserved animals from the Sakha Republic, Russia (Boeskorov et al., 2014; Gravendeel et al., 2014; Van Geel et al., 2014) (Table 1). The Oyogas Yar horse was identified as being most closely related to the extinct Lena horse, Equus lenensis, based on body size measurements (Boeskorov et al., 2018).

2.1.3. Pleistocene mammoth and horse

Six Pleistocene faecal samples were analysed, including five woolly mammoths (*Mammuthus primigenius* (Blumenbach, 1799)) and one Yukon horse (*Equus lambei* (Hay, 1917)). Four specimens were obtained from the republic of Sakha (Yakutia), Russia, including the Maly Lyakhovsky, Abyland, Adycha and Yukagir mammoths. The Cape Blossom mammoth sample (or Alaskan Late Glacial mammoth) was obtained from Cape Blossom, Alaska, USA, and the Yukon horse was obtained from Dawson City, Yukon, Canada. Faecal samples were taken directly from, or in close vicinity to the permafrost preserved animals, except for the Abyland, Adycha and Cape Blossom samples which were loose faeces. Validation of the faeces as being derived from woolly mammoth for the Yukagir, Maly Lyakhovsky, Cape Blossom and Yukon samples is based on previous studies (Grigoriev et al., 2017; Harington and Eggleston-Stott, 1996; van Geel et al., 2008, 2011b). The identities of the Adycha and Abyland samples were confirmed using Sanger DNA analyses (Supplementary Text S2).

Table 1. Overview of the samples used in this study including the existing and newly generated data, source of material and their age and collection locality. References from where the existing data was taken are [1] Galloway et al. (2012) [2] Boeskorov et al. (2014) [3] Gravendeel et al. (2014) [4] Van Geel et al. (2014) [5] van Geel et al. (2011b) [6] van Geel et al. (2008) [7] Harington and Eggleston-Stott (1996). ∗D = DNA, M = plant macrofossils, P = pollen. †DNA extract from previous study used.

Species	Name	Reference	Existing data∗	Newly generated data∗	Material	measured ^{14}C age BP	Locality
Caribou	Selwyn A (KfTe-1 surface)	[1]	P	D M	Faeces from ice patch	modern	Selwyn Mountains, NT, Canada
Caribou	Selwyn B (KfTe-1-C2-1)	[1]	M P	D	Faeces from ice patch	1,630 ± 40	Selwyn Mountains, NT, Canada
Caribou	Selwyn C (KfTe-1-C1-3)	[1]	M P	D	Faeces from ice patch	2,840 ± 40	Selwyn Mountains, NT, Canada
Horse	Oyogas Yar	[2,3]	D M P	D†	Faeces from colon	4,630 ± 35	N Sakha, Ust-Yana region, Russia
Bison	Yakutian	[2,4]	D M P	D†	Rumen	9,310 ± 45 9,295 ± 45	N Sakha, Chukchalakh Lake, Yana Mammoth reserve
Woolly mammoth	Cape Blossom	[5]	D M P	D	Faeces	12,300 ± 70	Kotzebue Sound, NW Alaska, USA
Woolly mammoth	Yukagir	[6]	D M P	D	Faeces from colon	18,680 ± 100	N Sakha, oxbow lake near Maxunuokha River, Russia
Woolly mammoth	Adycha	This study	–	D M P	Faeces	21,250 ± 100	N Sakha, Adycha River floodplain, Russia
Horse	Yukon	[7]	D M	D P	Faeces from intestine	26,280 ± 210	Last Chance Creek near Dawson City, Yukon, Canada
Woolly mammoth	Abyland	This study	–	D M P	Faeces	28,460 ± 160	N Sakha, Oguruoha River, Abyysky District, Russia
Woolly mammoth	Maly Lyakhovsky	This study	–	D M P	Faeces from stomach	28,610 ± 110	N Sakha, Maly Lyakhovsky Island, Russia

2.2. Radiocarbon dating

Radiocarbon dates of the caribou, horse, bison and Cape Blossom and Yukagir mammoth faeces were reported in previous publications (Boeskorov et al., 2014; Galloway et al., 2012; Gravendeel et al., 2014; Harington and Eggleston-Stott, 1996; van Geel et al., 2008, 2011b). The faecal samples of the Adycha, Abyland and Maly Lyakhovsky mammoths were dated at the AMS facility of the Center for Isotope Research of the University of Groningen (The Netherlands). The 14C ages are reported in BP, the conventional unit, and includes a correction for isotope fractionation and a defined half-life (Van der Plicht and Hogg, 2006). The 14C dates are calibrated into calendar ages using the presently recommended calibration curve IntCal20 (Reimer et al., 2020). The calibrated dates are reported in cal. BP, defined as calendar years relative to AD 1950 (Table S1).

2.3. Pollen and macrofossils

If available, pollen and macrofossil results were taken directly from published records (Table 1). Data was available for the Yukagir and Cape Blossom mammoths, the Yakutian bison, Oyogas Yar horse and two of the Selwyn caribou samples (Galloway et al., 2012; Gravendeel et al., 2014; van Geel et al., 2008, 2011b; Van Geel et al., 2014). For Selwyn caribou A, only a pollen analysis was available (Galloway et al., 2012). If multiple counts were present from different subsamples, these were averaged to obtain one pollen count per sample. Macrofossil results for the Yukon horse were generated by Paleotec Services, Canada. This sample was previously studied for its plant DNA using trnL by Willerslev et al. (2014).

Pollen and spores (hereafter 'pollen') counts and macrofossil analysis were performed for the faeces of the Abyland, Adycha and Maly Lyakhovsky mammoths, Yukon horse (only pollen) and Selwyn caribou A (only macrofossil). The method for pollen preparation followed Faegri and Iversen (1989). Samples for pollen and macrofossil analyses were taken from the core of the faeces. Microscopic analysis of pollen was done at 400X and 1000X magnification. Pollen identifications were based on Moore et al. (1991) and Beug (2004) and a pollen reference collection. For the preparation of macrofossils, Mauquoy and Van Geel (2007) was followed. Bryophyte specimens were identified using Lawton (1971), Crum et al. (1981) and Vitt and Buck (1992).

In pollen analysis, the use of 'types' is common to denote a group of taxa that produce pollen that cannot be identified to a lower taxonomic level using microscopic analysis. *Potentilla*-type pollen for example includes pollen from species of the genera *Potentilla, Comarum, Fragaria* and *Sibbaldia* (Reitsma, 1966), which are all part of the subtribe Fragariinae of the Rosaceae family. All 'type' identifications were therefore converted to their corresponding maximum taxonomic level so as to better compare them to the DNA and macrofossil data. Similarly, the commonly used Asteraceae pollen subdivision Tubuliflorae and

Liguliflorae were converted to Asteraceae subfamilies Asteroideae and Cichorioideae, respectively.

2.4. Molecular analysis: DNA extractions and primer selection

2.4.1. Molecular analysis: DNA extractions

All pre-PCR aDNA work (including subsampling) took place in the dedicated ancient DNA laboratory of Naturalis Biodiversity Center (Leiden, The Netherlands). We subsampled the faecal samples following recommendations of Cooper and Poinar (2000) and Wood and Wilmshurst (2016). Samples were UVC-irradiated for 5 min and the outer layer (±2 mm) removed with a clean scalpel. This process was repeated before taking three subsamples (±100 mg each) from the middle of the bisected samples.

The subsamples were ground in a Retsch CryoMill at −196 °C, before DNA was extracted separately for each subsample following the silica-based extraction protocol of Rohland and Hofreiter (2007), adjusted to the smaller volume of material used as described in Stech et al. (2011). DNA extracts from the three subsamples were then pooled together. To control cross-contamination, DNA extractions were carried out in batches of two to three samples with one extraction blank (excluding faecal material) included in each batch (in total five extraction blanks).

2.4.2. Molecular analysis: primer selection and DNA amplification

Amplification of chloroplast DNA was done using *trn*L intron P6 loop g and h primers (Taberlet et al., 2006) (Table S3). Nuclear ribosomal Internal Transcribed Spacer regions were amplified using plant-specific primer pairs for nrITS1 (ITS-p5/ITS-u2; Cheng et al., 2016) and nrITS2 (ITS-p3/ITS4; Cheng et al., 2016; White et al., 1990) as well as fungi-specific primer pair for nrITS2 (fITS7/ITS4; Ihrmark et al., 2012; White et al., 1990) to control for amplification of non-target DNA (Table S3).

A dual-indexing approach was applied using a set of unique primer-adapter combinations as described in Fadrosh et al. (2014). All DNA extracts were diluted 1:10, except for the Abyland and Cape Blossom mammoths, for which a 1:50 dilution was used. PCRs were carried out on a Bio-Rad C1000 Touch or Bio-Rad S1000 thermal cycler in 25 µl final volumes consisting of 15.4 µl nuclease-free ultrapure water, 1x Phire Green reaction buffer, 0.52 µM of each primer, 1.25 mM of dNTPs, 1 U Phire Hot Start II DNA Polymerase and 1 µl of the 1:10 or 1:50 diluted DNA sample template. Gradient PCR results were used to determine the optimum annealing temperature for each primer set. The following amplification protocol was used: a 30 s activation step at 98 °C, 40 cycles including 5 s at 98 °C, 5 s annealing at 55–60 °C (depending on primers used; Table S3) and 15 s elongation at 72 °C, plus a final extension step at 72 °C for 5 min.

In order to mitigate stochasticity of DNA results, three PCR replicates were used for all samples using a unique tag combination for each replicate. Coelogyne fimbriata (Orchidaceae), native to tropical SE Asia, was used as a positive control for each primer set. The resulting PCR products were pooled into two pools based on amplicon length: a pool containing the shorter trnL fragments and a pool containing the longer nrITS fragments. Equimolar pools were made after measuring DNA concentrations on a QIAXcel (Qiagen). The pools were purified using Agencourt AMPure XP beads (Beckman Coulter), with a 1:0.9 (nrITS) or 1:1 (trnL) ratio and quantified using an Agilent 2100 Bioanalyzer DNA High sensitivity chip. Illumina adapters were ligated onto the amplicons using TruSeq DNA Nano Library Preparation kit (Illumina, USA) and subsequently sequenced at the Norwegian Sequencing Center on an Illumina MiSeq v2 300 cycles (150 bp x 2) for the trnL fragments and an Illumina MiSeq v3 600 cycles (300 bp x 2) for the nrITS fragments.

2.5. Molecular analysis: DNA sequence analysis and filtering

2.5.1. Mammal DNA identification

The mitochondrial Sanger sequencing reads obtained from the Abyland and Adycha faeces were aligned and trimmed using BioEdit version 7.2.5 (Supplementary Text S2; Hall, 1999). A MegaBLAST search was performed to check the resulting consensus sequences against the NCBI nucleotide database, and only sequences resulting in percentage ID > 98% were kept.

2.5.2. NrITS sequences

The three pools of nrITS sequences (plant nrITS1 and nrITS2, fungal nrITS2) were analysed separately with a custom pipeline on the OpenStack environment of Naturalis Biodiversity Center through a Galaxy instance (Afgan et al., 2018). Paired-end reads were first merged with PEAR (Zhang et al., 2014) using the standard settings and discarding non-merged reads. Amplicons were subsequently demultiplexed using the linked adapters option in Cutadapt version 2.8 (Martin, 2011). Only sequences containing both unique sample tags and forward and reverse primers were kept. Primer sequences were subsequently removed from the sequences with Cutadapt, allowing a maximum error rate of 0.15 (i.e. 3 to 4 bases).

The sequences were quality filtered and trimmed using the PRINSEQ sequence filter/converter tool (Schmieder and Edwards, 2011), using a minimum mean quality score of 20 and removing any sequences shorter than 150 bp. Sequences were dereplicated and sorted by size in VSEARCH v2.14.2 (Rognes et al., 2016) and clustered into Operational Taxonomic Units (OTUs) using the unoise3 algorithm from USEARCH v11.0.667 (Edgar, 2016) with default settings, removing singletons and potential chimeras. OTUs were subsequently identified using a MegaBLAST search against the NCBI Genbank nucleotide database for

plant nrITS1 and nrITS2 (Benson et al., 2012), and the UNITE fungal nucleotide database for fungal nrITS2 (Nilsson et al., 2019). OTUs that matched at least 80% in coverage as well as identity to NCBI Genbank were kept. For final taxon identifications, a minimum of 80% identity recognition for family, 90% identity for genus and 97% for the species level was used. Sequences were further filtered in R (version 3.5.2) (R Core Team, 2013) to remove sequences with a lower number of reads from any of the samples than in negative controls (either extraction or PCR). This resulted in removal of suspected food contaminants including *Pisum sativum*, *Brassica rapa/napus* for nrITS1 and *Citrus* sp., *Cucumis sativus* and *Musa* sp. for plant nrITS2. For plant nrITS1 and nrITS2, the positive control was successfully amplified and the presence of *Coelogyne fimbriata* reads in the non-control samples was used to determine an OTU filtering threshold to correct for potential leakage. For nrITS2, this resulted in reduction of each sequence read count per replicate with 0.3%, while this value was 0.35% for nrITS1 and fungal nrITS2 (see Table S5.1 for full steps and read counts). Remaining replicates were merged while averaging the read counts per OTU. Finally, OTUs at species or genus level with the same taxonomic assignment were aggregated.

A curated arctic and boreal vascular plant and bryophyte database exists for trnL (see below), but not yet for nrITS. The plant nrITS results have therefore been carefully checked for their presence in the geographical areas where the faeces were collected. To this end, the Panarctic Flora (Elven et al., 2011), database of vascular plants of Canada (VASCAN) (Brouillet et al., 2010), GBIF (www.gbif.org) and the Plants of the World Online (POWO, 2019) were used (Boufford et al., 2016; Brouillet et al., 2010; Cody, 2000). This resulted in some aberrant records, such as non-boreal/tropical plants (e.g. *Celtis* sp. and *Pteroceltis* sp.) as well as some likely food contaminants (e.g. *Allium cepa*, *Lagenaria siceraria*) and these were manually removed (Supplementary Information S4). When many blast hits from different species with an equal BIT-score were found, the top 20 blast hits were manually checked for likely boreal species. When several species met this criterion, the last common ancestor of these hits was chosen. Fungal OTUs were assigned to functional groups (guilds) using FUNGuild (Nguyen et al., 2016).

2.5.3. TrnL sequences

The trnL sequences were analysed with the OBITools package (Boyer et al., 2016). OBITools is commonly used in ancient plant DNA studies with *trn*L as it allows direct assignment of sequences to taxa. The forward and reverse reads were assembled using illuminapairedend (min quality score of 40) and subsequently assigned to the corresponding samples using *ngsfilter* (only keeping sequences with a 100% tag match and allowing for a maximum of three mismatches with the primers). Using obiuniq, strictly identical sequences were merged, after which obigrep was used to remove singletons, sequences with ambiguous nucleotides and sequences shorter than 10 bp. Following Bellemain et al. (2013), *obiclean* was used to identify sequencing and amplification errors with a threshold ratio of 5%

for reclassification of sequences identified as 'internal' to their corresponding 'head' sequence. The resulting sequences were compared to two taxonomic databases using *ecotag*. The first priority was given to a local taxonomic reference library containing arctic and boreal vascular plant taxa and bryophytes (arctborbryo database; Soininen et al., 2015; Sønstebø et al., 2010; Willerslev et al., 2014). A second reference library based on the global EMBL database (release 137) was used for mitigation of missing taxonomic assignment due to species potentially lacking in the first database (see Table S5.2 for full steps and read counts). The computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

The resulting sequences were further filtered in R to remove sequences that had (a) < 100% identity match to the reference libraries, (b) < 10 reads per PCR repeat and (c) sequences with higher number of reads in negative controls compared to the samples. This process resulted in the removal of suspected contaminant sequences derived from modern food plants such as *Solanum* subgenus *Lycopersicon* and *Oryza* sp. as well as some potential true positives including the genera *Solidago, Trifolium* and *Helictochloa*. No *Coelogyne fimbriata* reads were recorded in the positive control for *trn*L, despite the presence of *C. fimbriata* sequences in the NCBI Genbank database (e.g. MK356212.1). The presence of *C. fimbriata* reads in the non-control samples to determine the MOTU filtering threshold (as was used for nrITS filtering) could therefore not be used. Instead, the maximum number of reads from the most abundant OTU (*Salix* sp.) in control samples was used, and accordingly each sequence read count per replicate was reduced with 1.0%. Remaining replicates were merged while averaging the read counts per OTU. Finally, OTUs at species or genus level with the same taxonomic assignment were aggregated.

Although this filtering resulted in losing potential true positives, these were only present in a low number of reads (<0.1% of the total number of reads). Furthermore, this relatively rigorous filtering allowed for removal of nearly all suspected false positives in the samples, and this was given preference over retaining as many true positives as possible (cf. Alsos et al., 2018). Remaining identifications were manually checked for suspected contaminants or taxa that were known not to occur in the arctic and boreal region. This process resulted in the removal of a few remaining suspected contaminants (Supplementary Information S4). This is a common problem in metabarcoding studies, and the taxa we identify are similar to those found in other studies (Chua et al., 2021; Van Geel et al., 2014; Willerslev et al., 2014).

2.6. Diet analysis and habitat types

The DNA reads were converted to relative read abundances to facilitate comparison with macrofossil and pollen data. When referring to 'diet' in this study from now on, we refer to the composition of the last meal consumed by the animals

studied here, as inferred through the multiproxy approach on the faecal samples. The taxon identifications were grouped into the major groups of graminoids (grasses, sedges, rushes), forbs, shrubs/deciduous trees, coniferous trees, mosses and lichens. Since pollen records are biased towards high pollen producers and show primarily a regional signal (Jorgensen et al., 2012), they cannot be used to reliably reconstruct the diet. The record of macrofossils is strongly influenced by the food choice of the animal during its last meal (Mol et al., 2006) and has been shown to largely overlap with aDNA results (Parducci et al., 2015). Therefore, to provide a visual representation of the last diets, the average values of the relative abundance of the macrofossil results and all available DNA results were taken.

Plant identifications from DNA, macrofossils and pollen that could be assigned to the species level were used to reconstruct the habitat types of the megafaunal last diets. Some genera that are typically found in specific habitats have also been included (e.g. *Eriophorum, Juncus* in wetlands and *Puccinellia* in saline meadows). Habitat types were identified using a combination of sources: efloras (Brach and Song, 2006), Kienast et al. (2005), Troeva et al. (2010), Janská et al. (2017), Axmanová et al. (2020) and references therein. Only the presence of taxa and not their abundance was used to reconstruct the habitats, since abundance of certain taxa is highly affected by the selective food choice of the animals and may not reflect the palaeovegetation (Ashastina et al., 2018). The taxa were divided into 13 habitat types, ranging from relatively dry (steppe) to very wet (wetland: marsh, bog, fen, swamp). The modern known habitat preferences for the plant species were used, and the resulting habitat types are compared to modern analogues. For the modern caribou (Selwyn caribou A), the habitat consists of boreal forest in low-elevation areas, found together with arctic-alpine tundra at high altitudes (Galloway et al., 2012).

3. Results

3.1. Mammal sample identity

Genetic analyses confirmed the identity of both the Abyland and Adycha samples as Mammuthus primigenius (woolly mammoth), with a 100% match in both cover and identity (Table S2). This was further supported by the shape and size of the faecal pellets.

3.2. Pollen and macrofossil recovery

3.2.1. Pollen

For seven mammals, the pollen records were taken from the published records while four were newly generated in this study (Tables S6.1 – S6.11). The Selwyn caribou samples studied by Galloway et al. (2012) showed a mixed pollen signal with trees (ranging from 25 to 30%, *Picea* sp., *Pinus* sp., *Alnus* sp. etc.) and forbs (34–40%, mostly *Artemisia* sp.) being the most abundant. Selwyn caribou A further

showed 33% shrubs (*Salix* sp. and *Betula* sp.) which were missing in Selwyn B, and rare (6%) in Selwyn C. Low amounts (<10%) of undifferentiated Poaceae as well as insect-dispersed pollen (e.g. *Asteraceae, Ericaceae, Polemonium* sp. and Rosaceae) were identified in all three caribou samples.

The Holocene Yakutian bison and Oyogas Yar horse had high amounts of undifferentiated Poaceae pollen (71% and 92%, respectively; Gravendeel et al., 2014; Van Geel et al., 2014). Cyperaceae was the second most abundant pollen type (4%) in the horse and also accounted for 6% in the bison sample. The bison further had a relatively high amount (9%) of Apiaceae pollen. Other pollen in both samples was derived from various shrubs (*Betula* sp. and *Salix* sp.) and forbs (e.g. Asteraceae, Plantaginaceae, Rosaceae). Tree-derived pollen (*Abies* sp., *Pinus* sp. and *Alnus* sp.) was present in both samples and made up 3–4% of the total.

The previously studied Yukagir and Cape Blossom mammoths showed abundant wind-dispersed pollen types consisting of Poaceae (both ~70%) and Artemisia sp. (16% and 7%, respectively; van Geel et al., 2008; van Geel et al., 2011b). The newly obtained pollen results from the three Pleistocene mammoths (Abyland, Adycha, Maly Lyakhovsky) as well as the Yukon horse were also dominated by Poaceae and *Artemisia* sp. (>85%). The only sample with a low *Artemisia* count (1%) was the Maly Lyakhovsky mammoth, which was for 97% dominated by Poaceae. Insect-dispersed pollen types were rare to very rare in all Pleistocene samples and were derived from many different families, e.g. Apiaceae, Brassicaceae, Caryophyllaceae and Papaveraceae. The only sample with coniferous tree derived pollen was the Adycha mammoth with 1% *Pinus* sp. pollen.

3.2.2. Macrofossils

Macrofossil analyses were taken from published records for eight samples and newly generated for three mammoths (Maly Lyakhovsky, Abyland and Adycha) as well as for Selwyn caribou A (Table S6.1 - S6.11). The macrofossils of the three Selwyn caribou samples showed a mixture of shrubs (genera Betula and Salix), lichen and mosses as the most dominant taxa, with grasses and forbs (e.g. Asteraceae, Caryophyllaceae) making up the remainder (Galloway et al., 2012). Selwyn C showed 44% lichen fragments.

The Yakutian bison faecal sample was dominated by vegetative remains of Poaceae and Cyperaceae (50%), wetland forbs (e.g. *Comarum palustre* and *Menyanthes trifoliata*) as well as *Salix* sp. and minor moss fragments (Van Geel et al., 2014). The Oyogar Yar horse sample was dominated by unidentified Cyperaceae remains and minor remains of Poaceae and several moss fragments (Gravendeel et al., 2014).

The previously studied macrofossils of the Yukagir mammoth faecal sample showed abundant poaceous remains together with *Salix* sp. and *Carex* sp. (van Geel et al., 2008). The herbaceous component was made up of plant remains from varying families, e.g. Asteraceae, Brassicaceae, Caryophyllaceae, Papaveraceae. Remains from several mosses were also identified, including *Drepanocladus*

aduncus, *Bryum* sp., *Entodon concinnus*. The Cape Blossom mammoth macrofossils consisted of over 90% *Carex* sp., followed by Poaceae and a herbaceous component consisting of e.g. *Minuartia rubella*, *Potentilla* sp. and *Cerastium/Silene* sp. (van Geel et al., 2011b). Graminoids dominated the newly obtained data of the three mammoths Abyland, Adycha and Maly Lyakhovsky. This included poaceous vegetative remains, in the case of Abyland combined with one *Carex* sp. fruit and for Maly Lyakhovsky with the remains of a variety of mosses (e.g. *Campylium stellatum*, *Cinclidium stygium*, *Drepanocladus* sp., *Warnstorfia sarmentosa*).

3.3. DNA

Illumina sequencing resulted in 20.4 M read pairs for trnL and 16.4 M read pairs for nrITS. After quality filtering and clustering, 11.7 M reads were retained for trnL, 2.1 M reads for plant nrITS1, 2.2 M reads for plant nrITS2 and 5.0 M reads for fungal nrITS2. TrnL and fungal nrITS2 was successfully amplified in all samples while plant nrITS1 and nrITS2 was obtained for all but the Yukon horse, Cape Blossom mammoth and Selwyn caribou C.

The plant specific primers for the nrITS marker effectively amplified plant taxa, where 63.4% (nrITS1) and 70.4% (nrITS2) of the total OTUs were assigned to green plants (Figure S15). Of the total OTUs, 3.8% and 7.3% were assigned to fungi, respectively. The remainder of the OTUs comprised green algae (Chlorophyta) and made up 6.6% of the total OTUs for nrITS1 and 19.4% for nrITS2. Across all samples, trnL produced 167 green plant OTUs, while 73 and 71 green plant OTUs were identified using plant nrITS1 and nrITS2, respectively (Tables S7 - S12). Per sample, trnL showed the highest number of green plant OTUs with on average 35.2 (range 12–74), while nrITS1 recovered on average 10.8 green plant OTUs (0–28) and nrITS2 12.5 (0–40) (Table S16). For the fungal nrITS2, 88.2% of the total OTUs were assigned to Fungi, 11.6% to Viridiplantae and 0.2% was unidentified, while showing on average 20.2 fungal OTUs per sample (range 7–38; Tables S16). Read or OTUs counts were not correlated to the age of the samples for any of the markers.

3.4. Comparison of pollen, macrofossils and DNA data

Across DNA, pollen and macrofossil datasets, 311 plant taxa including 146 species, 150 genera and 63 families were identified (Fig. 2; see Table S6.1-S6.11 for full recovered plant taxa information across all samples). With pollen analysis, 65 plant taxa were identified, while 84 plant and 5 lichen macrofossil taxa were found. DNA analysis resulted in 146 (trnL), 73 (nrITS1) and 71 (nrITS2) plant taxa. At all taxonomic levels, DNA analysis recovered the most unique plant taxa, with 16 families, 77 genera and 123 species (Fig. 2). However, unique taxa were also identified using both macrofossil (four families, 18 genera and 15 species) and pollen analysis (six families, 11 genera and one species). No species were recorded

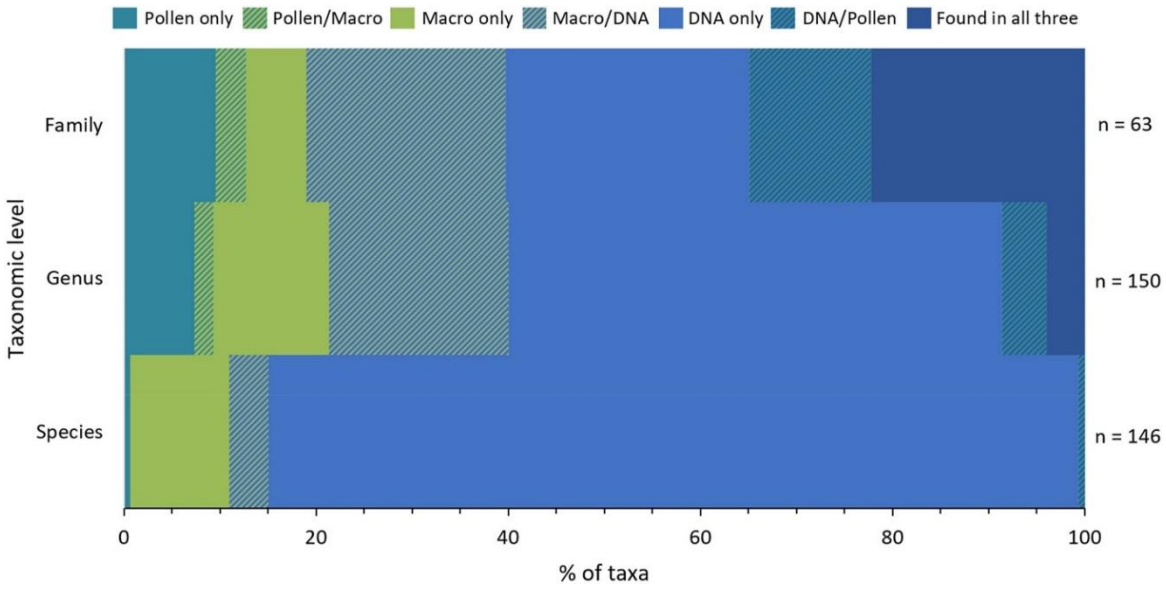


Figure 2. Percentage of identified plant taxa per proxy (pollen, macrofossil, DNA) at different taxonomic levels across all faecal samples studied here. Hatched areas represent overlap between two proxies. n = total number of taxa that was found in each specific taxonomic level.

across all three proxies, while six genera (*Androsace, Artemisia, Betula, Papaver, Rumex* and *Salix*) and 14 families were shared in the DNA, macrofossil and pollen data. The biggest overlap of proxies was found between DNA and macrofossil results at the genus level (29 genera), while there was little overlap between the pollen and macrofossil results (three genera and two families).

Pollen and macrofossils could be identified to species level in 3.1% and 24.7% of the recovered taxa, respectively. For the DNA markers, 44.8% of the OTUs were identified to species level for trnL, while this was 70.9% and 78.2% for nrITS1 and nrITS2, respectively (Table S7, S9, S11). To illustrate the differences in taxonomic resolution between the three proxies as well as between the DNA markers, results of three plant families (Poaceae, Asteraceae and Cyperaceae) that were common to abundant in all 11 faecal samples are shown in Table 2. Taxa from these three families were found using all three proxies. For plant families where pollen could only be identified to the family level, macrofossils could in several cases be identified to genera within those families, and in rare cases to species level (e.g., *Carex nardina* and *Carex dioica* in the Cyperaceae family). The nrITS marker could identify species for taxa where trnL results were only identifiable to genus or family level. An example of this is the identification of the species *Arctagrostis latifolia* (100% identity) and *Calamagrostis stricta* (99.7%; Poaceae) using nrITS1, while trnL identification was only possible to the subtribe level (Agrostidinae). Similarly, where trnL identified Asteraceae subfamily Anthemideae, the nrITS marker found the species *Artemisia scoparia* and *A. norvegica* (both 100% identity). Unique Poaceae species (*Koeleria asiatica, Festuca kolymensis*) and Asteraceae species and genera (*Artemisia gmelinii, Arnica, Saussurea*) were, however, also found using trnL and this pattern was found throughout the whole dataset (Table 2 and Table S7).

Table 2. All taxa recorded of three plant families (Poaceae, Asteraceae and Cyperaceae) that were common to abundant in all 11 faecal samples in DNA (trnL, nrITS1 and nrITS2), macrofossils and pollen analyses. The numbers represent the number of samples in which that specific taxon was found.

Family (subfamily)	Tribe	Subtribe	Genus (subgenus)	Species	trnL	nrITS1	nrITS2	Macro	Pollen
Poaceae								11	11
Pooideae	Bromeae		*Bromus*		4				
				B. pumpellianus	5				
	Hordeeae	Hordeinae	*Elymus*					1	
			Hordeum		3			1	
	Meliceae		*Glyceria*					1	
			Pleuropogon	*P. sabinei*	2				
	Poeae				6				
		incertae sedis		*A. fulva/D. fisheri*	2				
			Arctophila	*A. fulva*			4		
			Dupontia	*D. fisheri*		4	1		
		Agrostidinae			5				
			Arctagrostis					2	
				A. latifolia		1	3		
			Calamagrostis				3	2	
				C. stricta		1			
		Alopecurinae	*Alopecurus*					1	
				A. magellanicus		3	2		
		Aristaveninae	*Deschampsia*	*D. cespitosa*		3	3		
		Aveninae	*Koeleria*	*K. asiatica*	2				
		Coleanthinae	*Puccinellia*		2	2	1		
				P. tenuiflora/vahliana			2		
				P. vahliana		2			
		Loliinae	*Festuca*					2	
				F. altaica	3	1			
				F. kolymensis	3				
				F. ovina		1	2		
		Phalaridinae	*Hierochloe*					2	
		Poinae	*Poa*			1		3	
				P. arctica			4		
				P. glauca		2			
	Triticeae				4				
Asteraceae					3				
Asteroideae								1	10
	Anthemideae				6				
		Anthemidinae	*Tripleurospermum*	*T. maritimum*	1				
		Artemisiinae			4				
			Artemisia					2	11
				A. gmelinii	5				
				A. norvegica		2	1		
				A. scoparia		2	3		
	Astereae				3				
	Gnaphalieae				2				
			Antennaria					1	
	Madieae	Arnicinae	*Arnica*		2				
	Senecioneae	Tussilagininae	*Endocellion*	*E. sibiricum*		1	2		
			Tephroseris		1				
Carduoideae	Cardueae								2
		Carduinae	*Saussurea*		3				
Cichorioideae									5
Cyperaceae								4	10
Cyperoideae	Cariceae		*Carex*		5			6	
			Carex subg. *Carex*		1				
				C. aquatilis	3	2	2		
				C. microchaeta	2				
				C. nigra subsp. *juncea*		5			
				C. podocarpa		1			
				C. rostrata		2	1		
				C. vesicaria		1	1		
			Carex subg. *Euthyceras*		1			1	
				C. nardina				1	
			Carex subg. *Vignea*		3				
				C. chordorrhiza		1			
				C. dioica				1	
				C. duriuscula		1	1		
				C. lachenalii	1				
				C. maritima	2				
	Scirpeae		*Eriophorum*		3	1		3	
				E. angustifolium		3	3		

3.5. Diet analysis

High congruence between the quantitative results of the different DNA markers was found for the Selwyn A and B caribou samples, with a dominance of shrubs (87–98%; Salix, Betula and various ericaceous taxa) and low abundance of forbs, graminoids and mosses (Fig. 3a). In contrast, the macrofossil results indicated high abundance of mosses, graminoids and lichen with only low amounts of shrubs. The combined diet reconstruction - based on DNA and macrofossils only - showed ~75% shrubs with 10–15% mosses (Fig. 3b). Fungal nrITS2 results further identified low amounts of lichen, including Cladonia spp., Bryocaulon divergens and Stereocaulon saxatile (Table S13 – S14) that may have formed part of the caribou diet (0.3% of total fungal reads for Selwyn B and 0.1% for Selwyn A). For Selwyn caribou C, trnL showed a much higher amount of forbs (72%; mainly Asteraceae tribe Anthemideae and Sibbaldia procumbens) than the macrofossils (8%) or pollen (34%). The reconstructed diet differed from the other two caribou samples, consisting of 40% forbs and equal parts (15–20%) of shrubs (Salix), lichen and mosses.

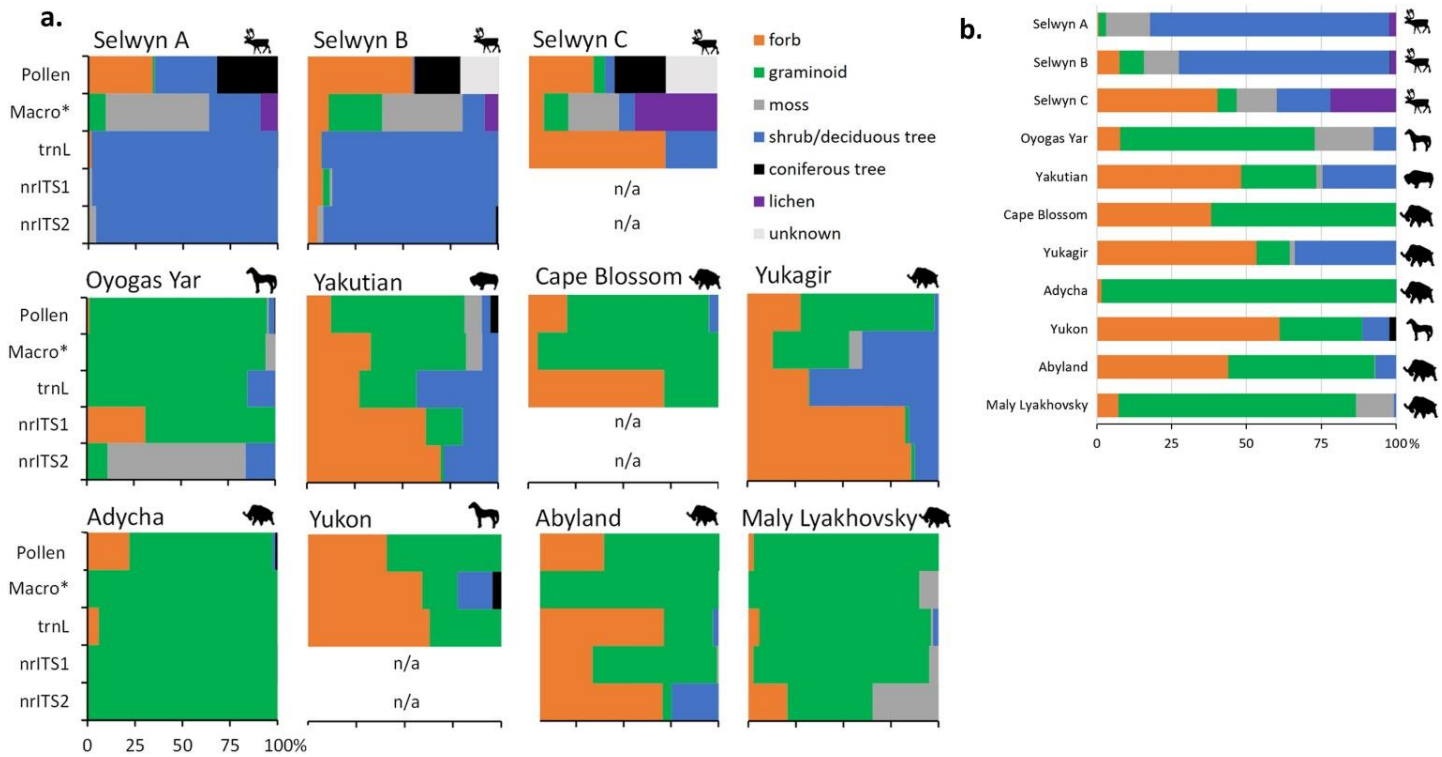


Figure 3. Diet reconstructions based on quantitative abundance of plant groups (forbs, graminoids, mosses, shrubs/deciduous trees, coniferous trees and lichens). a) Quantitative comparison of results from the different plant proxies used for all samples in this study. ∗ exact quantitative data from macrofossils was only present for the Selwyn caribou B and C. For all other samples, the semi-quantitative macrofossil results have been converted to quantitative measures for illustrative purposes. b) Reconstruction of the composition of the last diet by taking the average value of the relative abundance of macrofossil and all available DNA results.

Macrofossils of the Oyogas Yar horse were for >95% dominated by graminoids and this was also reflected in the trnL (85%) and nrITS1 (69%) data (mainly *Eriophorum* sp. and *Dupontia fisheri* respectively). The plant nrITS2 results, however, were dominated by mosses (73%). The diet reconstruction showed a dominance of graminoids (65%) with 20% mosses and equal amounts of shrubs and forbs (8%). The diet of the other, much older, Yukon horse contained a lower fraction of graminoids (28%) and, instead, was dominated by forbs (on average 60%; consisting of *Braya rosea* and Asteraceae tribe Anthemideae). Tree and shrub taxa were only identified in the macrofossil results for this sample. The Yakutian bison sample consisted on average of 48% forbs (mainly *Cicuta virosa*) and 25% each of graminoids (*Eriophorum, Carex*) and shrubs (*Salix*). The Adycha and Maly Lyakhovsky mammoth samples showed highly similar results from both proxies and the reconstructed diets consisted almost exclusively of graminoids (Fig. 3b). Graminoids in the Adycha sample consisted for >75% of Puccinellia sp. based on DNA analysis, while many species of Poaceae (including abundant *Deschampsia cespitosa* and *Alopecurus magellanicus*), as well as *Carex* sp. and *Eriophorum* sp. were found in the Maly Lyakhovsky sample. Mosses were found to be relatively abundant in this sample according to nrITS2 results (33%; mainly *Polytrichastrum alpinum*), while much lower percentages of mosses were found in nrITS1, trnL or macrofossil results.

The three other mammoth samples showed a higher contribution of forbs to their diet, often with the DNA results of the different markers showing one species dominating the assemblage. For the Abyland mammoth this dominant species was *Anemone patens*, while in the Yukagir mammoth sample *Myosotis alpestris* was abundant. The Yukagir mammoth was the only one of the mammoth samples showing relatively abundant (on average 34%) shrubs (*Salix*) in its diet. In the Cape Blossom mammoth, graminoids made up >75% of macrofossils, while the trnL results showed 28% graminoids, consisting mainly of *Carex*. In the *trn*L results forbs were abundant (71%) and consisted for the largest part of *Chamaenerion angustifolium* and Asteraceae tribe Anthemideae.

3.6. Habitat types

We combined species and genus-level plant identifications from all proxy results to reconstruct the habitats in which the last meals of the studied megafauna were consumed (Fig. 4; Table S17 for all plant species information).

Identified plant species in the Selwyn caribou A and B samples provided a range of habitats including wetland, woods and a large component of arctic-alpine tundra (e.g. *Arctous alpina*, *Anemone richardsonii*, *Carex podocarpa* and *Pyrola grandiflora*) along with taxa typical for mountainous/rocky habitats (e.g. *Rhodiola integrifolia*). The Selwyn caribou C sample similarly contained many species typical for arctic-alpine tundra but also included a large component of species typical for snow patches (e.g. *Ranunculus nivalis*, *Ranunculus pygmaeus*, *Oxyria digyna*).

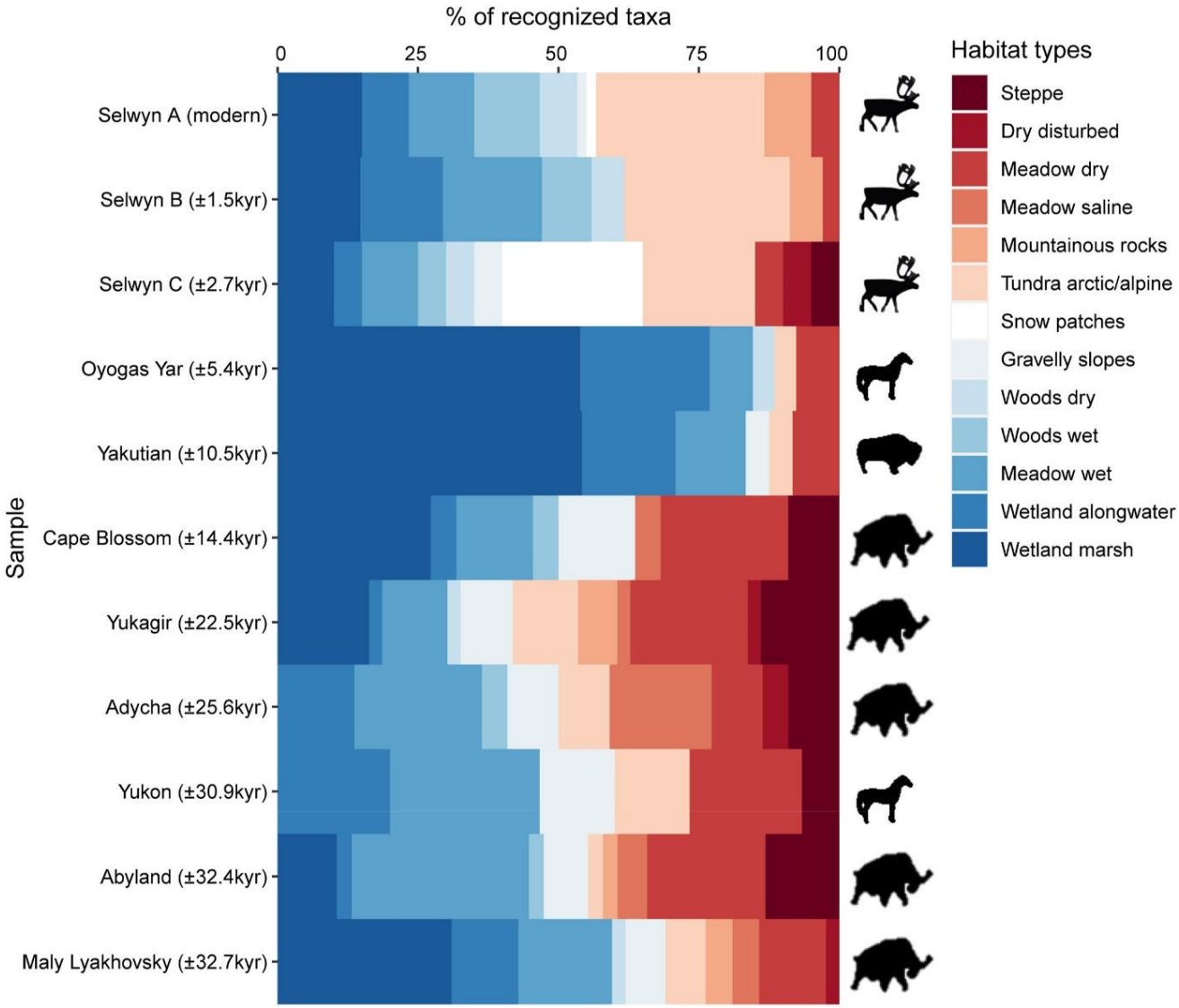


Figure 4. Habitat reconstruction of megafaunal species based on integrated (pollen, macrofossils, DNA) species and genus resolution data. The samples were sorted according to their age and the average calibrated age of each sample is indicated between brackets.

The reconstructed habitats of the Holocene Oyogas Yar horse and Yakutian bison consisted mainly of wetlands, including marshes and river/lake sides. For the Oyogas Yar horse this included *Eriophorum* sp., *Caltha palustris* and *Comarum palustre* typical for marshes and e.g. *Arctagrostis latifolia* and *Arctophila fulva/Dupontia fisheri* from water sides. The Yakutian bison showed numerous Carex species, *Menyanthes trifoliata*, *Epilobium palustre* and *Hippuris* sp., all indicative of marshy wetland conditions as well as e.g. *Endocellion sibiricum* and *Epilobium palustre* typically found along rivers or ponds.

The Cape Blossom and Maly Lyakhovsky mammoth samples also included wetland components, with in the case of Cape Blossom e.g. *Caltha palustris* and species of *Carex* and for Maly Lyakhovsky *Eriophorum* sp., *Caltha palustris* as well as several grass species (*Pleuropogon sabinei, Arctophila fulva*). Moss species in the

Maly Lyakhovsky mammoth further provided evidence of a wet, marshy environment (e.g. *Drepanocladus sordidus, Cratoneuron filicinum, Warnstorfia sarmentosa* and *Dicranum bonjeanii*). However, in contrast to the Holocene horse and bison, both these mammoth samples also included species indicative for dry meadows and, in the case of Cape Blossom, steppe (*Festuca kolymensis* and *Artemisia gmelinii*). Several true steppe species were also found in the Abyland mammoth (*Silene samojedorum, Carex duriuscula, Artemisia scoparia*) and Yukagir mammoth samples (e.g. *Eritrichium sericeum, Festuca kolymensis, Phlox hoodii*). Other taxa in both samples were indicative for dry meadows (e.g. *Anemone patens* and *Cerastium maximum* for Abyland and *Myosotis alpestris* and *Eremogone capillaris* for Yukagir). Furthermore for the Abyland mammoth, several species typical for wet meadows were identified (e.g. *Sanguisorba officinalis, Stellaria borealis*), while for the Yukagir mammoth a component of gravelly slopes and mountainous/rocky habitat was found (e.g. *Smelowskia alba, Oxytropis deflexa, Rhodiola rosea*). The Pleistocene Yukon horse also showed a last meal consisting of a mix of taxa from different habitats with species typically found in wet meadows and wetlands (*Alnus incana, Juncus alpinoarticulatus*) as well as dry meadow and steppe (*Bromus pumpellianus, Artemisia gmelinii*). The habitat for the Adycha mammoth consisted of meadows (e.g. *Deschampsia cespitosa, Bromus pumpellianus*) as well as a large component of saline meadow (*Puccinellia* sp.).

4. Discussion

4.1. Comparison of proxies

Out of the three proxies used in the present study (DNA, pollen and macrofossils), DNA recovered the highest number of unique taxa at all taxonomic levels (Fig. 2). This is likely caused by the large amount of vegetative remains in the faecal samples that could not be identified beyond the family or genus level using macrofossil or pollen analysis. DNA analysis does not depend on the season when plants carry seed, fruit or pollen and allows identification of many taxa to the species level irrespective of their developmental stage. We also used primers for multiple marker regions (*trn*L, nrITS1, nrITS2), each identifying unique taxa and increasing overall taxonomic resolution (Tables S7 - S12).

In comparison to pollen from sediments, pollen spectra from our faecal samples were not very diverse (Jorgensen et al., 2012; Parducci et al., 2015; Pedersen et al., 2013). This could be because lake sediments accumulate pollen over a much larger spatial and temporal scale than faeces do. We took all the samples for our analyses from the middle of the faeces and thus caught only a snapshot of airborne pollen (i.e., sticking on ingested vegetation), mixed with pollen coming from ingestion of inflorescences. The taxonomic overlap between pollen and DNA, as well as between pollen and macrofossils was surprisingly low, and we instead found the highest overlap between DNA and macrofossil results. This is likely because both of these proxies are providing a local signal (showing

the food choice of the animal) while the pollen analysis is influenced by accidental intake of pollen sticking to ingested vegetation as well as pollen from species producing high amounts of pollen (e.g. Jorgensen et al., 2012).

4.1.1. Metabarcoding detection gap

We use the term 'metabarcoding detection gap' here for taxa that were not retrieved in the DNA results (*trn*L or nrITS) but were present in the macrofossil and/or pollen records. In total, the metabarcoding detection gap consists of 12 families, 32 genera and 16 species (Fig. 2). Many of these taxa are very rare in the pollen or macrofossil counts, with most of them found in only one sample and in low abundance. For pollen this includes single identified spores and pollen of Botrychium sp. and Populus sp. in the Selwyn caribou samples, and *Epipactis* sp., *Persicaria* sp. and *Thalictrum* sp. in the mammoth samples. For such rare pollen grains it seems likely they were only present as pollen while being (very) rare in the consumed vegetation. A lysis step with mechanical bead beating is necessary to break the exine of pollen grains and release the inner DNA (Polling, 2021). Since these steps have not been used here, this could explain the absence of these taxa from the DNA results. On top of this, pollen contains very little DNA that is hard to amplify even if present in high numbers (Parducci et al., 2005). Similar to proxy comparison studies on lake sediments (e.g. Parducci et al., 2019), we find that DNA from pollen contributes very little to the total DNA signal in faeces.

There are also taxa that were found as pollen with high relative abundance, while being very rare or absent in the other proxies. This includes, for example, pollen of the family Pinaceae which account for up to 30% in the caribou samples. Pinaceae pollen is often overrepresented in pollen records from the (sub)Arctic because they are high pollen producers and their pollen is spread over large distances (Aario, 1940). The genus Artemisia reached up to 40% in some pollen records (Selwyn caribou B; Table S6.2), yet it is very rare in both DNA and macrofossil results. Unfortunately, using *trn*L, the genus Artemisia cannot be distinguished from other genera from the subfamily Anthemideae (*Anthemis, Achillea, Chrysanthemum, Tanacetum* etc.). This subfamily was relatively abundant in Selwyn caribou C, Cape Blossom mammoth and the Yukon horse, and it cannot be resolved whether these reads actually belong to Artemisia. Rare fragments of Artemisia in the macrofossil records were only recorded in the Yukon horse and Selwyn caribou C samples. Part of this discrepancy can be explained by differential preservation, since macrofossils of Artemisia such as seeds or fruits (achenes) deteriorate rapidly and are therefore rarely recovered (Anderson and Van Devender, 1991; Birks, 2007). Other studies on DNA metabarcoding of Pleistocene megafaunal faeces also found high amounts of Artemisia pollen but very low abundance with DNA or macrofossils from the same samples (e.g. Kolyma rhinoceros and Finish Creek mammoth; Willerslev et al., 2014). For caribou, where in all three samples Pinaceae and Artemisia pollen is common to abundant, it is furthermore known that they do not actively select Artemisia and avoid Pinaceae

(Denryter et al., 2017; Jung et al., 2015). These records are therefore interpreted as the results of accidental uptake of pollen sticking to selected plant taxa.

In the macrofossil data, we detected many taxa that were represented by one seed or plant fragment (e.g. *Antennaria* sp., *Draba* sp., *Sagina* sp., *Hedysarum* sp., *Lysimachia* sp.) and many of these are part of the metabarcoding detection gap. Furthermore, fragments of various mosses were exclusively found as macrofossils (e.g. *Calliergon* sp., *Plagiomnium* sp., *Rhizomnium* sp., *Thuidium* sp. and the spikemoss *Selaginella* sp.). It should be noted that DNA reference libraries are still far from complete, and this may be especially true for Arctic Russian moss species. Therefore, some of the species found as macrofossils may not be recoverable using DNA at this moment. One such example is the moss *Cinclidium stygium* for which no nrITS sequence is currently available in the NCBI Genbank. Apart from this, the expected amplicon size for bryophytes using the plant-specific nrITS primers in our study is > 500 bp (Cheng et al., 2016), which may cause some species to be missed due to the 600 bp restriction using Illumina sequencing. Furthermore, even though we applied a multi-locus approach, DNA primer mismatch in both *trn*L and nrITS could have occurred. Many *Selaginella* species for example show 5 mismatches in their DNA barcodes with the *trn*L-h as well as the ITS4 reverse primers used in this study. Lastly, DNA of plant fragments may have been simply too degraded to be amplified by any of the DNA markers.

4.1.2. Morphology detection gap

A 'morphology detection gap' is designated here as all taxa that are missing in either the pollen or macrofossil record but were found in the DNA results. In total, the morphology detection gap for the studied faecal samples consists of 16 families, 77 genera and 123 species (Fig. 2). The biggest factor contributing to many of the taxa only found as DNA is the higher taxonomic resolution that is achieved using DNA (although it depends on the percentage of identity used whether taxa identified by DNA are assignable to either, e.g., genus or species level). There are, however, a number of other factors that may determine the taxa in the morphology detection gap.

First, many taxa only found with DNA were very rare (<0.1% of the relative amount of reads) and only recorded in one sample. These taxa could have either been very minor diet items or taxa that were not targeted (i.e. accidental intake), which were present in such low quantities that they may have been missed with the macrofossil or pollen analyses. Accidental intake could also explain the presence of several species in the DNA results of the caribou samples of which the ingestion of high amounts may be toxic (e.g. *Pedicularis capitata, Oxytropis deflexa*; Denryter et al., 2017). Secondly, some plant taxa may be more affected by the digestive processes than other plant taxa, causing them to be unrecognizable as macrofossils while still being recoverable using DNA. Lastly, despite extensive reference collections for pollen and macrofossils, identification may still be somewhat subjective with regards to morphologically very similar taxa. This is less

the case for DNA using reference libraries that allow more objective identifications.

Taken together, this explains the abundance of some taxa in DNA results even though they were missing in the other proxies. One example is the willowherb family Onagraceae for which *Chamaenerion angustifolium* and *Epilobium palustre* were found in DNA of seven of the samples studied here. Rare Onagraceae pollen were only found in the Cape Blossom mammoth (van Geel et al., 2011b). Although pollen from insect-pollinated plants are always underrepresented in faecal samples, we identified abundant *Chamaenerion angustifolium* in the DNA results of the Cape Blossom sample. No macrofossil remains of Onagraceae were recorded in any of the samples, and this is likely because vegetative Onagraceae remains are very hard to recognize due to their ambiguous morphology (Anderson and Van Devender, 1991; Grímsson et al., 2012). Similarly, the forget-me-not family Boraginaceae is only recovered using DNA. It was especially abundant in the last meal of the Yukagir mammoth (*Myosotis alpestris* and *Eritrichium sericeum*). An additional species (*Mertensia paniculata*) was identified in the faecal samples of the caribou and the Cape Blossom mammoth, yet no remains of Boraginaceae were found in either pollen or macrofossil analyses of any sample. Pollen grains of members from this family are particularly small (5–7 μm) and could potentially be overlooked during analysis while vegetative macrofossil remains are hard to identify. Macrofossils of Boraginaceae and Onagraceae have not been recorded in any other mammoth faeces, even though they were recorded in high abundance in DNA data (e.g. Finish and Drevniy Creek mammoths as well as Yukagir bison; Willerslev et al., 2014). These examples show the added value of DNA analysis and indicate that vegetative plants of these families may likely have formed part of the diets of the studied megafauna.

4.1.3. Comparison of plant DNA markers

Our application of multiple DNA markers on megafaunal faecal samples reveals the added value of a multilocus approach. The three samples for which no plant nrITS results were obtained were of very different ages (±2.7, ±14.4 and ± 30.9 kyr BP), while older samples did produce plant nrITS amplicons (Abyland and Maly Lyakhovsky mammoths). While nrITS amplicons were found in all samples, for the three samples where no plant OTUs were found, these were all either derived from contamination, algae or fungi. Fragments of DNA up to 500 bp have been recovered from permafrost preserved sediments as old as 400 kyr (Lydolph et al., 2005). Therefore, it most likely depends on the conditions in which the specimens were preserved over time that determined whether or not these long fragments can be recovered. Some samples may have inadvertently been (partially) thawed at some stage, causing longer DNA fragments to be degraded, while the shorter and more stable *trn*L was not affected.

Most unique taxon identifications of the nrITS marker come from increased taxonomic resolution of several families and genera that show relatively low taxonomic resolution in the other proxies. This includes, for example, the genus *Carex* for which six unique species were found and the family Poaceae for which 11 unique species were identified with nrITS (Table 2). Furthermore, nrITS identified a larger variety of mosses than trnL, which is likely the result of the very short sequence length of the bryophyte P6 loop (±22 bp) obtained using the *trn*L g and h primers. These primers were not designed for bryophytes, and the recovered length often prevents sufficient taxonomic detail (Epp et al., 2012; Soininen et al., 2015). Nevertheless, many unique plant species were found using trnL, which could be the result of the more complete reference libraries available for *trn*L compared to nrITS. Many nrITS reference sequences in the NCBI Genbank database do not represent the complete marker region (e.g. *Pleuropogon sabinei* and *Ranunculus nivalis* with partial nrITS2 sequences) or are simply missing altogether because no reference sequences have been deposited yet. This is, for example, seen for species in the genus Puccinellia where not all Russian endemics have been sequenced (missing e.g. *Puccinellia manchuriensis, P. byrrangensis, P. jenisseiensis*), and this might also explain why we find *P. vahliana* (nowadays a western Arctic species) in nrITS results. Apart from that, the shorter and more stable *trn*L P6 loop produced results for the samples that did not produce any results from nrITS, which further explains the number of unique *trn*L identifications.

4.2. Diet analysis

The diet analysis of Selwyn A and B showed that shrubs are highly dominant in the summer diets of caribou, which is in agreement with known diets of summer foraging caribou that consists of deciduous shrubs along with reindeer lichen and fungi (Bergerud, 1972; Boertje, 1984). Lichen were observed using macrofossil and fungal nrITS2 analysis, and were also indirectly detected with plant DNA by the presence of lichen phycobionts in the plant nrITS2 results (e.g. *Asterochloris, Symbiochloris* and *Trebouxia* spp.), only found in the Selwyn caribou samples (Table S18). *Trebouxia* is the most common phycobiont in extant lichen, while *Asterochloris* is mainly associated with lichen of the families *Cladoniaceae* and Sterocaulaceae (Pino-Bodas and Stenroos, 2020). Both families were also identified using fungal nrITS2 (Table S13 – S14), providing further support that the caribou ate lichen. The diet of modern caribou is well studied and for many Arctic plant species it is known whether they are either “selected”, “neutral” or “avoided” based on observations of foraging caribou (Bergerud, 1972; Denryter et al., 2017). An average diet of modern caribou was found to consist of 78% selected, 15% neutral and 7% avoided species (Denryter et al., 2017). For Selwyn A and B, “selected” plant taxa made up >85% of relative abundance of all DNA markers, while “avoided” taxa made up <5% (Table S19). This is in contrast to macrofossil results that showed up to 21% avoided taxa, mainly from mosses. Selwyn caribou

C showed a large component of diet items that were of unknown (43%) and neutral diet preference (44%), with only minor (11%) selected plant taxa (Table S19). This points to a somewhat atypical summer diet for this caribou when compared to modern caribou preferences and may suggest a different vegetation composition in its habitat.

The diets of nearest living relatives for Holocene bison and horse are well studied. While horses are typical grazers nowadays with diets consisting >75% of graminoids (Mendoza and Palmqvist, 2008), this has not always been the case. Several studies have shown that prehistoric horses had mixed grass-browse diets, especially in winter when grasses were harder to access (Kaczensky et al., 2017; MacFadden et al., 1999). The diet of the Holocene Oyogas Yar horse (*Equus* cf. *lenensis*) is typical for a grazer, with the main component being identified as graminoids. The Pleistocene Yukon horse (*Equus lambei*), however, consumed mostly forbs. The season of death could not be determined for the Oyogas Yar horse (although it could be spring to summer due to relatively high amount of Cyperaceae pollen), while for the Yukon horse it was determined as winter (Harington and Eggleston-Stott, 1996; Harington, 2002). This could explain why grasses made up only 28% of the total diet for the Yukon horse (Fig. 3b). It is likely that snow covered much of the grass cover, forcing the horse to focus on other available dietary items or that grasses were simply less abundant or of lower nutritional value (Savage and Heller, 1947).

The now extinct steppe bison (*Bison priscus*) was closely related to modern bison (*Bison bison* (Linnaeus, 1758); Marsolier-Kergoat et al., 2015). While modern bison are often thought of as grazers feeding for the majority on graminoids, their summer diets are more variable, consisting on average of 44% grass, 38% forb, 16% shrubs and <2% sedge (Leonard et al., 2017). This is similar to the DNA results of the Yakutian bison studied here, where forbs and shrubs are important components. Pollen of undifferentiated Apiaceae (identified by nrITS as Cicuta virosa) were also relatively abundant in this sample (9%) indicating ingestion of inflorescences. This may indicate that the Yakutian bison had its last meal in summer and was a mixed feeder that did not rely solely on grasses. The 'warm season' (spring/summer) was also identified as the most likely season of death for the Yakutian bison by Van Geel et al. (2014) and Boeskorov et al. (2016). The >52 kyr old bison (*Bison* sp.) studied by Willerslev et al. (2014), similarly showed a high abundance of forbs and shrubs (80%), although no season of death was identified for this sample. Lastly, the abundance of poisonous *Cicuta virosa* (water hemlock) in nrITS, and also recognized to lower taxonomic resolution in *trn*L, pollen and macrofossils, possibly indicates that the Yakutian bison died of hemlock poisoning (Jacobson, 1915).

The last meals of the Maly Lyakhovsky and Adycha mammoths consisted almost exclusively of graminoids. Some of these grasses can grow to considerable size (75–100 cm) and may have provided sufficient nutritional value for mammoths (e.g. *Bromus pumpellianus, Deschampsia cespitosa, Dupontia fisheri*).

Furthermore, the genus *Puccinellia* which was identified as the main component in the Adycha mammoth last diet, includes several species that are commonly grown for hay making for cattle in modern day Yakutia, Russia (Gavril'eva, 2011). The other mammoths studied here had much lower relative amounts of graminoid DNA, or barely any in the case of the Yukagir mammoth. The last diet of the previously studied Mongochen mammoth as reconstructed using macrofossils consisted mainly of mosses, forbs and only minor grasses and shrubs while DNA results showed dominance of 98% graminoids (Kosintsev et al., 2012a; Willerslev et al., 2014). The authors suggested that the underrepresentation of graminoids in the mammoth faeces could be the result of the digestive processes breaking down the poaceous tissues, although this is not supported by our finding of graminoids being dominant in the other mammoth faeces. It does, however, hold for forbs which are underrepresented in macrofossil and pollen results as compared to our DNA data, which has also been found in previous studies (e.g., Kosintsev et al., 2012b; Willerslev et al., 2014). The last meals of the Abyland and Cape Blossom mammoths may not have consisted solely of graminoids as suggested by the macrofossil analysis, but supplemented with *Anemone patens* (Abyland) and various other forbs, while shrubs and *Chamaenerion angustifolium* were consumed by the Cape Blossom mammoth. The abundance of *Salix* sp. and Boraginaceae (Yukagir) provides further evidence for the diversity in mammoth diets.

Another potential explanation for the differing diets may be sought in the different seasons of death, which could be determined for three of the mammoth samples studied here. The season of death of Maly Lyakhovsky mammoth was determined as late summer to early autumn (Grigoriev et al., 2017), while for both Yukagir and Cape Blossom mammoths autumn to early spring was suggested (Mol et al., 2006; van Geel et al., 2011b). A recent study on molar enamel profiles found that mammoths may have had seasonally different diets, shifting between browse and grasses (Uno et al., 2020). Also in the previously published Mongochen mammoth that died mid-summer and for which DNA, pollen and macrofossil results were analysed, the last diet was interpreted to be dominated by graminoids (Kosintsev et al., 2012a; Willerslev et al., 2014). This limited amount of data suggests that warm season diets of mammoth may have been dominated by graminoids (Maly Lyakhovsky, Mongochen), while they relied on various other food sources in the cold season (Cape Blossom, Yukagir). However, more multiproxy data is needed to support this hypothesis.

In some of the faecal samples studied here, mosses were identified in abundance either in the macrofossils (Selwyn caribou A and B) or in DNA results (nrITS2; Oyogas Yar horse and Maly Lyakhovsky mammoth) while being nearly absent in the other proxies. The relative abundance of mosses in the macrofossils of the caribou faeces is probably the result of accidental ingestion when the caribou were foraging low on the ground for dwarf shrubs and lichens. The moss species that was abundantly found with nrITS2 in the Oyogas Yar and Maly

Lyakhovsky sample was *Polytrichastrum alpinum* which was detected only as rare fragments in the macrofossil remains of these samples. Potentially the primers used to amplify the nrITS2 region caused preferential amplification of this type of moss. Although abundant moss fragments have been identified in macrofossils from several mammoths (Kosintsev et al., 2012a, 2012b), and are sometimes found in caribou faeces (Denryter et al., 2017), they are unlikely to have formed a major part of the diet for any of the extinct and extant mammals studied here because of their low nutritional value.

4.3. Habitat types

The reconstructed habitat for Selwyn caribou A and B corresponds well with the known current habitat of these animals in the Selwyn Mountains in Northwest Territories, Canada. The habitat for these two samples consists of elements from both downslope boreal forest and its wetlands, along with upslope alpine tundra. It is important to note that the two most dominant diet items as identified by DNA (*Salix* and *Betula*), are not included in the habitat analysis because neither of them could be identified beyond the genus level. Species from these genera have varying habitat preferences and therefore the genus level identifications did not provide enough information to infer the habitat, the only exception being rare *Salix alaxensis* in Selwyn Caribou B which typically grows in forested habitat along streams and lakes (Boufford et al., 2016). The only Betula species found in the Selwyn Mountains are B. glandulosa (dwarf birch, shrub) and B. papyrifera (canoe birch, tree), with the dwarf birch being far more common (Galloway et al., 2012). The habitat reconstructed for Selwyn caribou C may indicate that the faeces in this sample was deposited by caribou that consumed a meal nearer to the ice patch.

When many megafauna species disappeared at the end of the Pleistocene, the Holocene vegetation shifted significantly to become a more waterlogged environment with mossy and shrub-dominated tundra and deciduous forests (Edwards et al., 2005; Guthrie, 2001). The habitats reconstructed for the Holocene horse and bison reflect this mesic environment. Previous studies on these samples, however, indicated dry steppe-like conditions based on pollen and macrofossils due to the abundance of Poaceae remains (Boeskorov et al., 2016; Gravendeel et al., 2014; Van Geel et al., 2014). However, here we find that the species composition of Poaceae for both samples included *Dupontia fisheri, Arctophila fulva* and *Arctagrostis latifolia*, all species typical for wetland habitats. Similar to the results for the Holocene Yakutian bison, modern bison (*Bison bison*) are known to prefer sedge marshes over other habitat types (Belanger et al., 2020 and references therein). Our results show that both horse and bison are not strictly graminoid grazers, but utilize wetlands in their habitat as well. This is also confirmed by the habitat reconstructed for the Pleistocene Yukon horse studied here, that showed a mixed environment of wetland and dryer meadows and steppe. Furthermore, a recent study on dental micro- and mesowear of horse and

steppe bison also found that both were likely mixed feeders, instead of obligate grazers (Kelly et al., 2021).

Mixed environments were also identified for the mammoth samples, although with varying degrees of wetland components. The oldest mammoth studied here, Maly Lyakhovsky, showed many species typical for a marshy environment. This is in contrast to the Abyland mammoth that was collected from the same geographic area (North Sakha republic, Russia) and of similar age, that showed a much larger steppe and dry meadow habitat. This relatively large steppe component was also found for the Yukagir mammoth, although for this mammoth it was mixed with many plants typically found on gravelly slopes and mountainous areas. This may indicate that mammoths may have been versatile in their diets, adapting to the various habitats that were available. This is further supported by the habitat reconstructed for the Adycha mammoth, which shows that saline meadows were present and utilized by mammoths as well. For the Cape Blossom mammoth, no nrITS results were obtained which hampers the habitat reconstruction. However, with the other proxies a habitat similar to the Maly Lyakhovsky mammoth was reconstructed, with marshy wetland and surrounding wet meadows, intermixed with steppe and dry meadow. The variety of diets obtained from different habitats supports the idea that the 'mammoth steppe' was a mosaic of habitats instead of a homogeneous vegetation type (e.g. Zazula et al., 2007). Furthermore, the specific plant species mixture identified for these mammoths is not found in any modern habitat type, pointing to non-analogue plant communities (Williams and Jackson, 2007). Our results also indicate that mammoths were not exclusively grazers, but rather opportunistic mixed-feeders.

5. Conclusions

We integrated multilocus plant DNA, macrofossil and pollen analysis to obtain detailed reconstructions of megafaunal diets and habitats. We found most plant species in faecal samples uniquely using DNA, some of which abundantly so. This could be because of the large number of vegetative plant remains in the faeces which have become unidentifiable for macrofossil analysis due to masticatory and digestive processes. Unique plant taxa were, however, also found using both macrofossil and pollen analysis. We further show that relatively long nrITS fragments can be amplified from faecal samples as old as 28,610 14C BP and that these help to increase species resolution for many plant families (e.g. Asteraceae, Cyperaceae and Poaceae) as well as mosses that could not be retrieved using *trn*L.

We could accurately reconstruct the known diet and habitat of modern and late Holocene caribou (i.e. abundant shrubs from an arctic alpine tundra) and extended this approach to Holocene and Pleistocene megafauna including horse, steppe bison and woolly mammoth. These reconstructions showed that the Holocene steppe bison and horse were not strict grazers but rather mixed feeders that were foraging in a marshy wetland environment. This result is in sharp

contrast with previous reconstructions that suggested dry steppe-like conditions for these samples. We further find that the five Pleistocene mammoths studied here had very different last meals obtained from a variety of habitats including wetland, wet meadow, gravelly slopes, saline meadow and steppe. This confirms the presence of a mosaic of habitats in the Pleistocene landscape often referred to as the 'mammoth steppe' that mammoths could fully exploit due to a high flexibility in their diet choice.

Author contributions

Marcel Polling: Conceptualization, Data Curation, Investigation, Methodology, Visualization, Formal analysis, Writing – original draft Anneke T.M. ter Schure: Software, Methodology, Visualization, Formal analysis, Writing - original draft Sanne Boessenkool: Methodology, Visualization, Writing - Review & Editing Bas van Geel: Conceptualization, Validation, Resources, Writing - Review & Editing Tom van Bokhoven: Methodology, Investigation Glen MacKay: Resources, Writing - Review & Editing Bram W. Langeveld: Investigation, Writing Original Draft María Ariza: Methodology, Visualization Hans van der Plicht: Investigation, Writing - Review & Editing Albert V. Protopopov: Resources, Writing - Review & Editing Alexei Tikhonov: Resources, Writing - Review & Editing Hugo de Boer: Funding acquisition, Supervision, Writing - Review & Editing Barbara Gravendeel: Conceptualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

Some of the faecal samples were kindly made available by Dick Mol, Russell Dale Guthrie and Grant Zazula. We like to thank Nathalie Van der Putten and Laurens Sparrius for their help with identifying the macrofossils of mosses and lichen, Marcel C. M. Eurlings, Bertie Joan van Heuven, Rob Langelaan and Elza Duijm for their help in the laboratory and sample handling, Audun Schrøder-Nielsen and Jarl Andreas Anmarkrud for their assistance with NGS library preparations in Oslo, Norway, Kevin Beentjes and Arjen Speksnijder for valuable comments on NGS data handling and Rune Halvorsen for his help in the data analysis. This research is part of the H2020 MSCA-ITN-ETN Plant. ID network, and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765000.

Data availability

All raw read data are available at the European Nucleotide Archive (ENA) with the study accession number PRJEB44352 (sample metadata, including sample names and primer-adapter sequences, is available in Table S20). Processed read data is available in the supporting information (Tables S7 - S14).

References

Aario, L. (1940). Waldgrenzen und subrezenten pollen-spektren in Petsamo Lappland. Annales Academiae Scientiarum Fenniae, Series A, 54(8), 1-120.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N. & Grüning, B.A. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids research, 46(W1), W537-W544.

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PloS one, 13(4), e0195403.

Anderson, R. S., & Van Devender, T. R. (1991). Comparison of pollen and macrofossils in packrat (Neotama) middens: A chronological sequence from the Waterman Mountains of southern Arizona, USA. Review of Palaeobotany and Palynology, 68(1-2), 1-28.

Ashastina, K., Kuzmina, S., Rudaya, N., Troeva, E., Schoch, W. H., Römermann, C., ... & Kienast, F. (2018). Woodlands and steppes: Pleistocene vegetation in Yakutia's most continental part recorded in the Batagay permafrost sequence. Quaternary Science Reviews, 196, 38-61.

Axmanová, I., Robovský, J., Tichý, L., Danihelka, J., Troeva, E., Protopopov, A., & Chytrý, M. (2020). Habitats of Pleistocene megaherbivores reconstructed from the frozen fauna remains. Ecography, 43(5), 703-713.

Belanger, R. J., Edwards, M. A., Carbyn, L. N., & Nielsen, S. E. (2020). Evaluating trade-offs between forage, biting flies, and footing on habitat selection by wood bison (Bison bison athabascae). Canadian Journal of Zoology, 98(4), 254-261.

Bellemain, E., Davey, M. L., Kauserud, H., Epp, L. S., Boessenkool, S., Coissac, E., ... & Brochmann, C. (2013). Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. Environmental microbiology, 15(4), 1176-1189.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. Nucleic acids research, 41(D1), D36-D42.

Bergerud, A. T. (1972). Food habits of Newfoundland caribou. The Journal of Wildlife Management, 913-923.

Beug, H. J. (2004). Leitfaden der pollenbestimmung für Mittelereuropa und angrenzende Gebiete, Verlag Dr. Fridrich Pfeil, München.

Birks, H. H. (2007). Plant macrofossil introduction. Encyclopedia of quaternary science, 3, 2266-2288.

Boast, A. P., Weyrich, L. S., Wood, J. R., Metcalf, J. L., Knight, R., & Cooper, A. (2018). Coprolites reveal ecological interactions lost with the extinction of New Zealand birds. Proceedings of the National Academy of Sciences, 115(7), 1546-1551.

Boertje, R. D. (1984). Seasonal diets of the Denali caribou herd, Alaska. Arctic, 161-165.

Boeskorov, G. G., Potapova, O. R., Mashchenko, E. N., Protopopov, A. V., Kuznetsova, T. V., Agenbroad, L., & Tikhonov, A. N. (2014). Preliminary analyses of the frozen mummies of mammoth (Mammuthus primigenius), bison (Bison priscus) and horse (Equus sp.) from the Yana-Indigirka Lowland, Yakutia, Russia. Integrative zoology, 9(4), 471-480.

Boeskorov, G. G., Potapova, O. R., Protopopov, A. V., Plotnikov, V. V., Agenbroad, L. D., Kirikov, K. S., ... & van der Plicht, J. (2016). The Yukagir Bison: The exterior morphology of a complete frozen mummy of the extinct steppe bison, Bison

priscus from the early Holocene of northern Yakutia, Russia. Quaternary International, 406, 94-110.

Boeskorov, G. G., Potapova, O. R., Protopopov, A. V., Plotnikov, V. V., Maschenko, E. N., Shchelchkova, M. V., ... & Tikhonov, A. N. (2018). A study of a frozen mummy of a wild horse from the Holocene of Yakutia, East Siberia, Russia. Mammal Research, 63(3), 307-314.

Boessenkool, S., McGlynn, G., Epp, L. S., Taylor, D., Pimentel, M., Gizaw, A., ... & Popp, M. (2014). Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. Conservation biology, 28(2), 446-455.

Boufford, D. E., Brouillet, L., Charron, M. T. M., Ford, B. A., Freeman, C. C., Gandhi, K. N., ... & Ickert

Bond, S. M. (2021). Flora of North America.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular ecology resources, 16(1), 176-182.

Brach, A. R., & Song, H. (2006). eFloras: New directions for online floras exemplified by the Flora of China Project. Taxon, 55(1), 188-192.

Brouillet, L., Coursol, F., Meades, S. J., Favreau, M., Anions, M., Bélisle, P., & Desmet, P. (2010). VASCAN, the database of vascular plants of Canada. Website http://data. canadensys.net/vascan/[accessed November 2020].

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Molecular ecology resources, 16(1), 138-149.

Chua, P. Y., Lammers, Y. Y., Menoni, E., Ekrem, T., Bohmann, K., Boessenkool, S., & Alsos, I. G. (2021). Molecular dietary analyses of western capercaillies (Tetrao urogallus) reveal a diverse diet. bioRxiv, 434346.

Chytrý, M., Horsák, M., Danihelka, J., Ermakov, N., German, D. A., Hájek, M., ... & Valachovič, M. (2019). A modern analogue of the Pleistocene steppe-tundra ecosystem in southern Siberia. Boreas, 48(1), 36-56.

Cody, W. J. (2000). Flora of the Yukon territory. NRC Research Press.

Cooper, A. & Poinar, H. N. (2000). Ancient DNA: do it right or not at all. Science, 5482(1139), 416.

Crum, H., Anderson, L. E., & Anderson, L. (1981). Mosses of Eastern North America (Vol. 1). Columbia University Press.

Denryter, K. A., Cook, R. C., Cook, J. G., & Parker, K. L. (2017). Straight from the caribou's (Rangifer tarandus) mouth: detailed observations of tame caribou reveal new insights into summer–autumn diets. Canadian Journal of Zoology, 95(2), 81-94.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19(1): 11-15.

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv, 081257.

Edwards, M. E., Brubaker, L. B., Lozhkin, A. V., & Anderson, P. M. (2005). Structurally novel biomes: a response to past warming in Beringia. Ecology, 86(7), 1696-1703.

Elven, R., Murray, D. F., Razzhivin, V. Y., & Yurtsev, B. A. (2011). Annotated checklist of the panarctic flora (PAF) vascular plants. Natural History Museum, University of Oslo.

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., ... & Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. Molecular ecology, 21(8), 1821-1833.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome, 2(1), p. 6.

Faegri, K., & Iversen, J. (1989). Textbook of pollen analysis (No. Ed. 4 by Faegri, K., Kaland, P. E., & Krzywinski, K.). John Wiley & Sons Ltd.

Galloway, J. M., Adamczewski, J., Schock, D. M., Andrews, T. D., MacKay, G., Bowyer, V. E., ... & Kutz, S. J. (2012). Diet and habitat of mountain woodland caribou inferred from dung preserved in 5000-year-old alpine ice in the Selwyn Mountains, Northwest Territories, Canada. Arctic, 59-79.

Gavril'eva, L. D. (2011). Communities of the class Artemisietea jacuticae Gogoleva et al. 1987 in alas meadows of Central Yakutia. Contemporary problems of ecology, 4(3), 339-345.

Gravendeel, B., Protopopov, A., Bull, I., Duijm, E., Gill, F., Nieman, A., ... & van Geel, B. (2014). Multiproxy study of the last meal of a mid-Holocene Oyogos Yar horse, Sakha Republic,

Russia. The Holocene, 24(10), 1288-1296.
Grigoriev, S. E., Fisher, D. C., Obadă, T., Shirley, E. A., Rountrey, A. N., Savvinov, G. N., ... & Tikhonov, A. N. (2017). A woolly mammoth (Mammuthus primigenius) carcass from maly Lyakhovsky island (new siberian islands, Russian Federation). Quaternary International, 445, 89-103.
Grímsson, F., Zetter, R., & Leng, Q. (2012). Diverse fossil Onagraceae pollen from a Miocene palynoflora of north-east China: early steps in resolving the phytogeographic history of the family. Plant systematics and Evolution, 298(3), 671-687.
Guthrie, R. D. (1990). Frozen fauna of the mammoth steppe, vol. 323 University of Chicago Press, p. 111-112.
Guthrie, R. D. (2001). Origin and causes of the mammoth steppe: a story of cloud cover, woolly mammal tooth pits, buckles, and inside-out Beringia. Quaternary science reviews, 20(1-3), 549-574.
Haarsma, A. J., Siepel, H., & Gravendeel, B. (2016). Added value of metabarcoding combined with microscopy for evolutionary studies of mammals. Zoologica Scripta, 45, 37-49.
Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Nucleic Acids Symp. Ser. (Vol. 41, pp. 95-98).
Harington, C. R. (2002). Yukon horse. Yukon Business, Tourism & Culture, Heritage Resources Unit.
Harington, C. R., & Eggleston-Stott, M. (1996). Partial carcass of a small Pleistocene horse from Last Chance Creek near Dawson City, Yukon. Current Research in the Pleistocene, 13, 105-107.
Hebert, P. D., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(1512), 313-321.
Hester, A. J., Bergman, M., Iason, G. R., & Moen, J. (2006). Impacts of large herbivores on plant community structure and dynamics. Large Herbivore Ecology, Ecosystem Dynamics and Conservation, Cambridge University Press, 11, 97.
Hofreiter, M., Poinar, H. N., Spaulding, W. G., Bauer, K., Martin, P. S., Possnert, G., & Pääbo, S. (2000). A molecular analysis of ground sloth diet through the last glaciation. Molecular Ecology, 9(12), 1975-1984.
Hofreiter, M., Betancourt, J. L., Sbriller, A. P., Markgraf, V., & McDonald, H. G. (2003). Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen
Province, southwest Argentina. Quaternary Research, 59(3), 364-378.
Hopkins, D. M. (1959). Cenozoic history of the Bering land bridge. Science, 129(3362), 1519-1528.
Hopkins, D. M., Matthews, J. V., Schweger, C. E. & Young, H. S. (Eds.). (1982). Paleoecology of Beringia. Academic Press.
Ihrmark, K., Bödeker, I., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., ... &
Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. FEMS microbiology ecology, 82(3), 666-677.
Jacobson, C. A. (1915). CICUTOXIN: THE POISONOUS PRINCIPLE IN WATER HEMLOCK (CICUTA). Journal of the American Chemical Society, 37(4), 916-934.
Janská, V., Jimenez-Alfaro, B., Chytrý, M., Divíšek, J., Anenkhonov, O., Korolyuk, A., ... & Culek, M. (2017). Palaeodistribution modelling of European vegetation types at the Last Glacial Maximum using modern analogues from Siberia: Prospects and limitations. Quaternary Science Reviews, 159, 103-115.
Johnson, C. N. (2009). Ecological consequences of Late Quaternary extinctions of megafauna. Proceedings of the Royal Society B: Biological Sciences, 276(1667), 2509-2519.
Jørgensen, T., Haile, J., Möller, P. E. R., Andreev, A., Boessenkool, S., Rasmussen, M., ... & Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Molecular Ecology, 21(8), 1989-2003.
Jung, T. S., Stotyn, S. A., & Czetwertynski, S. M. (2015). Dietary overlap and potential competition in a dynamic ungulate community in northwestern Canada. The Journal of Wildlife Management, 79(8), 1277-1285.
Kaczensky, P., Šturm, M. B., Sablin, M. V., Voigt, C. C., Smith, S., Ganbaatar, O., ... & Spasskaya, N. N. (2017). Stable isotopes reveal diet shift from pre-extinction to reintroduced Przewalski's horses. Scientific reports, 7(1), 1-9.
Kelly, A., Miller, J. H., Wooller, M. J., Seaton, C. T., Druckenmiller, P., & DeSantis, L. (2021). Dietary paleoecology of bison and horses on the

mammoth steppe of eastern Beringia based on dental microwear and mesowear analyses. Palaeogeography, Palaeoclimatology, Palaeoecology, 572, 110394.

Kienast, F., Schirrmeister, L., Siegert, C., & Tarasov, P. (2005). Palaeobotanical evidence for warm summers in the East Siberian Arctic during the last cold stage. Quaternary Research, 63(3), 283-300.

Kosintsev, P. A., Lapteva, E. G., Korona, O. M., & Zanina, O. G. (2012a). Living environments and diet of the Mongochen mammoth, Gydan Peninsula, Russia. Quaternary International, 276, 253-268.

Kosintsev, P. A., Lapteva, E. G., Trofimova, S. S., Zanina, O. G., Tikhonov, A. N., & Van der Plicht, J. (2012b). Environmental reconstruction inferred from the intestinal contents of the Yamal baby mammoth Lyuba (Mammuthus primigenius Blumenbach, 1799). Quaternary International, 255, 231-238.

Lawton, E. (1971). Moss flora of the Pacific Northwest. J Hattori Bot Lab, 1, 1-362.

Leonard, J. L., Perkins, L. B., Lammers, D. J., & Jenks, J. A. (2017). Are bison intermediate feeders? Unveiling summer diet selection at the northern fringe of historical distribution. Rangeland ecology & management, 70(4), 405-410.

CBOL plant Working Group, Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., ... & Duan, G. W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences, 108(49), 19641-19646.

Lozhkin, A. V., Anderson, P. M., Glushkova, O. Y., & Vazhenina, L. N. (2019). Late Quaternary environments on the far southwestern edge of Beringia. Quaternary Science Reviews, 203, 21-37.

Lydolph, M. C., Jacobsen, J., Arctander, P., Gilbert, M. T. P., Gilichinsky, D. A., Hansen, A. J., ... & Lange, L. (2005). Beringian paleoecology inferred from permafrost-preserved fungal DNA. Applied and environmental microbiology, 71(2), 1012-1017.

MacFadden, B. J., Solounias, N., & Cerling, T. E. (1999). Ancient diets, ecology, and extinction of 5-million-year-old horses from Florida. Science, 283(5403), 824-827.

Marsolier-Kergoat, M. C., Palacio, P., Berthonaud, V., Maksud, F., Stafford, T., Bégouën, R., & Elalouf, J. M. (2015). Hunting the extinct steppe bison (Bison priscus) mitochondrial genome in the Trois-Freres Paleolithic Painted Cave. PLoS One, 10(6), e0128267.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal, 17(1), 10-12.

Mauquoy, D., & Van Geel, B. (2007). Plant macrofossil methods and studies: mire and peat macros. In Encyclopedia of quaternary science (pp. 2315-2336). Elsevier Science.

Mendoza, M., & Palmqvist, P. (2008). Hypsodonty in ungulates: an adaptation for grass consumption or for foraging in open habitat? Journal of Zoology, 274(2), 134-142.

Mol, D., Shoshani, J., Tikhonov, A., Van Geel, B., Sano, S. I., Lazarev, P., ... & Agenbroad, L. D. (2006). The Yukagir Mammoth: brief history, 14C dates, individual age, gender, size, physical and environmental conditions and storage. Scientific Annals, School of Geology, Aristotle University of Thessaloniki, Special Volume, 98, 299-314.

Moore, P. D., Webb, J. A., & Collison, M. E. (1991). Pollen analysis. Blackwell scientific publications.

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., ... & Kennedy, P. G. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology, 20, 241-248.

Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., ... & Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic acids research, 47(D1), D259-D264.

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

Parducci, L., Väliranta, M., Salonen, J. S., Ronkainen, T., Matetovici, I., Fontana, S. L., ... & Suyama, Y. (2015). Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1660), 20130382.

Parducci, L., Alsos, I. G., Unneberg, P., Pedersen, M. W., Han, L., Lammers, Y., ... & Wohlfarth, B. (2019). Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from southern Sweden. Frontiers in Ecology and Evolution, 7, 189.

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., ... & Kjær, K. H. (2013). A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quaternary Science Reviews, 75, 161-168.

Pino-Bodas, R., & Stenroos, S. (2020). Global biodiversity patterns of the photobionts associated with the genus Cladonia (Lecanorales, Ascomycota). Microbial Ecology, 1-15.

POWO. (2019). Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew.

Rawlence, N. J., Lowe, D. J., Wood, J. R., Young, J. M., Churchman, G. J., HUANG, Y. T., & Cooper, A. (2014). Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. Journal of Quaternary Science, 29(7), 610-626.

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., ... & Talamo, S. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon, 62(4), 725-757.

Reitsma, T. J. (1966). Pollen morphology of some European Rosaceae. Acta Botanica Neerlandica, 15(2), 290-307.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, e2584.

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. Nature protocols, 2(7), 1756-1762.

Savage, D. A., & Heller, V. G. (1947). Nutritional qualities of range forage plants in relation to grazing with beef cattle on the Southern Plains Experimental Range

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics, 27(6), 863-864.

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., ... & Yoccoz, N. G. (2015). Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PloS one, 10(1), e0115335.

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., ... & Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Molecular Ecology Resources, 10(6), 1009-1018.

Stech, M., Kolvoort, E., Loonen, M. J. J. E., Vrieling, K., & Kruijer, J. D. (2011). Bryophyte DNA sequences from faeces of an arctic herbivore, barnacle goose (Branta leucopsis). Molecular ecology resources, 11(2), 404-408.

Swanson, D. K. (2006). Biogeographical evidence for the grass (Poaceae) species of Pleistocene Beringian lowlands. Arctic, 191-200.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., ... & Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research, 35(3), e14-e14.

Team, R. C. (2013). R: A language and environment for statistical computing.

Troeva, E. I., Isaev, A. P., Cherosov, M. M., & Karpov, N. S. (Eds.). (2010). The Far North:: Plant Biodiversity and Ecology of Yakutia (Vol. 3). Springer Science & Business Media.

Uno, K. T., Fisher, D. C., Wittemyer, G., Douglas-Hamilton, I., Carpenter, N., Omondi, P., & Cerling, T. E. (2020). Forward and inverse methods for extracting climate and diet information from stable isotope profiles in proboscidean molars. Quaternary International, 557, 92-109.

Van der Plicht, J., & Hogg, A. (2006). A note on reporting radiocarbon. Quaternary Geochronology, 1(4), 237-240.

van Geel, B., Aptroot, A., Baittinger, C., Birks, H. H., Bull, I. D., Cross, H. B., ... & van Tienderen, P. H. (2008). The ecological implications of a Yakutian mammoth's last meal. Quaternary Research, 69(3), 361-376.

van Geel, B., Fisher, D. C., Rountrey, A. N., van Arkel, J., Duivenvoorden, J. F., Nieman, A. M., ... & Gravendeel, B. (2011a). Palaeo-environmental and dietary analysis of intestinal contents of a mammoth calf (Yamal Peninsula, northwest Siberia). Quaternary Science Reviews, 30(27-28), 3935-3946.

van Geel, B., Guthrie, R. D., Altmann, J. G., Broekens, P., Bull, I. D., Gill, F. L., ... & Gravendeel, B. (2011b). Mycological evidence of coprophagy from the feces of an Alaskan Late Glacial mammoth. Quaternary Science Reviews, 30(17-18), 2289-2303.

Van Geel, B., Protopopov, A., Bull, I., Duijm, E., Gill, F., Lammers, Y., ... & Gravendeel, B. (2014). Multiproxy diet analysis of the last meal of an early Holocene Yakutian bison. Journal of Quaternary Science, 29(3), 261-268.

Vitt, D. H., & Buck, W. R. (1992). Key to the moss genera of North America north of Mexico.

University of Michigan Herbarium.
White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications, 18(1), 315-322.
Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., ... & Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47-51.
Williams, J. W., & Jackson, S. T. (2007). Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment, 5(9), 475-482.
Wood, J. R., & Wilmshurst, J. M. (2016). A protocol for subsampling Late Quaternary coprolites for multi-proxy analysis. Quaternary Science Reviews, 138, 1-5.
Zazula, G. D., Schweger, C. E., Beaudoin, A. B., & McCourt, G. H. (2006). Macrofossil and pollen evidence for full-glacial steppe within an ecological mosaic along the Bluefish River, eastern Beringia. Quaternary International, 142, 2-19.
Zazula, G. D., Froese, D. G., Elias, S. A., Kuzmina, S., & Mathewes, R. W. (2007). Arctic ground squirrels of the mammoth-steppe: paleoecology of Late Pleistocene middens (~ 24 000–29 450 14C yr BP), Yukon Territory, Canada. Quaternary Science Reviews, 26(7-8), 979-1003.
Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), 614-620.
Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin Iii, F. S. (2012). Mammoth steppe: a high-productivity phenomenon. Quaternary Science Reviews, 57, 26-45.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quascirev.2021.107084

Chapter 4

Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene

Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene

Anneke T.M.ter Schure, Manon Bajard, Kjetil Loftsgarden, Helge I.Høeg, Eirik Ballo, Jostein Bakke, Eivind W.N.Støren, Frode Iversen, Anneleen Kool, Anne K.Brysting, Kirstin Krüger, Sanne Boessenkool

Published in Quaternary Science Reviews 270, 107175, 2021

Highlights

- Integration of pollen, *sed*aDNA, geochemical, archaeological and climate data.
- Natural processes drove vegetation succession at Lake Ljøgottjern until Early Iron Age.
- Matching pollen and *sed*aDNA records show rapid intensification of agropastoralism.
- Deviations between pollen and *sed*aDNA reflect distance to farms since the Bronze Age.
- Major plant community shift coincided with growing human population in Early Iron Age.

Abstract

Uncovering anthropogenic and environmental drivers behind past biological change requires integrated analyses of long-term records from a diversity of disciplines. We applied an interdisciplinary approach exploring effects of human land-use and environmental changes on vegetation dynamics at Lake Ljøgottjern in southeastern Norway during the Holocene. Combined analysis of pollen and sedimentary ancient DNA (*sed*aDNA) metabarcoding of the sedimentary sequence of the lake describes the vegetation dynamics at different scales, and establishes a timeline for pastoral farming activities. We integrate this reconstruction with geochemical analysis of the sediments, climate data, archaeological evidence of local human settlement and regional human population dynamics.

Our data covering the last 10,000 years reveals consistent vegetation signals from pollen and *sed*aDNA indicating periods of deforestation connected to cultivation, matching the archaeological evidence. Multivariate analysis integrating the environmental data from geochemical and archaeological reconstructions with the vegetation composition indicates that the vegetation dynamics at Lake Ljøgottjern were primarily related to natural processes from the base of the core (in ca. 8000 BCE, Mesolithic) up to the Early Iron Age (ca. 500 BCE–550 CE), when agricultural activities in the region intensified. The pollen signal

reflects the establishment of a Bronze Age (ca. 1800–500 BCE) farm in the area, while subsequent intensification of pollen concentrations of cultivated plants combined with the first *sed*aDNA signals of cultivation and pastoralism are consistent with evidence of the establishment of farming closer to the lake at around 300 BCE. These signals also correspond to the intensification of agriculture in southeastern Norway in the first centuries of the Early Iron Age. Applying an interdisciplinary approach allows us to reconstruct anthropogenic and environmental dynamics, and untangle effects of human land-use and environmental changes on vegetation dynamics in southeastern Norway during the Holocene.

1. Introduction

Understanding the processes that shaped our modern ecosystems is important for explaining the role of anthropogenic and environmental factors in biological change. It is increasingly recognized that our modern ecosystems are the result of a long history of human-environment interactions (Boivin and Crowther, 2021; Boivin et al., 2016; Ellis, 2015; Ruddiman, 2003; Ruddiman et al., 2015; Williams et al., 2015), with human land-use as one of the major drivers of ecosystem change (Boivin et al., 2016; Ellis, 2015; Nelson et al., 2006; Vitousek et al., 1997). In turn, land-use transitions can be driven by cultural, ecological, and climatic shifts (Boivin et al., 2016; Ellis, 2015; Stephens et al., 2019), illustrating the need for integrated analysis of long-term records from a diversity of disciplines. Here, we combine multi-proxy analysis of pollen, *sed*aDNA and geochemical data with knowledge about the local human history as well as palaeodemographic and climatic changes at an inland lake site in southeastern Norway, an area where knowledge on the development of agrarian societies is scarce. We obtain a detailed reconstruction of the palaeoenvironment uncovering the anthropogenic and environmental drivers of biological change.

While in central Europe there were fully developed agrarian societies at around 4000 BCE, Scandinavia by that time was populated by hunter-gatherer-fisher groups and agricultural practices were not adopted until roughly 1500 years later (Bonsall et al., 2002; Krause-Kyora et al., 2013; Price, 2000). Most of what we currently know of the formation of anthropogenic landscapes in Scandinavia comes from pollen analyses and archaeological records, in particular from coastal areas. These records indicate regional differences in timing and rates of the transition from hunter-gatherer-fisher groups to fully developed agrarian societies (Hjelle et al., 2018; Robinson, 2003; Wieckowska-Lüth et al., 2018, 2017), suggesting anthropogenic landscapes were formed through a combination of both rapid introductions by migration processes as well as gradual change within indigenous populations. Ancient mitochondrial DNA analysis of human bones and teeth from coastal areas of Sweden and the Baltic Sea islands suggests Neolithic or post-Neolithic population replacement by migrating farmers (Malmström et al.,

2009), possibly driven by improved climatic conditions, allowing agrarian societies from central Europe to migrate northward, introducing crops and agricultural practices (Bonsall et al., 2002; Warden et al., 2017).

Archaeological evidence from Denmark and southern Sweden highlights differences between coastal and inland sites (Sørensen and Karg, 2014), with more rapid introductions of domesticated animals and cereal cultivation in inland sites compared to coastal sites. However, studies on the development of agrarian societies in inland Scandinavia are relatively scarce, possibly because mesolithic settlements were largely shore bound (Solheim and Persson, 2018; Wieckowska-Lüth et al., 2018). Coastal sites are also easier to detect, while bioarchaeological evidence from inland sites is limited, likely due to poorer preservation conditions (Sørensen and Karg, 2014). Within this limited knowledge from inland agrarian shifts, southeastern Norway is particularly underrepresented. Thus far, there are vegetation reconstructions based on pollen from coastal areas in southeastern Norway suggesting small-scale cereal cultivation and animal husbandry in the Early Neolithic (ca. 4000–3300 BCE; Wieckowska-Lüth et al., 2017), crop plant use in the Early Bronze Age (1800 BCE; Soltvedt and Henningsmoen, 2016) and more established cultivation at a later date depending on the region. Long-term settlement of the coastal regions of the Oslo-fjord during the Mesolithic and first part of the Neolithic (8500–2000 BCE; Solheim and Persson, 2018) is evident from archaeological remains and a relatively recent case study from this same region found temporal variation in the use of the coast and the directly adjacent landscape throughout the Mesolithic (ca. 8500–4000 BC; Wieckowska-Lüth et al., 2018). Pollen analyses of several inland sites in the Romerike region (northeast of Oslo) indicate a delayed introduction of cultivated plants to ca. 2000 BCE following indications of grazing activity from ca. 3000 BCE (Høeg, 1997).

Reconstruction of the extent, intensity, duration and biological consequences of early human land-use requires integrated analyses of long-term records from a diversity of disciplines, combining knowledge about the local human history as well as climatic and other environmental changes. Temporal changes in biological communities can be caused by biological processes such as succession and species dispersal, but these biological changes can be accelerated by environmental factors, e.g. temperature shifts, precipitation, or cultural changes such as through human-mediated introduction of cultivated plants and pastoral animals (Ellis, 2015; Nelson et al., 2006; Vitousek et al., 1997). Identification of the drivers of biological change can be difficult as direct comparison between sites and proxies of environmental change are not always possible, for example due to differences between age-depth models or in spatiotemporal resolution. Detailed multi-proxy reconstructions of the same palaeoenvironmental record are needed to allow comparison and validation of environmental proxies and direct inferences on the timing of environmental changes (Bajard et al., submitted). Lake sediments can be good archives of the palaeoenvironment, integrating biotic and abiotic information across the lake

catchment area with temporal stratification. Analysis of pollen and macrofossil remains from lake sediments have previously provided insight in past vegetation and landscape changes, and more recently, sedimentary ancient DNA (*sed*aDNA) analysis has proven valuable for vegetation reconstructions, providing a more local signal and at a higher taxonomic resolution than previously possible (Giguet-Covex et al., 2019; Parducci et al., 2017). Moreover, *sed*aDNA analyses are not limited to the detection of plants and have for instance also been applied to reconstruct pastoral activities (Bajard et al., 2020, 2017; Giguet-Covex et al., 2014) and uncover the effects of these activities, climate change and soil evolution on plant communities (Pansu et al., 2015). By combining pollen analysis with archaeological data, Wieckowska-Lüth et al. (2018) illustrate the value of an integrated approach, allowing the reconstruction of both the intensity and duration of past human land-use. Where archaeological evidence can give insight into societal changes, human settlement patterns and population dynamics, evidence of past presence of cultivated crops, pastoral animals and charcoal can be used to reconstruct a timeline of human presence and land-use changes. Geochemical analysis can further be used to reconstruct past abiotic changes, including physical and chemical weathering, climatic changes and sedimentation processes. Combining these lines of evidence enables identification of periods of biological change, as well as associated abiotic and human land-use changes and their relative timing, thereby allowing inferences about potential driving factors of biological change.

In this study we disentangled anthropogenic and environmental changes on the vegetation dynamics during the last 10,000 years by applying an interdisciplinary approach, focusing on an inland lake site located in southeastern Norway. Lake Ljøgottjern is of particular archaeological interest due to the rich archaeology surrounding the site, including the largest burial mound in Scandinavia: Raknehaugen (Skre, 1997). A new biological reconstruction based on pollen combined with *sed*aDNA metabarcoding describes the vegetation dynamics on different spatial scales, and establishes a timeline for pastoral and arable farming activities. We integrate this reconstruction with geochemical analysis of the same lake sediment record, published climate data and archaeological evidence of local human settlement and regional human population dynamics. Our interdisciplinary approach allows us to reconstruct natural vegetation dynamics during the Holocene, establish a timeline for mixed (pastoral and arable) farming activities, and correlate land-use and environmental changes to vegetation dynamics to uncover human-environment interactions at Lake Ljøgottjern.

2. Regional setting

Lake Ljøgottjern (60°8'54"N, 11°8'18"E, 185 m a.s.l., area: 1.8 hm^2) is located in the middle of the historic Romerike region in southeastern Norway (Fig. 1). It has a maximum depth of 18 m, no inlet or outlet of water, and a small catchment area

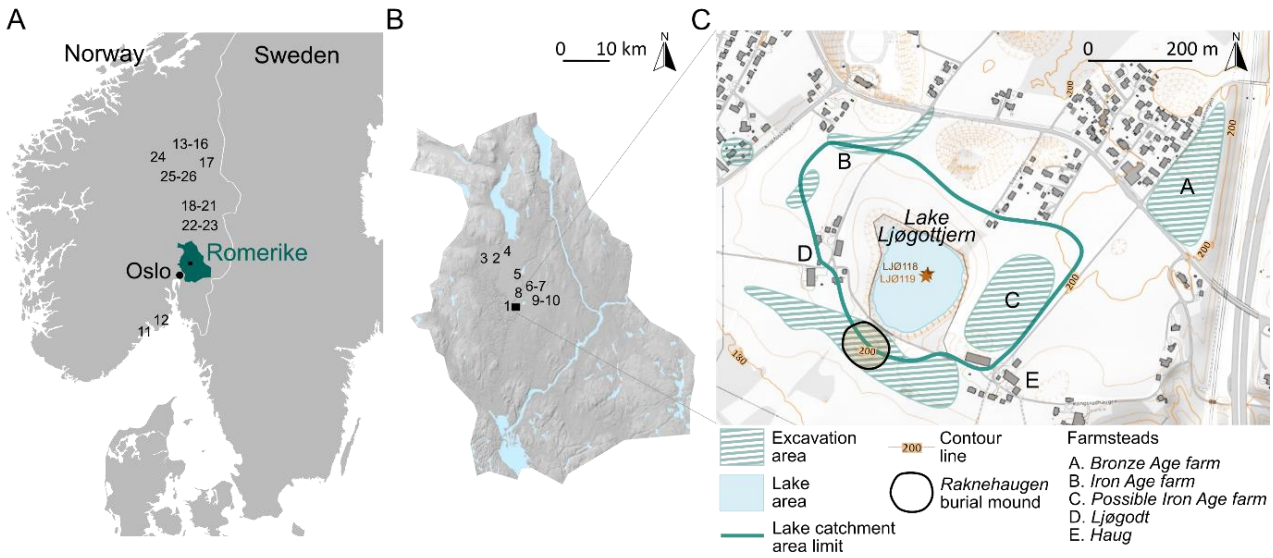


Figure 1. Location of the areas under study, with the historic Romerike region in southeastern Norway (A), Lake Ljøgottjern in the historic Romerike region (B), and a topographic map of the area surrounding Lake Ljøgottjern from www.kartverket.no (C), including the lake catchment area indicated with a solid blue line, stars indicating the location of the sediment cores in the lake, contour lines in solid orange, archaeological excavation areas indicated with hatching, and farmsteads of interest indicated with capital letters. Raknehaugen is a burial mound dated to ca. 550 CE (Skre, 1997) and indicated here with a black circle. Numbers in (A) and (B) indicate locations of pollen records mentioned in the text and correspond to: 1. Lake Ljøgottjern, described in this study, 2. Rud Øde, 3. Bjørkemosan, 4.Lybekkmosan, 5. Danielsetermyr, 6. Bånntjern in Ullensaker, 7. Svenskestutjern, 8. Skånetjern, 9. Myr ved Pinnebekk, 10. Myr ved Brenni, described by Høeg (1997); 11. Skogstjern, described by Wieckowska-Lüth et al. (2017); 12. Nordbytjern, described by Soltvedt and Henningsmoen (2016); 13. Bånntjern in Tolga, 14. Stortallsjøen, 15. Lensmannsvollen, 16. Kåsmyra, 17. Lille Sølensjøen, 18. Ottersmyra, 19. Dulpmoen, 20. Ulvehammeren, 21. Kilde, 22. Engelaug, 23. Hellemundsmyra, 24. Grimsdalen, 25. Skjerdingsfjell, and 26. Hirsjøen, described by Høeg (1996).

of 0.15 km^2. Lake Ljøgottjern was formed by a melting block of dead ice after the retreat of the Scandinavian Ice Sheet ca. 10,000 years ago (Longva and Thoresen, 1989). Sand, gravel and marine clay deposits from the last ice age combined with fluvial deposits from several rivers and the relatively flat geography of the area support the agriculture in the Romerike region. Built on the shore of the lake is the Raknehaugen burial mound, dated to ca. 551 CE (Skre, 1997) and possibly marking the centre of the petty kingdom of Romerike during the Migration Period (ca. 400–570 CE). Raknehaugen is the largest burial mound in Northern Europe at 15 m high, 77 m in diameter, consisting of three substantial layers of timber (pine and birch), with soil and sand from the surroundings. The earliest settlements in this area are dated to the Bronze Age (ca. 1800–500 BCE) and several farmsteads have been located around the lake during the Iron Age (ca. 550 BCE–1050 CE; Helliksen, 1997), while the farmsteads Haug and Ljøgot are known from medieval sources (Rygh, 1898, pp. 322–323), see Fig. 1C. Today, there is still continuous agriculture around Lake Ljøgottjern.

3. Materials and Methods

3.1. Sediment coring and sub-sampling

Lake Ljøgottjern was cored at the deepest point in November 2018 (core number LJØ118) and May 2019 (core number LJØ119). The corings were done from a raft using a modified piston corer equipped with a 110 mm diameter, 6 m long PVC tube (Nesje, 1992). Additionally, a 90 mm-Uwitec gravity corer was used to capture the sediment-water interface. The cores were cut into sections of ~140 cm in the field to facilitate transportation and were sealed immediately to reduce the risk of contamination. Cores were preserved in the dark and cold (~4°C) until opening at the Earth Surface Sediment Laboratory EARTHLAB at the University of Bergen. For each core, one half was used for geochemical analysis, while the other half was used for sediment sampling. Samples for C14 dating were taken from LJØ118, while *sed*aDNA and pollen samples were taken from the LJØ119 and the gravity core. A high sampling and thus temporal resolution was targeted for all analyses as further described below.

Half-cores were sub-sampled at 3 cm intervals for *sed*aDNA immediately upon opening, using 5 mL sterile disposable syringes following the protocol described by Epp et al. (2019). 1 cm^3 sub-samples were subsequently collected at 2 cm intervals for pollen and Non Pollen Palynomorphs (NPPs) analyses. *Sed*aDNA samples were transported to the ancient DNA facilities at the University of Oslo and kept at -20°C until DNA extraction.

3.2. Geochemical analysis and chronology

Both cores were scanned with an ITRAX (XRF) core scanner from COX analytics at the EARTHLAB as described by Bajard et al. (submitted; see also Appendix A.1) with a resolution of 200 µm for LJØ118 and 1000 µm for LJØ119. To obtain matching resolutions of 1 mm and correlate the two cores, Ca, Ti, K, Si, Fe, Mn, Inc and Coh data of LJØ118 were averaged every 5 measures. The depth of the cores were correlated based on the variations in Ti, Fe and Mn and visual observation of the sediment using QAnalyseries 1.4.2 (Kotov and Paelike, 2018). The chronology of the sediment sequence is based on six AMS (accelerator mass spectrometer) radiocarbon dates from plant macrofossils from LJØ118 and realized by the Tandem Laboratory at Uppsala University. The ^{14}C ages were calibrated using the IntCal20 calibration curve (Reimer et al., 2020). The age-depth model for the LJØ118 sequence was generated using R software (version 3.5.2; R Core Team, 2020) and the R code package 'Bacon' 2.4.3 (Blaauw and Christen, 2011). In the age model, the top of the core was set to the year of coring, i.e. 2018 CE. The chronology of LJØ119 was deduced from the LJØ118 age model.

XRF measurements of Ti were used as a proxy for terrestrial sediment influx into the lake, while the Inc/Coh ratio was used to estimate changes in the organic matter content of the lake sediment. Peaks of Ti can reflect erosion events such as heavy rainfall and floods (Bajard et al., submitted), or human activities (Arnaud et

al., 2016; Bajard et al., 2016). To facilitate statistical analyses, time resolutions of the Ti measurements and Inc/Coh ratio were matched to the resolution of the vegetational datasets from the pollen and the *sed*aDNA analyses using generalised additive models (GAM; Appendix A.1).

3.3. Pollen and charcoal analyses

We prepared 131 sediment sub-samples of 1 cm^3 as described in Faegri et al. (1989). For the period between 200 and 1250 CE, pollen samples were analysed in timesteps of ~18 years (every 1–2 cm) while timesteps were longer for the rest of the period covered by the core (30-300 years; 3–40 cm). Tablets of spores (*Lycopodium clavatum*) were added to each sample to calculate pollen and charcoal concentrations (Stockmarr, 1971). At least 500 pollen grains of terrestrial plants were identified and counted in each sample. Further analyses are based on pollen concentrations to allow comparison to the *sed*aDNA data, as both estimates are standardized to a fixed amount of sediment. Pollen terrestrial plant diversity was determined by Hill numbers with the *hill_taxa* function of the *HillR* R package (Li, 2018; R Core Team, 2020), where q=0, 1 and 2 were used to obtain the number of identified pollen types, Shannon index and inverse Simpson index. Charcoal preserved in the lake sediments can be used as an indicator of human presence. To include charcoal in further statistical analyses we matched the calibrated ages of the samples and the *sed*aDNA dataset using linear interpolation.

3.4. DNA analyses

3.4.1. DNA extraction and amplification

We extracted DNA from 40 samples covering timesteps of ~80 years for the period 200–1800 CE (approx. every 9 cm) and timesteps of around 450–500 years in the rest of the period covered by the core. Six extraction negative controls were included and DNA was extracted using the PowerSoil DNA Isolation Kit (Qiagen) following manufacturers' protocol with two modifications: extracts were lysed overnight leaving the PowerBead Tubes with C1 solution at 37 °C for 16–20 hours under constant rotation (Epp et al., 2019), and eluted in 100 µL elution buffer supplied with the kit. DNA extraction, PCR preparation and post-PCR work were carried out at the ancient DNA lab facilities of the University of Oslo.

Plant DNA was amplified using the *trn*L *g* and *h* primers that amplify the P6 loop of the *trn*L intron (Taberlet et al., 2007), while mammal DNA was amplified using the Mam-P007 primers (Giguet-Covex et al., 2014). Forward and reverse primers were tagged with a unique 8 or 9 bp barcode at the 5' end to allow for multiplexing as described by Voldstad et al. (2020), with each primer pair having the same tag. Six individually tagged PCR replicates were prepared for each sample and primer set and at least one PCR negative control was included per 12 replicates. Plant DNA amplifications were carried out following Alsos et al. (2016), halving the total reaction volume to 25 µL, but keeping the same volume of bovine

serum albumin (8 µg; BSA, Roche Diagnostic, Basel, Switzerland) and DNA (5 µL). Mammal DNA amplifications were carried out in the same amplification mixture as for the plants, with the addition of human blocker (Boessenkool et al., 2012) at a concentration of 2 µM to restrict amplification of human DNA when amplifying mammalian DNA. Both the plant and mammal PCR mixtures were denatured at 95 °C for 10 min, followed by 45 cycles of 30 s at 95 °C, 30 s at 50 °C, 1 min at 72 °C and a 10 min final elongation at 72 °C.

We evaluated amplification success by gel electrophoresis before equivolume pooling of PCR products based on PCR band strength (including those that showed no band) to a maximum volume of 500 µL and cleaned using MinElute Purification kit (Qiagen) following manufacturers' instructions. Concentrations of the cleaned products were measured on a Qubit 2.0 with the Qubit dsDNA HS kit (ThermoFisher). Purified products were then pooled before sequencing, while preventing overlap in demultiplex-tags, resulting in two pools (of 320 and 298 PCR replicates). Libraries were built from the two pools using the KAPA HyperPrep DNA kit (Roche) with Illumina Unique Dual Indexes and these were sequenced on separate lanes on the Illumina HiSeq 4000 platform (2 x 150-bp, paired-end) at the Norwegian Sequencing Centre.

3.4.2. DNA sequence analyses and filtering

We processed the *sed*aDNA sequence data using the OBITools package (http://metabarcoding.org/obitools/doc/index.html; Boyer et al., 2016). Assembling the forward and corresponding reverse reads was done using *illuminapairedend*, followed by sample assignment with *ngsfilter*. We removed reads with a quality score <40, <100% tag match, >3 mismatches with the primers, shorter lengths then expected (<8 bp), with a sequence that is present exactly once, and those containing ambiguous nucleotides. Amplification and sequencing errors were identified using *obiclean*, with a threshold ratio of 5% for reclassification of sequences identified as 'internal' to their corresponding 'head' sequence. Sequences that were identified only as 'internal' were removed. Finally, sequences were compared to their relevant taxonomic reference libraries using *ecotag*. The reference libraries were prepared by performing an in-silico PCR with the *ecoPCR* software (Ficetola et al., 2010) using the NCBI Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy).

For the plant identifications, we used two reference libraries. The first reference library (arctborbryo) contains 2289 unique sequences, with 815 Arctic (Sønstebø et al., 2010), and 835 boreal (Willerslev et al., 2014) vascular plant taxa and 455 bryophytes (Soininen et al., 2015). To mitigate erroneous or missing taxonomic assignments due to lacking references in the first library, we prepared a second reference library based on the global EMBL database (release 142, January 2020), containing 19,533 unique sequences from 14,327 plant taxa. Also for the mammal identifications we prepared a reference library based on the

global EMBL database (release 142, January 2020), containing 41,257 unique sequences from 24,068 mammal taxa.

In order to minimise any misidentifications, we filtered the identified sequences in R (version 3.5.2; R Core Team, 2020) using similar requirements as Alsos et al. (2018) removing: (1) sequences with higher occurrence (i.e. more reads) in negative controls than in samples, (2) sequences with a <100% (plants) or <98% (mammals) match, (3) sequences with <10 reads in a PCR replicate (plants) or in the total dataset (mammals), (4) PCR replicates with <100 reads in total, (5) sequences present in <2 PCR replicates of a sample, and (6) sequences with a mean read count > mean read count in the negative controls. In order to check and where possible narrow down some of the taxonomic identifications, the identified plant taxa were checked by botanists with extensive knowledge of the local flora. For the mammal dataset we removed genera that are not of interest for this study. Remaining unique sequences were designated as molecular operational taxonomic units (MOTUs). An overview of the filtering steps and their effect on the size of the dataset can be found in Appendix Table A.2.

To correct for the exponential increase in read counts during PCR, we log-transformed the filtered plant and mammal MOTU datasets (Giguet-Covex et al., 2019) and calculated relative read abundances (RRAs) to further facilitate comparison between samples. Both plant and mammal datasets were reduced by merging the PCR replicates, while calculating the number of positive replicates per MOTU and the mean RRA values. We assessed the quality of the samples by computing the summed read counts and the average read counts (+-SE; Appendix A.3). No significant relationship between summed or average read counts and the plant biodiversity estimates per replicate or per sample was found (Appendix A.4), indicating absence of a potential bias in biodiversity estimates due to differences in sample size. We found significant correlations between the number of positive replicates and the transformed read counts per MOTU (Appendix A.5) and based subsequent analyses on the more conservative positive replicates.

Terrestrial plant diversity was determined by Hill numbers with the *hill_taxa* function of the *HillR* R package (Li, 2018), using q=0, 1, and 2 to obtain the number of MOTUs, Shannon index and inverse Simpson index (Chao et al., 2014). The presence of DNA from livestock represents a proxy for pastoralism and we include presence/absence of livestock DNA in subsequent statistical analyses.

3.5. Palaeodemographic analysis

Radiocarbon dates from archaeological sites can be used as a proxy to estimate of the past population densities, accepting the basic premise of a relationship between quantities of radiocarbon dates and intensity of past population or activity (Freeman et al., 2018; Rick, 1987; Solheim and Iversen, 2019; Loftsgarden and Solheim, in press).

We analyzed 476 radiocarbon dates from archaeological contexts in the Romerike region using Summed Probability Distribution (SPD) analysis within the *rcarbon* package (Crema and Bevan, 2021). Dates were calibrated using the Intcal20 calibration curve (Reimer et al., 2020). The use of SPDs allows us to study long-term developments on a spatial and temporal level. The obtained SPDs were

subsequently matched to the calibrated ages of the pollen and DNA samples to facilitate statistical comparison.

3.6. Climate data

We obtained published multi-proxy surface temperature anomaly composites from the northern hemisphere (Kaufman et al., 2020) and the North Atlantic-Fennoscandian region (Sejrup et al., 2016), with proxies including oxygen isotopes, alkenone biomarkers, assemblages of pollen, chironomids, and diatoms among others. The Sejrup et al. (2016) temperature anomaly reconstruction composite is based on 81 proxy derived surface and near-surface summer temperature time series from 74 lake and marine sites in the North Atlantic and Fennoscandia (40E-40W, 58-80N) spanning the last 10,000 years. Kaufman et al. (2020) present a multi-method 100 year time-step reconstruction of global mean surface temperatures of the last 12,000 years based on the temperature 12k database of palaeo-temperature time series. From this publication we included the median values of the 60–90N northern hemisphere temperature composite.

General Additive Models (GAMs) allow flexible modelling of nonlinear relationships (Simpson, 2018), and were fitted to the temperature data using the *gam* function in the *mgcv* R package (Wood, 2018) to smooth and interpolate values and match the time resolutions of the pollen and DNA datasets. Further statistical analyses were performed with temperature data from both the Kaufman et al. (2020) and the Sejrup et al. (2016) data, but as results were similar we only present results including the Sejrup et al. (2016) data in the main text and figures (see Appendix A.6 for results including the Kaufman et al., 2020, data).

3.7. Statistical analyses

All statistical analyses were performed in R. To summarize the vegetation data, pollen and plant DNA taxa were assigned to one of 10 plant groups: anthropochores, apophytes, forbs, graminoids, aquatics, bryophytes, clubmosses, ferns, shrubs, trees, and remaining taxa were grouped to “other” (Appendices B.1–3). To visualize changes in abundance of plant taxa over time (i.e. pollen concentrations, number of positive DNA replicates) and mammal genera, we created stratigraphic plots with the *rioja* package (Juggins and Juggins, 2020; Appendices A.7–8). This package was also used for stratigraphically constrained cluster analysis using the CONISS algorithm (Grimm, 1987), identifying zones of similar terrestrial plant community composition (see Appendix A.9 for details). For a visual summary of the changes in abundance of plant groups over time, we calculated pollen fractions and DNA replicate fractions by summing the pollen concentrations and positive DNA replicates per plant group and dividing this by the total sum. To obtain estimates of the accumulation rate of taxa per plant group, we calculated the cumulative numbers of MOTUs and pollen types using the *specaccum* function of the *vegan* package (Oksanen et al., 2020).

To investigate plant community trajectories we used non-metric multidimensional scaling (NMDS), and to test how plant community trajectories relate to environmental change we applied a step-wise distance-based redundancy analysis (dbRDA) based on Bray-Curtis dissimilarities of Hellinger-transformed data. Where NMDS allows analysis of the total variation in the plant taxonomic composition, it is less suited for fitting environmental variables than dbRDA. The latter was specifically developed to test the significance of relationships between environmental variables and biological response data (Legendre and Anderson, 1999), but only analyses the variation explained by the environmental terms (Ramette, 2007). The environmental terms (i.e. Ti, Inc/Coh ratio, charcoal, presence/absence of livestock, ^{14}C SPD, temperature anomaly) were standardized with the *decostand* function, removing unwanted effects of different measurement units. For the dbRDA, we performed automatic stepwise model building combining the *ordiR2step* and *capscale* functions to obtain the best fitting model. We subsequently tested for statistical significance (p<.05) of the included environmental terms with an anova (999 permutations) and calculated the proportion of plant community variance explained by each term in the model. We also performed variation partitioning using the *varpart* function for assessing joint effects of the environmental terms included in the dbRDA model. Spearman's rank correlation coefficient was used to identify relationships between environmental terms and we applied the bonferroni method to correct p-values for multiple comparisons. For further details on these steps, see Appendices A.10–12.

3.8. Archaeological evidence

We collected and analysed existing archaeological and historical evidence on settlements and land-use around Lake Ljøgottjern. We used the *Askeladden* database of Norwegian archaeological sites and monuments (https://askeladden.ra.no/; hosted by the Directorate for Cultural Heritage, 2020) and the *UniMus* database of archaeological artefacts and samples (http://unimus.no; Universitetsmuseenes samlingsportaler, 2020). Further information on pre-modern farmsteads and farm territories around Lake Ljøgottjern was obtained from the standard works of *Norske Gaardnavne* (https://www.dokpro.uio.no/rygh_ng/rygh_felt.html), *Diplomatarium Norvegicum* (https://www.dokpro.uio.no/dipl_norv/diplom_felt.html) and public available enclosure maps dated 1890 and 1902 (https://wcarkiv.domstol.no/wcarkiv/kommunelist.wc?ID). This was further supplemented with information from literature on local history (Hagen, 1997, pp. 28–30; Johnsen, 1941, pp. 133–134; Nesten, 1951).

4. Results

4.1. Description of the sediment

4.1.1. Lithology and geochemistry

Ten different lithological facies were identified for the 535 cm of sediment of the Lake Ljøgottjern sediment sequence (Fig. 2). The bottom part (535–440 cm) presents fine regular light and dark laminations (<1 mm). From 440 to 391cm depth, the sediments are disturbed and oblique levels were found in LJØ118. This part of the sediment is also disturbed in LJØ119 with folded features. This section can be interpreted as a slump and outside of the continuous sedimentation. It was considered as an instantaneous deposit and removed from the age-depth modelling. Above this deposit, the sediment is composed of dark organic silt with layering until 350 cm. Below 350 cm, Fe intensities are high (123 kcps average) and show variability (s = 44 kcps). Above 350 cm, the dark organic silt does not present layering anymore and Fe declines suddenly, remaining close to 0 kcps until 280 cm, while the Inc/Coh ratio reaches a maximum in this facies. For the whole sequence, intensities of Ti and K are very similar. Below 280 cm, K and Ti are very low (close to 0 cps) except several synchronous high peaks. Above 280 cm, intensities of K and Ti are higher (averaging at 170 and 230 cps) and represent the main change in the sedimentation of the sequence. Fe increases again and the Inc/Coh ratio remains at a high level. The following facies present dark organic silt with orange level or thin laminations. The two uppermost facies are coarse silt and present several layers of different colors. Peaks of Ti and K in the top part of the sequence were related to historical floods as identified by Bajard et al. (submitted). See Appendix A.1 for an overview of all XRF results.

4.1.2. Chronology

The age-depth model of the sediment sequence is constrained by six ^{14}C-dated plant macrofossil samples. Details of the samples and calibrated ages are presented in Appendix Table A.1. The top part was set to the year of coring and verified with the deposit of the historical flood “Stor-ofsen” of 1789 (Bajard et al., submitted). The resulting age-model covers the last 9300 years (Fig. 2). The top part of the age-model is linear until 250 cm with a mean sedimentation rate of 1.25 mm.yr-1. There is a shift in the sedimentation rate between 250 cm (50 cal CE) and 375 cm (5350 cal BCE), which was set at 280 cm (350 cal BCE) in the age-model considering the major change in the lithology and geochemistry at this depth. The sedimentation rate of the 5350–350 cal BCE period is much lower, ca. 0.2 mm.yr-1. The instantaneous deposit identified in section 4.1.1. is dated to 5650 cal BCE. Below this deposit, the laminated sediment covers the period 7350–5650 cal BCE with a sedimentation rate of 0.5 mm.yr-1 in average.

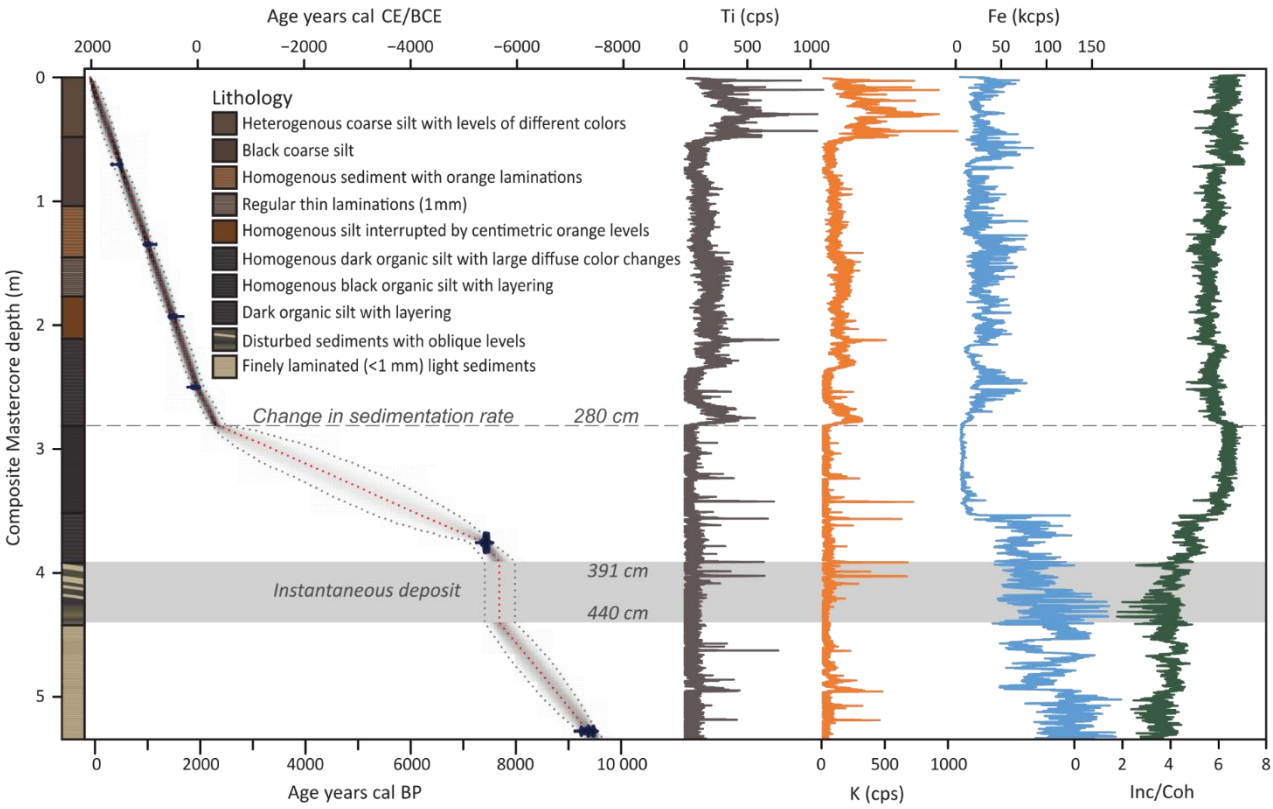


Figure 2. Age-depth model of the Lake Ljøgottjern sediment sequence (as represented with Bacon R package), including lithology and XRF geochemistry (Ti, K, Fe and Inc/Coh in counts per second). A change in the sedimentation rate was set at a depth of 280 cm based on the change in sediment and XRF geochemistry. We identified an instantaneous deposit of sediment at 391–440 cm depth.

4.2. Description of the biological data

4.2.1. Pollen

A total of 110 pollen types were counted in the pollen analysis, of which 65 plant families, 88 genera, and 12 species were identified. Pollen concentrations in samples from the instantaneous deposit were averaged to one concentration per pollen type, reducing the number of samples from 131 to 121. The most abundant plant groups in the pollen concentration dataset were trees (65%), followed by aquatics (17%), graminoids (7%) and anthropochores at 4% (Fig. 3A). The most abundant families were tree families Betulaceae (33%) and Pinaceae (28%), followed by green algae (Botryococcaceae) at 12%. Detailed pollen concentration diagrams can be found in Appendix Fig. A.7 and the corresponding data in Appendix Table B.1.

4.2.2. Plant and mammal DNA

For the first and second sequencing pools we obtained a total of 220,591,834 and 64,492,511 reads, respectively. After filtering, the plant DNA dataset consisted of 34,300,823 reads with on average 151,425 ± 20,349 reads per sample, corresponding to a total of 274 unique MOTUs (Fig. 3A, Appendix A.2–3). Taxonomic assignments of the plant sequences allowed us to identify 74 different plant families, 136 genera and 86 species (Appendices B.2–3).

The largest plant groups in relative log transformed read abundances were aquatic plants (37%), followed by forbs (non-graminoid herbaceous plants; 28%) and trees (20%). The same plant groups dominated when analysing the summed number of positive replicates, with the largest contribution from forbs (3978, 45%), followed by aquatic plants (1918, 22%) and trees (1191, 13%) (33%, 30% and 20%, respectively, when fractionated per sample as in Fig. 3A).

For the mammal DNA we obtained a total of 24,287,943 reads after filtering, corresponding to 119 unique MOTUs, from 3 different families and 5 genera, namely cattle (*Bos* sp.; 78 MOTUs), horses (*Equus* sp.; 11 MOTUs), pigs (*Sus* sp.; 1 MOTU), goats (*Capra* sp.; 3 MOTUs), and sheep (*Ovis* sp.; 26 MOTUs). Plant DNA was detected throughout the sediment core, while mammal DNA was only found in samples dating from ~200 BCE to 1800 CE. The number of read counts per sample in this time period averaged to 196,876 ± 60,890 (Appendices A.3 and B.4–5).

4.3. Plant identification by pollen versus *sed*aDNA

*Sed*aDNA detected a higher number of plant taxa and more taxa at a higher taxonomic level than was achieved by pollen. Of all detected plant families, 48 were detected by both methods, while 17 were unique to the pollen and 26 to the *sed*aDNA dataset, most of which were forbs (10), bryophytes (7) and aquatic plants (4), while plant families unique to the pollen dataset were primarily forbs (4), aquatics (4) and shrubs (3). Regarding plant genera, 52 were shared compared to 36 unique genera to the pollen dataset and 84 to the *sed*aDNA dataset. For the anthropochore and apophyte plant groups, containing mainly cultivated plant taxa and those benefiting from human disturbance, in total 17 pollen types were recorded compared to 15 MOTUs. For example, *Secale* was identified among the pollen but not detected by the *sed*aDNA despite its presence in the used reference database. S*ed*aDNA was able to identify both *Hordeum* and *Avena* as well as *Cannabis* and *Humulus*, while these pairs could not be distinguished from their pollen. For the forbs, 34 pollen types were recorded compared to 152 MOTUs from 21 and 31 plant families, respectively (Appendices B.1–3).

4.4. Plant community dynamics

4.4.1. Terrestrial vegetation zones

Two separate zones of similar terrestrial plant community composition were identified based on the plant *sed*aDNA dataset using CONISS clustering analysis, with the boundary between 580 and 230 cal BCE. Analysis of the pollen dataset identified 4 additional zones, though the broken stick analysis showed minor differences when adding the 3 last identified zones (Appendix Fig. A.9). Consequently, we identified one additional zone based on the pollen dataset, with the boundary between 440 and 450 CE. With the higher resolution of the pollen samples, the previously identified *sed*aDNA based zone was further delineated to between 430 and 320 cal BCE (Fig. 3B). NMDS ordination of pollen and *sed*aDNA data supported the results of the CONISS clustering analysis (Appendix Fig. A.10).

4.4.2. Vegetation zone 1 (ca. 8000–300 cal BCE)

Between 8000 and 7000 cal BCE, a relatively high rate of accumulation of new taxa was found, as cumulative numbers increased from 18 to 44 for the pollen and 9 to 40 for the plant DNA (Fig. 3A). Terrestrial plant biodiversity remained low throughout the 8000–300 cal BCE period with estimates <10 for pollen and <22 for the DNA MOTUs, averaging to 5.0 ± 0.2 (q=1) and 3.4 ± 0.2 (q=2) for pollen, and 11.4 ± 1.5 (q=1) and 10.7 ± 1.4 (q=2) for *sed*aDNA (Fig. 3B).

From 8000 to 3100 cal BCE, both the pollen and *sed*aDNA records were dominated by trees (93% and 64%). Recorded pollen types included *Betula* (52%), *Pinus* (24%), *Alnus* (12%), *Corylus* (5%), *Ulmus* (3%), and 3% of *Tilia*, *Quercus, Populus, Fraxinus* and *Fagus* combined. Pollen concentrations of shrubs were relatively stable, with an average of <1% of the terrestrial pollen concentration, and higher percentages (~2%) in the period between 7500 and 7000 cal BCE. *Sed*aDNA trends for shrubs were more variable, with percentages often at 0, but also peaking to 10, 17 and 43% at 7500, 2000 and 1600 cal BCE. At 2800 cal BCE, a distinct peak was found in the concentration of graminoid pollen, while in the DNA data a peak in graminoids was recorded later, at 2100 cal BCE. Forbs (<1% pollen, 15% DNA) and ferns (1% pollen, 9% DNA) were recorded throughout zone 1. No DNA of plants commonly associated with human settlements (apophytes and anthropochores) was found, while traces of pollen of these groups were found throughout this zone (<2%).

4.4.3. Vegetation zone 2 (ca. 300 cal BCE–450 cal CE)

A period of rapid increase in cumulative numbers of taxa could be distinguished between ca. 300 cal BCE and 450 cal CE for both pollen and *sed*aDNA records, as cumulative pollen types increased from 67 to 101, plant DNA MOTUs from 81 to 224 (Fig. 3A) and biodiversity estimates increased to averages of 8.1 ± 0.3 (q=1) and 4.9 ± 0.2 (q=2) for pollen, and 82.4 ± 16.4 (q=1) and 78.0 ± 15.8 (q=2) for *sed*aDNA (Fig. 3B).

We detected a reduction in tree pollen to 79% of the total pollen concentrations, and an increase in average abundance of mainly graminoids in the pollen record (from <1% to 9%) and forbs in the DNA record (from 15% to 62%). Concentrations of graminoid pollen fluctuated around 10% of the terrestrial pollen concentration, with notable reductions to below 5% at ~250 cal BCE and 450 cal CE, while graminoids in the *sed*aDNA record decreased in number of positive replicates throughout this zone, from 10% to 2%.

After 300 cal BCE, anthropochore and apophyte taxa first appeared in the *sed*aDNA record, simultaneously with increased abundances in the pollen record from 1.5% to 2% (apophytes) and 0.4% to 2.4% (anthropochores) and an increase in graminoids from <5% to >13% at 250 cal BCE. A decrease in relative pollen concentrations of apophytes and anthropochores was recorded at around 50 cal BCE, after which both groups returned to previous percentages, reaching 5% around 350 cal CE (apophytes) and 11% at ca. 450 cal CE (anthropochores).

4.4.4. Vegetation zone 3 (ca. 450–1800 cal CE)

Biodiversity increased to averages of 10.0 ± 0.3 (q=1) and 5.8 ± 0.2 (q=2) based on the pollen data, while estimates based on *sed*aDNA data decreased to 61.5 ± 5.3 (q=1) and 58.1 ± 5.0 (q=2) compared to the previous zone. Cumulative numbers of taxa increased until 1600 cal CE for the pollen, and 1800 cal CE for the plant *sed*aDNA (Fig. 3B).

Tree pollen concentrations continued to decrease until ~1300 cal CE, while still dominating the pollen record at an average of 68% of the terrestrial pollen concentration. In the *sed*aDNA record, trees averaged at 14% with some fluctuation, while forbs remained dominant at an average of 62% for this zone. The *sed*aDNA relative fractions of shrubs peaked at 1200, 1400 and 1600 cal CE, and pollen concentrations similarly indicate increased abundance of shrubs around 1200 and 1400 cal CE, but not at 1600 cal CE. Graminoid fractions remained relatively stable, with the exception of two marked decreases in both pollen and *sed*aDNA records at 1000 and 1350 CE.

The increased anthropochore pollen concentrations before ~450 CE were followed by a steady decrease to <2% in the 700–800 cal CE period, which was also apparent in the *sed*aDNA record. Both anthropochore and apophyte fractions in the *sed*aDNA dataset then showed a distinct increase towards ca. 850–1300 cal CE directly followed by decreased fractions. This pattern was matched by anthropochore pollen fractions, with a distinct peak (28%) after 1300 cal CE and a sharp decline (<4%) after 1350 cal CE, while apophyte pollen fractions fluctuated between ca. 2 and 6% throughout this zone.

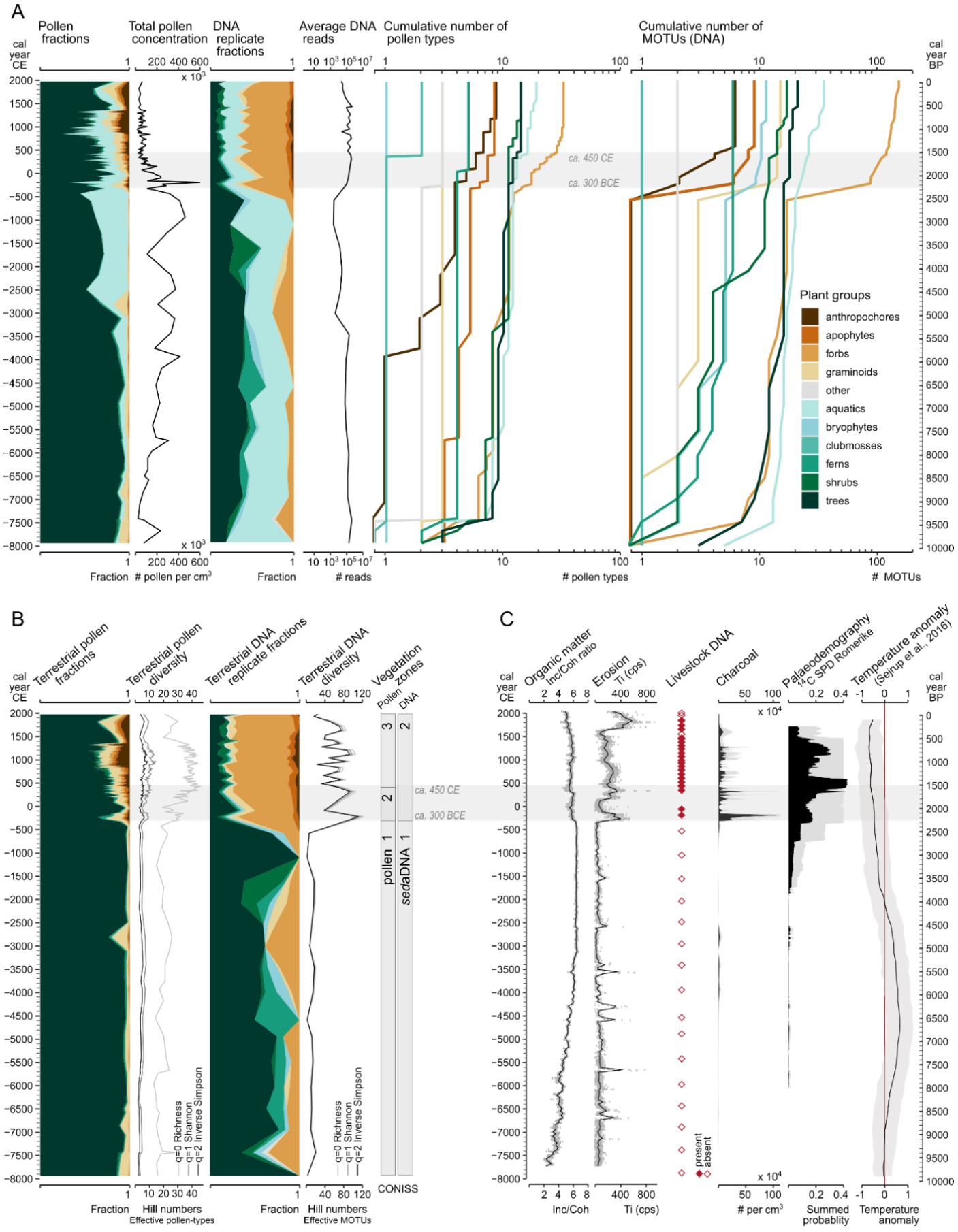

A
Pollen fractions
Total pollen concentration
DNA replicate fractions
Average DNA reads
Cumulative number of pollen types
Cumulative number of MOTUs (DNA)
cal year CE
cal year BP
ca. 450 CE
ca. 300 BCE
Plant groups
anthropochores
apophytes
forbs
graminoids
other
aquatics
bryophytes
clubmosses
ferns
shrubs
trees
Fraction
pollen per cm3
reads
pollen types
MOTUs
B
Terrestrial pollen fractions
Terrestrial pollen diversity
Terrestrial DNA replicate fractions
Terrestrial DNA diversity
Vegetation zones
Pollen
DNA
pollen
sedaDNA
CONISS
q=0 Richness
q=1 Shannon
q=2 Inverse Simpson
Hill numbers
Effective pollen-types
Effective MOTUs
C
Organic matter
Inc/Coh ratio
Erosion
Ti (cps)
Livestock DNA
present
absent
Charcoal
Palaeodemography
14C SPD Romerike
Temperature anomaly
(Seljup et al., 2016)
Inc/Coh
Ti (cps)
per cm3
Summed probablity
Temperature anomaly

Figure 3. Overview of the environmental data from Lake Ljøgottjern and the surrounding area. A) the total vegetation fractions of the anthropochores, apophytes, forbs, graminoids, other, aquatics, bryophytes, clubmosses, ferns, shrubs and trees in the pollen and the plant DNA datasets, the total pollen concentration per sample, the summed number of DNA reads per sample, and the cumulative number of taxa in the pollen dataset and the plant DNA dataset for each plant group. Grey shading indicates time periods of considerable cumulative increases in plant taxa. B) The terrestrial pollen fractions, terrestrial DNA replicate fractions and the terrestrial plant diversity determined by Hill numbers, with q=0, 1, and 2, representing the effective number of taxa (pollen-types or MOTUs) in the form of species richness, Shannon index and inverse Simpson index, respectively. Statistically different vegetation zones were determined through CONISS analysis, identifying ~400 cal BCE and 450 cal CE as boundaries between zones. C) An overview of the environmental variables, including data from the sediment core: organic matter content (Inc/Coh ratio), Ti, presence/absence of livestock DNA, and the charcoal concentration. For the Inc/Coh ratio and Ti, grey points represent all data points and black lines represent smoothed data values derived using a general additive model (GAM). The radiocarbon summed probability densities (^{14}C SPD) are from the Romerike region (Loftsgarden and Solheim, in press) and the median composite temperature anomaly ($\pm 1\sigma$) from 74 lake and marine sites in the North Atlantic-Fennoscandian region as determined by Sejrup et al. (2016).

4.5. Palaeodemographic dynamics

Most radiocarbon dates for the ^{14}C Summed Probability Distributions (SPD) of the Romerike region were dated to the Late Iron Age and the resulting curves largely match the charcoal concentrations in the Lake Ljøgottjern sediment (Fig. 3B; Appendices B.1 and B.6). Charcoal was detected from 7450 cal BCE onwards, throughout most of the sediment core (except for 3 samples) and generally in concentrations below $10x10^4$ pieces per cm^3. The ^{14}C SPD record starts at 6050 BCE and trace values (<0.01) were found between 6050 and 4000 cal BCE. Both records showed higher values around 2800 cal BCE, with peaks in charcoal between 4000 and 2800 cal BCE and peaks in ^{14}C SPD values between 2850 and 2650 cal BCE. Values of both proxies remained low in the subsequent period, until 2000 cal BCE when ^{14}C SPD values increased. Charcoal concentrations remained low until 600 cal BCE, when they increased and peaked at 200 cal BCE with a concentration of around $11x10^5$ pieces per cm^3. The peak in charcoal matches an increase in ^{14}C SPD for the Romerike region and a distinct short peak is evident in both records at around 300 cal CE (at 270 cal CE for ^{14}C SPD and 320 cal CE for charcoal). More stable high ^{14}C SPD values (>0.35) were recorded between 330 to 550 cal CE when they dropped to <0.2 at 600 cal CE while charcoal concentrations at Lake Ljøgottjern remained relatively stable. Another period of high ^{14}C SPD values (>0.25) was found between ca. 1050 and 1250 cal CE.

4.6. Traces of farming

4.6.1. Cultivated plant taxa

Isolated peaks of *Cannabis/Humulus*-type (hemp and hop), *Hordeum/Avena*-type (barley and oats), *Triticum* (wheat) and *Secale* (rye) pollen concentrations were identified in the period between ca. 4000 and 400 cal BCE. From between 400 and 250 cal BCE a more stable presence of all these pollen types was established. *Linum* (flax) pollen was detected occasionally from 900 cal CE onwards (Fig. 4).

No DNA from cultivated plants was recorded before 230 cal BCE, at 230 cal BCE *Hordeum* (barley) and *Triticum* (wheat) was found in 4 and 2 replicates out of 6 respectively. *Triticum* DNA was not detected in any other samples. The first

detection of *Hordeum* DNA matches a peak in *Hordeum/Avena*-type pollen at this time point, which was followed by a decrease in pollen concentration, whereas the *Hordeum* DNA was found in most PCR replicates up until 1350 cal CE. At 300 cal CE *Humulus* DNA and *Linum* DNA were first detected, followed by *Cannabis* at ca. 500 cal CE.

4.6.2. Livestock DNA

DNA from pastoral animals was first detected in the sample dated to 230 cal BCE from cattle, pig and horse (*Bos* sp., *Sus scrofa*, and *Equus* sp.; Fig. 4). From 230 cal BCE onwards, cattle DNA was continuously present, with exceptions at 1500 CE and after 1800 CE when no cattle DNA was detected. Between 230 cal BCE and 450 cal CE, horse DNA was detected in all samples, and sheep DNA (*Ovis* sp.) from 100 cal BCE onwards, while pig DNA was only found at 230 cal BCE. Horse and pig DNA were occasionally present in the 500–1350 cal CE period, whereas sheep DNA was continuously present between 750 and 1420 cal CE. DNA from goat (*Capra* sp.) was detected in two samples: 1110 and 1270 cal CE. At 1520 cal CE, no livestock DNA was found, however, DNA from sheep was present at 1630 cal CE, and both cattle and horse DNA were detected between 1630 and 1820 cal CE.

4.6.3. Archaeological evidence

The earliest settlements were dated to the Bronze Age (ca. 1800–500 BCE; A-ID 96260; Helliksen, 1997) representing one farmstead located north-east of Lake Ljøgottjern (Fig. 1C). A 1994–95 excavation identified seven houses and 105 hearths, with most radiocarbon dates falling within the 1500–200 BCE period (Helliksen, 1997). This farmstead was divided into several farms located closer to the lake during the Iron Age (ca. 550 BCE–1050 CE; A-ID 121551, 171660; Helliksen, 1997; Simonsen, 1997), with the settlement site just north of the medieval farmstead of Haug (A-ID 171660) dated to ca. 800–1050 CE. The farms of Haug (ca. 1390 CE) and Ljøgot (1514 CE) were mentioned in medieval sources (Rygh, 1898, pp. 322–323) and archaeological excavations, finds and historical maps confirm the location of the farmsteads close by the lake, as well as written sources indicating land-use for livestock and crop cultivation (Nesten, 1951).

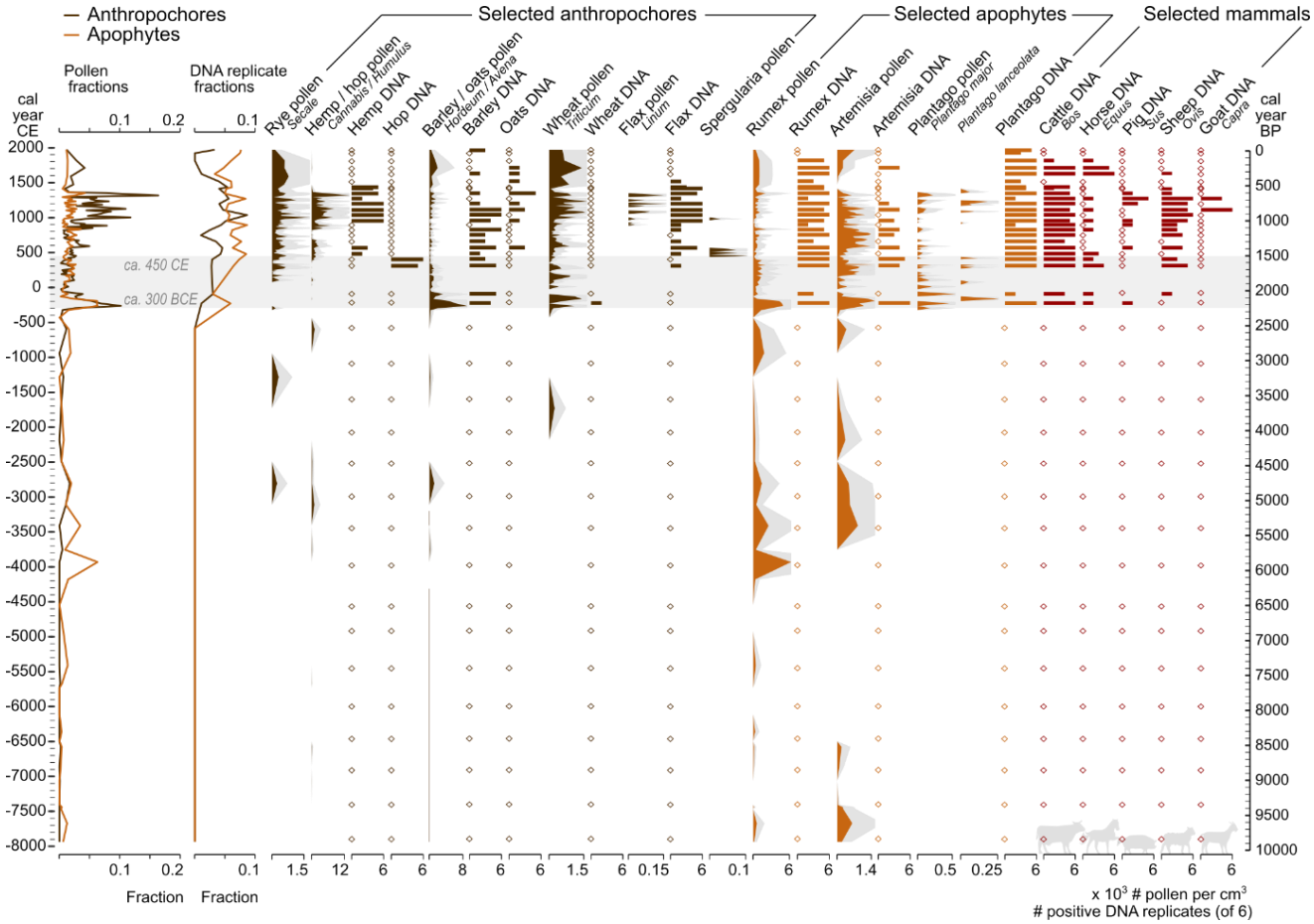


Figure 4. Anthropochore and apophyte fractions of terrestrial pollen fractions and terrestrial DNA replicate fractions and a selection of plant and mammal taxa associated with humans. Grey shading of pollen concentrations shows 3x exaggerated concentrations while the grey band across individual graphs indicates the time period of pollen zone 2 based on CONISS analysis between ca. 300 BCE and 450 CE.

4.7. Environmental terms related to plant community changes

Distance-based redundancy analysis (dbRDA) of the pollen data reveals that the temperature anomaly explains 5.1% of the variance in terrestrial plant composition (p=0.001), followed by organic matter content (Inc/Coh ratio 4%; p=0.001), presence/absence of livestock (2.4%; p=0.003), Ti (1.8%; p=0.017), and ^{14}C SPD (1.5%; p=0.043; Fig. 5A&B). The same analysis for the *sed*aDNA samples indicates presence/absence of livestock as explanation for 7% of the variance (p=0.001), followed by temperature (6%; p=0.006), organic matter content (5.5%; p=0.014) and Ti (3%; p=0.092; Fig. 5A&B; Appendix A.11). Variation partitioning indicates the highest joint effect of temperature anomaly and presence/absence of livestock on variation in plant community composition for both the pollen samples (31%) and the *sed*aDNA samples (28%). For *sed*aDNA samples, this is followed by the joint effect of presence/absence of livestock and Ti (15%). For pollen samples, the second highest joint effect is of temperature anomaly and ^{14}C SPD (21%; for detailed variation partitioning results see Appendix A.12).

DbRDA of subsections of the core shows that the variance in terrestrial plant assemblages in zone 1 (ca. 8000–300 cal BCE) are related to Ti (20%;

p=0.001), organic matter (16%; *p*=0.001) and temperature (16%, *p*=0.001) based on the *sed*aDNA data, compared to organic matter (27%; *p*=0.001) and temperature (9.3%; *p*=0.007) for the pollen data (Fig.5B, Appendix A.11). For zone 2 and 3 (ca. 300 cal BCE–1800 cal CE) temperature and ^{14}C SPD are found to best explain the variance in terrestrial plant assemblages based on both the *sed*aDNA (9.7%; *p*=0.008 and 8.4%; *p*=0.03) and pollen data (28%; *p*=0.001 and 3.9%; *p*=0.001), followed by Ti (3.1%; *p*=0.004) only for the pollen data.

Temperature anomaly data is highly correlated with sample age (r_s=.93, *p*=0.000, N=120) and we found similar relationships for these variables with other environmental terms (Fig. 5C, Appendix A.13). Significant relationships are found between sample age and plant diversity, livestock presence/absence, charcoal, and Ti for both the *sed*aDNA and pollen datasets. Palaeodemographic trends (^{14}C SPD) are significantly correlated with sample age at the time resolution of the pollen data, but not of the *sed*aDNA data. Positive relationships between plant diversity, livestock, charcoal, ^{14}C SPD and Ti are evident for both datasets. Organic matter (Inc/Coh ratio) is not correlated with other environmental terms when considering the entire core, however, in the period from ca. 8000–300 cal BCE they show a significant relation with ^{14}C SPD (r_s=.6, *p*=0.02, N=120) and sample age (r_s=-.95, *p*=0.000, N=120).

5. Discussion

In this study we applied an interdisciplinary approach to uncover anthropogenic and environmental drivers of biological change during the Holocene at Lake Ljøgottjern. Analysis of pollen, *sed*aDNA, geochemical and archaeological data enabled us to reconstruct the palaeoenvironmental dynamics and establish a timeline of cultivation and pastoralism at Lake Ljøgottjern for the last 10,000 years while analysing driving factors of biological change. We argue for an integrated approach in the reconstruction of palaeoenvironmental dynamics to better understand past and present human-environment interactions.

5.1. Stratigraphic integrity of the sediment record

We investigated individual and combined analyses of the proxies to infer the stratigraphic integrity of the core for the last 10,000 years. The lithological description and XRF analysis show an abrupt change in the sedimentation rate at ca. 280 cm depth dated to ca. 350 cal BCE (Fig. 2) concurrent with the stratigraphic zonation based on CONISS analysis of pollen and *sed*aDNA, between 430 and 320 cal BCE. Stratigraphic zonation and the separation of NMDS clusters are consistent between pollen and plant *sed*aDNA analysis (Appendices A.9 and A.10), and direct comparison of plant taxa presence between pollen and *sed*aDNA records revealed consistent distributions of plant DNA, strongly indicating that disturbance or leaching do not impact the sedimentary record of Lake Ljøgottjern. This is evident in overall patterns of pollen and plant *sed*aDNA fractions (Fig. 3A and B) as well as specific plant taxa (Fig. 4, Appendices A.7 and A.8), thereby confirming the stratigraphic integrity of the record. Additionally, the pattern in the number of positive replicates for *Cannabis* DNA over time matches the dynamics in the

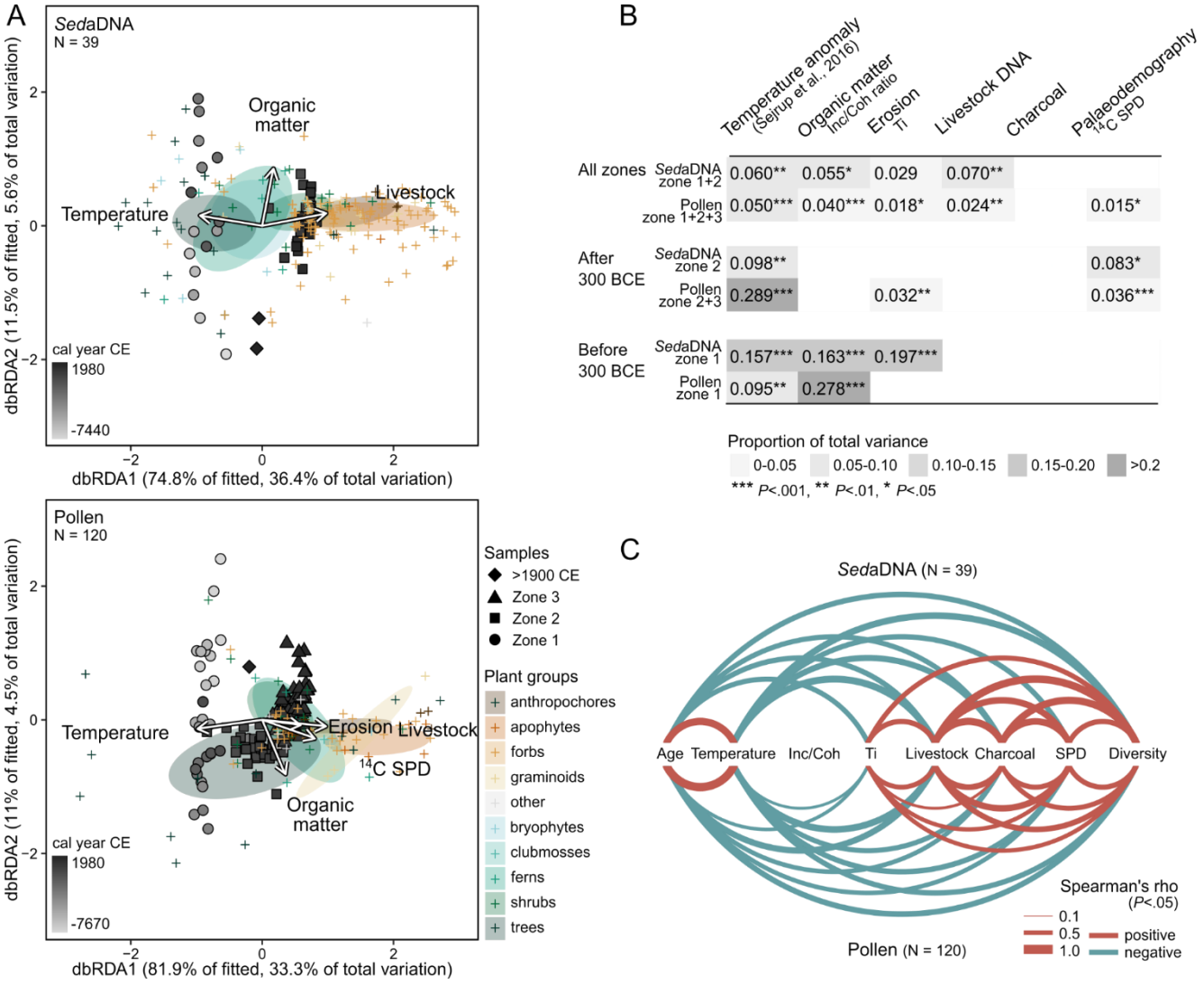


Figure 5. Summary of automatic stepwise ordinations of plant assemblages using distance-based redundancy analysis (dbRDA). Environmental terms included in dbRDAs: surface temperature anomaly data from Sejrup et al. (2016), organic matter corresponds to the Inc/Coh ratio determined by XRF analysis, Ti is measured by XRF analysis, livestock corresponds to presence/absence of livestock DNA, charcoal corresponds to the charcoal concentrations determined by palynological analysis, and ^{14}C SPD corresponds to radiocarbon summed probability distributions, a proxy for palaeodemography. A) SedaDNA (top) and pollen (bottom) dbRDA ordination sample scores (filled symbols) and plant taxa scores (plus symbols) coloured per plant group. Ellipses indicate the standard error of the mean per plant group. Arrows indicate significant environmental terms. B) Proportions of variance explained per environmental term of the total variance in plant assemblages for six automatic stepwise dbRDAs based on: samples from all zones combined, samples from after 300 BCE, and samples from before 300 BCE for sedaDNA and pollen datasets. Blank cells indicate that the environmental term was not included in the model as inclusion did not result in a better model (increased adjusted R2). C) Spearman's rank order correlations for environmental terms included in the dbRDAs of sedaDNA (top) and pollen (bottom) data, with addition of the sample age (cal year BP) and diversity (Hill q=2; inverse Simpson). Line thickness indicates the strength of the correlation in Spearman's rho value and only significant correlations are shown according to Bonferroni adjusted p-values correcting for multiple tests (p<0.05).

concentrations of *Cannabis/Humulus*-type pollen, with peaks in both records at 950–1200 cal CE and 1350–1450 cal CE, even though pollen is an unlikely source of chloroplast DNA (Parducci et al., 2017; Sjögren et al., 2017). The apparent absence of DNA leaching is consistent with other lake *sed*aDNA records (Alsos et al., 2020, 2018; Epp et al., 2015). Moreover, plant DNA was detected throughout the sediment core, but mammal DNA was only found in samples dating from 230 cal BCE to 1800 cal CE, indicating that these sequences are unlikely to be a result

of contamination. The lithology, geochemistry and the biological data therefore support the stratigraphic integrity of the analysed sequence and the authenticity of the *sed*aDNA.

5.2. Source of pollen and *sed*aDNA in Lake Ljøgottjern

Both pollen and *sed*aDNA data are affected by source productivity and taphonomic processes of dispersal, transfer, deposition and preservation (Giguet-Covex et al., 2019; Prentice, 1985). Primary sources of animal DNA are urine and faeces and it has been proposed that scattered distributions of animals can result in non-detection of DNA, while enclosures or folds within a lake catchment area can represent a "point source", concentrating the supply of mammal DNA to the lake sediments (Giguet-Covex et al., 2019, 2014). *Sed*aDNA of plants similarly originates from the lake catchment area and is of local provenance (Alsos et al., 2018; Giguet-Covex et al., 2019), whereas pollen may originate from a wide area (Birks and Bjune, 2010). The relevant source area of pollen (RSAP) is dependent on the relative pollen productivity (RPP) of plant taxa and correlated to the size of the lake, with records from small lakes (50 m radius) consisting for 30–45% of pollen originating from within 300–400 m from the lake edge, or 600–800 m for lakes with a 250 m radius (Sugita, 1994). Estimations of RSAP values for small-size lakes (25-250 m radius) in southern Scandinavia fall within ca. 900–2500 m distance from the lake centre, with RSAP values of ca. 1000 m for lakes of 100 m radius in southern Sweden (Sugita, Gaillard & Broström, 1999), within ca. 1000–2500 m for simulation tests using a 50 m lake radius (Hellman, Bunting & Gaillard, 2009), and ca. 900–1100 m for lakes of a 37–247 m (mean: 125 m) radius in western Norway (Hjelle & Sugita, 2012). For Lake Ljøgottjern, with a radius of ca. 60–80 m, this implies a RSAP of at least 900 m. We found many pollen grains from wind-pollinated species (e.g. trees and graminoids), with higher RPPs compared to plants with other forms of reproduction, corroborating that their pollen can come from a much larger distance (Birks and Bjune, 2010; Hjelle & Sugita, 2012; Sugita et al., 1999). On the other hand, *sed*aDNA originates from within the lake catchment, 60 to maximum 230 m distance from the lake edge (Fig. 1C). The vegetation changes inferred from these two proxies therefore reflect environmental dynamics at different scales.

5.3. Palaeoenvironmental history

5.3.1. Ecological succession (ca. 8000–300 cal BCE)

Pollen and *sed*aDNA analyses from the earliest period covered by the core (ca. 8000–4000 cal BCE) revealed low terrestrial plant diversity throughout this period, despite the high rate of plant taxa accumulation found between 8000 and 7000 cal BCE, suggesting a turnover of plant taxa. This turnover is associated with an increase in organic matter content during this period which can reflect an increase in lake productivity or the development of the soils in the lake catchment. The earliest stage of the vegetation record is dominated by *Pinus*, adapted to cold conditions, and *Betula*, a known pioneer. Pollen analysis further revealed the presence of other known pioneer taxa, including *Artemisia*, *Filipendula* and *Rumex* already from the oldest sample (7930 cal BCE). Taxa that are more dependent on

fertile soils came into the record at a later time, such as *Urtica* at 5670 cal BCE (Behre, 1981). We also found some *Cannabis/Humulus*-type pollen already at 7000 cal BCE, which are likely from wild *Humulus lupulus* L. as other anthropochore pollen (*Hordeum/Avena*-type, *Triticum* and *Secale*) were not recorded until 3800–1300 cal BCE.

Climate proxy data from Sejrup et al. (2016) showed an increase in temperature between ca. 7000 and 5000 cal BCE, indicating a transition to the Holocene Thermal Maximum, a period ca. 5000–2200 cal BCE characterised by warm and dry summers in the Northern Hemisphere (Antonsson and Seppä, 2007; Wanner et al., 2011). We accordingly found increased concentrations of thermophilic trees, including *Ulmus*, *Corylus*, *Alnus*, and from ca. 4500 cal BCE also *Quercus* and *Tilia*. Ti remained low, with peaks indicating a number of erosion events during the Holocene Thermal Maximum. Similar frequent flood events during the warm early and middle Holocene in southern Norway have been identified as intense summer rainstorms based on the sedimentary composition (Støren et al. 2016). Erosion rates can affect the representation of the catchment vegetation in the *sed*aDNA record (Giguet-Covex et al., 2019). Indeed, Ti variations in zone 1 of the sediment core (ca. 8000–300 cal BCE) were found to be significantly related to the variations in plant communities as recovered by *sed*aDNA analysis, but not by pollen analysis. In conclusion, we recognize that the increase in temperature, but also the accumulation of organic matter, are related to the variation in plant communities from pollen and *sed*aDNA analysis and are likely driving factors in ecological succession during this period up to 300 cal BCE at Lake Ljøgottjern.

5.3.2. Early human land-use? (ca. 4000–300 cal BCE)

Many vegetation reconstructions from southeastern Norway (such as from Rud Øde, Danielsetermyr, and Skogstjern; Fig. 1A-B) show an overall progression from the possible presence of hunter-gatherers recorded in the charcoal records, through the establishment of pastoral farming to either mixed farming or cereal cultivation (Høeg, 1996, 1996; Wieckowska-Lüth et al., 2017), as is the case for Lake Ljøgottjern (this study). The exact timing of these transitions differs between localities. Pastoralism can be inferred indirectly through presence of plant taxa that are favoured by the presence of livestock, through animal faeces, trampling and selection through consumption. Such taxa include nitrophilous and ruderal species, e.g. *Rumex* sp.*, Urtica* sp., *Chenopodium* sp. and *Plantago* sp. (Behre, 1981). Over half of the locations indicated in Fig. 1 show the first indications of pastoral farming concurrent with the Early Neolithic (ca. 4000–3300 BCE; Skogstjern, Sveskestutjern, Danielsetermyr, Bånntjern in Tolga, and Ljøgottjern) and Middle Neolithic (ca. 3300–2400 BCE; Rud Øde, Bånntjern in Ullensaker, Skånetjern, Stortallsjøen, Lensmannsvollen, Kåsmyra, Hellemundsmyra, Skjerdingsfjell, and Hirsjøen), while other locations record the first signs of pastoralism anytime between ca. 2000 BCE and 1800 CE (Høeg, 1997; Høeg, 1996; Wieckowska-Lüth et al., 2017; Fig. 1A-B). In this study, we found peaks in concentrations of several of these taxa reflected in increased apophyte fractions in the periods of ca. 5500 cal BCE, 4000–1700 cal BCE, and 950–500 cal BCE (Fig. 4), matching other pollen records from the Romerike region that showed some indications for grazing

already in the Early Neolithic (4000–3300 BCE; Danielsetermyr and Svenskestutjern) and Middle Neolithic (3300–2400 BCE; Rud Øde, Bånntjern in Ullensaker, and Skånetjern). The absence of *sed*aDNA of pastoral animals at Lake Ljøgottjern during this time could be due to scattered distribution or pastoral activities taking place outside of the lake catchment area (Giguet-Covex et al., 2019).

Early detection of anthropochore pollen (e.g. *Cannabis/Humulus*-type, *Hordeum/Avena*-type and *Secale*) between ca. 3800 and 2500 cal BCE at Lake Ljøgottjern matches peaks in charcoal. The new high resolution pollen record further showed an increase in concentrations of graminoids between 4000 and 2500 cal BCE and the *sed*aDNA record indicated a higher diversity in forbs as well as more positive replicates for *Carex* (sedge, a graminoid), both indicating a more open landscape around the lake during this time. The temperature anomaly composite from Sejrup et al. (2016) suggested a still relatively high but decreasing temperature anomaly between 4000 and 2000 cal BCE, which was followed by decreased concentrations in thermophilic trees (*Ulmus*, *Corylus, Alnus, Quercus* and *Tilia*) marking the end of the Holocene Thermal Maximum. These findings are largely concurrent with the earliest indications of cereal cultivation in southeastern Norway in the Middle and Late Neolithic (ca. 3300–1750 BCE; Danielsetermyr, Bånntjern in Ullensaker, Skånetjern, Dulpmoen, Engelaug, Kjerdingsfjell, and Hirsjøen; Høeg, 1996, 1997; Fig. 1A-B), although at Lake Skogstjern the occurrence of *Plantago lanceolata*-type and cereal pollen are both dated to ca. 3650–3400 cal BCE (Wieckowska-Lüth et al., 2017). With the exception of Danielsetermyr, pollen records from the Romerike region and further north indicate a delayed introduction of pastoralism (ca. 3000 BCE) and cultivation (ca. 2500–2000 BCE). Within the Romerike region, the northeastern sites do not attest cereal cultivation before ca. 200 BCE (Høeg, 1997).

For most of the discussed pollen records in southeastern Norway, indicators of pastoralism and cultivation were not continuous from the first detection (Høeg, 1996, 1997) and intensification of farming activities is mostly evident from the Early Iron Age (from 500 BCE–550 CE), consistent with the general intensification of agriculture in southeastern Norway (Mjærum, 2020; Myhre, 2002; Solberg, 2000). Also at Lake Ljøgottjern, we found little evidence for human presence between ca. 2500 and 2000 cal BCE as ^{14}C SPD and concentrations of charcoal and anthropochore pollen returned to values close to zero. ^{14}C SPD values increased again after 2000 cal BCE, followed by indications of cereal cultivation with for example *Triticum* pollen at 1700 cal BCE and *Hordeum/Avena* and *Secale* pollen at 1300 cal BCE, and increased charcoal concentrations from 600 cal BCE coinciding with detection of *Cannabis/Humulus* pollen. These patterns, visible in the pollen record but not in the *sed*aDNA, are in accordance with the archaeological evidence of a Bronze Age settlement at approximately 500 m north-east from the lake, dated to ca. 1800–500 BCE (A-ID96260; Fig. 1C).

5.3.3. Intensification of agriculture and pastoralism (ca. 300 cal BCE–present)

From ca. 300 cal BCE taxonomic diversity at Lake Ljøgottjern increased rapidly (Fig. 3A-B), most notable in the *sed*aDNA record in which crops and livestock appeared for the first time at 230 cal BCE. From this time onwards, also plant taxa

associated with pastoral farming (*Rumex* sp.*, Urtica* sp., *Chenopodium* sp. and *Plantago* sp.; Behre, 1981) were consistently present in both *sed*aDNA and pollen records. Moreover, increases in Ti and charcoal concentrations were observed (Fig. 3C), suggesting more erosion through anthropogenic activities such as pastoralism and the development of an agricultural landscape (Giguet-Covex et al., 2019, 2014). The simultaneous decrease in organic matter despite pastoral activities during this period is likely caused by the increased erosion. Overall, all these proxies indicate intensified human activity, consistent with pollen records from the Romerike region that indicate widespread grazing activity in the region from the Early Iron Age (Fig. 3, 500 BCE–550 CE; Høeg, 1997) and the general intensification of agriculture in southeastern Norway (Mjærum, 2020; Myhre, 2002; Solberg, 2000). At Lake Ljøgotjern we can trace, from the *sed*aDNA and Ti records, the abrupt changes in the local environment resulting from the establishment of one or multiple farms in close proximity to the lake, as evident by several archaeological sites and finds (A-ID 121551, 171659, 171592, 173761, 172010 and 180779; Fig. 1C).

During the past 2000 years we can recognise periods of decreased versus intensified human activity. During ca. 150 cal BCE to 100 cal CE reduced pollen diversity, lower numbers of detected anthropochore taxa, decreased charcoal concentrations and ^{14}C SPD values as well as lower Ti suggest a reduction in the nearby human population or at least in their activities in the local surroundings. A subsequent peak in charcoal concentrations and ^{14}C SPD at 300 cal CE matches the first detection of *Humulus* and *Linum* in the *sed*aDNA record. Between ca. 300 and 600 cal CE we found relatively high concentrations of apophyte pollen, coinciding with detection of cattle, horse and sheep DNA, indicating pastoral activities. A change in the pollen record at 450 cal CE (detected by the CONISS analyses) is driven by the first detection of *Spergularia* pollen and increased concentrations of anthropochore pollen, coinciding with the first finding of *Cannabis* DNA, suggesting the intensification of agricultural practices. Also ^{14}C SPD values were high between 330 and 550 cal CE, while temperature anomalies were relatively low (Sejrup et al., 2016). A second period of relatively high ^{14}C SPD values was identified ca. 1050–1250 cal CE, corresponding with a continued presence of livestock DNA (from cattle, pigs, sheep and goats), anthropochore DNA (*Linum, Hordeum* and *Cannabis*), and high concentrations of anthropochore pollen during the Medieval warm period (ca. 800–1300 CE). A settlement site just north of the medieval farmstead of Haug is dated to this period (ca. 800–1050 CE; A-ID 171660; Helliksen, 1997; Fig. 1C) and may be related to this increase. Medieval sources mention the farms of Haug (ca. 1390 CE) and Ljøgot (1514 CE) with locations of their farmsteads close to the lake (Rygh, 1898, pp. 322–323; Fig. 1C). *Sed*aDNA, pollen and written records (Nesten, 1951) all indicate mixed farming activities with cattle, sheep and cultivation of cereals, oats, hemp and flax during this time. At 1520 cal CE, however, no livestock DNA was detected and anthropochore DNA reduced to oats and flax (Fig. 4). Anthropochore pollen fractions peaked again around 1700–1800 cal CE, matching the returned presence of livestock between 1630 and 1820 cal CE while little evidence of anthropochores was present in the *sed*aDNA record at this time, suggesting a focus on pastoral farming within the lake catchment area of specifically cattle and horse. Modern samples (>1800 cal CE)

showed limited plant diversity, no presence of livestock and indicate a focus on the cultivation of *Hordeum* in close vicinity of the lake reflected by its presence in the *sed*aDNA record, while *Triticum* was only found in the pollen record and therefore likely cultivated outside of the lake catchment area. These results reflect modern (mono) cultivation with no or reduced alternation in land use, and animals possibly receiving water from established small ponds as identified on historical maps.

Between 300 cal BCE and 1800 cal CE temperature explained most of the variation in plant composition followed by palaeodemography (Fig. 5B). During this period, ^{14}C SPD values increased and peaked earlier at ca. 330–550 cal CE and 1050–1250 cal CE, matching high detection of DNA from pastoral animals and increased apophyte and anthropochore concentrations. Temperature anomaly decreased in Scandinavia and was especially low around ca. 450 CE and 1250–1650 CE (Sejrup et al., 2016), with the first cold period matching the change in the pollen record marking the subsequent intensification of agricultural activities, while the 1250–1650 CE cold period showed decreasing trends in anthropochore pollen and *sed*aDNA fractions, suggesting decreased agricultural activities during colder periods. In summary, we found that the variation in surface temperature anomaly and ^{14}C SPD relate to the variation in plant communities from pollen and *sed*aDNA analysis and observed several periods of intensified mixed farming activities, likely driven by increasing human population densities.

5.4. The role of anthropogenic and environmental factors

Previous studies have emphasized the long-term role of humans in shaping our modern ecosystems through their use of land (Boivin and Crowther, 2021; Ellis, 2015; Stephens et al., 2019; Williams et al., 2015), with the timing of human land-use transitions likely reflecting cultural, ecological and climatic shifts (Boivin et al., 2016). In this study, estimates of human population density, pastoralism, surface temperature, erosion and organic matter all correlated significantly with changes in plant communities during the Holocene, especially at ca. 300 cal BCE (Fig. 5A&B). However, the importance of each of these factors changed over time and most of the included environmental terms correlated with sample age (Fig. 5C). Moreover, analyses of subsections of the data from specific time periods (i.e. pollen and *sed*aDNA zones; Fig. 3) affected results (Fig. 5B), indicating the importance of an analytical basis for the choice of time period to include in these types of analyses.

Climate played an important role in early postglacial succession of plants, together with soil maturation, competition for light and plant migration patterns (Antonsson and Seppä, 2007). Surface temperature is a climatic parameter for which reliable long-term proxies exist and is related not only to vegetation change, but also to human land-use (Warden et al., 2017). Indeed, surface temperature and the accumulation of organic matter were strongly related to the variation in plant communities at Lake Ljøgottjern from ca. 8000 to 300 cal BCE. Significant relationships between surface temperature anomaly and the change in plant community were found in the separate vegetation zones as well as throughout the entire studied period of the last 9000 years (Fig. 5A&B). The strong role of temperature in the vegetation dynamics, also illustrated by increased pollen

concentrations of thermophilic tree taxa during the Holocene Thermal Maximum, is in agreement with the general ecological understanding that temperature is one of the main limiting factors for the northern limits of temperate tree taxa (Jackson et al., 2009; Woodward et al., 2004). Other studies quantifying the role of anthropogenic and environmental factors similarly indicate climate as the major driver of vegetation composition throughout the Holocene and especially before the introduction of agriculture (Kuosmanen et al., 2018; Marquer et al., 2017; Reitalu et al., 2013). We further identified significant negative relationships between temperature and all of the environmental changes associated with human presence, i.e. ^{14}C SPD, charcoal, diversity, and increased erosion (Fig. 5C). No clear statistical relationships between these environmental changes were found within the separate time periods (8000–300 cal BCE and 300 cal BCE–1800 cal CE; Appendix A.12).

In the more recent period, ca. 300 cal BCE–1800 cal CE, variations in plant communities at Lake Ljøgottjern were related to surface temperature, ^{14}C SPD, and Ti, with evidence of more intense farming during periods of high population density based on the intensification of anthropochore pollen, which were followed by periods of lower temperatures. Some variation in the presence of pastoral animals was observed, but as we focused on the presence and absence of livestock as a conservative estimate of pastoral activities at Lake Ljøgottjern, inferences about the intensity of pastoral activities are limited. Nevertheless, plant community composition and presence/absence of livestock were significantly related when including all zones in the stepwise dbRDA (Fig. 5B). However, this relationship was not evident within individual zones. Similarly, many correlations between pairs of environmental terms, particularly relating to human activities (charcoal, ^{14}C SPD, livestock and Ti), were only found when including both 8000–300 cal BCE and 300 cal BCE–1800 cal CE time periods, strongly suggesting that the transition between the two vegetation zones was primarily a result of anthropogenic impacts. The onset of strong human impacts on vegetation through the intensification of agriculture in the wider area of southeastern Norway has been dated to the Early Iron Age (500 BCE–550 CE; Høeg, 1997; Mjærum, 2020; Myhre, 2002). Similar increased roles of anthropogenic factors on plant communities during the late Holocene have been previously found for Sweden and Finland (Kuosmanen et al., 2018) associated with growing human population size, and for Estonia (Reitalu et al., 2013) and Europe (including Scandinavia and the Baltic; Marquer et al., 2017) associated with the establishment and expansion of agriculture.

6. Conclusions

Past human land-use has had a major impact on our ecosystems and knowledge about the role of cultural, ecological and climatic factors in driving these changes is important for understanding how our modern ecosystems were shaped. Distinguishing anthropogenic and environmental drivers of biological change requires integrated analyses of long-term records from a diversity of disciplines, combining knowledge about the local human history as well as climatic and other environmental changes. Using evidence from high-resolution pollen, *sed*aDNA and

geochemical analysis from the same lake sediment core, combined with archaeological evidence of local human settlement and regional population dynamics we were able to obtain a detailed reconstruction of complex anthropogenic and environmental dynamics affecting the vegetation at Lake Ljøgottjern for the last 10,000 years of the Holocene.

Although the exact timing of transitions to pastoralism and cereal cultivation differ between localities, the timing of these transitions at Lake Ljøgottjern are concurrent with multiple other locations in southeastern Norway. Pollen and *sed*aDNA analyses reflected vegetation at different spatial scales matching the archaeological evidence of the establishment of Bronze Age (ca. 1800–500 BCE) and Iron Age (ca. 550 BCE–1050 CE) farms at varying distances from the lake. Together, they provide a detailed timeline of cultivation and pastoralism. Our statistical analyses demonstrate that vegetation changes were primarily related to natural processes during most of the Holocene (ca. 8000–300 cal BCE), up until the Early Iron Age (ca. 500 BCE–550 CE), when human population densities in the region started increasing. Periods of decreased versus intensified human activity could be distinguished in the 300 cal BCE–1800 cal CE time period, with evidence of more intense farming during periods of high population density and possibly related to surface temperatures. This led to a rapid shift in plant communities, presence of livestock, increased erosion and high charcoal concentrations. Together, the integrated multi-proxy results from our study show the complex relations between environmental changes, facilitate the understanding of the coupled dynamics of climate, soils, human activities, and vegetation during the Holocene, and specifically highlight the importance of anthropogenic activities in long-term shaping of plant communities.

Author contributions

Conceptualisation ATS, MB, KK, SB, KL, FI; coring JB, ES, MB, EB; core subsampling MB, EB, SB, ATS; geochemical analysis MB, EB; sedaDNA and bioinformatic analysis ATS; pollen analysis HH; age-depth modelling MB; summed probability distribution analysis KL; archaeological analysis FI; validation of sedaDNA identifications AK, AB; statistical analysis and writing original draft ATS; review and editing SB, KK, MB, KL, FI. All co-authors commented on the manuscript.

Acknowledgements

This work contributes to the FRIPRO Toppforsk project VIKINGS "Volcanic Eruptions and their Impacts on Climate, Environment, and Viking Society in 500–1250 CE", funded by the Research Council of Norway (project number 275191). The two coring campaigns and the XRF geochemistry were conducted within the National infrastructure EARTHLAB (NRC 226171) at the University of Bergen. This work was supported by the Research Council of Norway Centres of Excellence CEED project 223272. Financial support was provided by the Faculty of Mathematical and Natural Sciences, University of Oslo. We thank the Norwegian Sequencing centre for their advice and DNA sequencing. DNA analyses were

performed on the Saga computer cluster, owned by the University of Oslo and Uninett/Sigma2, and operated by the Department of Research Computing at USIT, University of Oslo.

Data availability

All raw read data are available at the European Nucleotide Archive (ENA) under study accession number PRJEB44988. Scripts for read data processing are available at https://github.com/terschure/dataprocessing_LJO. All processed data are available in the supporting information of this article.

References

Alsos, I. G., Sjögren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., ... & Pedersen, M. W. (2016). Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene, 26(4), 627-642.

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PloS one, 13(4), e0195403.

Alsos, I. G., Sjögren, P., Brown, A. G., Gielly, L., Merkel, M. K. F., Paus, A., ... & Van Der Bilt, W. G. (2020). Last Glacial Maximum environmental conditions at Andøya, northern Norway; evidence for a northern ice-edge ecological "hotspot". Quaternary Science Reviews, 239, 106364.

Antonsson, K., & Seppä, H. (2007). Holocene temperatures in Bohuslän, southwest Sweden: a quantitative reconstruction from fossil pollen data. Boreas, 36(4), 400-410.

Arnaud, F., Poulenard, J., Giguet-Covex, C., Wilhelm, B., Révillon, S., Jenny, J. P., ... & Sabatier, P. (2016). Erosion under climate and human pressures: An alpine lake sediment perspective. Quaternary Science Reviews, 152, 1-18.

Bajard, M., Bakke, J., Støren, E., Ballo, E., Høeg, H., Iversen, F., Loftsgarden, K., Svensen, H. & Krüger, K. (Submitted). Climate variability controlled the pre-Viking society development during the Late Antiquity in Eastern Norway.

Bajard, M., Poulenard, J., Sabatier, P., Bertrand, Y., Crouzet, C., Ficetola, G. F., ... & Arnaud, F. (2020). Pastoralism increased vulnerability of a subalpine catchment to flood hazard through changing soil properties. Palaeogeography, Palaeoclimatology, Palaeoecology, 538, 109462. https://doi.org/10.1016/j.palaeo.2019.109462

Bajard, M., Poulenard, J., Sabatier, P., Etienne, D., Ficetola, F., Chen, W., ... & Arnaud, F. (2017). Long-term changes in alpine pedogenetic processes: Effect of millennial agro-pastoralism activities (French-Italian Alps). Geoderma, 306, 217-236. https://doi.org/10.1016/j.geoderma.2017.07.005

Bajard, M., Sabatier, P., David, F., Develle, A. L., Reyss, J. L., Fanget, B., ... & Arnaud, F. (2016). Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. The Holocene, 26(3), 350-364.

Behre, K. E., & KE, B. (1981). The interpretation of anthropogenic indicators in pollen diagrams. Pollen et Spores 23, 225–245.

Birks, H. H., & Bjune, A. E. (2010). Can we detect a west Norwegian tree line from modern samples of plant remains and pollen? Results from the DOORMAT project. Vegetation history and archaeobotany, 19(4), 325-340.

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian analysis, 6(3), 457-474.

Boessenkool, S., Epp, L. S., Haile, J., Bellemain, E. V. A., Edwards, M., Coissac, E., ... & Brochmann, C. (2012). Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. Molecular ecology, 21(8), 1806-1815. https://doi.org/10.1111/j.1365-

294X.2011.05306.x

Boivin, N., & Crowther, A. (2021). Mobilizing the past to shape a better Anthropocene. Nature Ecology & Evolution, 5(3), 273-284. https://doi.org/10.1038/s41559-020-01361-4

Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., ... & Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. Proceedings of the National Academy of Sciences, 113(23), 6388-6396. https://doi.org/10.1073/pnas.1525200113

Bonsall, C., Macklin, M. G., Anderson, D. E., & Payton, R. W. (2002). Climate change and the adoption of agriculture in north-west Europe. European journal of Archaeology, 5(1), 9-23.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular ecology resources, 16(1), 176-182.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological monographs, 84(1), 45-67.

Crema, E. R., & Bevan, A. (2021). Inference from large sets of radiocarbon dates: software and methods. Radiocarbon, 63(1), 23-39.

Ellis, E. C. (2015). Ecology in an anthropogenic biosphere. Ecological Monographs, 85(3), 287-331. https://doi.org/10.1890/14-2274.1

Epp, L. S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., ... & Brochmann, C. (2015). Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. Quaternary Science Reviews, 117, 152-163.

Epp, L. S., Zimmermann, H. H., & Stoof-Leichsenring, K. R. (2019). Sampling and Extraction of Ancient DNA from Sediments. Methods in molecular biology (Clifton, NJ), 1963, 31-44.

Faegri, K., Kaland, P. E., & Krzywinski, K. (1989). Textbook of pollen analysis (No. Ed. 4). John Wiley & Sons Ltd.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., ... & Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. BMC genomics, 11(1), 1-10. https://doi.org/10.1186/1471-2164-11-434

Freeman, J., Byers, D. A., Robinson, E., & Kelly, R. L. (2018). Culture process and the interpretation of radiocarbon data. Radiocarbon, 60(2), 453-467.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P. J., Griggo, C., Gielly, L., ... & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature communications, 5(1), 1-7. https://doi.org/10.1038/ncomms4211

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., ... & Arnaud, F. (2019). New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. Scientific reports, 9(1), 1-21. https://doi.org/10.1038/s41598-019-50339-1

Grimm, E. C. (1987). CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. Computers & geosciences, 13(1), 13-35.

Hagen, A. (1997). Gåten om kong Raknes grav: hovedtrekk i norsk arkeologi. Cappelen.

Helliksen, W. (1997). Gård og utmark på romerrike 1100 f. Kr.-1400 e. Kr: Gardermoprosjektet. Universitetets Oldsaksamling.

Hjelle, K. L., & Sugita, S. (2012). Estimating pollen productivity and relevant source area of pollen using lake sediments in Norway: how does lake size variation affect the estimates?. The Holocene, 22(3), 313-324.

Hjelle, K. L., Halvorsen, L. S., Prøsch-Danielsen, L., Sugita, S., Paus, A., Kaland, P. E., ... & Midtbø, I. (2018). Long-term changes in regional vegetation cover along the west coast of southern Norway: The importance of human impact. Journal of Vegetation Science, 29(3), 404-415. https://doi.org/10.1111/jvs.12626

Hellman, S., Bunting, M. J., & Gaillard, M. J. (2009). Relevant source area of pollen in patchy cultural landscapes and signals of anthropogenic landscape disturbance in the pollen record: a simulation approach. Review of Palaeobotany and Palynology, 153(3-4), 245-258.

Høeg, H. I. (1996). Pollenanalytiske undersøkelser i" Østerdalsområdet" med hovedvekt på Rødsmoen, Åmot i Hedmark. Universitetets oldsaksamling.

Høeg, H. I. (1997). Pollenanalytiske undersøkelser på Øvre Romerike: Ullensaker og Nannestad, Akershus fylke: Gardermoprosjektet.

Universitetets oldsaksamling.
Jackson, S. T., Betancourt, J. L., Booth, R. K., & Gray, S. T. (2009). Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. Proceedings of the National Academy of Sciences, 106(Supplement 2), 19685-19692.
Johnsen, G. (1941). Slekten Koren. Bokstuas forlag, Oslo.
Juggins, S. (2020). rioja: Analysis of Quaternary Science Data. R package version 0.9-26, https://cran.r-project.org/package=rioja.
Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P. S., ... & Davis, B. (2020). Holocene global mean surface temperature, a multi-method reconstruction approach. Scientific data, 7(1), 1-13. https://doi.org/10.1038/s41597-020-0530-7
Kotov, S., & Paelike, H. (2018, December). QAnalySeries-a cross-platform time series tuning and analysis tool. In AGU Fall Meeting Abstracts (Vol. 2018, pp. PP53D-1230).
Kuosmanen, N., Marquer, L., Tallavaara, M., Molinari, C., Zhang, Y., Alenius, T., ... & Seppä, H. (2018). The role of climate, forest fires and human population size in Holocene vegetation dynamics in Fennoscandia. Journal of Vegetation Science, 29(3), 382-392.
Krause-Kyora, B., Makarewicz, C., Evin, A., Flink, L. G., Dobney, K., Larson, G., ... & Nebel, A. (2013). Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. Nature communications, 4(1), 1-7. https://doi.org/10.1038/ncomms3348
Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological monographs, 69(1), 1-24.
Li, D. (2018). hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. Journal of Open Source Software, 3(31), 1041.
Loftsgarden, K. & Solheim, S. (in press). Uncovering population dynamics in southeast Norway from 1300 BC to 1500 AD using summed radiocarbon probability distributions.
Longva, O., & Thoresen, M. K. (1989). The age of the Hauerseter delta. Norsk geologisk tidsskrift, 69, 131-134.
Malmström, H., Gilbert, M. T. P., Thomas, M. G., Brandström, M., Storå, J., Molnar, P., ... & Willerslev, E. (2009). Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. Current biology, 19(20), 1758-1762. https://doi.org/10.1016/j.cub.2009.09.017
Marquer, L., Gaillard, M. J., Sugita, S., Poska, A., Trondman, A. K., Mazier, F., ... & Seppä, H. (2017). Quantifying the effects of land use and climate on Holocene vegetation in Europe. Quaternary Science Reviews, 171, 20-37.
Mjærum, A. (2020). The emergence of mixed farming in eastern Norway. Agricultural History Review, 68(1), 1-21.
Myhre, B. (2002). Landbruk, landskap og samfunn 4000 f. Kr.–800 e. Kr. Norges landbrukshistorie I, 400f. Kr.-135 e. Kr. Jorda blir levevei. Det Norske Samlaget, Oslo, 11-213.
Nelson, G. C., Bennett, E., Berhe, A. A., Cassman, K., DeFries, R., Dietz, T., ... & Zurek, M. (2006). Anthropogenic drivers of ecosystem change: an overview. Ecology and Society, 11(2).
Nesje, A. (1992). A piston corer for lacustrine and marine sediments. Arctic and alpine research, 24(3), 257-259.
Nesten, H. J. L. (1951). Ullensaker: en bygdebok (Vol. 3). Grøndahl & Søns Boktrykkeri.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan
Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., ... & Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. Molecular ecology, 24(7), 1485-1498. https://doi.org/10.1111/mec.13136
Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924-942. https://doi.org/10.1111/nph.14470
Prentice, I. C. (1985). Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. Quaternary Research, 23(1), 76-86.
Price, T. D. (2000). The introduction of farming in northern Europe. Europe's first farmers, 260-300.
R Core Team (2020). R: A language and environment for statistical computing (manual). Vienna,

Austria.
Ramette, A. (2007). Multivariate analyses in microbial ecology. FEMS microbiology ecology, 62(2), 142-160.
Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., ... & Talamo, S. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon, 62(4), 725-757.
Reitalu, T., Seppä, H., Sugita, S., Kangur, M., Koff, T., Avel, E., ... & Veski, S. (2013). Long-term drivers of forest composition in a boreonemoral region: the relative importance of climate and human impact. Journal of Biogeography, 40(8), 1524-1534.
Rick, J. W. (1987). Dates as data: an examination of the Peruvian preceramic radiocarbon record. American Antiquity, 52(1), 55-73.
Robinson, D. E. (2003). Neolithic and Bronze Age agriculture in southern Scandinavia–recent archaeobotanical evidence from Denmark. Environmental Archaeology, 8(2), 145-165. https://doi.org/10.1179/env.2003.8.2.145
Ruddiman, W. F. (2003). The anthropogenic greenhouse era began thousands of years ago. Climatic change, 61(3), 261-293.
Ruddiman, W. F., Ellis, E. C., Kaplan, J. O., & Fuller, D. Q. (2015). Defining the epoch we live in. Science, 348(6230), 38-39. https://doi.org/10.1126/science.aaa7297
Rygh, O. (1898). Norske gaardnavne: Akershus amt (1898) (Vol. 2). WC Fabritius.
Sejrup, H. P., Seppä, H., McKay, N. P., Kaufman, D. S., Geirsdóttir, Á., de Vernal, A., ... & Andrews, J. T. (2016). North Atlantic-Fennoscandian Holocene climate trends and mechanisms. Quaternary Science Reviews, 147, 365-378.
Simonsen, M. F. (1997). Boplasspor. Ljøgot, 137/1, Ullensaker kommune, Akershus. Rapport Arkeologisk utgravning.
Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. Frontiers in Ecology and Evolution, 6, 149.
Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Merkel, M. K. F., ... & Alsos, I. G. (2017). Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytologist, 213(2), 929-941.
Skre, D. (1997). Raknehaugen. En empirisk loftsrydding. Viking, 60, 7-42.
Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., ... & Yoccoz, N. G. (2015). Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PloS one, 10(1), e0115335. https://doi.org/10.1371/journal.pone.0115335
Solberg, B. (2000). Jernalderen i Norge: ca. 500 f. Kr.-1030 e. Kr. Cappelen akademisk forlag.
Solheim, S., & Persson, P. (2018). Early and mid-Holocene coastal settlement and demography in southeastern Norway: Comparing distribution of radiocarbon dates and shoreline-dated sites, 8500–2000 cal. BCE. Journal of Archaeological Science: Reports, 19, 334-343. https://doi.org/10.1016/j.jasrep.2018.03.007
Solheim, S., & Iversen, F. (2019). The mid-6th century crises and their impacts on human activity and settlements in southeastern Norway. Ruralia, 423-434.
Soltvedt, E. C., & Henningsmoen, K. E. (2016). Agricultural and household activities in Vestfold, Southeast Norway, as illustrated by pollen data and the charred remains of crops and wild plants. Environmental Archaeology, 21(1), 11-30.
Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., ... & Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Molecular Ecology Resources, 10(6), 1009-1018. https://doi.org/10.1111/j.1755-0998.2010.02855.x
Sørensen, L., & Karg, S. (2014). The expansion of agrarian societies towards the north–new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. Journal of Archaeological Science, 51, 98-114. https://doi.org/10.1016/j.jas.2012.08.042
Stephens, L., Fuller, D., Boivin, N., Rick, T., Gauthier, N., Kay, A., ... & Ellis, E. (2019). Archaeological assessment reveals Earth's early transformation through land use. Science, 365(6456), 897-902. https://doi.org/10.1126/science.aax1192
Stockmarr, J. A. (1971). Tabletes with spores used in absolute pollen analysis. Pollen spores, 13, 615-621.
Sugita, S. (1994). Pollen representation of vegetation in Quaternary sediments: theory and method in patchy vegetation. Journal of ecology, 881-897.
Sugita, S., Gaillard, M. J., & Broström, A. (1999). Landscape openness and pollen records: a simulation approach. The Holocene, 9(4), 409-

421.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., ... & Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic acids research, 35(3), e14-e14. https://doi.org/10.1093/nar/gkl938

Universitetsmuseenes samlingsportaler (2020). UniMus: Universitetsmuseenes samlingsportaler. http://www.unimus.no

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. Science, 277(5325), 494-499.

Voldstad, L. H., Alsos, I. G., Farnsworth, W. R., Heintzman, P. D., Håkansson, L., Kjellman, S. E., ... & Eidesen, P. B. (2020). A complete Holocene lake sediment ancient DNA record reveals long-standing high Arctic plant diversity hotspot in northern Svalbard. Quaternary science reviews, 234, 106207. https://doi.org/10.1016/j.quascirev.2020.106207

Wanner, H., Solomina, O., Grosjean, M., Ritz, S. P., & Jetel, M. (2011). Structure and origin of Holocene cold events. Quaternary Science Reviews, 30(21-22), 3109-3123.

Warden, L., Moros, M., Neumann, T., Shennan, S., Timpson, A., Manning, K., ... & Damsté, J. S. (2017). Climate induced human demographic and cultural change in northern Europe during the mid-Holocene. Scientific reports, 7(1), 1-11. https://doi.org/10.1038/s41598-017-14353-5

Wieckowska-Lüth, M., Kirleis, W., & Doerfler, W. (2017). Holocene history of landscape development in the catchment of Lake Skogstjern, southeastern Norway, based on a high-resolution multi-proxy record. The Holocene, 27(12), 1928-1947. https://doi.org/10.1177/0959683617715691

Wieckowska-Lüth, M., Solheim, S., Schülke, A., & Kirleis, W. (2018). Towards a refined understanding of the use of coastal zones in the Mesolithic: New investigations on human–environment interactions in Telemark, southeastern Norway. Journal of Archaeological Science: Reports, 17, 839-851. https://doi.org/10.1016/j.jasrep.2017.12.045

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., ... & Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47-51. https://doi.org/10.1038/nature12921

Williams, M., Zalasiewicz, J., Haff, P. K., Schwägerl, C., Barnosky, A. D., & Ellis, E. C. (2015). The anthropocene biosphere. The Anthropocene Review, 2(3), 196-219.

Wood, S. (2018). Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. R package version, 1-8.

Woodward, F. I., Lomas, M. R., & Kelly, C. K. (2004). Global climate and the distribution of plant biomes. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359(1450), 1465-1476.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quascirev.2021.107175.

A.1. Geochemistry and radiocarbon ages

XRF geochemistry and chronology methods

Both cores were scanned with an ITRAX (XRF) core scanner from COX analytics at the EARTHLAB with a resolution of 200 µm for LJØ118 and 1000 µm for LJØ119. The surface of the cores was cleaned, smoothed and covered with an ultra-thin transparent film to avoid contamination and desiccation of the sediment. We used different settings for optimizing the sensitivity for the most interesting geochemical elements. These settings were adjusted to 35 kV and 30 mA for 10 s with a Mo tube to detect Ca, Ti, K, Si, Fe, Mn. To obtain matching resolutions of 1 mm and correlate the two cores, Ca, Ti, K, Si, Fe, Mn, Inc and Coh data of LJØ118 were averaged every 5 measures. The depth of the cores were correlated based on the variations in Ti, Fe and Mn and visual observation of the sediment using QAnalyseries 1.4.2 (Kotov and Paelike, 2018). As pollen and *sed*aDNA samples were taken from the LJØ119 core, we focus on the geochemical results for LJØ119 in this study (Fig. A.1).

The chronology of the sediment sequence is based on six AMS (accelerator mass spectrometer) radiocarbon dates from plant macrofossils from LJØ118 and realized by the Tandem Laboratory at Uppsala University (Table A.1). The ^{14}C ages were calibrated using the IntCal20 calibration curve (Reimer et al., 2020). The age-depth model for the LJØ118 sequence was generated using R software (version 3.5.2; R Core Team, 2020) and the R code package 'Bacon' 2.4.3 (Blaauw and Christen, 2011). In the age model, the top of the core was set to the year of coring, i.e., 2018 CE. The chronology of LJØ119 was deduced from the LJØ118 age model.

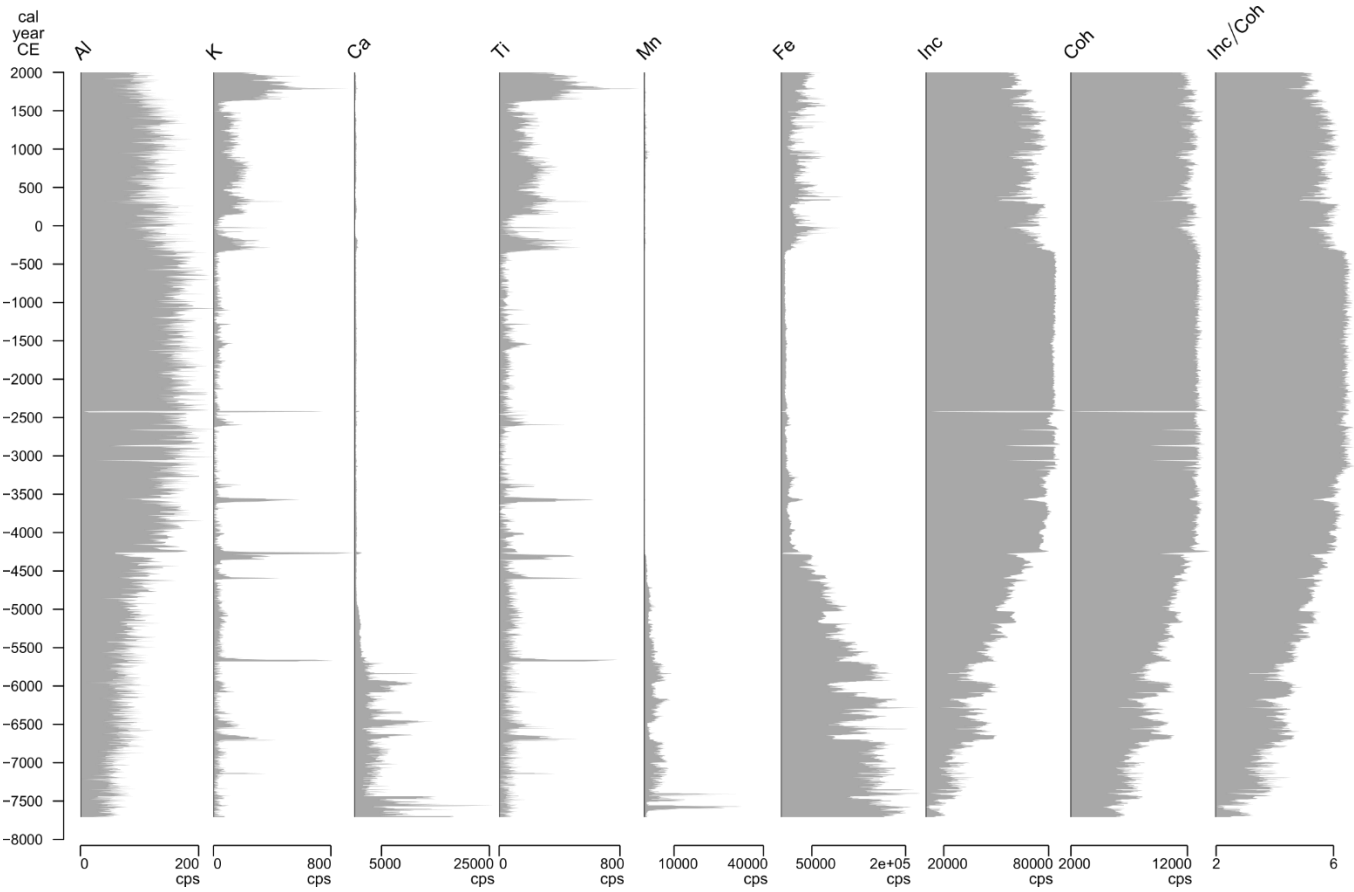


Fig. A.1. XRF measurements of Ca, Ti, K, Si, Fe, Mn, Inc and Coh in counts per seconds of the LJØ119 sediment core by an ITRAX core scanner from COX analytics.

Table A.1. Radiocarbon ages for the Lake Ljøgottjern sediment sequence. The ^{14}C ages were calibrated using the IntCal20 calibration curve (Reimer et al., 2020).

Lab number	Sample name	Core depth (cm)	Sample type	$\delta^{13}C$‰ V-PDB	^{14}C age BP	Min. age (cal BP)	Max. age (cal BP)	Min. age (cal CE)	Max. age (cal CE)	Mean age (cal CE)
Ua-60967	LJP118-I/IV-21,5-22cm	70.25	Leaf/bark	-23,6	431 ± 30	452	526	1498	1424	1461
Ua-60968	LJP118-I/IV-86-86,5	134.75	Grass and some leaves	-27,9	1115 ± 32	956	1071	994	879	936.5
Ua-60969	LJP118-II/IV-12cm	193	Aquatic plant	-26,8	1623 ± 31	1406	1547	544	403	473.5
Ua-60970	LJP118-II/IV-69cm	250	Leaves/bark, seed	-24,8	1970 ± 31	1827	1949	123	1	62
Ua-61898	LJP118w-3av4 54-55cm	375,5	Pieces of leaves	-28,9	6455 ± 34	7309	7431	-5359	-5481	-5420
Ua-61899	LJP118w-4av4 86,5-87cm	527,75	Bark	-26,8	8288 ± 35	9194	9424	-7244	-7474	-7359

Interpolation method

A generalised additive model (GAM) was used to smooth and interpolate values for generating geochemistry time series matching the resolution of the pollen and DNA datasets. In this process, the response variable for analysis (i.e. Ti and Inc/Coh ratio) was modelled as a function of time (cal year BP) using the *gam*-function of the *mgcv* package (version 1.8-35; Simpson, 2018; Wood, 2018) in R. We chose an adaptive smoother as Ti and Inc/Coh ratio series include periods of rapid change and periods of more gradual change. Settings for *k* varied to find the best model fit, evaluated with *gam.check.* We obtained interpolated values using the *predict.gam*-function.

A.2. DNA filtering steps

Table A.2. Number of sequence reads and unique sequences remaining after each filtering step for the two sequencing pools. We processed the *sed*aDNA sequence data using the OBITools package (http://metabarcoding.org/obitools/doc/index.html; Boyer et al., 2016) and filtered the identified sequences in R (version 3.5.2; R Core Team, 2020).

			Sequencing pool 1		*Sequencing pool 2*	
Filtering steps	Program	Target group (reference database)	Total reads	Unique sequences	Total reads	Unique sequences
Raw reads			220 591 834		64 492 511	
Pairwise alignment	*Illumina-pairedend*		220 591 834		64 492 511	
Assignment to samples	*ngsfilter*	plants	51 781 661		31 320 055	
		mammals	103 795 998		23 091 660	
Removal of reads with count =1 & < 8 bp length & merging identical reads	*obigrep & obiuniq*	plants	50 438 161	230 810	30 972 513	97 833
		mammals	87 674 496	647 930	15 492 318	134 593
Matching the reference database	*ecotag & obigrep*	plants (arctborbryo)	50 438 161	230 810	30 972 513	97 833
		plants (embl)	50 433 435	229 905	30 964 838	96 985
		mammals (embl)	87 674 496	647 930	15 492 318	134 593
Identification & removal of PCR/sequencing errors	*obiclean*	plants (arct)	48 012 330	143 172	29 810 852	97 833
		plants (embl)	48 008 064	142 610	29 803 609	96 985
		mammals (embl)	83 348 865	453 823	15 015 242	81 575
Merging the plant results from the two reference databases	R	plants	48 012 348	143 190	29 810 881	51 821
Removal of sequences that are not identified to the target group or without family attribute	R	mammals	28 543 298	59 527	2 099 879	1 370
Removal of sequences with maximum abundance in negative controls	R	plants	47 682 752	143 045	29 766 311	51 613
		mammals	24 781 219	49 066	1 894 232	1 177
Removal of sequences with < 100% match (plants) or 98% match (mammals)	R	plants	23 993 421	652	10 681 977	577
		mammals	24 531 589	551	1 886 507	116
Removal of sequences with < 10 reads in a PCR repeat	R	plants	23 983 890	358	10 674 392	347
		mammals	24 530 641	156	1 886 380	32
Removal of PCR repeats with < 100 reads in total	R	plants	23 983 745	358	10 674 365	347
		mammals	24 530 420	156	1 886 230	32
Removal of sequences present in < 2 PCR repeats & sequences with mean read count > mean read count in blanks	R	plants	23 751 263	223	10 633 638	232
		mammals	23 025 817	122	1 846 149	23
Merging the pools	R	plants	34 300 780	274	-	-
		mammals	24 871 966	139	-	-

A.3. DNA read counts per sample

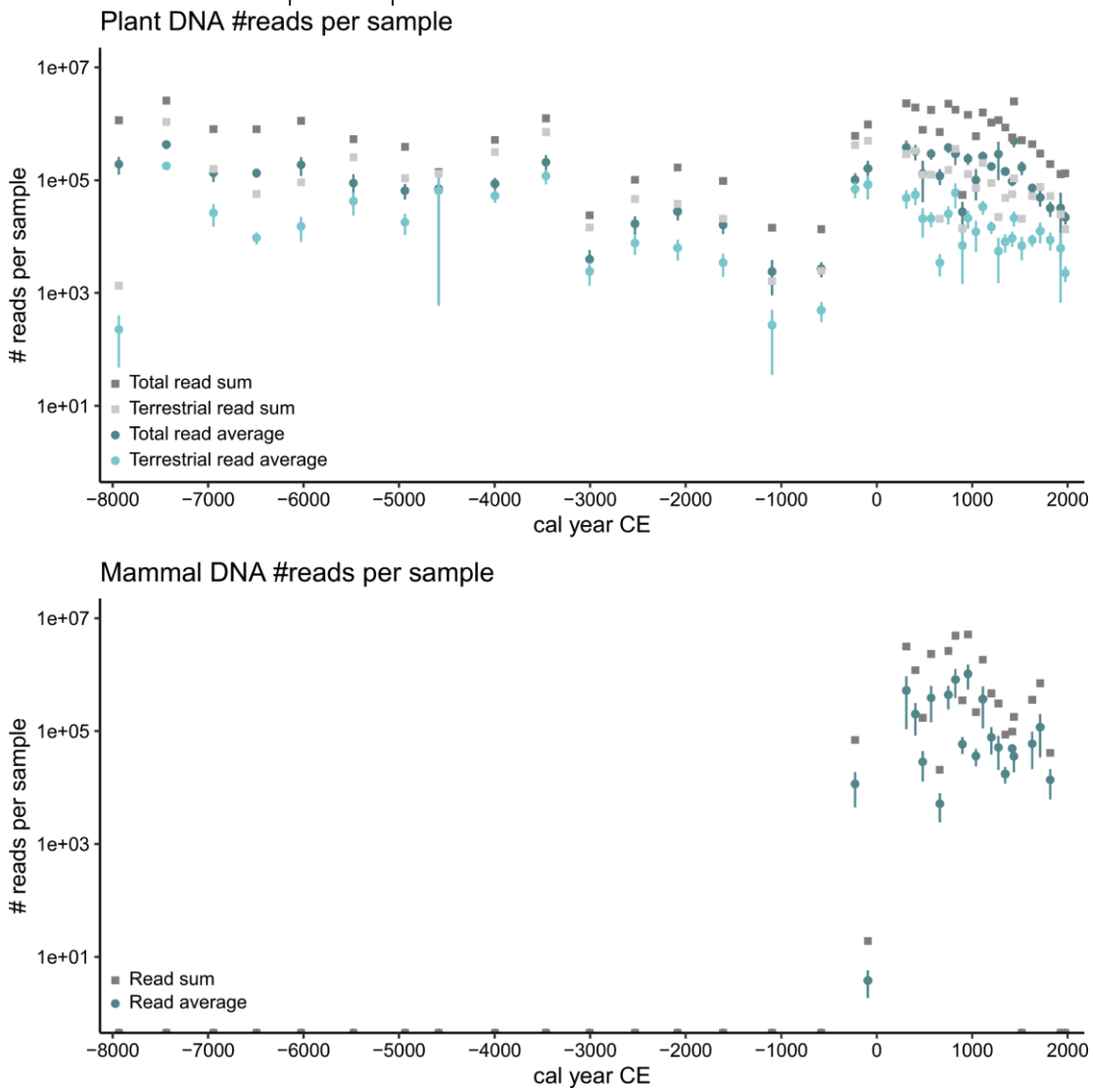


Fig. A.3. Summed and average read counts per sample for the total and the terrestrial plant *sed*aDNA (top) and mammal *sed*aDNA (bottom) datasets, with bars indicating the standard error from the mean.

A.4. DNA read counts and terrestrial biodiversity

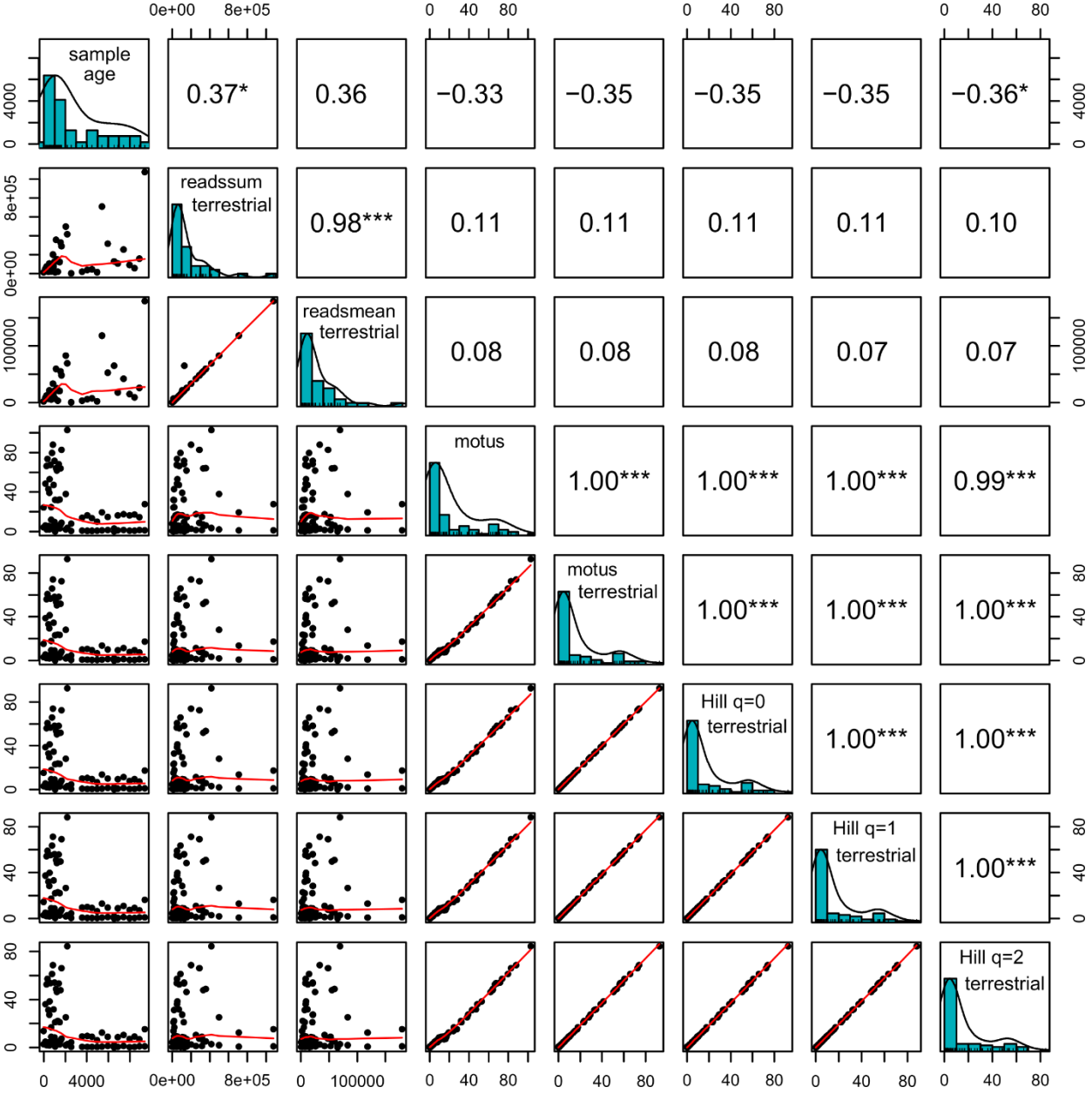


Fig. A.4. Spearman rank order correlations between terrestrial plant *sed*aDNA sample age (cal year BP), total number of sequence reads, average number of sequence reads, number of molecular operational taxonomic units (MOTUs), and Hill biodiversity numbers where q=0 corresponds to taxonomic richness, q=2 to Shannon diversity index, and q=2 to inverse Simpson index. P-values were adjusted with the Bonferroni method and significant correlations are indicated with *** *p*-value < 0.001, ** *p*-value < 0.01 and * *p*-value < 0.05.

A.5. Correlations between DNA reads and replicates

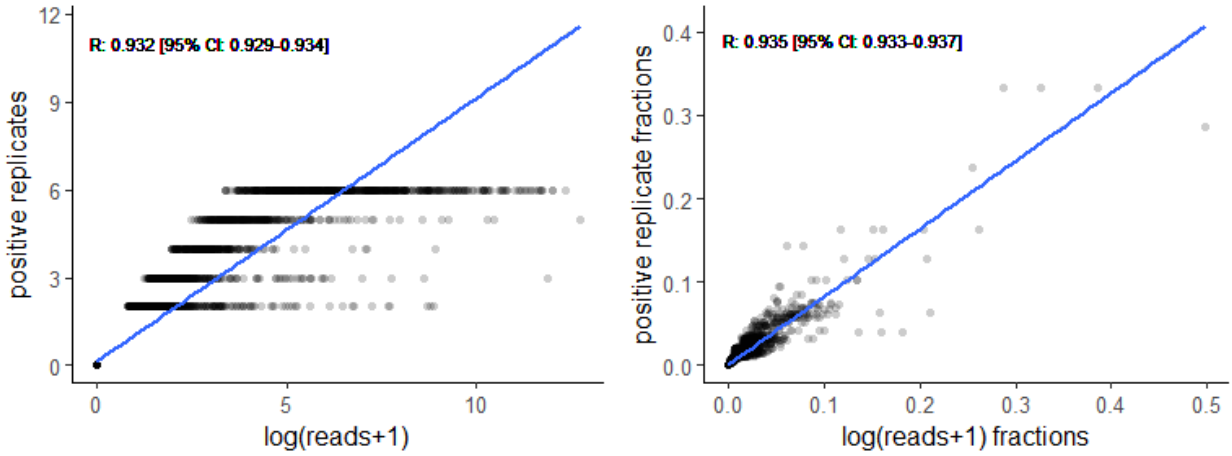


Fig. A.5. Pearson correlations between the log-transformed read counts and the number of positive replicates per plant molecular operational taxonomic unit (MOTU; r=0.932 [95% CI: 0.929-0.934]; left) and those standardized to sample fractions (r=0.935 [95% CI: 0.933-0.937]; right) of the total plant *sed*aDNA dataset (N=2185). Blue lines represent the calculated Pearson correlations.

Table A.5. Pearson correlations between the log-transformed read counts and positive replicates per molecular operational taxonomic unit (MOTU) in the plant *sed*aDNA dataset.

MOTU	N	R	*P*-value	MOTU	N	R	*P*-value	MOTU	N	R	*P*-value
GH000072	10	0.945	4.31E-15***	GH000015	9	0.991	1.17E-29***	GH000116	1	1	0.00E+00***
GH000073	2	1	5.50E-70***	GH000016	1	1	0.00E+00***	GH000115	1	1	0.00E+00***
GH000166	15	0.948	1.18E-15***	GH000018	17	0.973	5.49E-21***	GH000188	2	0.997	1.84E-40***
GH000087	8	0.980	2.14E-23***	GH000020	4	0.998	3.21E-41***	GH000195	4	0.971	1.86E-20***
GH000085	16	0.985	6.71E-26***	GH000021	4	0.986	2.21E-26***	GH000124	4	0.995	1.03E-34***
GH000044	1	1	0.00E+00***	GH000023	3	0.995	1.47E-35***	GH000121	1	1	0.00E+00***
GH000277	2	0.998	1.75E-41***	GH000024	1	1	0.00E+00***	GH000122	24	0.928	5.93E-13***
GH000278	4	0.972	1.18E-20***	GH000027	23	0.946	2.25E-15***	GH000125	1	1	0.00E+00***
GH000017	10	0.996	1.82E-36***	GH000028	1	1	0.00E+00***	GH000126	20	0.965	9.58E-19***
GH000019	6	0.996	5.99E-36***	GH000029	9	0.959	1.73E-17***	GH000127	2	1	1.17E-55***
GH000038	2	0.998	1.02E-43***	GH000030	26	0.915	1.07E-11***	GH000128	1	1	0.00E+00***
GH000146	2	0.993	5.81E-32***	GH000031	7	0.994	1.62E-33***	GH000163	1	1	0.00E+00***
GH000147	24	0.951	3.80E-16***	GH000032	17	0.929	4.27E-13***	GH000051	20	0.933	1.24E-13***
GH000264	22	0.953	1.64E-16***	GH000033	3	0.992	1.13E-30***	GH000158	1	1	0.00E+00***
GH000229	12	0.978	9.30E-23***	GH000034	7	0.983	9.25E-25***	GH000050	2	0.998	1.10E-40***
GH000162	22	0.967	2.07E-19***	GH000007	15	0.981	1.34E-23***	GH000052	12	0.989	2.51E-28***
GH000206	1	1	0.00E+00***	GH000035	3	0.977	2.42E-22***	GH000046	2	0.993	4.07E-32***
GH000205	12	0.977	2.18E-22***	GH000036	24	0.936	6.25E-14***	GH000047	19	0.941	1.40E-14***
GH000091	5	0.909	3.80E-11***	GH000025	1	1	0.00E+00***	GH000159	23	0.983	1.09E-24***
GH000090	3	0.987	1.13E-26***	GH000189	1	1	0.00E+00***	GH000049	14	0.981	6.92E-24***
GH000257	2	0.995	2.44E-34***	GH000069	4	0.998	7.88E-42***	GH000059	23	0.921	2.94E-12***
GH000089	1	1	0.00E+00***	GH000071	4	0.997	9.07E-40***	GH000045	18	0.935	7.19E-14***
GH000088	1	1	0.00E+00***	GH000130	1	1	0.00E+00***	GH000157	19	0.982	1.96E-24***
GH000057	7	0.985	8.07E-26***	GH000132	1	1	0.00E+00***	GH000104	1	1	0.00E+00***
GH000259	4	0.992	1.26E-30***	GH000129	1	1	0.00E+00***	GH000111	2	1	6.19E-64***
GH000155	5	0.995	5.52E-35***	GH000131	2	0.998	1.46E-42***	GH000095	5	1	4.52E-57***
GH000156	3	0.993	1.19E-31***	GH000055	1	1	0.00E+00***	GH000102	8	0.958	3.04E-17***
GH000053	28	0.967	2.65E-19***	GH000169	13	0.979	6.66E-23***	GH000103	10	0.99	7.69E-29***
GH000056	7	0.990	2.11E-29***	GH000039	2	0.989	2.01E-28***	GH000170	11	0.986	2.20E-26***
GH000150	38	0.798	5.18E-05***	GH000040	1	1	0.00E+00***	GH000105	1	1	0.00E+00***
GH000151	38	0.544	1.00E+00	GH000041	1	1	0.00E+00***	GH000185	15	0.958	2.05E-17***
GH000148	3	0.979	5.57E-23***	GH000037	1	1	0.00E+00***	GH000186	2	0.998	2.09E-41***
GH000149	1	1	0.00E+00***	GH000273	1	1	0.00E+00***	GH000075	2	0.999	2.34E-53***
GH000145	6	0.999	1.40E-52***	GH000272	1	1	0.00E+00***	GH000113	3	0.983	1.22E-24***
GH000077	4	0.996	1.75E-36***	GH000258	1	1	0.00E+00***	GH000237	3	0.999	7.98E-48***
GH000084	4	0.988	1.11E-27***	GH000200	14	0.99	6.28E-29***	GH000238	18	0.992	6.83E-31***
GH000063	30	0.979	8.07E-23***	GH000199	5	0.922	2.14E-12***	GH000239	8	0.983	1.50E-24***
GH000265	33	0.843	6.62E-07***	GH000120	11	0.987	6.98E-27***	GH000240	5	0.983	7.40E-34***

GH000060	18	0.961	7.24E-18***	GH000201	9	0.992	1.46E-30***	GH000080	5	0.99	5.22E-29***
GH000061	21	0.927	7.56E-13***	GH000198	1	1	0.00E+00***	GH000081	9	0.993	1.64E-31***
GH000117	28	0.960	1.19E-17***	GH000202	1	1	0.00E+00***	GH000062	3	0.999	5.59E-52***
GH000118	4	0.997	6.09E-38***	GH000203	18	0.927	6.56E-13***	GH000043	7	0.981	6.57E-24***
GH000070	10	0.976	4.78E-22***	GH000197	20	0.961	6.36E-18***	GH000078	2	0.999	5.99E-45***
GH000123	1	1	0.00E+00***	GH000193	10	0.954	1.34E-16***	GH000079	15	0.941	1.20E-14***
GH000048	23	0.968	1.57E-19***	GH000097	18	0.973	5.64E-21***	GH000082	13	0.957	3.18E-17***
GH000058	22	0.925	1.25E-12***	GH000209	7	0.985	1.42E-25***	GH000083	14	0.903	1.19E-10***
GH000196	1	1	0.00E+00***	GH000210	1	1	0.00E+00***	GH000086	9	0.987	1.27E-26***
GH000182	1	1	0.00E+00***	GH000065	2	0.992	2.29E-31***	GH000076	3	1	7.10E-56***
GH000119	18	0.909	3.88E-11***	GH000252	2	0.998	6.08E-41***	GH000042	10	0.971	2.30E-20***
GH000194	2	0.997	3.03E-38***	GH000253	17	0.978	1.09E-22***	GH000260	1	1	0.00E+00***
GH000228	2	1	8.12E-55***	GH000249	9	0.989	1.74E-28***	GH000261	19	0.967	2.44E-19***
GH000236	1	1	0.00E+00***	GH000250	2	0.989	2.89E-28***	GH000282	1	1	0.00E+00***
GH000232	2	0.979	4.25E-23***	GH000251	2	1	2.09E-88***	GH000064	16	0.966	6.13E-19***
GH000280	3	0.988	9.06E-28***	GH000230	2	0.991	3.51E-30***	GH000215	3	0.999	1.80E-50***
GH000281	2	0.977	2.12E-22***	GH000179	2	1	1.88E-62***	GH000211	9	0.973	8.32E-21***
GH000283	1	1	0.00E+00***	GH000190	1	1	0.00E+00***	GH000212	1	1	0.00E+00***
GH000279	2	0.998	3.63E-44***	GH000183	7	0.984	4.67E-25***	GH000208	1	1	0.00E+00***
GH000002	5	0.987	8.45E-27***	GH000164	13	0.976	8.57E-22***	GH000213	1	1	0.00E+00***
GH000233	1	1	0.00E+00***	GH000175	1	1	0.00E+00***	GH000216	9	0.969	6.37E-20***
GH000234	1	1	0.00E+00***	GH000176	4	1	1.91E-60***	GH000204	11	0.992	6.20E-31***
GH000235	1	1	0.00E+00***	GH000178	2	0.997	2.69E-38***	GH000214	4	0.987	7.81E-27***
GH000263	1	1	0.00E+00***	GH000177	20	0.938	2.99E-14***	GH000217	1	1	0.00E+00***
GH000231	2	0.996	8.41E-37***	GH000174	1	1	0.00E+00***	GH000218	8	0.998	4.96E-43***
GH000247	1	1	0.00E+00***	GH000114	1	1	0.00E+00***	GH000219	13	0.953	1.87E-16***
GH000248	12	0.993	3.41E-32***	GH000181	4	0.998	3.27E-43***	GH000220	13	0.947	2.11E-15***
GH000152	2	0.982	3.78E-24***	GH000098	15	0.979	8.39E-23***	GH000221	6	0.988	1.41E-27***
GH000245	8	0.983	1.27E-24***	GH000099	17	0.977	3.09E-22***	GH000074	4	0.989	6.26E-28***
GH000242	7	0.991	5.70E-30***	GH000192	1	1	0.00E+00***	GH000184	4	0.994	4.04E-33***
GH000246	7	0.986	1.55E-26***	GH000191	16	0.978	1.14E-22***	GH000161	10	0.992	4.93E-31***
GH000241	16	0.998	7.76E-43***	GH000270	13	0.984	3.20E-25***	GH000168	13	0.972	1.04E-20***
GH000244	4	0.988	2.70E-27***	GH000271	5	0.991	3.75E-30***	GH000167	3	0.999	6.34E-50***
GH000243	4	1	8.07E-56***	GH000269	1	1	0.00E+00***	GH000172	5	0.976	6.69E-22***
GH000256	1	1	0.00E+00***	GH000187	4	0.973	7.42E-21***	GH000171	32	0.95	6.68E-16***
GH000112	4	0.970	5.46E-20***	GH000180	1	1	0.00E+00***	GH000173	38	0.904	1.03E-10***
GH000106	2	0.998	6.94E-43***	GH000133	6	0.976	7.48E-22***	GH000165	15	0.976	7.26E-22***
GH000101	1	1	0.00E+00***	GH000134	18	0.969	9.66E-20***	GH000094	1	1	0.00E+00***
GH000100	8	0.991	3.03E-30***	GH000135	1	1	0.00E+00***	GH000096	1	1	0.00E+00***
GH000207	16	0.988	6.94E-28***	GH000136	1	1	0.00E+00***	GH000153	1	1	0.00E+00***
GH000054	1	1	0.00E+00***	GH000137	1	1	0.00E+00***	GH000092	4	0.998	5.72E-41***
GH000255	1	1	0.00E+00***	GH000138	1	1	0.00E+00***	GH000093	12	0.989	4.68E-28***
GH000254	1	1	0.00E+00***	GH000139	10	0.949	1.07E-15***	GH000067	20	0.961	5.28E-18***
GH000010	17	0.985	1.08E-25***	GH000140	4	0.949	7.81E-16***	GH000068	1	1	0.00E+00***
GH000026	2	0.999	5.44E-45***	GH000141	1	1	0.00E+00***	GH000066	36	0.977	2.76E-22***
GH000011	7	0.972	1.39E-20***	GH000142	1	1	0.00E+00***	GH000109	1	1	0.00E+00***
GH000014	20	0.965	6.24E-19***	GH000143	15	0.969	6.43E-20***	GH000110	33	0.918	6.18E-12***
GH000005	16	0.953	2.03E-16***	GH000144	1	1	0.00E+00***	GH000107	2	0.966	5.33E-19***
GH000004	2	0.989	2.00E-28***	GH000267	11	0.959	1.78E-17***	GH000108	26	0.965	6.43E-19***
GH000006	1	1	0.00E+00***	GH000276	6	0.99	1.84E-29***	GH000262	1	1	0.00E+00***
GH000008	15	0.944	4.41E-15***	GH000266	20	0.928	4.78E-13***	GH000154	4	0.998	1.41E-42***
GH000009	5	0.989	3.14E-28***	GH000268	2	1	5.23E-68***	GH000160	9	0.989	1.71E-28***
GH000012	10	0.978	1.42E-22***	GH000274	23	0.826	3.78E-06***				
GH000013	15	0.98	2.748E-24***	GH000275	7	0.991	8.19E-30***				

* P-value < .05
** P-value < .01
*** P-value < .001

A.6. Temperature anomaly data

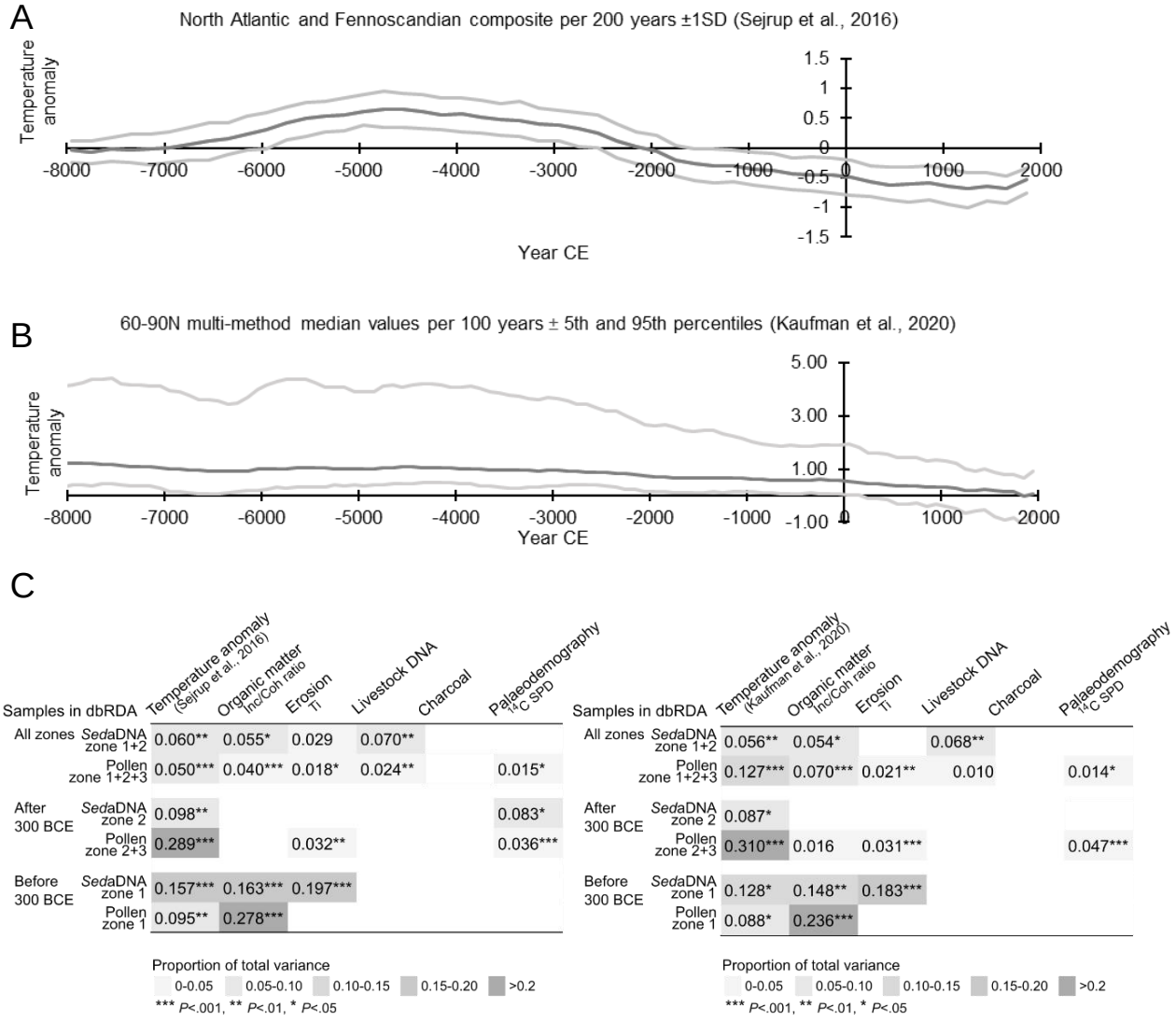


Fig. A.6. Surface temperature anomaly composites for A) the North Atlantic and Fennoscandia from Sejrup et al. (2016) and for B) the 60-90N northern hemisphere as presented by Kaufman et al. (2020). Grey lines around the temperature anomalies indicate the standard deviation (Sejrup et al., 2016) or the 5th and 95th percentiles (Kaufman et al., 2020). C) Stepwise distance-based redundance analyses (dbRDA; see Appendix A.11 for details) were performed separately for subsets of the pollen and *sed*aDNA data and using either the Sejrup et al. (2016; left) or the Kaufman et al. (2020; right) surface temperature anomaly composite. The conditional proportion of variance explained by the temperature anomaly composite from the Kaufman et al. (2020) publication for the dbRDA of all pollen samples is higher than that from the Sejrup et al. (2016) publication and some differences were found in the proportions of variance explained by the included environmental terms. However, the overall patterns of the dbRDA analyses using the different temperature anomaly composites are very similar.

Interpolation method

General Additive Models (GAMs) allow flexible modelling of nonlinear relationships (Simpson, 2018), and were fitted to the surface temperature anomaly data using the *gam* function in the *mgcv* R package (version 1.8-35; Wood, 2018) to interpolate values and match the time resolutions of the pollen and DNA. In this process, the temperature anomaly was modelled as a function of time (cal year BP) with a thin-plate regression spline using the *gam*-function in the *mgcv* package (Simpson, 2018; Wood, 2018) in R. Settings for *k* varied to find the best model fit, evaluated with *gam.check*, and we obtained interpolated values using the *predict.gam*-function.

A.7. Pollen stratigraphic plots

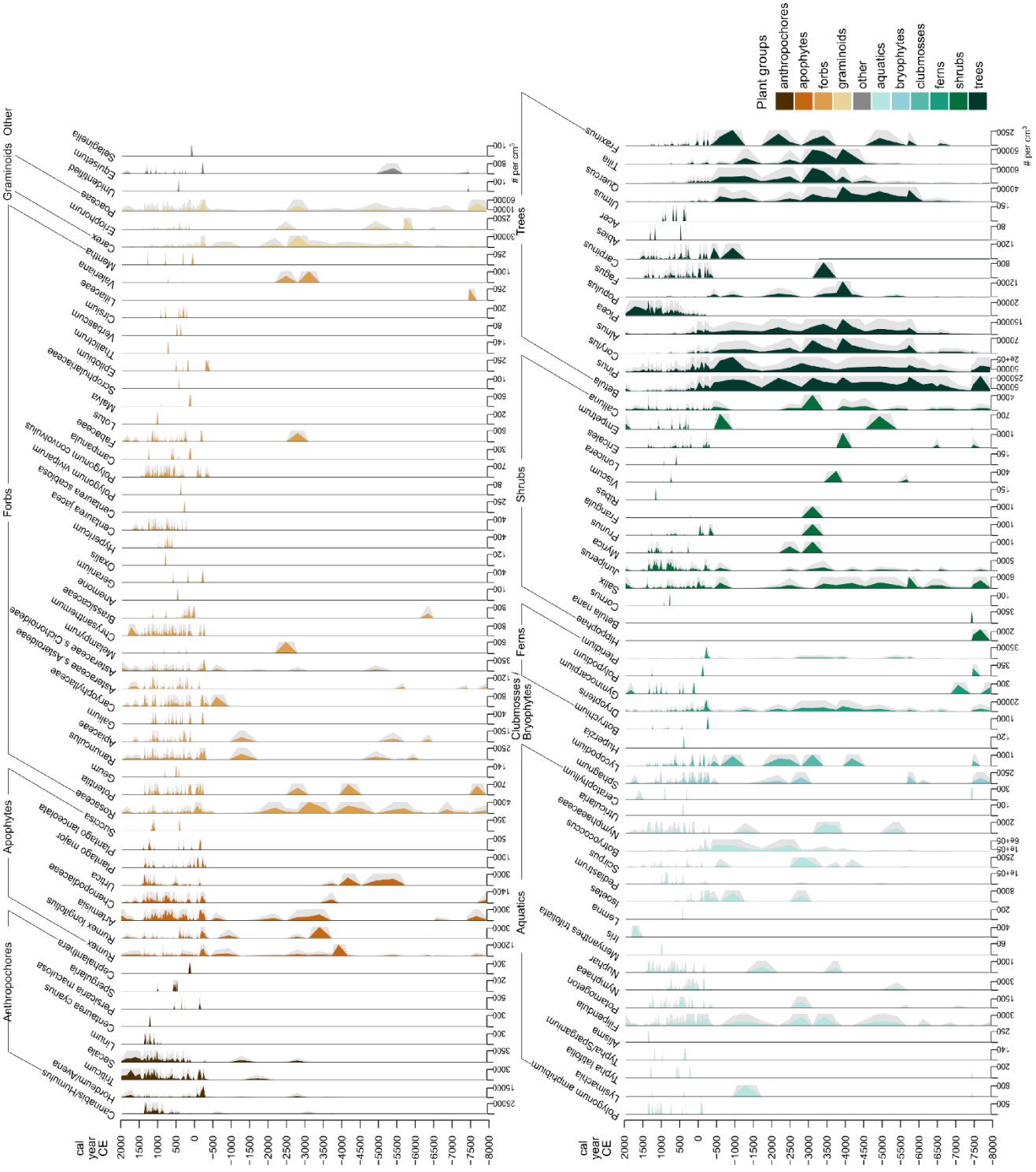


Fig. A.7. Pollen concentrations arranged and coloured by plant group. Grey areas indicate 3x exaggerated values.

A.8. SedaDNA stratigraphic plots

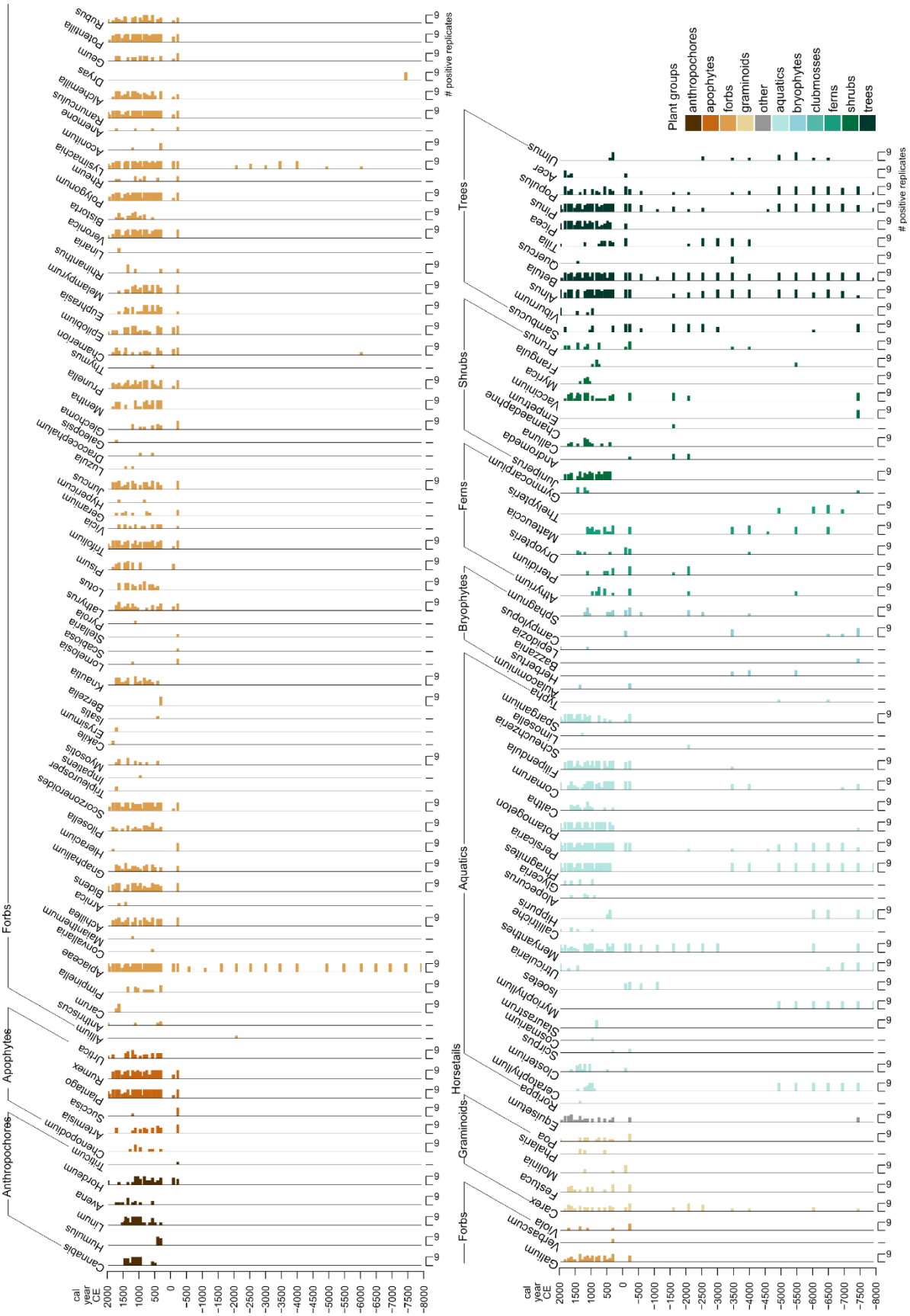


Fig. A.8. Number of plant *sed*aDNA positive PCR replicates per genus arranged and coloured by plant group.

A.9. CONISS analysis

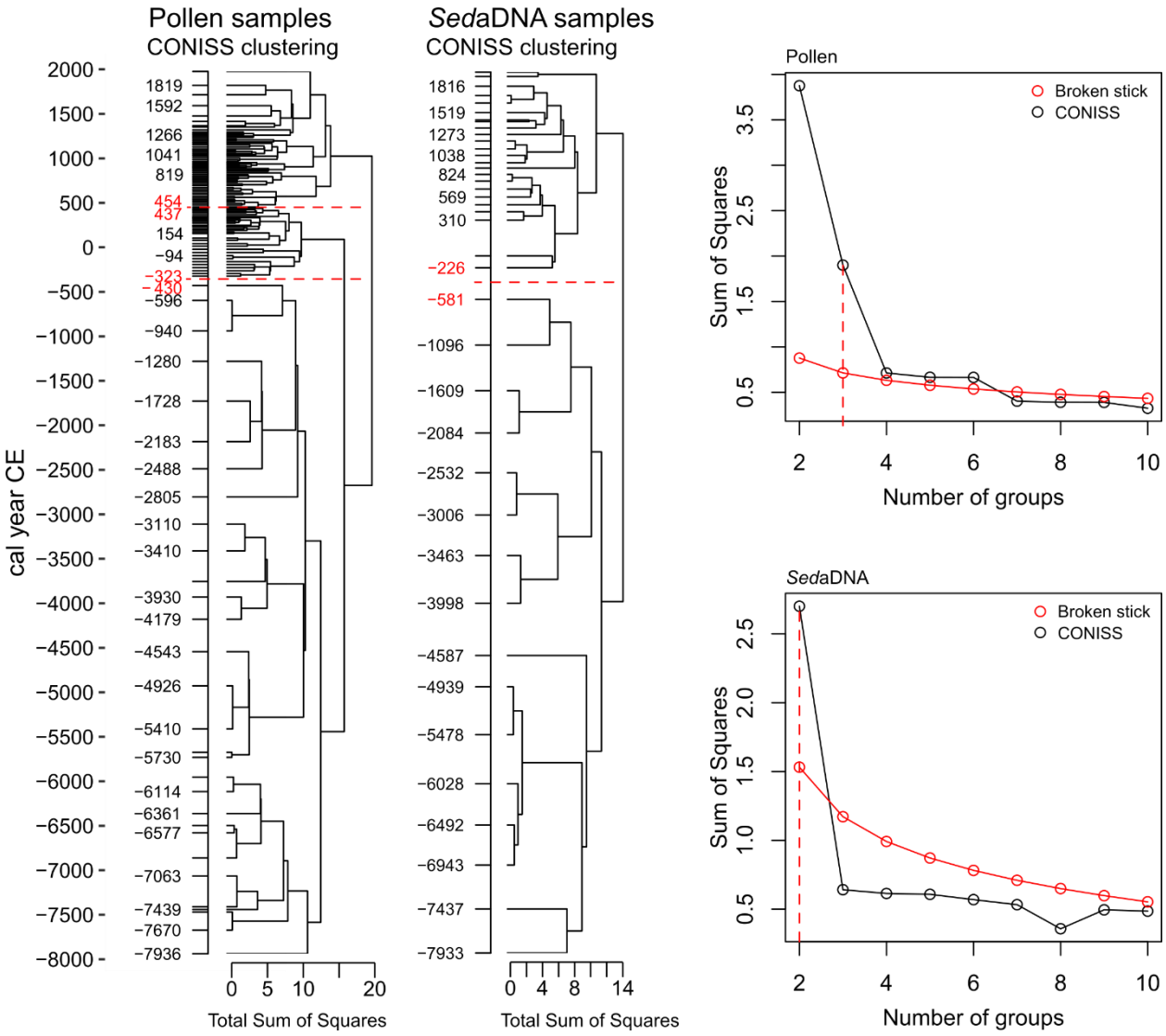


Fig. A.9. CONISS stratigraphically constrained cluster analysis based on the terrestrial vegetation data from pollen and *sed*aDNA analyses (left) and broken stick graphs showing the significant number of clusters found (right). Red dotted lines indicate significant clusters, and red-coloured cal year CE values indicate the age of the samples at the edges of the clusters. Pollen concentrations and *sed*aDNA positive replicate values were standardised to sample fractions using the "total" option in the *decostand*-function from the *vegan* package. We used *chclust*-function of the *rioja* package (Juggins and Juggins, 2020) with the CONISS algorithm posed by Grimm (1987) and Bray-Curtis distances. Based on the pollen data we found 6 significantly different groups, but as differences between the CONISS estimates and broken stick estimates were small for the additional 3 groups, we decided to focus only on the first 3 identified groups.

A.10. Non-Metric Multidimensional Scaling (NMDS)

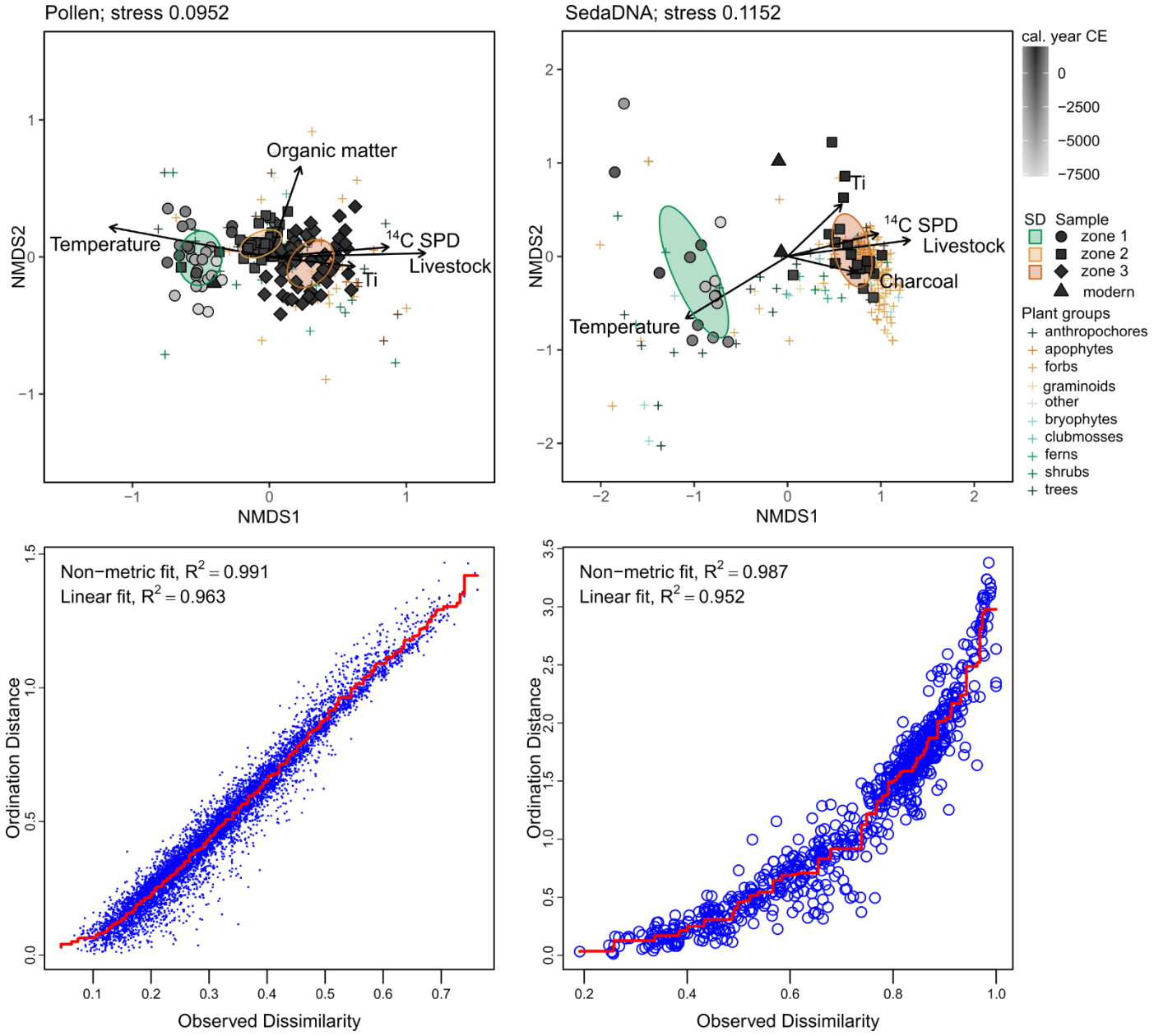


Fig. A.10. Non-metric multi-dimensional scaling ordinations of terrestrial plant pollen concentrations (left) and *sed*aDNA positive replicates (right) based on Bray-Curtis distances using the *metaMDS* function of the *vegan* package (Oksanen et al., 2020). Subdivision of samples into zones was done with CONISS analysis and we define modern samples as those from >1850 cal CE. Ellipses indicate the standard deviation of the mean for each zone. Plant taxa scores are indicated with '+'-signs and are coloured per plant group. Environmental terms were fitted to the ordination using the *envfit*-function with 999 permutations and only significant terms ($P < .05$) are shown here. Temperature corresponds to the temperature anomaly data published by Sejrup et al. (2016). Organic matter content corresponds to Inc/Coh ratio trends determined by XRF analysis. Ti was similarly determined by XRF analysis. Charcoal concentrations were determined through palynological analysis. ^{14}C SPD are radiocarbon summed probability distributions reflecting palaeodemographic trends. Livestock corresponds to the presence and absence of *sed*aDNA from pastoral animals.

A.11. Distance-based Redundancy Analysis (dbRDA)

We performed stepwise distance-based redundancy analysis (dbRDA) based on Bray-Curtis dissimilarities of Hellinger-transformed terrestrial pollen and plant *sed*aDNA data. The environmental terms (i.e. Ti, Inc/Coh ratio, charcoal, presence/absence of livestock, ^{14}C SPD, temperature anomaly) were standardized prior to analysis with the "standardize" method in the *decostand* function in *vegan* (Oksanen et al., 2020), removing unwanted effects of different measurement units. We performed automatic stepwise model building combining the *capscale* and *ordiR2step* functions to obtain the best fitting model. *OrdiR2step* builds the model so that it maximizes adjusted R^2 and we increased the number of permutations for this process to 999. As input, we used a minimal model with only one environmental term as a starting point (we tested several to identify one that is consistently present in the resulting models) and for scope we used a full model with all of the environmental terms, both built using the *capscale* function. We tested for statistical significance of the included environmental terms with an anova (999 permutations) and calculated the total proportion and the conditional proportion of plant community variance explained by each term.

In R code:

```
set.seed(42)
# preparing the minimal and full model
dbRDAmin = capscale(data ~ Temperature, env_std, dist="bray", na.action=na.exclude)
dbRDAfull = capscale(data ~ . , env_std, dist="bray", na.action=na.exclude)
dbRDA = ordiR2step(dbRDAmin, scope = formula(dbRDAfull), na.rm=TRUE, permutations = 999)

# is the model significant?
anova(dbRDA)
dbRDA_a = anova(dbRDA, by="axis", perm.max=999) # test axes for significance
dbRDA_t = anova(dbRDA, by="terms", permu=999) # test environmental terms for significance

# calculating the total proportion of variance explained per environmental term
# this includes the shared variation also explained by other terms
dbRDA_t = anova(dbRDA, by="terms", permu=999)
dbRDA_t$SumOfSqs / with(dbRDA, tot.chi)

# calculating the conditional proportion of variance explained per environmental term
# this excludes the shared variation also explained by other terms thus representing the individual fraction
dbRDA_m = anova(dbRDA, by="margin", permu=999)
dbRDA_m$SumOfSqs / with(dbRDA, tot.chi)
```

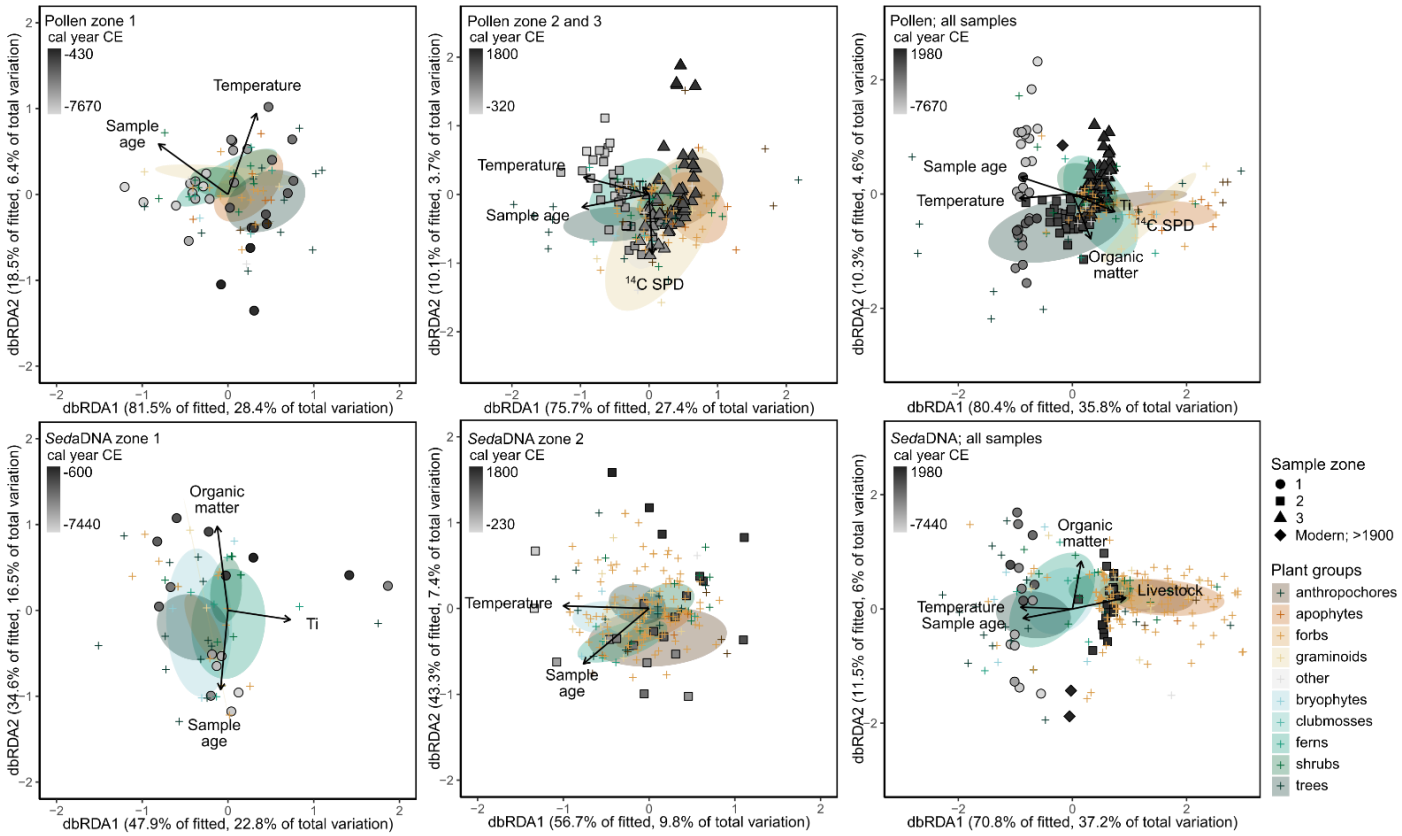


Fig. A.11.1. Pollen (top) and *sed*aDNA (bottom) dbRDA ordination sample scores (filled symbols) and plant taxa scores (plus symbols; coloured per plant group) of models including the sample age, illustrating close correlations between sample age and temperature especially when including all samples (right; see also Table A.11 and Appendix A.12). Ellipses indicate the standard error of the mean per plant group. Arrows indicate significant environmental terms. Environmental terms included in dbRDAs: temperature corresponds to the temperature anomaly data from Sejrup et al. (2016), organic matter corresponds to the Inc/Coh ratio determined by XRF analysis, Ti is measured by XRF analysis, livestock corresponds to presence/absence of livestock DNA, charcoal corresponds to the charcoal concentrations determined by palynological analysis, and ^{14}C SPD corresponds to radiocarbon summed probability distributions, a proxy for palaeodemography.

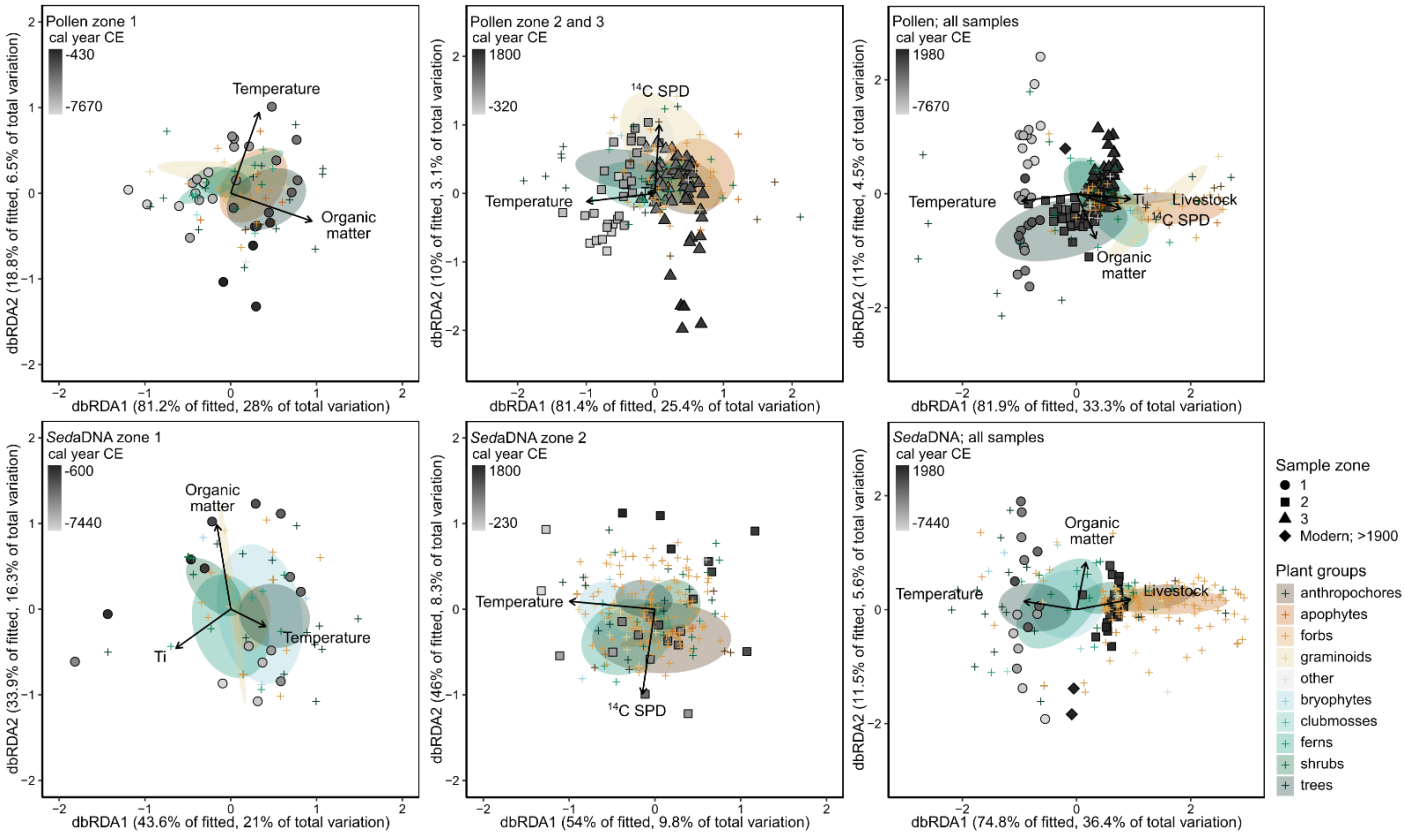


Fig. A.11.2. Pollen (top) and *sed*aDNA (bottom) dbRDA ordination sample scores (filled symbols) and plant taxa scores (plus symbols; coloured per plant group) of models excluding the sample age. Ellipses indicate the standard error of the mean per plant group. Arrows indicate significant environmental terms. Environmental terms included in dbRDAs: temperature corresponds to the temperature anomaly data from Sejrup et al. (2016), organic matter corresponds to the Inc/Coh ratio determined by XRF analysis, Ti is measured by XRF analysis, livestock corresponds to presence/absence of livestock DNA, charcoal corresponds to the charcoal concentrations determined by palynological analysis, and ^{14}C SPD corresponds to radiocarbon summed probability distributions, a proxy for palaeodemography.

Table. A.11. Proportions of variance explained per environmental term of the total variance in plant assemblages for 12 automatic stepwise dbRDAs based on: samples from all zones combined, samples from after 300 cal BCE, and samples from before 300 cal BCE for *sed*aDNA and pollen datasets. DbRDAs were performed both including and excluding (indicated with a cross) sample age in cal BP. Other environmental terms included in dbRDAs: temperature corresponds to the temperature anomaly data from Sejrup et al. (2016), organic matter corresponds to the Inc/Coh ratio determined by XRF analysis, Ti is measured by XRF analysis, livestock corresponds to presence/absence of livestock DNA, charcoal corresponds to the charcoal concentrations determined by palynological analysis, and ^{14}C SPD corresponds to radiocarbon summed probability distributions, a proxy for palaeodemography. Blank cells indicate that the environmental term was not included in the model as inclusion did not result in a better model (increased adjusted R^2). Grey shading shows proportions of variance categories and significant values are indicated in bold; asterisks indicate significance with '***' $p \leq 0.001$, '**' $p \leq 0.01$ and '*' $p \leq 0.05$.

Samples used in dbRDA			Sample age (cal. BP)	Temperature (Sejrup et al., 2016)	Organic matter (Inc/Coh ratio)	Ti	Livestock	Charcoal	^{14}C SPD
All zones	*Sed*aDNA		0.040*	0.034*	0.055**	0.026	0.085***		
		P-value	*0.028*	*0.048*	*0.007*	*0.090*	*0.001*		
			X	0.060**	0.055*	0.029	0.070**		
		P-value		*0.006*	*0.014*	*0.092*	*0.005*		
	Pollen		0.044***	0.014*	0.047***	0.021**	0.010		0.016*
		P-value	*0.001*	*0.042*	*0.001*	*0.010*	*0.100*		*0.028*
			X	0.050***	0.040***	0.018*	0.024**		0.015*
		P-value		*0.001*	*0.001*	*0.025*	*0.007*		*0.044*
> 300 cal BCE	*Sed*aDNA		0.074	0.084*					
		P-value	*0.078*	*0.033*					
			X	0.098**					0.083*
		P-value		*0.008*					*0.030*
	Pollen		0.043***	0.020*	0.018*	0.029**			0.021*
		P-value	*0.001*	*0.026*	*0.037*	*0.002*			*0.021*
			X	0.289***		0.032**			0.036***
		P-value		*0.001*		*0.004*			*0.001*
< 300 cal BCE	*Sed*aDNA		0.153**		0.150**	0.165***			
		P-value	*0.003*		*0.002*	*0.001*			
			X	0.157***	0.163***	0.197***			
		P-value		*0.001*	*0.001*	*0.001*			
	Pollen		0.276***	0.149***					
		P-value	*0.001*	*0.001*					
			X	0.095**	0.278***				
		P-value		*0.007*	*0.001*				

Conditional proportion of variance

0-0.05 | 0.05-0.10 | 0.10-0.15 | 0.15-0.20 | >0.2

A.12. Variation partitioning

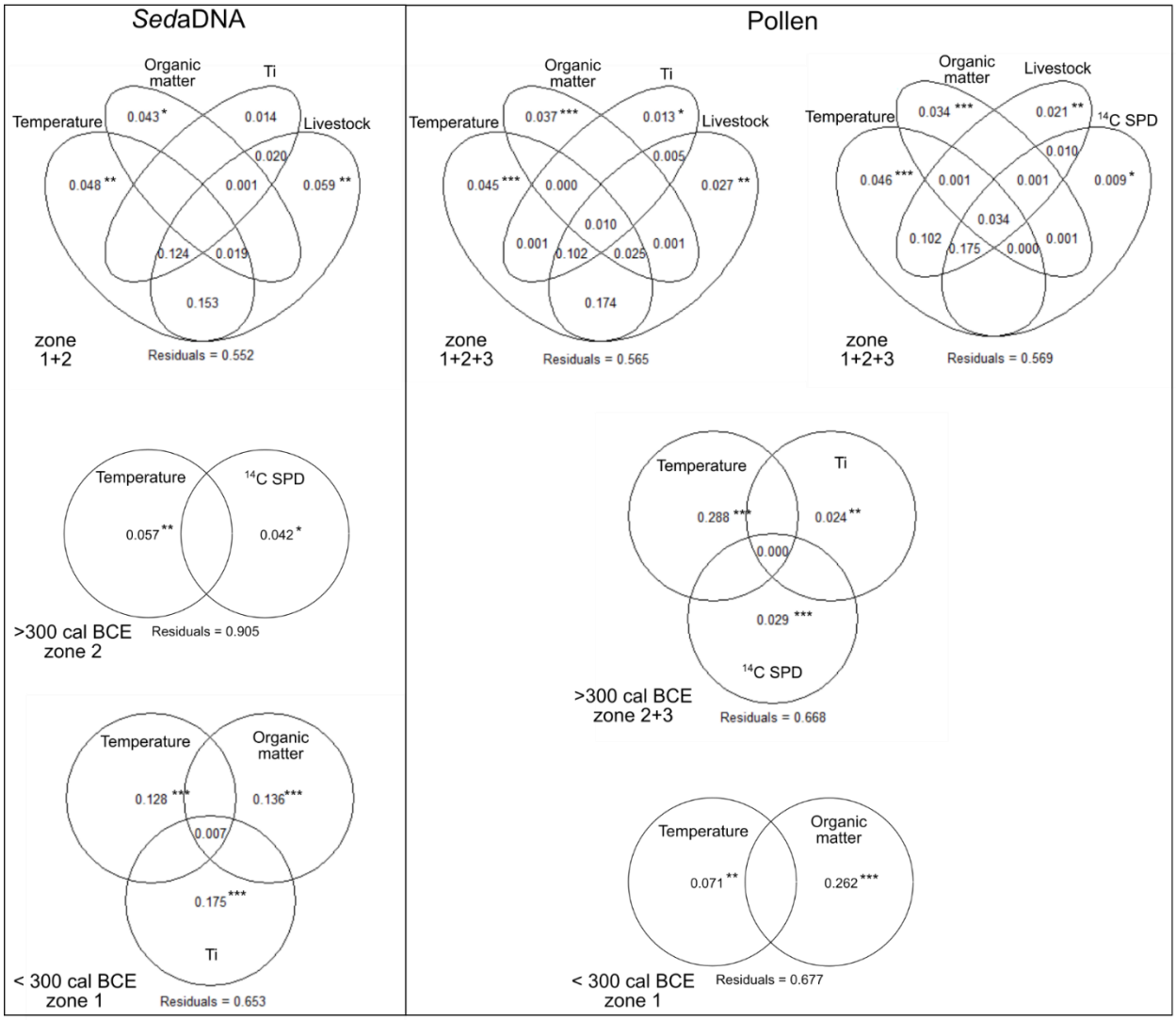


Fig. A.12. Partitioning of the proportions of variance explained per environmental term of the total variance in plant assemblages based on: samples from all zones combined, samples from after 300 cal BCE, and samples from before 300 cal BCE for *sed*aDNA (left) and pollen (right) datasets. Only environmental terms included in the automated stepwise dbRDAs were used in the variation partitioning (see Appendix A.11 for details). The number of included environmental terms for variation partitioning is limited to 4, therefore two separate variation partitioning analyses were performed for pollen samples from zone 1, 2 and 3 combined. Environmental terms included: temperature corresponds to the temperature anomaly data from Sejrup et al. (2016), organic matter corresponds to the Inc/Coh ratio determined by XRF analysis, Ti is measured by XRF analysis, livestock corresponds to presence/absence of livestock DNA, and ^{14}C SPD corresponds to radiocarbon summed probability distributions, a proxy for palaeodemography. Values below 0 are not shown. Asterisks indicate significance with '***' $p \leq 0.001$, '**' $p \leq 0.01$ and '*' $p \leq 0.05$ based on dbRDA analysis (see Appendix A.11).

A.13. Correlations between environmental terms

Table A.13.1. Spearman rank correlations between environmental terms at the time resolution of the pollen data (top-right; N=120) and the *sed*aDNA data (bottom-left; N=39) for the entire research period (ca. 7700 cal BCE-1975 cal CE). P-values were adjusted with the Bonferroni method and significant correlations are indicated in bold with grey shading.

All samples		Sample age (cal. BP)	Organic matter (Inc/Coh ratio)	Ti	Temperature (Sejrup et al., 2016)	Charcoal	^{14}C SPD	Livestock	Diversity (HillN2)	NMDS1	NMDS2
Sample age (cal BP)			-0.1	-0.39	0.93	-0.51	-0.5	-0.69	-0.63	-0.9	0.12
	P.adj		*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*
Organic matter (Inc/Coh ratio)		-0.07		-0.18	-0.09	0.24	0.08	0.03	0.16	0.04	0.31
	P.adj	*1*		*1*	*1*	*0.42*	*1*	*1*	*1*	*1*	*0.04*
Ti		-0.57	-0.43		-0.32	0.54	0.47	0.54	0.48	0.36	-0.02
	P.adj	*0.01*	*0.31*		*0.01*	*0*	*0*	*0*	*0*	*0*	*1*
Temperature (Sejrup et al., 2016)		0.81	-0.14	-0.38		-0.56	-0.6	-0.74	-0.59	-0.92	0.12
	P.adj	*0*	*1*	*0.84*		*0*	*0*	*0*	*0*	*0*	*1*
Charcoal		-0.57	0.21	0.49	-0.68		0.57	0.65	0.55	0.57	0.12
	P.adj	*0.01*	*1*	*0.07*	*0*		*0*	*0*	*0*	*0*	*1*
^{14}C SPD		-0.29	0.31	0.24	-0.59	0.75		0.71	0.58	0.55	0
	P.adj	*1*	*1*	*1*	*0*	*0*		*0*	*0*	*0*	*1*
Livestock		-0.59	0.03	0.59	-0.77	0.8	0.72		0.63	0.7	0.03
	P.adj	*0*	*1*	*0*	*0*	*0*	*0*		*0*	*0*	*1*
Diversity (HillN2)		-0.61	0.04	0.56	-0.73	0.8	0.59	0.84		0.61	-0.3
	P.adj	*0*	*1*	*0.01*	*0*	*0*	*0*	*0*		*0*	*0.07*
NMDS1		-0.62	-0.05	0.6	-0.76	0.8	0.58	0.84	0.95		-0.03
	P.adj	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*		*1*
NMDS2		-0.47	-0.08	0.27	-0.47	0.06	0.12	0.16	0.02	0.04	
	P.adj	*0.1*	*1*	*1*	*0.13*	*1*	*1*	*1*	*1*	*1*	

Table A.13.2. Spearman rank correlations between environmental terms at the time resolution of the pollen data (top-right; N=90) and the *sed*aDNA data (bottom-left; N=22) for zone 2 & 3 (ca. 300 cal BCE-1800 cal CE). P-values were adjusted with the Bonferroni method and significant correlations are indicated in bold with grey shading.

Zone 2 & 3 (ca. 300 cal BCE – 450 cal CE)		Sample age (cal. BP)	Organic matter (Inc/Coh ratio)	Ti	Temperature (Sejrup et al., 2016)	Charcoal	^{14}C SPD	Livestock	Diversity (HillN2)	NMDS1	NMDS2
Sample age (cal BP)			0.05	0.25	0.93	0.03	0	-0.18	-0.37	-0.9	0.33
	P.adj		*1*	*0.89*	*0*	*1*	*1*	*1*	*0.02*	*0*	*0.06*
Organic matter (Inc/Coh ratio)		0.09		-0.31	-0.03	0.27	-0.14	-0.23	-0.03	-0.03	0.01
	P.adj	*1*		*0.13*	*1*	*0.48*	*1*	*1*	*1*	*1*	*1*
Ti		0.16	-0.53		0.34	0.21	0.07	-0.09	0.14	-0.24	0.01
	P.adj	*1*	*0.47*		*0.04*	*1*	*1*	*1*	*1*	*1*	*1*
Temperature (Sejrup et al., 2016)		0.69	-0.33	0.6		-0.04	-0.15	-0.18	-0.29	-0.87	0.26
	P.adj	*0.02*	*1*	*0.14*		*1*	*1*	*1*	*0.25*	*0*	*0.57*
Charcoal		0.42	0.33	-0.03	0.18		0.08	-0.04	0.13	0.07	0.05
	P.adj	*1*	*1*	*1*	*1*		*1*	*1*	*1*	*1*	*1*
^{14}C SPD		0.61	0.16	-0.02	0.2	0.43		0.19	0.24	0.08	-0.15
	P.adj	*0.12*	*1*	*1*	*1*	*1*		*1*	*1*	*1*	*1*
Livestock		0.26	-0.12	0.36	0.26	0.02	0.22		0.15	0.15	-0.1
	P.adj	*1*	*1*	*1*	*1*	*1*	*1*		*1*	*1*	*1*
Diversity (HillN2)		0.08	0.11	0.28	0.11	0.31	0.15	0.15		0.31	-0.77
	P.adj	*1*	*1*	*1*	*1*	*1*	*1*	*1*		*0.13*	*0*
NMDS1		-0.03	0.23	0.3	-0.09	0.31	0.17	0.15	0.85		-0.16
	P.adj	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*0*		*1*
NMDS2		-0.51	-0.07	-0.33	-0.59	-0.34	-0.23	-0.29	-0.75	-0.5	
	P.adj	*0.74*	*1*	*1*	*0.17*	*1*	*1*	*1*	*0*	*0.86*	

Table A.13.3. Spearman rank correlations between environmental terms at the time resolution of the pollen data (top-right; N=29) and the *sed*aDNA data (bottom-left; N=15) for zone 1 (ca. 8000-300 cal BCE). P-values were adjusted with the Bonferroni method and significant correlations are indicated in bold with grey shading.

Zone 1 (ca. 8000 - 300 cal BCE)		Sample age (cal. BP)	Organic matter (Inc/Coh	Ti	Temperature (Sejrup et al.,	Charcoal	^{14}C SPD	Livestock	Diversity (HillN2)	NMDS1	NMDS2
	Sample age (cal BP)		-0.95	0.39	0.05	-0.35	-0.7	NA	-0.52	0.25	-0.7
	P.adj		*0*	*1*	*1*	*1*	*0*		*0.13*	*1*	*0*
	Organic matter (Inc/Coh ratio)	-0.95		-0.46	0.01	0.3	0.6	NA	0.55	-0.27	0.71
	P.adj	*0*		*0.44*	*1*	*1*	*0.02*		*0.07*	*1*	*0*
	Ti	0.62	-0.69		0.17	-0.21	-0.23	NA	-0.21	-0.1	-0.39
	P.adj	*0.46*	*0.17*		*1*	*1*	*1*		*1*	*1*	*1*
	Temperature (Sejrup et al., 2016)	0.37	-0.3	0.43		0.35	-0.14	NA	0.53	-0.25	0.38
	P.adj	*1*	*1*	*1*		*1*	*1*		*0.12*	*1*	*1*
	Charcoal	-0.36	0.47	-0.17	0.4		0.1	NA	0.34	-0.11	0.44
	P.adj	*1*	*1*	*1*	*1*		*1*		*1*	*1*	*0.6*
	^{14}C SPD	-0.69	0.64	-0.35	-0.27	0.06		NA	0.31	0.14	0.44
	P.adj	*0.16*	*0.37*	*1*	*1*	*1*			*1*	*1*	*0.58*
	Livestock	NA	NA	NA	NA	NA	NA		NA	NA	NA
	P.adj										
	Diversity (HillN2)	0.19	-0.11	-0.12	0.03	0.01	-0.52	NA		-0.19	0.92
	P.adj	*1*	*1*	*1*	*1*	*1*	*1*			*1*	*0*
	NMDS1	0.7	-0.66	0.13	0.25	-0.09	-0.77	NA	0.65		-0.15
	P.adj	*0.13*	*0.25*	*1*	*1*	*1*	*0.03*		*0.3*		*1*
	NMDS2	-0.07	-0.02	0.16	-0.47	-0.57	0.26	NA	-0.26	-0.46	
	P.adj	*1*	*1*	*1*	*1*	*0.9*	*1*	*NA*	*1*	*1*	

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian analysis, 6(3), 457-474.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular ecology resources, 16(1), 176-182. https://doi.org/10.1111/1755-0998.12428

Grimm, E. C. (1987). CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. Computers & geosciences, 13(1), 13-35.

Juggins, S. (2020). rioja: Analysis of Quaternary Science Data. R package version 0.9-26, https://cran.r-project.org/package=rioja.

Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P. S., ... & Davis, B. (2020). Holocene global mean surface temperature, a multi-method reconstruction approach. Scientific data, 7(1), 1-13. https://doi.org/10.1038/s41597-020-0530-7

Kotov, S., & Paelike, H. (2018, December). QAnalySeries-a cross-platform time series tuning and analysis tool. In AGU Fall Meeting Abstracts (Vol. 2018, pp. PP53D-1230).

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2020). vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

R Core Team (2020). R: A language and environment for statistical computing (manual). Vienna, Austria.

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., ... & Talamo, S. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon, 62(4), 725-757.

Sejrup, H. P., Seppä, H., McKay, N. P., Kaufman, D. S., Geirsdóttir, Á., de Vernal, A., ... & Andrews, J. T. (2016). North Atlantic-Fennoscandian Holocene climate trends and mechanisms. Quaternary Science Reviews, 147, 365-378.

Simpson, G. L. (2018). Modelling palaeoecological time series using generalised additive models. Frontiers in Ecology and Evolution, 6, 149.

Wood, S. (2018). Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. R package version, 1-8.

Chapter 5

SedaDNA metabarcoding as a tool for assessing prehistoric plant use at Upper Palaeolithic cave site Aghitu-3, Armenia

*Sed*aDNA metabarcoding as a tool for assessing prehistoric plant use at Upper Palaeolithic cave site Aghitu-3, Armenia

Anneke T.M. ter Schure[1], Angela A. Bruch[2], Andrew W. Kandel[3], Boris Gasparyan[4], Rainer W. Bussmann[5], Anne K. Brysting[1], Hugo J. de Boer[6], Sanne Boessenkool[1]

Manuscript

[1] Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway.
[2] The Role of Culture in Early Expansions of Humans, Heidelberg Academy of Sciences and Humanities, Senckenberg Research Institute, Frankfurt, Germany.
[3] The Role of Culture in Early Expansions of Humans, Heidelberg Academy of Sciences and Humanities, University of Tübingen, Tübingen, Germany.
[4] Institute of Archaeology and Ethnography, National Academy of Sciences of the Republic of Armenia, Yerevan, Armenia.
[5] Department of Ethnobotany, Institute of Botany, Ilia State University, Tbilisi, Georgia.
[6] Natural History Museum, University of Oslo, Oslo, Norway.

Abstract

Caves provide a unique opportunity to study past human activity, as they mitigate the full effect of the elements on the organic materials (e.g. bones, faeces and plant remains) concentrated by humans and animals. By analysing DNA preserved in cave sediments, we can identify organisms in the absence of any visible remains. This method opens up new ways to study past plant use and complements more traditional proxies such as pollen, which may come from plants growing far beyond the surroundings of a cave. Aghitu-3 Cave contains a 15,000-year long record (from ~39,000-24,000 cal BP) of Upper Palaeolithic human settlement and environmental variability in the Armenian Highlands. Finds from the cave include stone artefacts, faunal remains, bone tools, shell beads, charcoal and pollen, among others. We applied *sed*aDNA metabarcoding to the Aghitu-3 sedimentary sequence to obtain a temporal reconstruction of plant assemblages concentrated by humans and animals during the Upper Palaeolithic. Our results reveal the presence of many useful taxa, including those that are known to be used for food, medicine, dyes, repellents, or fibers. Our *sed*aDNA results support and complement palynological evidence while increasing taxonomic resolution. Assessment of plant use based on plant presence in cave sediments is inherently indirect, and we found the majority of Armenian plant taxa to be of potential use. Despite these limitations, we found *sed*aDNA to more accurately reflect periods of human occupation than pollen, and presence of useful taxa in layers associated with human occupation suggesting human exploitation of plant resources. With this study we show that ancient DNA metabarcoding of cave sediments provides novel insights into the plant use of prehistoric humans.

1. Introduction

Plants played an integral role in prehistoric human life, providing food, medicine, raw materials and fuel. However, current knowledge about early human plant use is limited and primarily based on scarce visible plant remains. Plant remains are perishable, but if preserved and identifiable they can provide a wealth of information about the palaeoenvironment and the human past. Particularly analyses of pollen and macrofossils have been widely used for vegetation reconstructions, as well as inferences on past landscapes, changes in climate, and human occupation (Edwards et al., 2015; Gaillard et al., 2010; Trondman et al., 2015). Evidence for the Palaeolithic human diet from plant fragments, phytoliths, microfossils, and biomarkers from food preparation tools and dental calculus has recently shifted the assumption of a largely animal-based diet to one including a range of different plants (Power et al., 2018; Revedin et al., 2010; Wadley et al., 2020). DNA, chemical biomarkers, and starch grains extracted from dental calculus at several Neanderthal sites suggest ingestion of plant foods (Hardy et al., 2012; Power et al., 2018; Salazar-García et al., 2013; Weyrich et al., 2017). Starch grains recovered from ~30,000-year-old grinding tools at several locations in Europe suggests widespread vegetal food processing by hunter-gatherers, and possibly the production of flour (Revedin et al., 2010). Plants also provided key raw materials and fuel for fires (Albert et al., 2000; Allué et al., 2012; Goldberg et al., 2012), and were possibly used for medicine (Hardy, 2019; for reviews on Palaeolitic human plant use see Hardy, 2018; Shipley & Kindscher, 2016), and even for early production of flax fibers/textiles (Kvavadze et al., 2009) at Dzudzuana in Georgia. Nevertheless, survival of identifiable plant remains is rare at Palaeolithic sites and our current knowledge about human plant-use during this time period is defined and limited by the degradation of plant remains, the findability of the remains, and the availability of methods and technologies for plant detection and identification.

Developments in sedimentary ancient DNA (*sed*aDNA) analysis methods and protocols have allowed the recovery of DNA from a variety of sedimentary contexts, including cave sediments (Haile et al., 2007; Hofreiter et al., 2003; Willerslev, 2003). With caves mitigating the full effects of the elements on organic materials such as bones, faeces and plant remains concentrated by humans and animals, they provide a unique opportunity to study not only past human presence (Slon et al., 2017; Vernot et al., 2021; Zavala et al., 2021; Zhang et al., 2020), but also past environments and human activities. Here, we assess the potential of *sed*aDNA metabarcoding for reconstructing prehistoric plant assemblages from cave sediments by applying these methods to ca. 39-24,000-year-old human occupation layers from Aghitu-3 Cave in the Armenian Highlands.

Aghitu-3 represents one of a handful of stratified Upper Palaeolithic sites in the Republic of Armenia. The Armenian Highlands together with the southern Caucasus was a major route for early humans as they migrated out of Africa, with

the earliest evidence of this migration present in the Georgian site of Dmanisi (Ferring et al., 2011; Gabunia & Vekua, 1995). This is a topographically diverse region, and the Armenian Highlands today support a shrubby steppe vegetation that, based on palynological evidence, appeared to have persisted throughout the Pleistocene, with phases of expanded forest-steppe during more humid periods in response to climate oscillations (Joannin et al., 2010; Ollivier et al., 2010).

Through analysis of DNA preserved in cave sediments we can identify organisms in the absence of any visible remains, as illustrated by studies that successfully retrieved *sed*aDNA in caves originating from animals, including birds and mammals, as well as hominins (Haile et al., 2007; Hofreiter et al., 2003; Slon et al., 2017; Vernot et al., 2021; Willerslev, 2003; Zhang et al., 2020). Recent *sed*aDNA analyses of cave sites in Europe, Siberia, and the Tibetan Plateau have reconstructed Neanderthal and Denisovan population history (Slon et al., 2017; Vernot et al., 2021; Zavala et al., 2021; Zhang et al., 2020), focusing on the traces of mammal DNA, particularly from hominin origin. Previous *sed*aDNA studies on cave sites in New Zealand, Australia, and North America demonstrate the possibility to recover and identify plant DNA and reconstruct plant assemblages from >20,000 years ago (Haile et al., 2007; Haouchar et al., 2014; Seersholm et al., 2020; Willerslev, 2003). Moreover, a recent study at Solkota Cave in Georgia successfully retrieved up to 84,000 year old mammal and plant DNA from stalagmites (Stahlschmidt et al., 2019), attesting to the unique conditions of preservation in caves.

To assess the potential of *sed*aDNA metabarcoding for reconstructing prehistoric plant assemblages in caves associated with human occupation, we combine plant *sed*aDNA metabarcoding with previous pollen evidence from Aghitu-3 Cave in southern Armenia. Aghitu-3 Cave contains a 15,000 year long record (from ~39-24,000 cal BP) of Upper Palaeolithic human settlement and environmental variability in the Armenian Highlands. Previous remains at Aghitu-3 included pollen and non-pollen palynomorphs (NPPs), such as spores, algal, and fungal remains, wood and other undifferentiated plant remains present throughout the stratigraphy (Kandel et al., 2017) and implying good preservation of organic remains. Human occupation layers at Aghitu-3 yielded many finds, including stone artefacts, combustion features, and faunal remains, separated by archaeologically sterile layers with few finds (Bertacchi et al., 2021; Kandel et al., 2017). The stratification of rich layers interspersed with sterile layers creates an opportunity to test whether identified plant remains may have been concentrated by past humans and animals. Moreover, through the analysis of modern uses of the recovered plant taxa we can provide insight into possible prehistoric plant use at Aghitu-3.

2. Materials and Methods

2.1. Study site

Aghitu-3 Cave (39.5138° N, 46.0822° E, 1601 m a.s.l.) is situated 115 m above the Vorotan River in the Syunik Province of southern Armenia (Fig. 1A). With dimensions of 11 m depth, 18 m width, and 6 m height, it is the largest of several caves along the base of a basalt massif, whose source is an eruption from nearby Mount Bugdatapa between 126,000 and 111,000 years ago (Ollivier et al., 2010). The Vorotan River has cut through the basalt, forming a valley that constitutes a corridor of movement for people as well as game (Kandel et al., 2014). Archaeological excavations at Aghitu-3 started in 2009 and yielded many stone artefacts, faunal remains, bone tools, shell beads, charcoal, and pollen (Bertacchi et al., 2021; Kandel et al., 2017, 2014). Twelve geological horizons were originally identified, correlating with seven archaeological horizons (AHs), including three horizons containing Upper Palaeolithic (UP) finds, AH III, VI and VII. Study of the lithic and faunal assemblages has suggested that the cave was used as a shelter for short term stays for hunting (Taller et al., 2018). As the upper horizons are mixed with Holocene debris, we focus on the intact Pleistocene stratigraphy of AHs III-VII, dated to ~39,000-24,000 cal BP (Kandel et al., 2017).

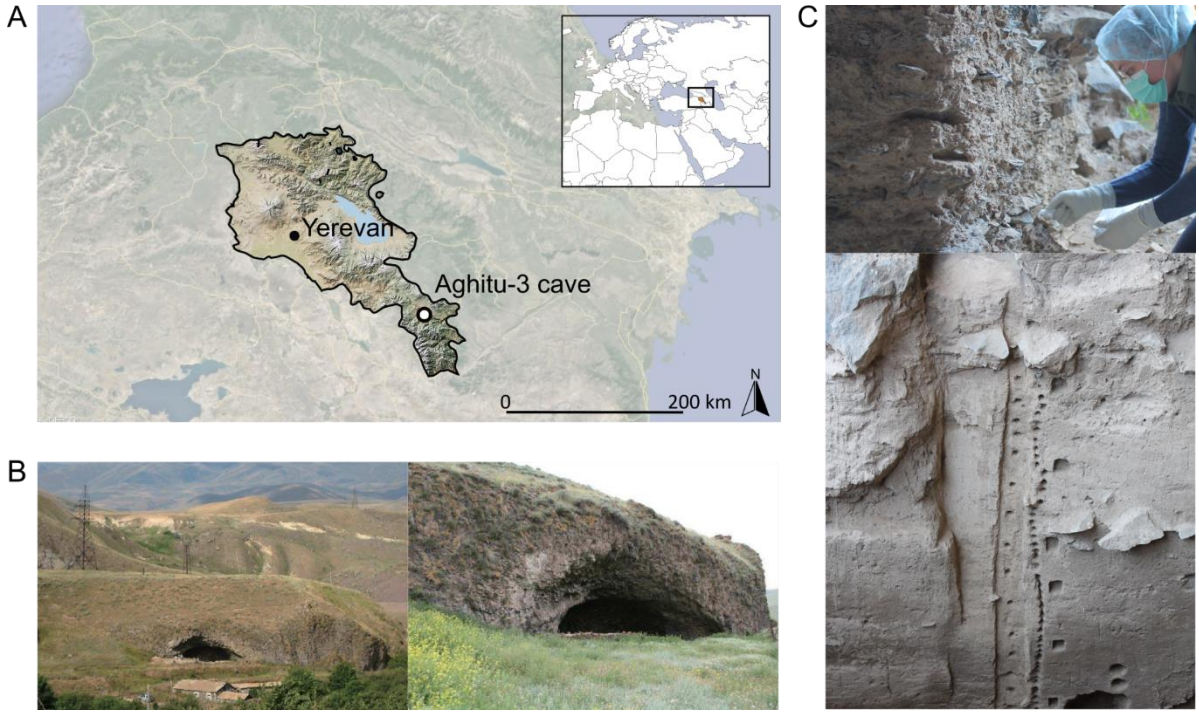


Figure 1. Aghitu-3 Cave. A) The cave's location in the Republic of Armenia. B) View of the Aghitu basalt massif looking west into the valley of the Vorotan River (left) and a closer view looking west into the cave (right). C) *Sed*aDNA sampling in between the basalt plates of AH III (top) and a section of the sediment profile in AH VI in the excavation pit (bottom). Photos by A. Kandel.

2.2. Sediment sampling for *sed*aDNA

The Aghitu-3 excavation profile (>5.5 m) consisted mainly of dry silt interspersed with basalt debris from the cave ceiling. *Sed*aDNA samples were collected in September 2019 over the full depth of the profile. AH III contains high amounts of platy basalt (Fig. 1C), so samples were taken from accessible sediments in between. In the lower horizons, we did not sample close to basalt boulders and avoided including small rocks.

The cave sedimentary sequence was sampled while wearing personal protective equipment (hairnets, facemasks, double gloves) and using sterile instruments. Sampling tubes and scalpel holders were UV irradiated in preparation of the sampling, and scalpel holders were cleaned with bleach in between each sample to avoid cross-contamination of DNA. First, 1-2 cm of surface material was removed. Then two consecutive cuts were made with sterile scalpels prior to sample collection in a sample tube, following the procedure described by Epp et al. (2019). Samples were kept cool during transport to the ancient DNA facilities at the University of Oslo and kept there at -20°C until DNA extraction.

2.3. DNA analyses

2.3.1. DNA extraction and amplification

DNA was extracted from 25 samples and 3 negative controls at the dedicated ancient DNA laboratory at the Institute for Biosciences, University of Oslo. We extracted DNA from 90-190 mg of sediment using the sediment extraction protocol as described by Rohland et al. (2018) with silica spin columns and binding buffer D. We increased the elution buffer volume to 100 µL.

Plant chloroplast DNA was amplified using the *trnL-gh* primers that target the P6 loop from the trnL-intron (Taberlet et al., 2007). These primers are specifically designed for amplification of degraded plant DNA and are widely used in plant *sed*aDNA studies. Both primers were tagged with a unique 8 or 9 bp barcode at the 5' end to allow for multiplexing as described by Voldstad et al. (2020). We conducted six PCR replicates per sample and extraction negative control, and included 30 PCR negative controls. DNA amplifications were carried out in a final volume of 25 µL containing 5 µL of DNA extract and 0.2 µM of each primer. The amplification mixture further contained 2U of AmpliTaq Gold DNA Polymerase with Buffer II (Applied Biosystems), 1 x Buffer II, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 0.32 µg/µL of bovine serum albumin (BSA, Roche Diagnostic). The mixture was denatured at 95 °C for 10 min, followed by 45 cycles of 30 s at 95 °C, and 30 s at 50 °C, 1 min at 72 °C and a 10 min final elongation at 72 °C.

PCR products were visualised with agarose gel electrophoresis. Equal volumes of PCR products with similar band intensity were mixed and purified using the MinElute Purification kit (Qiagen GmbH, Hilden, Germany), including PCR products that did not show a band. Concentrations of the pools with purified

PCR products were measured on a Qubit 2.0 with the Qubit dsDNA HS kit (ThermoFisher), and pools were subsequently combined in an equimolar superpool. Libraries were built from the resulting superpool using the KAPA HyperPrep DNA kit (Roche) and sequenced on the Illumina HiSeq 4000 at the Norwegian Sequencing Centre.

2.3.2. DNA sequence analyses and filtering

Initial sequence data processing was done using the OBITools package (http://metabarcoding.org/obitools/doc/index.html; Boyer et al., 2016). Forward and corresponding reverse reads were assembled using *illuminapairedend*, followed by sample assignment with *ngsfilter*. We removed reads with a quality score <40, <100% tag match, >3 mismatches with the primers, shorter lengths than expected (<8 bp), singletons and those containing ambiguous nucleotides. Amplification and sequencing errors were identified using *obiclean*, with a threshold ratio of 5% for reclassification of sequences identified as 'internal' to their corresponding 'head' sequence. Finally, sequences were compared to a taxonomic reference library using *ecotag*. The reference library was prepared using the global EMBL sequence database (release 142) and the NCBI Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy) by performing an in-silico PCR with the *ecoPCR* software (Ficetola et al., 2010).

To minimise the chance of misidentification, we filtered the identified sequences in R (version 3.5.2) to further remove: (1) sequences occurring in negative controls, (2) sequences with a percentage identity <97.5%, (3) read counts <10 reads per sequence in a PCR replicate, and (4) PCR replicates with <10 reads. As the used reference database contains limited data from the Caucasus region, taxonomic annotations were checked to assess the likeliness of occurrence in the region at the time. Additional filtering steps were performed removing sequences occurring in a single PCR replicate over the entire dataset, sequences with a percentage identity <98%, without family-annotation or that were deemed unlikely. Samples consisting of <2 positive replicates after filtering were considered empty and their read counts set to 0. Remaining unique DNA sequences were designated as molecular operational taxonomic units (MOTUs) and different MOTUs of the same species or genus were later summed together for downstream analyses. An overview of the filtering steps and their effect on the size of the dataset, as well as a list of manually removed taxa can be found in supplemental materials (Appendix Table A.1).

The filtered sequence data was corrected for the amount of input material (Appendix Table A.2) and subsequently log-transformed to correct for the exponential increase in read counts during PCR amplification. PCR replicates were merged, while calculating the number of positive replicates as well as the mean log-transformed read counts for each sample.

2.4. Pollen data

Sediment samples for palynological analysis (including pollen) were obtained along the sediment profile walls during excavations in 2011 and 2013 (Kandel et al., 2017). Per sample, 50 g of sediment was processed using 30% HCl and 30% HF to remove carbonates and silica. Tablets of spores (*Lycopodium*) were added to each sample to calculate pollen concentrations (Stockmarr, 1971). Sodium polytungstate (SPT) was used to separate particle fractions by gravity. The samples were sieved with a 6 µm filter and centrifuged with SPT to separate particle fractions. Two strew-mounted slides were prepared from each residue with glycerin gelatin as a mounting medium. The resulting palynological data have previously been used for qualitative analysis of the vegetation and environmental conditions at Aghitu-3 (Kandel et al., 2017). In this study, we focus on the pollen data.

2.5. Plant use assessment

For the assessment of plant use by humans we selected all taxa from the pollen and *sed*aDNA record that were identified to genus or species. From this list we limited our analyses to those species that are included in the Flora of Armenia (Takhtajan, 1954-2009). For taxa determined at genus level (DNA and pollen) or as pollen types, which both represent several species, we used only Armenian species for the compilation of information. For example, out of 472 accepted species of *Artemisia*, 23 occur in Armenia and these 23 species have been assessed for evaluating the potential plant use of the genus *Artemisia* in this study. The taxonomic status of these Armenian species was checked with the Plants of the World Online database (POWO; http://www.plantsoftheworldonline.org/) for their validity, and potential synonyms.

Potential uses of plants were identified using the ROCEEH project database PlantBITES (Altolaguirre et al., 2021, and unpublished data), the Plants For A Future database (https://pfaf.org; Fern, 1997), the Useful Temperate Plants database (http://temperate.theferns.info/; Fern, 2019), as well as botanical and ethnobotanical publications (e.g. Fleischhauer et al., 2016; Rivera et al., 2012). In cases where no information was available, the taxon was assumed not to have any uses (see Table 1). We evaluated the following categories of use: edibility, medicinal, dye, repellent and other (e.g. string making, thatching, strewing, weaving). Edibility was further divided into whether parts of the respective plant can be eaten in larger amounts, i.e. serve as bulk food, or if the plant is rather used as a flavouring (condiment, herb, spice, tea). For each taxon we scored its use per category whether it is useful or not. For taxa identified to genus level and comprising several species, we scored the genus as being useful if one or several of the Armenian species listed are known to be useful. In cases where all Armenian species of a genus were considered useful, this was noted in the table. If none of

the Armenian species of this genus had any reported use, the genus was recorded as not useful, the respective usefulness of the genus was given as 'no'.

2.6. Statistical analyses

All statistical analyses were performed in R (version 4.1.0; R Core Team, 2021). To visualize changes in abundance of plant taxa over time, we created stratigraphic plots with the *rioja* package (Juggins & Juggins, 2020; Appendix Fig. A.3).

To investigate plant assemblages, we used non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities, allowing the visualization of dissimilarities in composition across samples. A stress level of <0.2 for NMDS is considered acceptable for this purpose (Clarke, 1993). We grouped samples according to their archaeological horizons and tested for differences between groups by performing a perMANOVA with 999 permutations using the *adonis*-function in the *vegan* package (Oksanen et al., 2020). We tested the perMANOVA assumption of homogeneity of multivariate dispersions using the *permutest* function with 999 permutations.

To facilitate comparison between *sed*aDNA and pollen diversity and relating these to plant-use information, we reduced the *sed*aDNA dataset from MOTUs to species and genera, summing the different MOTUs of the same species or genus together. This reduced dataset was used to determine the taxonomic richness and the inverse Simpson index using Hill numbers (q = 0 and q = 2) with the *hill_taxa* function of the *HillR* package (Li, 2018). We used the non-parametric Mann-Whitney-Wilcoxon test to identify differences in plant diversity among archaeological horizons , available under the function *wilcox_test* from the *rstatix*-package (Kassambara, 2021) with standard settings.

3. Results

3.1. Brief description of the data

3.1.1. Pollen

Twenty-seven of the 38 previously analysed palynological samples contained pollen, and from these 39 pollen taxa were identified, covering 28 families and 25 plant genera, with seven species identifications corresponding to species occurring in Armenia. Total pollen concentrations average to 278 ± 75 grains per gram of sediment (excluding empty samples; Fig. 2A). Herbs were the most abundant plant group in the pollen dataset with 26 pollen types accounting for more than 70% of the total pollen concentration. Trees, represented by 12 pollen types, accounted for 17.5%, followed by grasses at 12%. Detailed pollen concentration diagrams can be found in Appendix Fig. A.3 and the corresponding data in Appendix Table B.1.

3.1.2. *Sed*aDNA

We obtained 168.8 million raw DNA sequence reads. After filtering, the DNA dataset consisted of 3.3 million reads from 21 samples, leaving four samples that had no sequences remaining after filtering. Read counts per sample, corrected for the amount of input material, averaged to 262 ± 32 log-transformed reads per mg of sediment (excluding empty samples; Fig. 2A). The filtered DNA dataset contained 128 unique MOTUs, corresponding to 10 plant families, 20 genera, and six species (Appendix Table B.2). Herbs account for 116 MOTUs and more than 93% of the log-transformed read abundance, followed by grasses with 11 MOTUs at approximately 6% and trees with 1 MOTU at <0.5%.

3.2. Recovery of plant taxa

3.2.1. Comparison of pollen and *sed*aDNA

Across pollen and *sed*aDNA datasets, a total of 32 plant families, 43 genera and 14 species were identified, with an overlap between the two datasets of six plant families (19%), two genera (5%), and no overlap in detected species. Regarding the plant families, 22 are unique to the pollen dataset, including Betulaceae and Pinaceae, and four plant families are unique to the *sed*aDNA dataset (Boraginaceae, Fabaceae, Lamiaceae, Polygonaceae). The six plant families occurring in both datasets are Asteraceae, Brassicaceae, Chenopodiaceae, Fagaceae, Plumbaginaceae, and Poaceae. Asteraceae (55%) represents the most abundant plant family in the pollen record, while Brassicaceae (44%) dominates the *sed*aDNA dataset, followed by Fabaceae (21%). The plant genera detected in both datasets are wormwood (*Artemisia* spp.) and oak (*Quercus* spp.). Another 23 genera were uniquely detected by pollen compared to 18 by *sed*aDNA analysis.

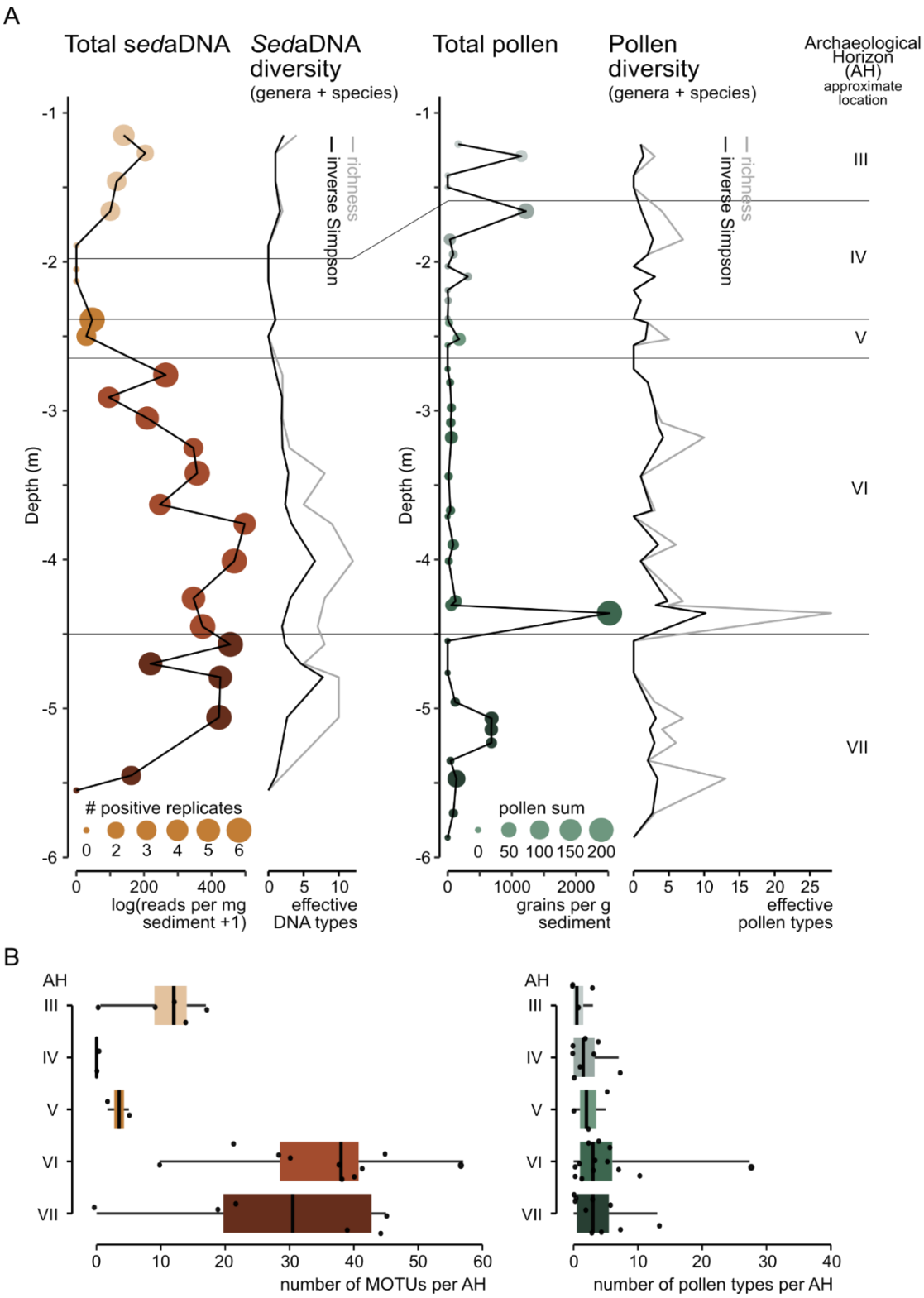


Figure 2. Plant diversity per Archaeological Horizon (AH) based on *sed*aDNA and pollen analysis. A) The total concentrations of plant *sed*aDNA and pollen per sample and the inverse Simpson index of the recovered plant genera and species with respect to depth. The size of the circles indicates for *sed*aDNA, the number of positive replicates kept after filtering, and for pollen, the total pollen sum. B) The total number of *sed*aDNA MOTUs (left) and pollen taxa (right) detected per layer illustrated with boxplots (centre line: median, box limits: upper and lower quartiles, whiskers: extremes of the data). A total of 25 *sed*aDNA samples (AH III, n = 5; IV, n = 2; V, n = 2; VI, n = 10; VII, n = 6) and 38 pollen samples (AH III, n = 4; IV, n = 8; V, n = 3; VI, n = 13; VII, n = 10) were compared.

3.2.2. Plant assemblages across archaeological horizons

Among the analysed samples, 21 *sed*aDNA and 27 pollen samples contained identifiable plant taxa and were used to investigate the plant assemblages. Among them, four *sed*aDNA and two pollen samples were from AH III (~29-24,000 cal BP), five pollen samples were from AH IV (~30-29,000 cal BP), while no *sed*aDNA was recovered from the two samples taken from this AH. Further, two *sed*aDNA and three pollen samples were from AH V (~32-30,000 cal BP), 10 *sed*aDNA and 10 pollen samples were from AH VI (~36-32,000 cal BP), and five *sed*aDNA and seven pollen samples were from AH VII (~39-36,000 cal BP). The NMDS showed visual separation between some of these groups of samples, while others overlap (Fig. 3). PerMANOVA analyses on Bray-Curtis dissimilarities derived from both datasets indicated significant differences between plant assemblages among AHs (*sed*aDNA adonis $F_{3,20} = 2.53$, $r^2 = 0.31$, $p \leq .001$; pollen adonis: $F_{4,26} = 1.70$, $r^2 = 0.24$, $p \leq .01$). The assumption of homogeneity of dispersion among AHs was supported by nonsignificant permutest results (*sed*aDNA: $p = .158$, pollen: $p = .352$). Post hoc pairwise perMANOVAs on subsets of the *sed*aDNA data indicated statistically significant (Bonferroni-adjusted $p < .05$) differences between AH III and VI, between AH III and VII, and between sets of AHs, VII-VI and V-IV, and VII-VI and III. The same analysis for the pollen dataset revealed a significant difference between the AH VII-VI set and III, but no significant differences in pairwise comparisons of individual AHs according to Bonferroni-adjusted *p*-values. Based on unadjusted $p = .05$ only the difference between AH III and VII proved significant. Similarly, no significant differences in number of plant taxa between individual AHs were found for either *sed*aDNA or pollen, while grouping together AH VII and VI, and AH V and IV revealed a significant decrease in number of MOTUs (Mann–Whitney $U = 3$, $p = .021$; see Appendix Table A.4).

The number of plant taxa were highest in the lower archaeological horizons, with AH VII having a median of 30.5 MOTUs (IQR = 19.75-42.75, n = 6) and three pollen types (IQR = 0.5-5.5, n = 10) and AH VI having a median of 38 MOTUs (IQR = 28.5-40.75, n = 10) and three pollen types (IQR = 1-6, n = 13; Fig. 2A). No significant differences were found in diversity between these horizons.

In AH VII, the pollen record was dominated by Asteraceae, while Brassicaceae dominated the *sed*aDNA record with five genera (*Brassica* sp., *Isatis* sp., *Lepidium* sp., *Raphanus* sp., and *Sterigmostemum* sp.). Both records indicate presence of these families, as well as Chenopodiacaeae, identified by *sed*aDNA as mainly *Atriplex* sp. and *Chenopodium* spp.. Other plant families detected by both records are Poaceae (*Alopecurus* sp. and other unidentified grasses) and Fagaceae (i.e. *Quercus* sp., oak). Specific to the pollen record in AH VII were traces of other trees, being *Pinus* sp. and *Acer* sp., and some herbs (Apiaceae, Caryophyllaceae, Cyperaceae, *Convolvulus* sp.), and the gymnosperm shrub *Ephedra* sp., while the *sed*aDNA record counted another 10 genera.

In AH VI, one pollen sample (sample 41, depth = -4.26) stands out with higher pollen counts (n = 218; Fig. 2A) and higher diversity than others, containing

28 out of the 38 identified pollen types detected throughout the Aghitu-3 sediment sequence. Particularly *Phragmites* sp. and Asteraceae pollen were abundant in this sample. Asteraceae remained present in the pollen record throughout AH VI, although in decreasing concentrations with decreasing depth. In the *sed*aDNA record, Brassicaceae and Chenopodiaceae remained abundant, while Fabaceae (particularly *Astragalus* sp.) showed much higher presence than in AH VII.

From AH VI to AH V, the median number of plant taxa declined to 3.5 MOTUs (IQR = 2.75-4.25, n = 2) or 2 pollen taxa (IQR = 1-3.5, n = 3). Detected plant taxa in AH V included Asteraceae and Poaceae in both *sed*aDNA and pollen, Fabaceae and Lamiaceae only in the *sed*aDNA record, and Chenopodiaceae, walnut (*Juglans regia*), oak (*Quercus* sp.) and pine (*Pinus* sp.) only in the pollen. In AH IV, *sed*aDNA detected no plant MOTUs and the median number of pollen taxa was 1.5 (IQR = 0-3.25, n = 8). Asteraceae pollen concentrations increased compared to AH V and pine pollen was continuously detected. Other pollen finds included Ericaceae, Chenopodiaceae, Brassicaceae, and low concentration of walnut (*Juglans regia*) at the bottom of this AH.

In AH III, the median number of MOTUs increased to 12 (IQR = 9-14, n = 5), while pollen taxa further decreased to a median of 0.5 (IQR = 0-1.5, n = 4). A high concentration of *Artemisia* sp. pollen was found, as well as *Pinus* sp. and Chenopodiaceae. Regarding *sed*aDNA, this AH was dominated by Fabaceae, particularly *Astragalus* sp., but also some *Pisum* sp. Further finds included several Brassicaceae (*Brassica nigra, Brassica oleracea*, and *Raphanus raphanistrum*) and one taxon from the legumes family (Fabaceae: *Pisum* sp.).

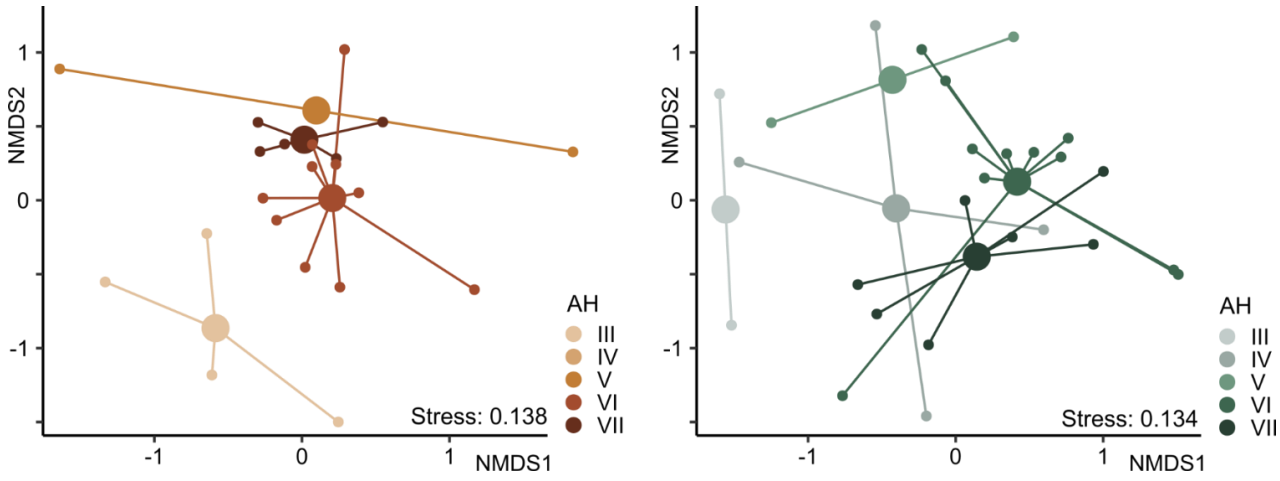


Figure 3. *Sed*aDNA (left) and pollen (right) NMDS ordination of Bray–Curtis dissimilarity of samples (small circles) and averages per archaeological horizon (big circles). The stress levels of 0.138 and 0.134 are under the cut-off value of 0.2 as posed by Clarke (1993) to indicate an interpretable ordination. One pollen sample outlier (sample 11 from AH VI, depth = -2.26, NMDS1 = -4.06, NMDS2 = -0.31, containing only *Juglans* sp. pollen) was excluded from visualisation here, but included in the statistical analysis. Plant assemblages among archaeological horizons were significantly different for both *sed*aDNA (adonis $F_{3,20}$ = 2.53, r^2 = 0.31, $p \leq .001$) and pollen (adonis $F_{4,26}$ = 1.70, r^2 = 0.24, $p \leq .01$).

3.3. Potential use of recovered plant taxa

3.3.1. Useful plant data

The potential use of the detected plant taxa was determined for those identified to genus and species level, reducing the *sed*aDNA dataset to 21 taxa, and the pollen dataset to 24 taxa. With the focus on Armenian species, most of these taxa were considered useful in at least one, and many of them in multiple use categories (i.e. bulk food, flavouring, medicinal, dye, repellent, other), including 16 of the 21 taxa identified with *sed*aDNA, and 20 of the 24 taxa identified with pollen analysis. Those considered not-useful include: *Alopecurus myosuroides*, *Lappula* sp., *Alyssum* sp., *Sterigmostemum* sp., and *Hedysarum* sp.. In all five cases this is due to a lack of sufficient information available rather than being based on ethnobotanical evidence.

3.3.2. Edible plants

A total of 15 edible plant taxa were detected by *sed*aDNA analysis, nine of which categorized as bulk foods, and six usable as flavouring (Fig. 4). Pollen analysis detected 18 edible plant taxa; six bulk food and 12 flavouring. The bulk food taxa detected by *sed*aDNA were primarily found in AH VII, VI and III and included several members of the Brassicaceae (e.g. *Brassica nigra*, *Brassica oleracea*, *Raphanus raphanistrum* and *Isatis* sp.) and Chenopodiaceae (i.e. *Atriplex* sp., and *Chenopodium* sp.) as well as *Rumex scutatus*, *Cirsium* sp., and *Pisum* sp.. The pollen record showed a high richness in bulk food in one particular sample (depth = -4.26) including *Beta vulgaris*-type pollen and those of *Phragmites* sp. and *Typha* sp., while other pollen samples included one bulk food taxon at most, from either *Celtis* sp., *Juglans* sp., or *Acer* sp.. Plant taxa that could be used for flavouring were more consistently present in the pollen record, and were also detected in many *sed*aDNA samples throughout the sediment sequence.

3.3.3. Other uses

All of the 16 useful plant taxa detected with *sed*aDNA were categorized as medicinal compared to 19 out of the 21 useful plant taxa detected through pollen, and both records indicate consistent presence of these taxa throughout the sedimentary sequence (Fig. 4).

Plant taxa possibly used for dye were similarly detected throughout the pollen record, but limited to AH VII and VI in the *sed*aDNA record. Those detected by *sed*aDNA included *Rumex scutatus*, *Lithospermum* sp., *Isatis* sp., and *Chenopodium* sp., while the pollen record distinguished eight different taxa, of which four trees (two *Quercus*-types, *Pinus* sp., and *Acer* sp.) with particularly *Pinus* sp. being present throughout most of the sediment sequence.

The category of repellents encompassed five taxa detected with *sed*aDNA and three detected with pollen. These include several of the taxa that were also categorised as bulk foods and/or flavouring (*Brassica* spp., *Artemisia* sp.,

Chenopodium sp., and *Quercus* sp.), and *sed*aDNA from these categories showed similar patterns (Fig. 4). Repellent taxa found by pollen analysis were *Acorus calamus* sp., *Artemisia* sp., and *Matricaria matricarioides*.

Ten taxa were assigned to the category of other uses, of which two were found with both proxies (*Quercus* sp. and *Artemisia* sp.), two others uniquely with *sed*aDNA (*Cirsium* sp. and *Isatis sp.*), and eight with pollen. Several of the pollen taxa belonging to this category are trees (*Pinus* sp., *Quercus* sp., and *Betula* sp.), others were grasses (*Phragmites* sp.) and herbs (*Acorus calamus* sp., *Artemisia* sp., *Convolvulus* sp. and *Typha* sp.).

Table 1. List of Armenian plant genera and species documented by *sed*aDNA and pollen from Aghitu-3 Cave and their assignments to plant-use categories based on information in the ROCEEH project database PlantBITES (Altolaguirre et al., 2021, and unpublished data), the Plants For A Future database (https://pfaf.org; Fern, 1997), the Useful Temperate Plants database (http://temperate.theferns.info/; Fern, 2019), as well as botanical and ethnobotanical publications (e.g. Fleischhauer et al., 2016; Rivera et al., 2012). The taxonomic status of Armenian plant taxa was checked with the Plants of the World Online database (POWO; http://www.plantsoftheworldonline.org/).

Assigned Armenian taxa	N species in Armenia (POWO accepted)	Proxy	Bulk food	Flavouring	Medicinal	Dye	Repellent	Other	Other uses:
Herbs									
Acoraceae									
Acorus calamus L.		Pollen	-	√	√	-	√	√	basketry, strewing, thatching, weaving;
Asteraceae									
Achillea pseudoaleppica Hausskn. ex Hub.-Mor.		DNA	–[2]	–[2]	√	–[2]	–[2]	–[2]	
Artemisia sp.	24(23)	Both	-	√	√	-	√	√	strewing;
Centaurea sp.	88	Pollen	-	-	-	√	-	-	
Cirsium sp.	22(11)	DNA	√	-	√	-	-	√	tinder;
Matricaria matricarioides (Less.) Porter ex Britt.		Pollen	-	√	√	-	√	-	
Boraginaceae									
Lappula sp.	8(6)	DNA	–[2]	–[2]	–[2]	–[2]	–[2]	–[2]	
Lithospermum officinale L.		DNA	-	√	√	√	-	-	
Brassicaceae									
Alyssum sp.	24(15)	DNA	-	-	–[2]	–[2]	–[2]	–[2]	
Brassica sp.	6(5)		√[1]	-	√	-	√	-	
Brassica nigra (L.) W.D.J. Koch		DNA	√	-	√	-	√	-	
Brassica oleracea (L.)		DNA	√	-	√	-	√	-	
Isatis sp.	13(12)	DNA	√	√	√	√	-	√	preservative;
Lepidium sp.	12(12)	DNA	-	√	√	–[2]	–[2]	–[2]	
Raphanus raphanistrum L.		DNA	√	-	√	-	-	-	
Sterigmostemum sp.	2(2)	DNA	–[2]	–[2]	–[2]	–[2]	–[2]	–[2]	
Caprifoliaceae									
Knautia arvensis type	2(2)	Pollen	-	√	√	-	-	-	
Chenopodiaceae									
Atriplex sp.	12(11)	DNA	√	-	√	-	-	-	
Beta vulgaris type	5	Pollen	√	-	√	-	-	-	
Chenopodium sp.	12(11)	DNA	√[1]	-	√	√	√	-	
Convolvulaceae									
Convolvulus arvensis type	2	Pollen	-	√	√	√	-	√	string;
Ephedraceae									
Ephedra distachya type	1	Pollen	-	√	√	-	-	-	
Fabaceae									
Astragalus sp.	144	DNA	-	√	√	-	-	-	
Hedysarum sp.	8	DNA	–[4]	–[4]	-	-	-	-	
Onobrychis sp.	25(23)	DNA	–[3]	√[3]	–[3]	–[3]	–[3]	–[3]	
Pisum sativum L.		DNA	√	-	√	-	-	-	

Assigned Armenian taxa	N species in Armenia (POWO accepted)	Proxy	Bulk food	Flavouring	Medicinal	Dye	Repellent	Other	Other uses:
Malvaceae									
Malva neglecta type	3	Pollen	√	-	√	√	-	√	toothbrush;
Polemoniaceae									
Polemonium sp.		Pollen	-	-	√	-	-	-	
Polygonaceae									
Rumex scutatus		DNA	√	-	√	√	-	-	
Ranunculaceae									
Thalictrum sp.	6	Pollen	-	√	√	-	-	-	
Typhaceae									
Typha sp.	10	Pollen	√	-	√	-	-	√	tinder, thatching, weaving, stuffing;
Viscaceae									
Viscum album L.		Pollen	-	√	√	-	-	-	
Grasses									
Poaceae									
Alopecurus myosuroides Huds.		DNA	-[2]	-[2]	-[2]	-[2]	-[2]	-[2]	
Phragmites australis (Cav.) Trin. ex Steud.		Pollen	√	-	√	√	-	√	thatching, string, weaving;
Trees									
Betulaceae									
Betula sp.	3(2)	Pollen	-	√[1]	√	-	-	√	drinking vessels;
Carpinus sp.	3	Pollen	-	-	√	√	-	-	
Buxaceae									
Buxus sempervirens L.		Pollen	-	-	√	-	-	-	
Cannabaceae									
Celtis sp.	2	Pollen	√[1]	-	-[2]	-	-	-	
Fagaceae									
Quercus sp.	6(4)	DNA	-	√	√	-	√	√	gum, basketry, tannin
Quercus frainetto type	4	Pollen	-	√	√	√	-	√	basketry;
Juglandaceae									
Juglans regia L.		Pollen	√	-	-[2]	-	-	-	
Pinaceae									
Pinus sp.	6(4)	Pollen	-	√[1]	√	√	-	√	string, tannin, pitch, turpentine;
Salicaceae									
Salix sp.	14	Pollen	√	-	√	√	√	√	string, resin, adhesive, basketry, stuffing, tannin;
Sapindaceae									
Acer sp.	7	Pollen	√	-	√	√	-	-	

[1] True for all Armenian species

[2] No information found, designated as not useful

[3] No information on Armenian species, information on *Onobrychis viciifolia* was used

[4] Some Chinese and American species are reported to be edible, but those known to occur in Armenia are not

4. Discussion

We combined evidence of plant presence from *sed*aDNA metabarcoding and pollen data from Aghitu-3 Cave with information on known plant uses in order to assess the potential of *sed*aDNA metabarcoding for reconstructing prehistoric plant assemblages in caves associated with human occupation. Pollen and *sed*aDNA analysis enabled us to reconstruct the plant assemblages, while the plant assessment of usability provided insight into possible human plant use at Aghitu-3 during the Upper Palaeolithic.

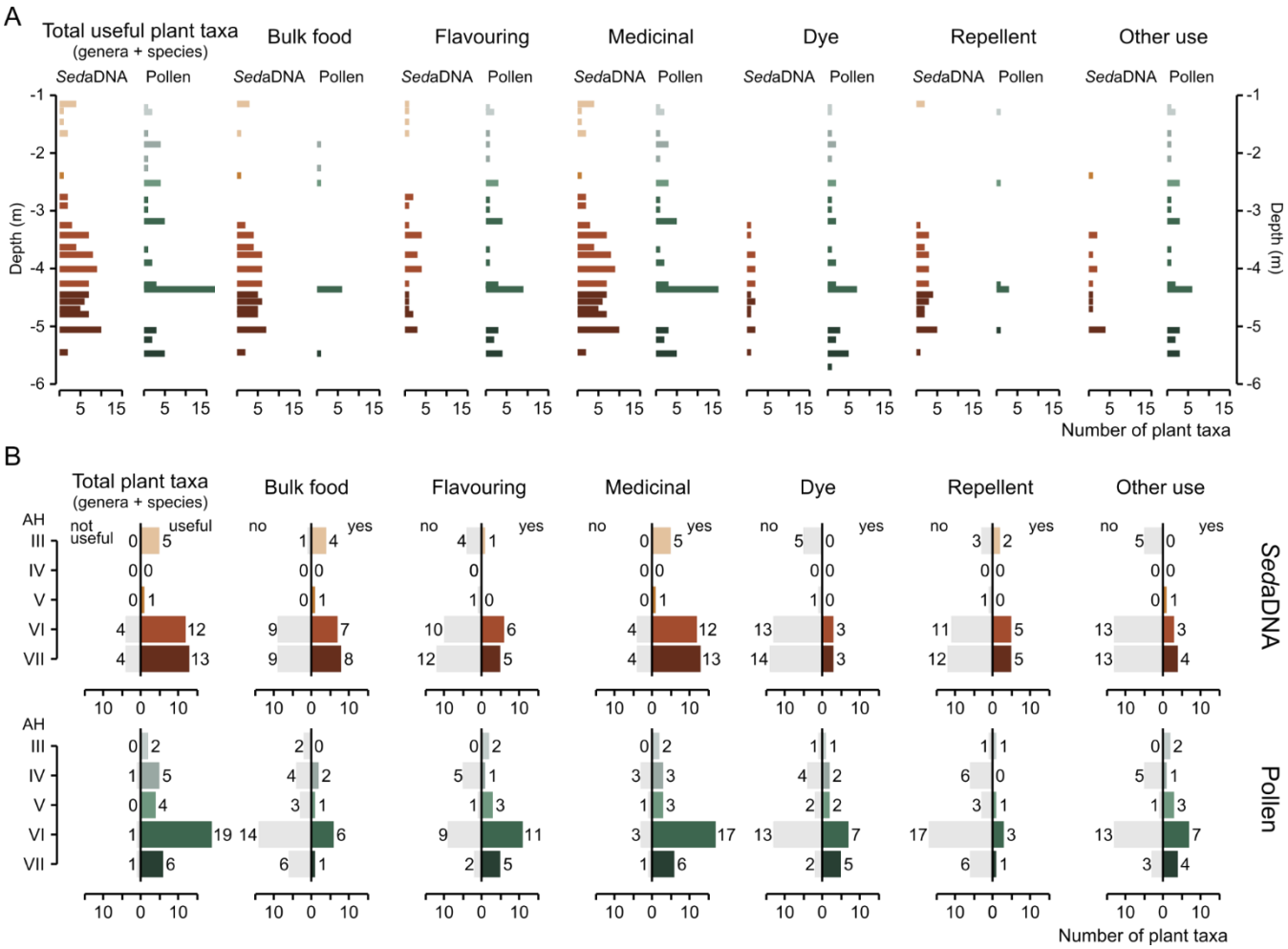


Figure 4. Useful plants per Archaeological Horizon (AH) based on reduced *sed*aDNA and pollen datasets to genus-species-level identifications. A) The number of useful plant taxa per category and proxy (*sed*aDNA and pollen) over dept The number of useful and not-useful plant taxa per category, summarized per AH for *sed*aDNA (top) and pollen (bottom) colouring of the bars is indicative of the AH, with darker colours for deeper horizons.

4.2. Taphonomy and recovery of plant *sed*aDNA and pollen

Plant *sed*aDNA was recovered throughout the Aghitu-3 Cave sequence, consistent with the numerous organic finds throughout the stratigraphy as reported by Kandel et al. (2017). By contrast, the recovery of pollen appeared more variable and overall pollen numbers were low (Fig. 2A). The pathways through which vegetation travels from outside the cave to the cave floor vary, including transportation by wind, by animals (for nests, burrows, on their bodies, or in faecal matter), or by humans (Hunt & Fiacconi, 2018). Alternatively, plant remains could come directly from vegetation growing at the cave entrance (Hunt & Fiacconi, 2018). Aghitu-3 is a relatively wide and shallow cave (11 m deep, 18 m wide, and 6 m high), not ruling out the possibility of plant remains being deposited in the cave without animal or human agency. Comparing the different botanical proxies

and integrating these with other evidence from the site facilitates further inferences about the origin of the identified plant diversity.

4.2.1. Sources and transfer of pollen and sedaDNA

Pollen and *sed*aDNA records are differentially affected by source productivity and taphonomic processes of dispersal, transfer and deposition (Alsos et al., 2018; Giguet-Covex et al., 2019; Niemeyer et al., 2017), and both proxies signify different records. *Sed*aDNA is generally considered of local origin (Alsos et al., 2018; Jørgensen et al., 2012), whereas pollen (especially from wind-pollinated plants, such as many trees and grasses) may originate from a wide area as they are distributed regionally through the air (Birks & Bjune, 2010; Parducci et al., 2018). Moreover, plant DNA in sediments can originate from most plant parts, but is unlikely to source from pollen due to the robustness of the pollen wall and the low amount of DNA in pollen (Clarke et al., 2020; Parducci et al., 2017; Sjögren et al., 2017). The *sed*aDNA record at Aghitu-3 indicates relatively high plant DNA concentrations throughout AH VII and VI, whereas pollen concentrations are low (Fig. 2A). The Asteraceae family dominates the pollen record, and although several known wind-pollinator taxa were found, (including grasses, oak, and pine) their relatively low concentrations suggest wind-transportation of plant remains into the cave has been limited. The combination of low pollen concentrations with high *sed*aDNA abundance may indicate periods where animals or humans actively brought plant remains into the cave.

The potential of plants growing at the Aghitu-3 Cave entrance or inside the cave would have probably been related to occupancy, with more chances for plants to establish during periods of reduced occupation. Moreover, remains of these plants in the cave deposits would likely be detected by both pollen and *sed*aDNA simultaneously. Comparison of the pollen and *sed*aDNA records revealed six families and two genera in common. Of the two genera, oak seems an unlikely pioneer plant of cave environments, whereas *Artemisia* species are described as preferring no-shade to semi-shade conditions (Fern, 1997). With the cave opening towards the northeast, the cave receives only morning sun. We further found few instances of simultaneous detection of plant families by *sed*aDNA and pollen. Brassicaceae was recorded throughout AH VII and VI by *sed*aDNA, but occurred only in single pollen samples of AH VII, VI and III. Both records found Chenopodiaceae in AH VII and VI, but only the pollen record also found this family in AH IV and III. Fagaceae was found throughout the pollen record, and only in two DNA samples. Plumbaginaceae was recorded with *sed*aDNA throughout AH VI, but only in one pollen sample in AH VI. Lastly, Poaceae as well as Asteraceae were present in both pollen and *sed*aDNA records in AH VII, VI and V, however, Asteraceae was highly abundant in the pollen record in AH IV, but absent from the *sed*aDNA record of that layer. Overall, we see few layers where both records overlap. In contrast to the assumed higher potential of plants to grow at the cave entrance or inside the cave during periods of reduced occupation, we found the

layers indicating overlap among proxies to be those associated with increased human occupation, and overall lower abundance and diversity of plants in the archaeologically sterile layers of AH VI and V. AH VI and V were also reported to contain fewer other botanical remains compared to the other layers, as evident from previous palynological analysis, including investigation of non-pollen palynomorphs (Kandel et al., 2017). Overall, these findings suggest a limited influence on the pollen and *sed*aDNA records of plants growing at the cave entrance or inside the cave.

4.2.2. Post-depositional preservation

Post-depositional factors impacting pollen and *sed*aDNA could produce differences in the recovery of plant data by the two proxies. Pollen walls are highly robust, while *sed*aDNA is known to degrade significantly as a result of temperature fluctuations, low pH, high oxygen and water content, decreasing the amount of detectable DNA over time (Willerslev & Cooper, 2005). DNA is better preserved in sediments with a high mineral content and at low temperatures. Minerals can bind to and protect DNA, while low temperatures thermally stabilize DNA against chemical degradation (Torti et al., 2015). Desiccated dry sediments can also strongly decrease the effects of hydrolysis, breaking down and damaging DNA. The relatively dry cave environment of Aghitu-3 has preserved many organic remains (Bertacchi et al., 2021; Kandel et al., 2017), and the sediments themselves are composed mainly of silt, clay minerals, volcanic ash and basalt debris from the cave ceiling (Kandel et al., 2014).

When there is a more or less constant input of plant remains through time, as for many lake sediments, *sed*aDNA recovery is usually lower in older sediments due to degradation processes (Parducci et al., 2017). At Aghitu-3, we indeed observe reduced *sed*aDNA concentrations in the lowest two samples. However, in the lower archaeological horizons (AH VII and VI) overall DNA concentrations were higher compared to the younger layers (AH V, IV, and the lower part of AH III) without obvious evidence of changes in the sediment composition. Most of the sedimentary sequence consists of clayey silt, with somewhat more clay in AH VI compared to other layers, and only in AH VII does this grade with depth to silty sand. Remains obtained from palynological analyses were observed throughout the sediment sequence and other organic remains found in the cave sediments include charcoal and faunal remains from birds, fish, micromammals and macrofauna (Bertacchi et al., 2021; Kandel et al., 2017). The abundance patterns of these remains along the archaeological horizons indicate an association with human occupation, with more organic remains found in archaeological horizons that also contained high densities of lithics (AH VII, VI and III) compared to those considered almost sterile (AH IV, V). Macrofaunal remains (mainly goat and horse) were detected with the lowest abundance in AH V and IV, followed by AH VII and high abundance in AH VI and III, with anthropogenic modifications observed on several bones from these layers as well as evidence for accumulation by carnivores

in AH VI. Bone preservation was reported to be moderate in AH VI and V, while faunal remains in AH III were well-mineralised (Kandel et al., 2017). The macrofaunal abundance pattern matches that of the lithics as well as the micro-charcoal concentrations, and together with the rare presence of carnivore remains and lack of coprolites suggests human occupation in AH VII, VI and III. In contrast, the faunal remains from small mammals (lagomorphs and rodents) showed high rates of deposition especially in AH V and IV, associated with a reduced presence of humans at the cave. In sum, the abundance of organic remains throughout the sediment sequence suggests good preservation conditions, with patterns of *sed*aDNA matching those of human associated finds while contrasting with micro-mammal presence.

4.3. Reconstructed plant assemblages across archaeological horizons

Our analysis of pollen and plant *sed*aDNA in the sediment record of Aghitu-3 Cave revealed a diverse plant record with an abundance of Brassicaceae, Fabaceae, Chenopodiaceae (supported by the *sed*aDNA), and Asteraceae (supported by the pollen data). The diversity of plant taxa matched the total abundance patterns, with highest diversity in the lowest archaeological find horizons (AH VII and VI) and lower diversity in the almost sterile levels AH V and IV. For the pollen record, plant diversity also remained low in AH III, the level with the highest archaeological record, whereas for the *sed*aDNA record, plant diversity increased in AH III (Fig. 2). From the plant families that were detected by both proxies, the *Brassicaceae* family could not be further specified by pollen, while *sed*aDNA enabled further distinction to genus and species levels. Overall, pollen and *sed*aDNA records showed little overlap in detected taxa.

The higher diversity in plant assemblages in AH VII, VI and for *sed*aDNA also in AH III, matches changes in the palaeoclimate previously reconstructed from micromammal and pollen data as well as the nature of the sediments, and is in accordance with the global climatic trend for the time period covered (Lisiecki & Raymo, 2005). Conditions were warm and humid during the deposition of AH VII and especially VI (~36-32 ka cal BP), while there was a cooler and humid climate during AH V (~32-30 ka cal BP), a further cooling trend during deposition of AH IV (~30-29 ka cal BP) and a cold period in the lower part of AH III (~29-26 ka cal BP) followed by warmer conditions during the upper part of AH III (~26-24 ka cal BP). Especially higher humidity supports the expansions of woody plants and a diversification of the landscape, as is the case today in more humid parts of the Southern Caucasus (Gulisashvili et al., 1975). Together with warmer conditions during AH VII, VI and the upper part of AH III, this change in environmental conditions likely facilitated establishment of a higher diversity of plants in the vicinity of the cave.

The reconstructed plant assemblages differed significantly among AHs, as supported by NMDS and perMANOVA. This stratification in assemblages was stronger in the *sed*aDNA compared to the pollen, showing that the plant

assemblages reconstructed using *sed*aDNA are more specific to the AHs than those from the pollen. This may reflect an artefact of the nature of the *sed*aDNA data, with MOTUs providing a higher resolution than pollen types. Alternatively, this may indicate that the *sed*aDNA more closely reflects the delineation into archaeological horizons of human occupation alternating with sterile layers, where the presence of more or different plants resulted from increased human activity. The patterns in the reduced *sed*aDNA dataset (species and genera only) support an association with human occupation in terms of diversity (Fig. 2A). The abundance and composition of the *sed*aDNA record match inferences from previous finds at Aghitu-3, with lithic densities indicating low settlement intensity in AH VII and VI, no significant occupation in AH V and IV, followed by high lithic density in AH III, mirrored by similar patterns in the presence of combustion features and evidence for animal exploitation (Bertacchi et al., 2021; Kandel et al., 2017).

4.4. Potential plant use at Aghitu-3

Plant use assessment based on modern use data limits inferences to our current relation with plants, supplemented by traditional knowledge that has persisted. A further limitation of the assessment of plant use based on plant presence in cave sediments is that this evidence is inherently indirect. To imply human plant use, the presence of useful taxa should correlate with evidence of increased human occupation. Despite the number of recovered plant taxa being higher in layers associated with human occupation, no clear distinction between sterile and human occupation layers was found in the proportion of usable plant taxa. This can be explained by the very limited recovery of taxa in sterile layers, as well as our finding that the majority of recovered plants were useful in at least one of the defined categories. From the plant taxa that were recovered at Aghitu-3, we were able to obtain information about their potential use for all (43 taxa) but three plant taxa, which were assumed not to be useful, making a total of five taxa found not to be useful across all categories (Table 1).

Within edible plants, we found the highest number of plant taxa that could be used as bulk food in AHs VII and VI, primarily from the vegetable families of Brassicaceae (*Brassica nigra*, *Brassica oleracea*, *Raphanus raphanistrum*, *Isatis* sp.) and Chenopodiaceae (*Atriplex* sp., and *Chenopodium* sp.). The Brassicaceae family, also known as the mustard or cabbage family, includes many of our modern winter vegetables. *Brassica oleracea* is also known as wild cabbage and its modern cultivars include several common foods such as broccoli, kale, Brussels sprouts, and cauliflower. Seeds of the black mustard (*Brassica nigra*) are currently used as spice, but also the rest of the plant is edible, although it should not be eaten in large amounts due to its high content of glucosinolate (Fern, 1997; Fleischhauer et al., 2016). Similarly, all parts of the wild radish (*Raphanus raphanistrum*) are edible. Many *Isatis* species are also edible, though they have a very bitter taste (Fern, 1997). The Amaranth family (Chenopodiaceae) was represented by saltbush

(*Atriplex* sp.) and goosefoots (*Chenopodium* sp.), and species of both genera can be eaten like spinach. Other detected bulk foods in AH VII and VI included maple (*Acer* sp.), French sorrel (*Rumex scutatus*), a member of the buckwheat family mainly used in salads, plume thistles (*Cirsium* sp.) including cultivated and uncultivated species, and *Pisum* sp. (recently renamed to *Lathyrus*; Govaerts et al., 2021), which was likely garden pea (*Lathyrus oleraceus*). Of the described plant taxa present in AH VII and VI, *Cirsium* sp. was also detected in AH V, while several of the Brassicaceae as well as the likely garden pea, were also found in AH III. One pollen sample in AH VI (at -4.26 m) with high plant richness further included *Beta vulgaris*-type pollen, likely from common beet, as well as pollen from reed grasses (*Phragmites* sp.), and *Typha* sp.. In Armenia, only one species of *Phragmites* is an accepted taxon in the POWO database: *Phragmites australis*, of which all parts are edible, particularly the stems have a high sugar content and the roots are rich in starch. *Typha* has many common names including bulrush, reed, and cattail, and the underground storage organs (USOs) are especially rich in starch as is also the case for *Acorus calamus*. Evidence of preserved starch grains on grinding stones indicates USOs of *Typha* were eaten and processed for flour in Europe and the Russian plain up to 30,000 years ago (Revedin et al., 2010). Some of these detected food plants also had other potential uses. For example, the seed fluff of some species of plume thistles (*Cirsium* sp.) has been reported to be used as tinder (Fern, 1997).

The majority of the recovered plants was found to be of potential medicinal use, several are known repellents and other potential uses included strewing (i.e. the scattering of fragrant or insect repellent plants), basketry, tinder, and string (Table 1). Four out of the 43 plant taxa recovered at Aghitu-3 can be used for making string, and archaeological finds at Aghitu-3 Cave include a bone eyed needle recovered from AH III (Kandel et al., 2017), suggesting the manufacturing of clothing. The oldest reported dyed textile fibres to date were found in Dzudzuana Cave, Georgia, made from wild flax and dated to approximately 30 ka years ago (Kvavadze et al., 2009). 11 out of the 43 plant taxa recovered at Aghitu-3 can be used for dyeing. For example, *Isatis tinctoria* is a species also known as dyer's woad, for the indigo dye that can be produced from the leaves. Common reed (*Phragmites australis*) could similarly be used to make dye from the flowers (light green), but has multiple other uses due to the sturdiness of the stems and leaves. These include the making of constructions and fences, thatching, as well as making mats, nets and rope (Fern, 1997).

5. Conclusions

Survival of identifiable plant remains is rare at Palaeolithic sites and our current knowledge about human plant-use during this period is defined and limited by the degradation and recovery of plant remains and the availability of methods and technologies for plant detection and identification. Cave environments can mitigate the full effect of the elements on the organic materials concentrated by animals and humans. Using evidence from *sed*aDNA and pollen, we were able to reconstruct the plant assemblages of the 15,000 year long sediment record of Aghitu-3 Cave.

We found that *sed*aDNA can provide higher taxonomic resolution for families of herbaceous plants than pollen, e.g. Brassicaceae, while for some taxa pollen provide higher taxonomic resolution (especially for woody plants) making the two methods highly complementary. However, *sed*aDNA more accurately reflects periods of human occupation than pollen, as pollen sources from a larger geographic distance and therefore more strongly represents the environments surrounding the cave and are less likely to be transported into the cave through human agency. Assessment of the potential use of the recovered plant taxa reveals that most of them could have been used for food, flavouring, medicine, and/or other technical purposes, documenting the high potential of the past environments to provide a diverse array of useful plants. Their presence in layers associated with human occupation suggests anthropogenic transport and exploitation of plant resources. This study represents the first application of *sed*aDNA analysis on Armenian cave sediments and allowed for more specific inferences about potential plant use of prehistoric humans in the Armenian Highlands.

Author contributions

Conceptualisation HB, SB, AK, ATS; design SB, AK, AAB, ATS; excavation BG, AK; sampling AK, ATS; *sed*aDNA and bioinformatic analysis ATS; pollen analysis AAB; validation of *sed*aDNA identifications AAB, RB, AKB; plant use assessment AAB; statistical analysis and writing original draft ATS. All co-authors commented on the manuscript.

Acknowledgements

The excavation and research programs at Aghitu were funded by The Gefoeller Foundation (USA), the Armenian National Academy of Sciences, and the Heidelberg Academy of Sciences and Humanities. Further financial support was provided by the Faculty of Mathematical and Natural Sciences, University of Oslo. We thank the Norwegian Sequencing Centre for their advice and DNA sequencing. DNA analyses were performed on the Saga computer cluster, owned by the University of Oslo and Uninett/Sigma2, and operated by the Department of Research Computing at USIT, University of Oslo.

References

Albert, R. M., Weiner, S., Bar-Yosef, O., & Meignen, L. (2000). Phytoliths in the Middle Palaeolithic deposits of Kebara Cave, Mt Carmel, Israel: Study of the plant materials used for fuel and other purposes. Journal of Archaeological Science, 27(10), 931–947.

Allué, E., Cabanes, D., Solé, A., & Sala, R. (2012). Hearth functioning and forest resource exploitation based on the archeobotanical assemblage from Level J. In High resolution archaeology and Neanderthal behavior (pp. 373–385). Springer.

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PLOS ONE, 13(4), e0195403. https://doi.org/10.1371/journal.pone.0195403

Altolaguirre, Y., Schulz, M., Gibert, L., & Bruch, A. A. (2021). Mapping Early Pleistocene environments and the availability of plant food as a potential driver of early Homo presence in the Guadix-Baza Basin (Spain). Journal of Human Evolution, 155, 102986.

Bertacchi, A., Gasparyan, B., Gruwier, B., Rivals, F., & Kandel, A. W. (2021). Upper Paleolithic animal exploitation in the Armenian Highlands: The zooarchaeology of Aghitu-3 Cave. Quaternary International, 587–588, 400–414. https://doi.org/10.1016/j.quaint.2020.04.029

Birks, H. H., & Bjune, A. E. (2010). Can we detect a west Norwegian tree line from modern samples of plant remains and pollen? Results from the DOORMAT project. Vegetation History and Archaeobotany, 19(4), 325–340.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Clarke, C. L., Alsos, I. G., Edwards, M. E., Paus, A., Gielly, L., Haflidason, H., Mangerud, J., Regnéll, C., Hughes, P. D. M., Svendsen, J. I., & Bjune, A. E. (2020). A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. Quaternary Science Reviews, 247, 106564. https://doi.org/10.1016/j.quascirev.2020.106564

Edwards, K. J., Fyfe, R. M., Hunt, C. O., & Schofield, J. E. (2015). Moving forwards? Palynology and the human dimension. Journal of Archaeological Science, 56, 117–132.

Epp, L. S., Zimmermann, H. H., & Stoof-Leichsenring, K. R. (2019). Sampling and extraction of ancient DNA from sediments. In Ancient DNA (pp. 31–44). Springer.

Fern, K. (1997). Plants for a future: Edible & useful plants for a healthier world. Permanent Publications. https://pfaf.org/

Fern, K. (2019). Useful temperate plants. Http://Temperate.Theferns.Info.

Ferring, R., Oms, O., Agustí, J., Berna, F., Nioradze, M., Shelia, T., Tappen, M., Vekua, A., Zhvania, D., & Lordkipanidze, D. (2011). Earliest human occupations at Dmanisi (Georgian Caucasus) dated to 1.85–1.78 Ma. Proceedings of the National Academy of Sciences, 108(26), 10432–10436.

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. BMC Genomics, 11(1), 434. https://doi.org/10.1186/1471-2164-11-434

Fleischhauer, S. G., Guthmann, J., & Spiegelberger, R. (2016). Enzyklopädie Essbare Wildpflanzen. AT Verlag.

Gabunia, L., & Vekua, A. (1995). A plio-pleistocene hominid from Dmanisi, East Georgia, Caucasus. Nature, 373(6514), 509–512.

Gaillard, M.-J., Sugita, S., Mazier, F., Trondman, A.-K., Broström, A., Hickler, T., Kaplan, J. O., Kjellström, E., Kokfelt, U., Kuneš, P., & others. (2010). Holocene land-cover reconstructions for studies on land cover-climate feedbacks. Climate of the Past, 6(4), 483–499.

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A. L., David, F., Taberlet, P., Brisset, E., Guiter, F., Sinet, R., & Arnaud, F. (2019). New insights on lake sediment DNA from the catchment: Importance of taphonomic and analytical issues on the record quality. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-50339-1

Goldberg, P., Dibble, H., Berna, F., Sandgathe, D., McPherron, S. J., & Turq, A. (2012). New

evidence on Neandertal use of fire: Examples from Roc de Marsal and Pech de l'Azé IV. Quaternary International, 247, 325–340.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Scientific Data, 8(1), 1–10.

Gulisashvili, V., Makhatadze, L., & Prilipko, M. (1975). Rastitelnost Kavkaza (Vegetation of the Caucasus). Nauka, Moskva.

Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M. T. P., Nielsen, R., Munch, K., Ho, S. Y. W., Shapiro, B., & Willerslev, E. (2007). Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a Factor? Molecular Biology and Evolution, 24(4), 982–989. https://doi.org/10.1093/molbev/msm016

Haouchar, D., Haile, J., McDowell, M. C., Murray, D. C., White, N. E., Allcock, R. J. N., Phillips, M. J., Prideaux, G. J., & Bunce, M. (2014). Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–Holocene cave deposit on Kangaroo Island, South Australia. Quaternary Science Reviews, 84, 56–64. https://doi.org/10.1016/j.quascirev.2013.11.007

Hardy, K. (2018). Plant use in the Lower and Middle Palaeolithic: Food, medicine and raw materials. Quaternary Science Reviews, 191, 393–405.

Hardy, K. (2019). Paleomedicine and the use of plant secondary compounds in the Paleolithic and Early Neolithic. Evolutionary Anthropology: Issues, News, and Reviews, 28(2), 60–71.

Hardy, K., Buckley, S., Collins, M. J., Estalrrich, A., Brothwell, D., Copeland, L., García-Tabernero, A., García-Vargas, S., De La Rasilla, M., Lalueza-Fox, C., & others. (2012). Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. Die Naturwissenschaften, 99(8), 617–626.

Hofreiter, M., Mead, J. I., Martin, P., & Poinar, H. N. (2003). Molecular caving. Current Biology, 13(18), R693–R695. https://doi.org/10.1016/j.cub.2003.08.039

Hunt, C. O., & Fiacconi, M. (2018). Pollen taphonomy of cave sediments: What does the pollen record in caves tell us about external environments and how do we assess its reliability? Quaternary International, 485, 68–75.

Joannin, S., Cornée, J.-J., Münch, P., Fornari, M., Vasiliev, I., Krijgsman, W., Nahapetyan, S., Gabrielyan, I., Ollivier, V., Roiron, P., & others. (2010). Early Pleistocene climate cycles in continental deposits of the Lesser Caucasus of Armenia inferred from palynology, magnetostratigraphy, and 40Ar/39Ar dating. Earth and Planetary Science Letters, 291(1–4), 149–158.

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N. H., Andersen, K., Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Molecular Ecology, 21(8), 1989–2003. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Juggins, S., & Juggins, M. S. (2020). Package 'rioja.'

Kandel, A. W., Gasparyan, B., Allué, E., Bigga, G., Bruch, A. A., Cullen, V. L., Frahm, E., Ghukasyan, R., Gruwier, B., Jabbour, F., Miller, C. E., Taller, A., Vardazaryan, V., Vasilyan, D., & Weissbrod, L. (2017). The earliest evidence for Upper Paleolithic occupation in the Armenian Highlands at Aghitu-3 Cave. Journal of Human Evolution, 110, 37–68. https://doi.org/10.1016/j.jhevol.2017.05.010

Kandel, A. W., Gasparyan, B., Nahapetyan, S., Taller, A., & Weissbrod, L. (2014). The upper paleolithic settlement of the Armenian Highlands. Modes de Contacts et de Déplacements Au Paléolithique Eurasiatique—Modes of Contact and Displacements during the Eurasian Palaeolithic, Actes Du Colloque International de La Commission, 8, 39–62.

Kassambara, A. (2021). rstatix: Pipe-friendly framework for basic statistical tests [Manual]. https://CRAN.R-project.org/package=rstatix

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., & Meshveliani, T. (2009). 30,000-year-old wild flax fibers. Science (New York, N.Y.), 325(5946), 1359–1359.

Li, D. (2018). HillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. Journal of Open Source Software, 3(31), 1041.

Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography, 20(1).

Niemeyer, B., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzschuh, U. (2017). A comparison of sedimentary DNA and pollen from lake sediments in recording vegetation composition at the Siberian treeline. Molecular Ecology Resources, 17(6), e46–e62. https://doi.org/10.1111/1755-0998.12689

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). vegan: Community ecology package [Manual]. https://CRAN.R-project.org/package=vegan

Ollivier, V., Nahapetyan, S., Roiron, P., Gabrielyan, I., Gasparyan, B., Chataigner, C., Joannin, S., Cornée, J.-J., Guillou, H., Scaillet, S., & others. (2010). Quaternary volcano-lacustrine patterns and palaeobotanical data in southern Armenia. Quaternary International, 223, 312–326.

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924–942. https://doi.org/10.1111/nph.14470

Parducci, L., Nota, K., & Wood, J. (2018). Reconstructing past vegetation communities using ancient DNA from lake sediments. In Paleogenomics (pp. 163–187). Springer.

Power, R. C., Salazar-García, D. C., Rubini, M., Darlas, A., Harvati, K., Walker, M., Hublin, J.-J., & Henry, A. G. (2018). Dental calculus indicates widespread plant use within the stable Neanderthal dietary niche. Journal of Human Evolution, 119, 27–41.

R Core Team. (2021). R: A language and environment for statistical computing [Manual]. https://www.R-project.org/

Revedin, A., Aranguren, B., Becattini, R., Longo, L., Marconi, E., Lippi, M. M., Skakun, N., Sinitsyn, A., Spiridonova, E., & Svoboda, J. (2010). Thirty thousand-year-old evidence of plant food processing. Proceedings of the National Academy of Sciences, 107(44), 18815–18819.

Rivera, D., Matilla, G., Obón, C., & Alcaraz, F. (2012). Plants and Humans in the Near East and the Caucasus: Ancient and Traditional Uses of Plants as Food and Medicine, a Diachronic Ethnobotanical Review (Vols. 1–2). Ediciones de la Universidad de Murcia.

Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nature Protocols, 13(11), 2447–2461. https://doi.org/10.1038/s41596-018-0050-5

Salazar-García, D. C., Power, R. C., Serra, A. S., Villaverde, V., Walker, M. J., & Henry, A. G. (2013). Neanderthal diets in central and southeastern Mediterranean Iberia. Quaternary International, 318, 3–18.

Seersholm, F. V., Werndly, D. J., Grealy, A., Johnson, T., Early, E. M. K., Lundelius, E. L., Winsborough, B., Farr, G. E., Toomey, R., Hansen, A. J., & others. (2020). Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. Nature Communications, 11(1), 1–10.

Shipley, G. P., & Kindscher, K. (2016). Evidence for the paleoethnobotany of the neanderthal: A review of the literature. Scientifica, 2016.

Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Merkel, M. K. F., Fonville, T., & Alsos, I. G. (2017). Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytologist, 213(2), 929–941.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. Science, 356(6338), 605–608. https://doi.org/10.1126/science.aam9695

Stahlschmidt, M. C., Collin, T., Fernandes, D., Bar-Oz, G., Belfer-Cohen, A., Gao, Z., Jakeli, N., Matskevich, Z., Meshveliani, T., Pritchard, J., & others. (2019). Ancient mammalian and plant DNA from late quaternary stalagmite layers at Solkota Cave, Georgia. Scientific Reports, 9(1), 1–10.

Stockmarr, J. (1971). Tabletes with spores used in absolute pollen analysis. Pollen Spores, 13, 615–621.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14–e14. https://doi.org/10.1093/nar/gkl938

Takhtajan, A. L. (1954). Flora Armenii (Vols. 1–11). Institute of Botany of the Armenian National Academy of Sciences.

Taller, A., Gasparyan, B., & Kandel, A. W. (2018).

Living on the Edge: The Earliest Modern Human Settlement of the Armenian Highlands in Aghitu-3 Cave. In Y. Nishiaki & T. Akazawa (Eds.), The Middle and Upper Paleolithic Archeology of the Levant and Beyond (pp. 119–131). Springer Singapore. https://doi.org/10.1007/978-981-10-6826-3_9

Torti, A., Lever, M. A., & J, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Marine Genomics, 24 Pt 3, 185–196. https://doi.org/10.1016/j.margen.2015.08.007

Trondman, A.-K., Gaillard, M.-J., Mazier, F., Sugita, S., Fyfe, R., Nielsen, A. B., Twiddle, C., Barratt, P., Birks, H. J. B., & Bjune, A. E. (2015). Pollen-based quantitative reconstructions of Holocene regional vegetation cover (plant-functional types and land-cover types) in Europe suitable for climate modelling. Global Change Biology, 21(2), 676–697.

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., & others. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science, 372(6542).

Voldstad, L. H., Alsos, I. G., Farnsworth, W. R., Heintzman, P. D., Håkansson, L., Kjellman, S. E., Rouillard, A., Schomacker, A., & Eidesen, P. B. (2020). A complete Holocene lake sediment ancient DNA record reveals long-standing high Arctic plant diversity hotspot in northern Svalbard. Quaternary Science Reviews, 234, 106207. https://doi.org/10.1016/j.quascirev.2020.106207

Wadley, L., Backwell, L., d'Errico, F., & Sievers, C. (2020). Cooked starchy rhizomes in Africa 170 thousand years ago. Science, 367(6473), 87–91.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V., & others. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature, 544(7650), 357–361.

Willerslev, E. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Willerslev, E., & Cooper, A. (2005). Ancient DNA. Proceedings. Biological Sciences / the Royal Society, 272(1558), 3–16. https://doi.org/10.1098/rspb.2004.2813

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, 565, 399–403. https://doi.org/10.1038/s41586-021-03675-0

Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., Jacobs, Z., Dai, Q., Massilani, D., & others. (2020). Denisovan DNA in late pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science, 370(6516), 584–587.

Supplemental data

Supplementary information consisting of primary data and raw analyses are not included due to the size of tables. Access to these supplemental materials is provided through an online GitHub repository:
https://github.com/terschure/dung-dirt-dna

A.1. DNA filtering steps

Table A.1. Overview of the DNA sequence read counts and number of unique sequences per filtering step.

Filtering step	Program	Total reads	Unique sequences
Raw reads		168 787 958	
Pairwise alignment	*illuminapairedend*	168 787 958	
Assignment to samples	*ngsfilter*	137 558 235	
Removal of reads with count = 1 and < 8 bp length & merging identical reads	*obigrep & obiuniq*	130 010 462	338 885
Matching the reference database	*ecotag & obigrep*	130 000 975	338 182
Identification & removal of PCR/sequencing errors	*obiclean*	125 211 062	174 208
Removal of sequences also present in negative controls	R	68 191 867	151 204
Removal of sequences with < 97.5% ID match	R	41 461 019	1010
Removal of sequence read counts < 10 reads per PCR repeat	R	41 456 965	250
Removal of sequences occurring in < 2 PCR repeats over the entire dataset	R	36 828 760	135
Removal of sequences with < 98% ID match, without family annotation, or deemed unlikely by a taxonomic expert	R	33 259 880	129
Setting read counts to 0 for samples with < 2 positive replicates	R	33 069 116	129
Repeated removal of sequences occurring in < 2 PCR repeats over the entire dataset	R	33 022 936	128
Merging PCR replicates while averaging the read counts per sample	R	9 346 840	128

A.2. Correction of read counts

Table A.2. Calculation of the correction ratio for the calculation of the read counts per extracted mg of sediment. Correction ratio = mg extracted sediment * fraction in PCR * fraction in cleanup *(fraction in pool / dilution factor).

sample_s_replicate	depth	mg extracted	fraction in PCR	fraction in cleanup	dilution factor pool	volume for pool	fraction in pool	correction ratio
AGH03E1_s_1	-1.66	110	0.05	0.034	1	8.54	0.097	0.018416
AGH03E1_s_4	-1.66	110	0.05	0.042	1	5.78	0.066	0.015072
AGH03E1_s_5	-1.66	110	0.05	0.042	1	5.78	0.066	0.015072
AGH05E1_s_3	-1.46	120	0.05	0.033	1	12.00	0.136	0.027291
AGH05E1_s_4	-1.46	120	0.05	0.043	10	10.27	0.117	0.003045
AGH05E1_s_6	-1.46	120	0.05	0.043	10	10.27	0.117	0.003045
AGH07E1_s_4	-1.27	90	0.05	0.100	10	6.59	0.075	0.003372
AGH07E1_s_5	-1.27	90	0.05	0.042	1	5.78	0.066	0.012331
AGH08E1_s_2	-1.15	160	0.05	0.033	1	12.00	0.136	0.036388
AGH08E1_s_3	-1.15	160	0.05	0.033	1	12.00	0.136	0.036388
AGH08E1_s_4	-1.15	160	0.05	0.043	10	10.27	0.117	0.004061
AGH08E1_s_5	-1.15	160	0.05	0.043	10	10.27	0.117	0.004061
AGH08E1_s_6	-1.15	160	0.05	0.042	1	6.47	0.074	0.024542
AGH09E1_s_2	-5.55	100	0.05	0.033	1	12.00	0.136	0.022743
AGH10E1_s_3	-5.45	130	0.05	0.077	10	24.32	0.277	0.013825
AGH10E1_s_4	-5.45	130	0.05	0.042	1	5.78	0.066	0.017812
AGH10E1_s_6	-5.45	130	0.05	0.125	10	3.17	0.036	0.002933
AGH12E1_s_1	-5.06	110	0.05	0.077	1	1.59	0.018	0.007629
AGH12E1_s_2	-5.06	110	0.05	0.033	1	12.00	0.136	0.025017
AGH12E1_s_3	-5.06	110	0.05	0.091	10	5.95	0.068	0.003384
AGH12E1_s_4	-5.06	110	0.05	0.043	10	10.27	0.117	0.002792
AGH12E1_s_5	-5.06	110	0.05	0.043	10	10.27	0.117	0.002792
AGH12E1_s_6	-5.06	110	0.05	0.042	1	6.47	0.074	0.016873
AGH14E1_s_2	-4.79	100	0.05	0.077	10	24.32	0.277	0.010635
AGH14E1_s_3	-4.79	100	0.05	0.077	10	24.32	0.277	0.010635
AGH14E1_s_4	-4.79	100	0.05	0.125	10	3.17	0.036	0.002256
AGH14E1_s_5	-4.79	100	0.05	0.125	10	3.17	0.036	0.002256
AGH14E1_s_6	-4.79	100	0.05	0.100	10	6.59	0.075	0.003747
AGH17E1_s_2	-4.7	90	0.05	0.077	10	24.32	0.277	0.009571
AGH17E1_s_3	-4.7	90	0.05	0.077	10	24.32	0.277	0.009571
AGH17E1_s_4	-4.7	90	0.05	0.042	1	5.78	0.066	0.012331
AGH17E1_s_5	-4.7	90	0.05	0.100	10	6.59	0.075	0.003372
AGH17E1_s_6	-4.7	90	0.05	0.042	1	5.78	0.066	0.012331
AGH19E1_s_1	-4.57	110	0.05	0.091	10	5.95	0.068	0.003384
AGH19E1_s_2	-4.57	110	0.05	0.033	1	12.00	0.136	0.025017
AGH19E1_s_3	-4.57	110	0.05	0.091	10	5.95	0.068	0.003384
AGH19E1_s_4	-4.57	110	0.05	0.043	10	10.27	0.117	0.002792
AGH19E1_s_5	-4.57	110	0.05	0.143	10	3.26	0.037	0.002911
AGH19E1_s_6	-4.57	110	0.05	0.043	10	10.27	0.117	0.002792
AGH21E1_s_1	-4.45	150	0.05	0.091	10	5.95	0.068	0.004614
AGH21E1_s_2	-4.45	150	0.05	0.033	1	12.00	0.136	0.034114
AGH21E1_s_3	-4.45	150	0.05	0.091	10	5.95	0.068	0.004614
AGH21E1_s_4	-4.45	150	0.05	0.043	10	10.27	0.117	0.003807
AGH21E1_s_5	-4.45	150	0.05	0.043	10	10.27	0.117	0.003807
AGH21E1_s_6	-4.45	150	0.05	0.043	10	10.27	0.117	0.003807
AGH23E1_s_2	-4.26	110	0.05	0.077	10	24.32	0.277	0.011698
AGH23E1_s_3	-4.26	110	0.05	0.034	1	8.54	0.097	0.018416
AGH23E1_s_4	-4.26	110	0.05	0.100	10	6.59	0.075	0.004122
AGH23E1_s_5	-4.26	110	0.05	0.100	10	6.59	0.075	0.004122
AGH23E1_s_6	-4.26	110	0.05	0.042	1	5.78	0.066	0.015072
AGH25E1_s_1	-4.01	130	0.05	0.077	1	1.59	0.018	0.009017
AGH25E1_s_2	-4.01	130	0.05	0.077	1	1.59	0.018	0.009017
AGH25E1_s_3	-4.01	130	0.05	0.077	1	1.59	0.018	0.009017
AGH25E1_s_4	-4.01	130	0.05	0.143	10	3.26	0.037	0.003440

AGH25E1_s_5	-4.01	130	0.05	0.043	10	10.27	0.117	0.003299
AGH25E1_s_6	-4.01	130	0.05	0.043	10	10.27	0.117	0.003299
AGH27E1_s_2	-3.76	130	0.05	0.077	10	24.32	0.277	0.013825
AGH27E1_s_3	-3.76	130	0.05	0.077	10	24.32	0.277	0.013825
AGH27E1_s_4	-3.76	130	0.05	0.042	1	5.78	0.066	0.017812
AGH27E1_s_5	-3.76	130	0.05	0.100	10	6.59	0.075	0.004871
AGH27E1_s_6	-3.76	130	0.05	0.042	1	5.78	0.066	0.017812
AGH28E1_s_2	-3.63	190	0.05	0.077	1	1.59	0.018	0.013178
AGH28E1_s_3	-3.63	190	0.05	0.077	1	1.59	0.018	0.013178
AGH28E1_s_4	-3.63	190	0.05	0.042	1	6.47	0.074	0.029144
AGH28E1_s_5	-3.63	190	0.05	0.043	10	10.27	0.117	0.004822
AGH30E1_s_1	-3.42	130	0.05	0.033	1	12.00	0.136	0.029566
AGH30E1_s_2	-3.42	130	0.05	0.091	10	5.95	0.068	0.003999
AGH30E1_s_3	-3.42	130	0.05	0.077	1	1.59	0.018	0.009017
AGH30E1_s_4	-3.42	130	0.05	0.043	10	10.27	0.117	0.003299
AGH30E1_s_5	-3.42	130	0.05	0.143	10	3.26	0.037	0.003440
AGH30E1_s_6	-3.42	130	0.05	0.143	10	3.26	0.037	0.003440
AGH32E1_s_2	-3.25	180	0.05	0.034	1	8.54	0.097	0.030135
AGH32E1_s_3	-3.25	180	0.05	0.077	10	24.32	0.277	0.019143
AGH32E1_s_5	-3.25	180	0.05	0.125	10	3.17	0.036	0.004061
AGH34E1_s_1	-3.05	100	0.05	0.091	10	5.95	0.068	0.003076
AGH34E1_s_2	-3.05	100	0.05	0.077	1	1.59	0.018	0.006936
AGH34E1_s_3	-3.05	100	0.05	0.077	1	1.59	0.018	0.006936
AGH34E1_s_4	-3.05	100	0.05	0.143	10	3.26	0.037	0.002646
AGH34E1_s_6	-3.05	100	0.05	0.042	1	6.47	0.074	0.015339
AGH36E1_s_2	-2.91	90	0.05	0.034	1	8.54	0.097	0.015068
AGH36E1_s_3	-2.91	90	0.05	0.034	1	8.54	0.097	0.015068
AGH36E1_s_4	-2.91	90	0.05	0.125	10	3.17	0.036	0.002030
AGH36E1_s_5	-2.91	90	0.05	0.125	10	3.17	0.036	0.002030
AGH38E1_s_1	-2.76	100	0.05	0.091	10	5.95	0.068	0.003076
AGH38E1_s_2	-2.76	100	0.05	0.033	1	12.00	0.136	0.022743
AGH38E1_s_3	-2.76	100	0.05	0.033	1	12.00	0.136	0.022743
AGH38E1_s_4	-2.76	100	0.05	0.042	1	6.47	0.074	0.015339
AGH38E1_s_5	-2.76	100	0.05	0.042	1	6.47	0.074	0.015339
AGH38E1_s_6	-2.76	100	0.05	0.043	10	10.27	0.117	0.002538
AGH40E1_s_4	-2.5	110	0.05	0.100	10	6.59	0.075	0.004122
AGH40E1_s_5	-2.5	110	0.05	0.125	10	3.17	0.036	0.002481
AGH40E1_s_6	-2.5	110	0.05	0.042	1	5.78	0.066	0.015072
AGH41E1_s_1	-2.39	140	0.05	0.091	10	5.95	0.068	0.004307
AGH41E1_s_2	-2.39	140	0.05	0.077	1	1.59	0.018	0.009710
AGH41E1_s_3	-2.39	140	0.05	0.077	1	1.59	0.018	0.009710
AGH41E1_s_4	-2.39	140	0.05	0.043	10	10.27	0.117	0.003553
AGH41E1_s_5	-2.39	140	0.05	0.143	10	3.26	0.037	0.003705
AGH41E1_s_6	-2.39	140	0.05	0.043	10	10.27	0.117	0.003553
AGH43E1_s_1	-2.13	130	0.05	0.033	1	12.00	0.136	0.029566
AGH43E1_s_6	-2.13	130	0.05	0.143	10	3.26	0.037	0.003440
AGH44E1_s_4	-2.05	90	0.05	0.100	10	6.59	0.075	0.003372

A.3. Plant stratigraphic plots

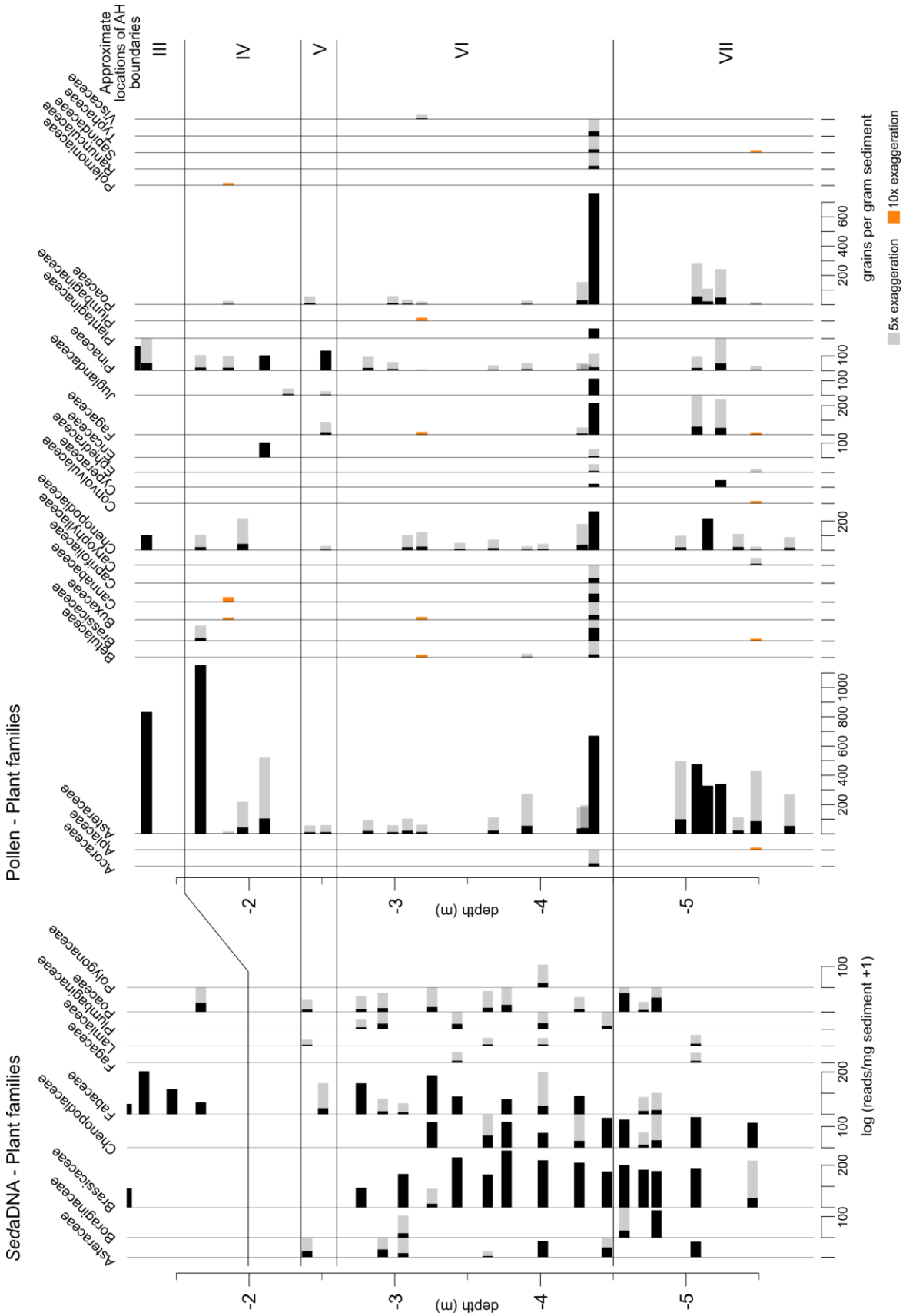


Figure A.3.1. *Seda*DNA concentrations (left; in log (reads per mg sediment +1)) and pollen concentrations (right; in grains per gram sediment) per plant family.

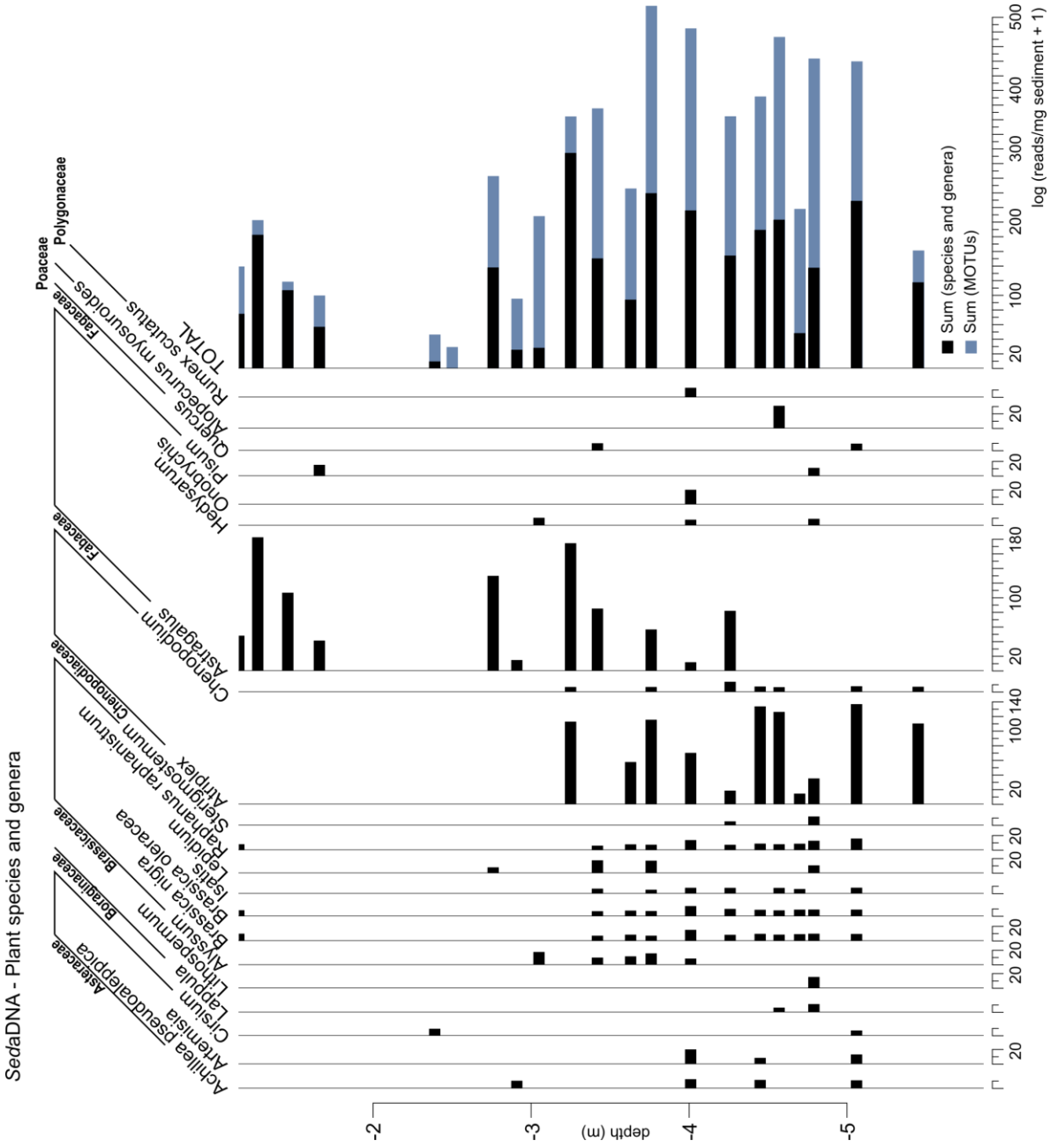


Figure A.3.2. *Sed*aDNA concentrations (log (reads per mg sediment +1)) per plant species or genus.

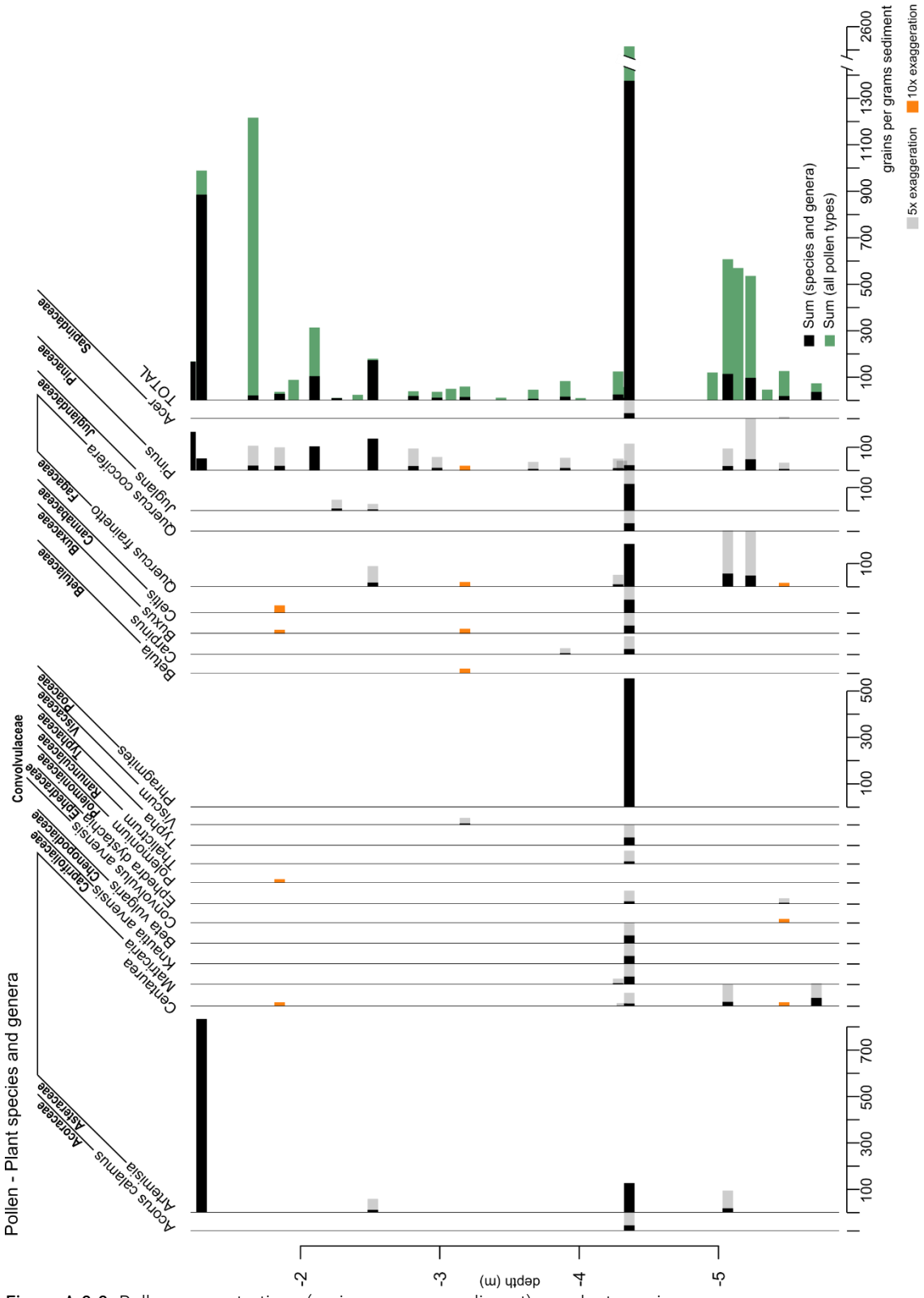
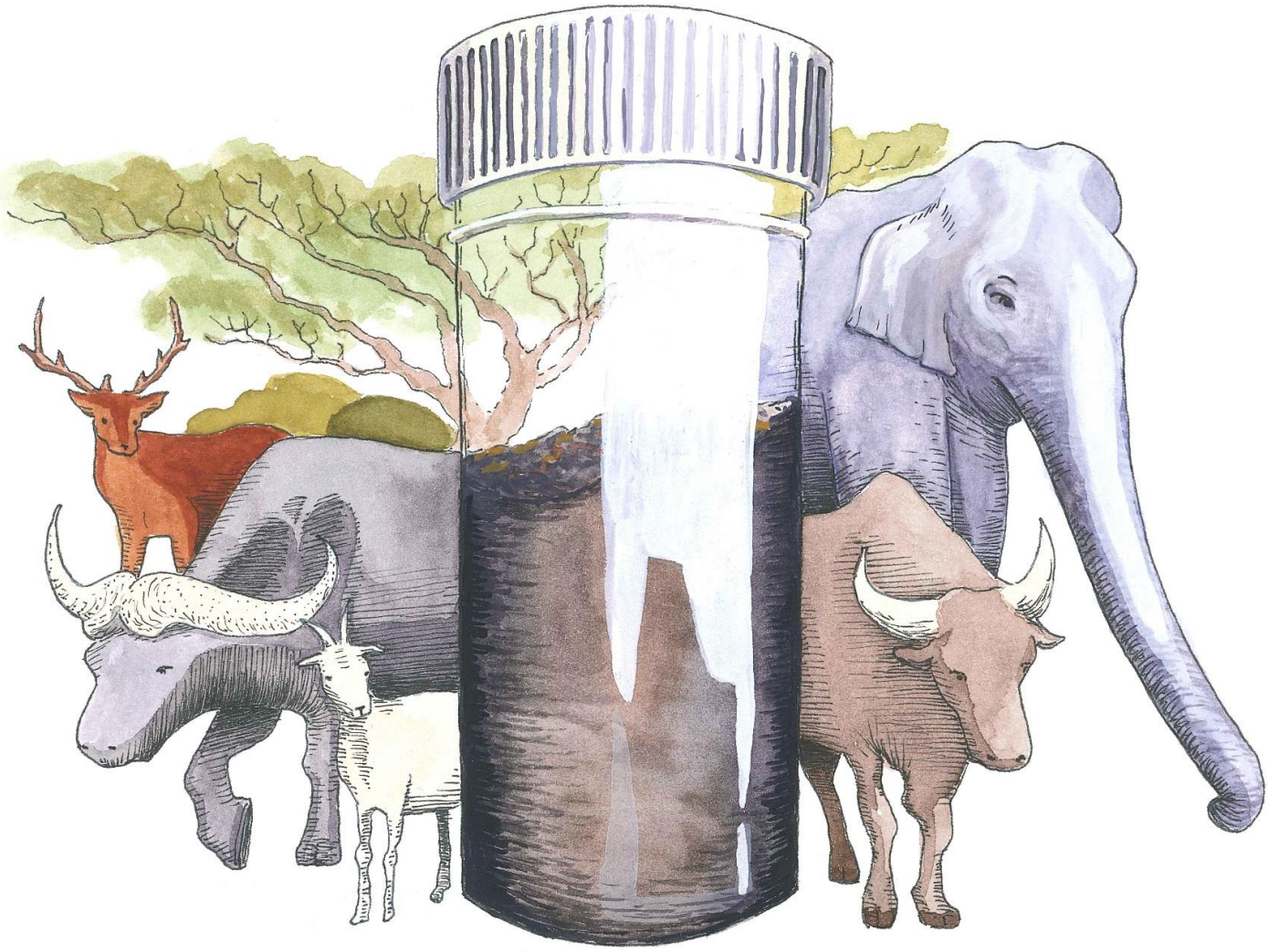


Figure A.3.3. Pollen concentrations (grains per gram sediment) per plant species or genus.

A.4. Statistical differences in plant richness between Archaeological Horizons

Table A.4. Results from the non-parametric Mann-Whitney-Wilcoxon test to identify differences in plant richness (*sed*aDNA MOTUs or pollen taxa) among archaeological horizons (AH VII, VI, V, IV, III). *P*-values were adjusted using the Holm-method.

proxy	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
DNA	III	IV	5	2	9	0.16	0.8	ns
DNA	III	V	5	2	8	0.381	0.884	ns
DNA	III	VI	5	10	3	0.008	0.084	ns
DNA	III	VII	5	6	4.5	0.067	0.471	ns
DNA	IV	V	2	2	0	0.221	0.884	ns
DNA	IV	VI	2	10	0	0.041	0.365	ns
DNA	IV	VII	2	6	1	0.124	0.744	ns
DNA	V	VI	2	10	0	0.041	0.365	ns
DNA	V	VII	2	6	2	0.286	0.884	ns
DNA	VI	VII	10	6	34.5	0.664	0.884	ns
DNA	III	IV-V	5	4	17	0.105	0.105	ns
DNA	III	VI-VII	5	16	7.5	0.008	0.021	*
DNA	IV-V	VI-VII	4	16	3	0.007	0.021	*
pollen	III	IV	4	8	12	0.536	1	ns
pollen	III	V	4	3	4	0.582	1	ns
pollen	III	VI	4	13	11	0.097	0.969	ns
pollen	III	VII	4	10	11	0.217	1	ns
pollen	IV	V	8	3	11	0.917	1	ns
pollen	IV	VI	8	13	34.5	0.214	1	ns
pollen	IV	VII	8	10	31	0.44	1	ns
pollen	V	VI	3	13	14	0.499	1	ns
pollen	V	VII	3	10	12	0.667	1	ns
pollen	VI	VII	13	10	71	0.731	1	ns
pollen	III	IV-V	4	11	16	0.457	0.457	ns
pollen	III	VI-VII	4	23	22	0.104	0.312	ns
pollen	IV-V	VI-VII	11	23	91.5	0.198	0.396	ns



Discussion

Discussion

In the introduction, I mentioned how the current biodiversity crisis, as well as impacts of anthropogenic climate change on the biosphere, highlight the urgency of large-scale biodiversity monitoring for conservation, as well as the need to understand current and past biological interactions and processes in light of environmental change. I also introduced eDNA metabarcoding as a method that allows for the simultaneous identification of many organisms in an environmental sample and so enables biodiversity monitoring without the need to sight or sample the actual organisms. Chapter 1 provided further background to certain aspects of the methodology and the studies in chapters 2 to 5 illustrated how eDNA metabarcoding can provide unprecedented insight into past and present biodiversity and human-environment interactions. Here I will summarize and discuss the results of these chapters and reflect on their implications and remaining challenges of and opportunities for eDNA methods for reconstructing past and present biodiversity. The discussion is split up into three sections. 'Dung: reconstructing diets' covers chapter 2 and 3, with a focus on diets and further ecological inferences. 'Dirt: reconstructing human-environment interactions' covers chapter 4 and 5, assessing human and environmental drivers of past vegetation change, as well as prehistoric plant use using cave sediments. Finally, 'DNA: a window to things past' points towards the future of eDNA methods by offering concluding remarks.

Dung: reconstructing diets

Environmental DNA metabarcoding of modern and ancient faecal samples allowed non-invasive monitoring of wildlife in chapter 2, and unprecedented insight into the diets of extinct megafauna in chapter 3, as well as multiple further inferences based on the composition of the reconstructed diets. This included the analysis of dietary niches and their partitioning for biomonitoring purposes (chapter 2), as well as the reconstruction of ancient habitats (chapter 3).

Dietary niche partitioning in a wildlife sanctuary (Chapter 2)

Analysis of modern faecal samples from herbivores living in the Malai Mahadeshwara Wildlife Sanctuary enabled dietary niche partitioning analysis and revealed dietary overlap, particularly between cattle and several wild herbivores including sambar deer, Indian hare and Asian elephant (ter Schure, Pillai, et al., 2021). Using relative read abundances obtained through eDNA metabarcoding as proposed by Deagle et al. (2019), the reconstructed diets matched well with feeding guild assignments.

Dietary niche width was estimated by calculating the plant diversity in the samples, which was found to be lower for specialised feeders (bonnet macaque,

Macaca radiata; cattle, *Bos taurus indicus*) compared to generalists (domestic goat, *Capra hircus*; water buffalo, *Bubalus bubalis*; sambar deer, *Rusa unicolor*). The exception being Asian elephants (*Elephas maximus*), considered mixed feeders but found to have a narrow niche, based on a majority of grasses in their faeces. This was likely due to seasonal effects as our samples were all from the wet season and Asian elephants have been suggested to shift between feeding on more woody plants in the dry season, and more grasses in the wet season (Ahrestani et al., 2012; Baskaran et al., 2010; Sukumar, 2006). The narrow dietary niches of bonnet macaque and Asian elephant could make them vulnerable to environmental changes (Clavel et al., 2011; Devictor et al., 2010).

Dietary niche partitioning was visualised using non-metric multidimensional scaling, and estimated using two complementary metrics: Bray-Curtis dissimilarity index, and Pianka's niche overlap index. Both metrics indicated substantial dietary niche overlap between multiple pairs of the studied domestic and wild herbivores. Particularly the diets of cattle and domestic goats overlapped considerably with sambar deer, Indian hare and Asian elephant. When food resources are limited, the overlap in dietary niches could result in resource competition. Furthermore, we found indications of direct human impacts on wildlife, including the collection of plants from the wild for food, and changes in land-use, with reports of wild herbivores raiding croplands.

Overall, chapter 3 illustrates the potential of eDNA metabarcoding for reconstructing the diets of herbivores and estimating dietary niche overlap. This allowed the identification of several species that could be vulnerable to environmental changes, particularly bonnet macaque and Asian elephant, pointed out the need for cross-seasonal sampling, and provided a starting point for continued monitoring of the herbivores living in the Malai Mahadeshwara Wildlife Sanctuary.

Diets and habitats of extinct and extant megafauna (Chapter 3)

Dietary reconstructions using eDNA metabarcoding of living herbivores can generally be compared to information of known diets. However, this was certainly not the case for the woolly mammoth, or other extinct megafauna that were the topic of chapter 3. Here, a multiproxy approach was employed to reconstruct the diets of woolly mammoth, horse and steppe bison, combining visual analysis of the plant remains, i.e. macrofossil and pollen analysis, with DNA metabarcoding using several sets of primers (Polling et al., 2021). The applied approach was validated using Holocene samples from the extant caribou, with known diets and habitats.

Comparisons between the plant taxa recovered by pollen, macrofossils and DNA analyses indicated that recovery of plant taxa from the faecal samples was highest using DNA analyses. This can be explained by the combination of metabarcoding primers used (i.e. *trn*L, nrITS1, and nrITS2), each increasing the number of detected taxa. Another likely explanation is that plant DNA in faeces can originate from most plant parts and is independent of the season when plants

carry seed, fruit or pollen. We accordingly found high overlap between results from DNA and macrofossil analyses, while both datasets showed limited overlap with the results from pollen analysis. Pollen analysis was likely influenced by accidental ingestion (i.e. airborne pollen potentially sticking to ingested vegetation), ingestion of flowering inflorescences, as well as plants that are known high pollen producers (e.g. Pinaceae and other wind-pollinated species; Aario, 1940; Birks & Bjune, 2010; van Geel et al., 2014).

Within the used DNA metabarcoding primers, the nrITS primers obtained higher taxonomic resolutions compared to the trnL primers for several plant families, including Asteraceae, Cyperaceae, and Poaceae, as well as for bryophytes. However, the data obtained using trnL also includes many plant taxa that were not found using nrITS. These differences could be caused by differences in the available reference sequences for taxonomic identification, with more complete reference libraries available for *trn*L, and many incomplete reference sequences for nrITS. Moreover, the shorter and more stable trnL P6 loop (10-143 bp; Taberlet et al., 2007) can have a limited obtainable taxonomic resolution compared to the longer nrITS region targeted here (~300-500 bp; Cheng et al., 2016), but is thereby also less influenced by degradation processes damaging and fragmenting the DNA.

Analysis of faecal samples from caribou indicated an abundance of arctic alpine shrubs and, based on comparison with known diets of modern caribou (e.g. Denryter et al., 2017), indicates an accurate reconstruction, validating the used approach. The diet of prehistoric horse was found to be variable across analysed samples, possibly indicating a more mixed diet than known for modern horses, as also found in previous studies (Kaczensky et al., 2017; MacFadden et al., 1999) and possibly due to seasonal effects with decreased access to grasses in winter. The diet of steppe bison was similarly found to be more mixed compared to modern bison (a close relative of the extinct steppe bison), with an abundance of grass, but also forbs and shrubs as important components. Results for woolly mammoth diets indicated a different composition for each of the analysed samples, with a dominance of grasses in some, presence of mosses and forbs, shrubs or even willow (*Salix* sp.) in others. This suggests very different last meals across the studied woolly mammoths.

Based on the ecology of the obtained dietary items, inferences could be made on the habitat of the megafauna under study. Holocene steppe bison and horse were both likely foraging in marshy wetland environments, based on the presence of typical wetland species in these samples. For the woolly mammoth the diversity in diets also indicated a variety of habitats, including wetland, wet meadow, gravelly slopes, saline meadow, and steppe, suggesting a mosaic of habitats in the Pleistocene 'mammoth steppe'.

Moving beyond diet items

Fundamental to establishing the (near) true diet of a species is the estimation of the relative biomass of the individual diet items. In chapter 3, comparison of traditional proxies for diet analysis with results from eDNA metabarcoding revealed that analysis of pollen was likely influenced by accidental ingestion and ingestion of flowering inflorescences, with macrofossil and eDNA results presenting similar and presumably more accurate reconstructions. Chapter 3 further found an underrepresentation of forbs in macrofossil and pollen results as compared to results obtained through DNA analysis, likely resulting from the effects of digestion. This has also been reported in previous studies (Kosintsev et al., 2012; Willerslev et al., 2014) and suggests that eDNA metabarcoding can provide more reliable dietary reconstructions compared to visual analysis of faecal samples. However, DNA methods cannot distinguish between different plant parts, which is particularly of interest when aiming to assess potential dietary competition (as in Chapter 2) and other limitations include the degradation of DNA through digestion as well as biases introduced during the processing of the samples in the lab (Pompanon et al., 2012). This can result in false identifications as well as false non-detections, and the choice of primers can be an important factor in this, as evident from the differences in results obtained using *trn*L and nrITS primers in Chapter 3.

Sequence read counts as obtained through eDNA metabarcoding first need to be analysed and filtered, removing bias and contamination, to obtain sequence data that can be interpreted. However, some debate exists on how to proceed from there. Counts of the obtained DNA sequences have the potential to quantify the DNA present in the sample, but biological and technical biases can affect these counts (Elbrecht & Leese, 2015). The conversion of read counts to presence/absence of diet items presents a more conservative approach, but limits the ecological interpretation, and dietary reconstructions based on this presence/absence data can be prone to overestimate the importance of food items consumed in small quantities (Deagle et al., 2019). In chapter 2, we presented the results of dietary analysis and niche partitioning analysis based on relative read abundances (i.e. the proportional representation of each plant sequence in a sample) and also performed these analyses on presence/absence data (see the supplemental information for chapter 2). We found similar results for both datasets with slight differences in the overlap/differentiation between herbivore species, possibly a result of the presence of rare diet items weighing higher in presence/absence data (Deagle et al., 2019). Also the choice of metric for estimating niche overlap was found to be of influence, illustrated by discrepancies between Pianka's niche overlap index and results from the NMDS based on Bray-Curtis dissimilarities. This was similarly a likely result of rare diet items in the data, as Pianka's niche overlap index is prone to over- as well as underestimations particularly with datasets containing many zeros (Rödder & Engler, 2011), suggesting that estimates based on Bray-Curtis dissimilarities are more accurate.

Results from chapters 2 and 3 illustrate that eDNA metabarcoding is a useful tool for the reconstruction of herbivore diet from modern as well as ancient faecal samples. Agreement of the obtained reconstructions with known diets in both studies indicates the accuracy of this approach. Ecological inferences from diet reconstructions heavily rely on the choice of methods as discussed in this section. The studies described in chapters 2 and 3 show that diet items recovered from analysis of faecal samples can be used for interpretations beyond the diet of the herbivores under study, with indications for potential competition in chapter 2, and inferences about the habitats of extinct megafauna in chapter 3.

Dirt: reconstructing human-environment interactions

Analysis of ancient DNA from sediments uncovered human-induced vegetation change in chapter 4, as well as the potential of plant *sed*aDNA metabarcoding for unearthing prehistoric human plant-use in chapter 5. Both of these chapters included a large archaeological component and illustrate the importance of interdisciplinary collaborations for integrating various types of evidence, thus making a stronger case for reconstructions of past human-environment interactions.

Environmental drivers of vegetation change (Chapter 4)

The 10,000-year long lake sediment record of Lake Ljøgottjern revealed consistent vegetation signals from pollen and *sed*aDNA indicating vegetation changes related to climate as well as cultivation, and matching archaeological evidence (ter Schure, Bajard, et al., 2021). To untangle the different environmental changes related to vegetation change, an interdisciplinary approach was taken, combining inferences from plant and mammal *sed*aDNA metabarcoding, pollen analysis, geochemical analysis, climate data, as well as archaeological evidence of local human settlement and regional human population dynamics.

Comparisons between pollen and *sed*aDNA detection of plant taxa showed higher recovery of plant taxa by sedaDNA, especially of forbs. Furthermore, several cultivated taxa that could not be distinguished from their pollen, could be identified by *sed*aDNA, such as *Hordeum, Avena, Cannabis*, and *Humulus*. Non-metric multidimensional scaling (NMDS) as well as stratigraphically constrained cluster analysis (CONISS) distinguished between time periods with different plant communities, and identified significant changes around 300 cal BCE and 450 cal CE. These changes were associated with abrupt changes in the presence of plant taxa, particularly the appearance of anthropochores (plant taxa distributed by humans including many crops) and apophytes (plant taxa growing in disturbed lands) in the *sed*aDNA record, and increased pollen concentrations of these plant groups.

Evidence from cultivated crops as well as mammal DNA in the lake sediment record allowed the reconstruction of a timeline of cultivation and pastoralism around the lake. The pollen record indicated cultivation of crops (e.g. rye and wheat) during the Bronze Age (ca. 1800-500 BCE), while evidence for cultivation in *sed*aDNA remained absent until the Early Iron Age (ca. 500 BCE-550 CE) when DNA from several food crops, as well as pastoral animals was recorded. *Sed*aDNA in lake sediments originates from the lake catchment area (60-230 m from the edge of Lake Ljøgottjern), whereas the pollen present in the Lake Ljøgottjern sediment record have a likely relevant source area of > 900 m distance from the lake centre. Archaeological evidence indicates a Bronze Age farm located at approximately 400 m distance from the lake, and one or two Iron Age farms within the lake catchment area. The noted deviations between the pollen and *sed*aDNA records thereby reflect the distance to farms as established through archaeological analysis.

Integrated statistical analysis of the analysed environmental proxies with the vegetation data allowed assessment of the relative role of anthropogenic and environmental factors on the variation in plant communities. Distance-based redundancy analysis as well as variation partitioning analysis pointed towards a primary effect of temperature on the vegetation dynamics from the base of the lake sediment core (ca. 8000 BCE, Mesolithic) up to the Early Iron Age (ca. 500 BCE-550 CE) and associated erosion and organic matter content. For the period after ca. 300 BCE to the present, a significant correlation was also found for estimates of increased human population densities, with evidence of increased agricultural activities. These results also match the general intensification of agriculture in southeastern Norway in the first centuries of the Early Iron Age, as evident from other pollen records in this area. Combined analysis of pollen, *sed*aDNA, geochemistry, climate data and archaeological evidence in chapter 4 has thereby allowed not only the distinction between establishment of individual farms around Lake Ljøgottjern, but also the distinction between past drivers of vegetation change in the region.

Assessing prehistoric plant use using sedaDNA from caves (Chapter 5)

*Sed*aDNA and pollen analysis of the Upper Palaeolithic sediment sequence (~39,000-24,000 cal BP) from Aghitu-3 Cave revealed relatively good recovery of plant DNA, as well as many plant taxa that could have been used by prehistoric humans during their stays at the cave.

Overall pollen concentrations were low with some variation, while presence of *sed*aDNA was more consistent, particularly in deeper sediments (~2.7-5.2 m depth, corresponding to Archaeological Horizon, AH, VII and VI) as well as the top section of the sequence (the upper part of AH III), where relatively high concentrations of DNA were recovered. As DNA degradation takes place over time, a trend of lower recovery of DNA over depth is common for lake sediments (Parducci et al., 2017). However, lake sediments generally have a more or less

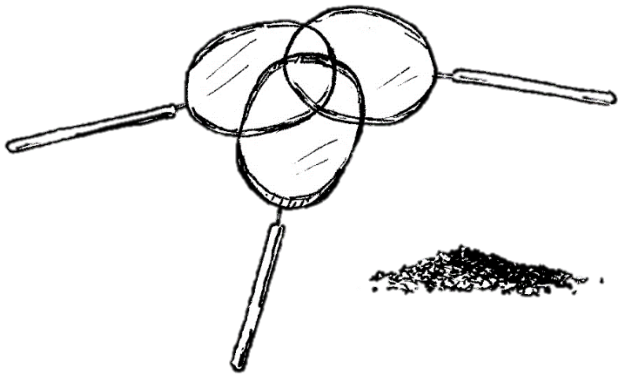
constant input of plant remains, whereas the input of plant remains in cave sediments is likely more stochastic and dependent on concentration of plant remains by humans or animals. No obvious evidence of changes in the sediment composition were established that could explain the relatively high abundance of *sed*aDNA in AH VII, VI and the upper part of AH III. Pollen concentrations were likely low as a result of limited wind-transport of plant remains into the cave, as well as being less likely to have been transported into the cave through human agency.

Comparison of the abundance patterns of *sed*aDNA with previous finds suggested an association with human occupation, as particularly high densities of archaeological finds (mostly lithics), as well as macrofaunal remains (several with anthropogenic modifications), combustion features and micro-charcoal concentrations were reported for AH VII, VI and III (Bertacchi et al., 2021; Kandel et al., 2017). AH IV and V, on the other hand, were reported to yield few finds, contrasted by high numbers of remains from small mammals (lagomorphs and rodents), associated with a reduced presence of humans at the cave. Statistical analysis and visual representation of the plant assemblages and plant richness (NMDS combined with perMANOVA analysis and Mann-Whitney-Wilcoxon tests) further pointed towards a higher correspondence to the pre-defined archaeological horizons of sedaDNA-based plant assemblages compared to pollen-based reconstructions. Overall, these findings suggest an association of *sed*aDNA plant assemblages with human occupation.

Assessment of the potential use of the recovered plants indicated that the majority could be used for food, flavouring, medicine, and/or other technical purposes such as basketry, tinder, string and dye. Previous archaeological finds at Aghitu-3 Cave include a bone eyed needle recovered from AH III (Kandel et al., 2017), suggesting manufacturing of clothing. 30,000-year old dyed textile fibres recovered from a cave in the neighbouring country of Georgia (Kvavadze et al., 2009) further illustrates the use of plants for fibre and dye by prehistoric humans. The integration of evidence from *sed*aDNA, pollen and other finds at Aghitu-3 Cave allowed inferences about potential plant use of prehistoric humans in the Armenian Highlands.

On the importance of triangulation and integration

Reconstructions of palaeoenvironments using *sed*aDNA can be challenging due to the prolonged exposure of the DNA to degradation processes. These are related to the properties of the sediments as well as the environment surrounding them, including temperature, pH, oxygen levels, water content, and mineral content (Giguet-Covex et al., 2019; Torti et al., 2015). Chapter 1 explained how the degraded nature of *sed*aDNA influences research design choices of studies employing these methods, including the selection of sediments that provide good conditions for DNA preservation. Permafrost, as a result of its anaerobic conditions, neutral pH and subzero temperatures, has allowed the amplification of



*sed*aDNA of ca. 400 thousand years old (Willerslev, 2003; Willerslev et al., 2004). Many palaeoecological studies that employ *sed*aDNA metabarcoding use lake sediments as they are waterlogged and anaerobic, and represent a continuous sequence of environmental change integrating biotic and abiotic information across the lake catchment area (see Capo et al., 2021, for a recent comprehensive review). In chapter 4, this approach was employed for the reconstruction of the vegetation changes around Lake Ljøgottjern during the Holocene, whereas in chapter 5, it was the absence of water that allowed detection of *sed*aDNA from the Pleistocene cave sediments at Aghitu-3.

Combined analysis of multiple proxies in chapters 3 through 5 have shown how different palaeovegetation proxies can come from different sources and are differentially affected by taphonomic processes, and as a result detect different plant taxa, thereby providing complementary records. Pollen records are limited to flowering plants, and can be biased as a result of the influence of high pollen-producing plants and pollen, especially from wind-pollinated plants, can come from a wide region (Birks & Bjune, 2010), while both macrofossils and *sed*aDNA generally originate from local sources (Jørgensen et al., 2012; Parducci et al., 2017). On the other hand, macrofossils can prove too degraded to identify (chapter 3), and *sed*aDNA methods are similarly affected by degradation of the remains, but can be further hindered by contamination and technological biases. Each of these proxies therefore has its own limitations and combined analysis is advisable.

Comparison of the *sed*aDNA record from Lake Ljøgottjern (chapter 4) with a pollen-based reconstruction from the same sediment record showed that the source area for pollen and *sed*aDNA were different, with pollen coming from beyond the lake catchment, and the archaeological evidence from farms around the lake provided the context to explain the differences in these records, with the establishment of the farms matching the detection of cultivated plants. Integration of evidence from multiple lines of evidence also proved valuable for the assessment of *sed*aDNA as a tool for reconstructing human plant use in chapter 5, where we found that *sed*aDNA more accurately reflected periods of human occupation than pollen.

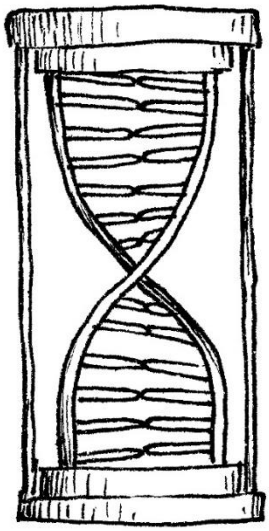
The study of human-environment interactions requires the integration of environmental records with evidence of human activities, which can be a complex process. In interdisciplinary collaborations such as those presented in chapter 4 and 5, and to some extent in chapter 2, good communication is very important. Topics can become increasingly complex as new methods are developed and

misunderstandings are likely to occur. Environmental DNA metabarcoding, combined with statistical approaches as well as clear visualisation of the obtained results, such as those employed particularly in chapter 2 and 4, provide a way forward, untangling the complexity of biodiversity in all its dimensions.

DNA: a window to things past

The current biodiversity crisis and the impacts of anthropogenic climate change on the biosphere highlight the urgency of large-scale biodiversity monitoring for conservation, as well as the need to understand both current and past biological interactions and processes in light of environmental change. This thesis aimed to untangle some of these complex interactions and processes by applying eDNA metabarcoding analysis to ancient and modern faecal and sediment samples.

Trace DNA in faecal and sediment samples allow the detection of species and biodiversity monitoring without the need to sight or sample the actual organisms. Current advances in metagenomics, sequencing all of the DNA present in a sample, prove promising (Pedersen et al., 2016), as this can overcome the bias introduced with PCR amplification in DNA metabarcoding (Bellemain et al., 2010). However, this approach can be costly and requires a high sequencing depth. Developments towards target capture can provide the answer. This approach, using baits to enrich the concentration of the taxa of interest in a shotgun approach, has recently been applied to *sed*aDNA samples (Murchie et al., 2020; Schulte et al., 2020; Seeber et al., 2019). However, the low cost and easy to standardise eDNA metabarcoding approach will remain potent. Chapter 2 proves the usefulness of eDNA metabarcoding, as a tool for the simultaneous identification of many organisms in an environmental sample, for monitoring current endangered species. Application of these methods to palaeo records (chapters 3-5) can provide a window to the biological past, including biological interactions, past effects of climate change and human activities.



References

Aario, L. (1940). Waldgrenzen und subrezenten pollen-spektren in Petsamo Lappland. Annales Academiae Scientiarum Fenniae, Series A, 54(8), 1–120.

Ahrestani, F. S., Heitkönig, I. M. A., & Prins, H. H. T. (2012). Diet and habitat-niche relationships within an assemblage of large herbivores in a seasonal tropical forest. Journal of Tropical Ecology, 28(4), 385–394. https://doi.org/10.1017/S0266467412000302

Baskaran, N., Balasubramanian, M., Swaminathan, S., & Desai, A. A. (2010). Feeding ecology of the asian elephant elephas maximus linnaeus in the nilgiri biosphere reserve, southern india. Journal of the Bombay Natural History Society, 107(1), 3.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. BMC Microbiology, 10, 189. https://doi.org/10.1186/1471-2180-10-189

Bertacchi, A., Gasparyan, B., Gruwier, B., Rivals, F., & Kandel, A. W. (2021). Upper Paleolithic animal exploitation in the Armenian Highlands: The zooarchaeology of Aghitu-3 Cave. Quaternary International, 587–588, 400–414. https://doi.org/10.1016/j.quaint.2020.04.029

Birks, H. H., & Bjune, A. E. (2010). Can we detect a west Norwegian tree line from modern samples of plant remains and pollen? Results from the DOORMAT project. Vegetation History and Archaeobotany, 19(4), 325–340.

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A. G., Clarke, C. L., Crump, S. E., Debroas, D., Englund, G., Ficetola, G. F., Garner, R. E., … Parducci, L. (2021). Lake Sedimentary DNA Research on Past Terrestrial and Aquatic Biodiversity: Overview and Recommendations. Quaternary, 4(1), 6. https://doi.org/10.3390/quat4010006

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. Molecular Ecology Resources, 16(1), 138–149.

Clavel, J., Julliard, R., & Devictor, V. (2011). Worldwide decline of specialist species: Toward a global functional homogenization? Frontiers in Ecology and the Environment, 9(4), 222–228.

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? Molecular Ecology, 28(2), 391–406. https://doi.org/10.1111/mec.14734

Denryter, K. A., Cook, R. C., Cook, J. G., & Parker, K. L. (2017). Straight from the caribou's (Rangifer tarandus) mouth: Detailed observations of tame caribou reveal new insights into summer–autumn diets. Canadian Journal of Zoology, 95(2), 81–94.

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., Venail, P., Villéger, S., & Mouquet, N. (2010). Defining and measuring ecological specialization. Journal of Applied Ecology, 47(1), 15–25. https://doi.org/10.1111/j.1365-2664.2009.01744.x

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—Sequence relationships with an innovative metabarcoding protocol. PloS One, 10(7), e0130324.

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A. L., David, F., Taberlet, P., Brisset, E., Guiter, F., Sinet, R., & Arnaud, F. (2019). New insights on lake sediment DNA from the catchment: Importance of taphonomic and analytical issues on the record quality. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-50339-1

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N. H., Andersen, K., Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Molecular Ecology, 21(8), 1989–2003. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Kaczensky, P., Šturm, M. B., Sablin, M. V., Voigt, C. C., Smith, S., Ganbaatar, O., Balint, B., Walzer, C., &

Spasskaya, N. N. (2017). Stable isotopes reveal diet shift from pre-extinction to reintroduced Przewalski's horses. Scientific Reports, 7(1), 1–9.

Kandel, A. W., Gasparyan, B., Allué, E., Bigga, G., Bruch, A. A., Cullen, V. L., Frahm, E., Ghukasyan, R., Gruwier, B., Jabbour, F., Miller, C. E., Taller, A., Vardazaryan, V., Vasilyan, D., & Weissbrod, L. (2017). The earliest evidence for Upper Paleolithic occupation in the Armenian Highlands at Aghitu-3 Cave. Journal of Human Evolution, 110, 37–68. https://doi.org/10.1016/j.jhevol.2017.05.010

Kosintsev, P. A., Lapteva, E. G., Trofimova, S. S., Zanina, O. G., Tikhonov, A. N., & Van der Plicht, J. (2012). Environmental reconstruction inferred from the intestinal contents of the Yamal baby mammoth Lyuba (Mammuthus primigenius Blumenbach, 1799). Quaternary International, 255, 231–238.

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., & Meshveliani, T. (2009). 30,000-year-old wild flax fibers. Science (New York, N.Y.), 325(5946), 1359–1359.

MacFadden, B. J., Solounias, N., & Cerling, T. E. (1999). Ancient diets, ecology, and extinction of 5-million-year-old horses from Florida. Science (New York, N.Y.), 283(5403), 824–827.

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., MacPhee, R., Froese, D., & Poinar, H. (2020). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quaternary Research, 1–24. https://doi.org/10.1017/qua.2020.59

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924–942. https://doi.org/10.1111/nph.14470

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kj, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. Nature, 537(7618), 45–49. https://doi.org/10.1038/nature19085

Polling, M., ter Schure, A. T., van Geel, B., van Bokhoven, T., Boessenkool, S., MacKay, G., Langeveld, B. W., Ariza, M., van der Plicht, H., Protopopov, A. V., & others. (2021). Multiproxy analysis of permafrost preserved faeces provides an unprecedented insight into the diets and habitats of extinct and extant megafauna. Quaternary Science Reviews, 267, 107084.

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing: NGS DIET ANALYSIS. Molecular Ecology, 21(8), 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Rödder, D., & Engler, J. O. (2011). Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. Global Ecology and Biogeography, 20(6), 915–927.

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2020). Hybridization capture of larch (Larix Mill) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13311

Seeber, P. A., McEwen, G. K., Löber, U., Förster, D. W., East, M. L., Melzheimer, J., & Greenwood, A. D. (2019). Terrestrial mammal surveillance using hybridization capture of environmental DNA from African waterholes. Molecular Ecology Resources, 19(6), 1486–1496. https://doi.org/10.1111/1755-0998.13069

Sukumar, R. (2006). A brief review of the status, distribution and biology of wild Asian elephants Elephas maximus. International Zoo Yearbook, 40(1), 1–8.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14–e14. https://doi.org/10.1093/nar/gkl938

ter Schure, A., Bajard, M., Loftsgarden, K., Høeg, H., Ballo, E., Bakke, J., Støren, E., Iversen, F., Kool, A., Brysting, A., & others. (2021). Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene. Quaternary Science Reviews, 270, 107175.

ter Schure, A., Pillai, A. A., Thorbek, L., Bhavani Shankar, M., Puri, R., Ravikanth, G., de Boer, H. J., & Boessenkool, S. (2021). EDNA

metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary. Environmental DNA, 3(3), 681–696.

Torti, A., Lever, M. A., & J, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Marine Genomics, 24 Pt 3, 185–196. https://doi.org/10.1016/j.margen.2015.08.007

van Geel, B., Protopopov, A., Bull, I., Duijm, E., Gill, F., Lammers, Y., Nieman, A., Rudaya, N., Trofimova, S., Tikhonov, A. N., & others. (2014). Multiproxy diet analysis of the last meal of an early Holocene Yakutian bison. Journal of Quaternary Science, 29(3), 261–268.

Willerslev, E. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vesterg, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Br, K. A., Yoccoz, N., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47–51. https://doi.org/10.1038/nature12921

Willerslev, E., Hansen, A. J., R, R., Brand, T. B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D., & Cooper, A. (2004). Long-term persistence of bacterial DNA. Current Biology, 14(1), R9–R10. https://doi.org/10.1016/j.cub.2003.12.012