Biodiversity Changes In Siberia Between Quaternary Glacial And Interglacial Stages
–
Exploring The Potential Of *seda*DNA

Jérémy Courtin

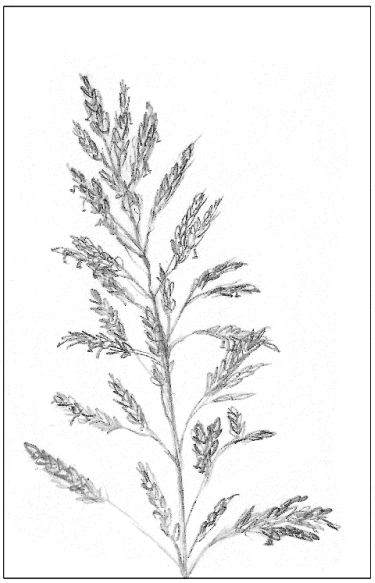
Univ.-Diss.

zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin "Genetik"

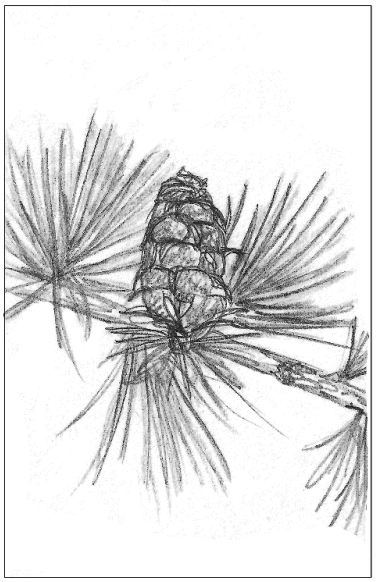
eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universität Potsdam
und am
Alfred-Wegener-Institut
Helmholtz-Zentrum für Polar- und Meeresforschung

Potsdam, April 2023

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-59584
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-595847

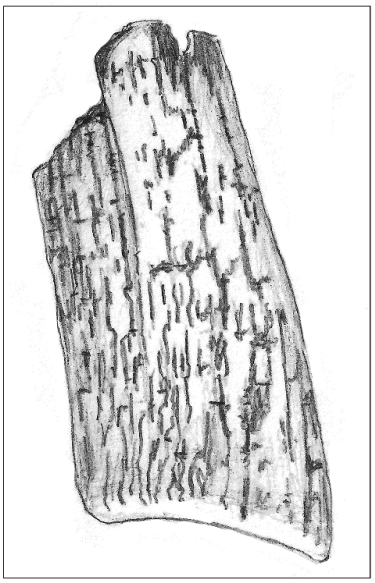


Bromus pumpellianus



Larix gmelinii

Rhododendron dauricum



Mammuthus primigenius

Hauptbetreuerin:	Prof. Dr. Ulrike Herzschuh (Universität Potsdam)
Weitere Gutachter:	Prof. Dr. Peter Heintzman (The Arctic University Museum, Tromsø, Norway) Prof. Dr. Susanne Liebner (Universität Potsdam)

TABLE OF CONTENTS

Acknowledgements i

Summary iii

Zusammenfassung iv

1 General introduction 1

1.1 A changing world 1

1.1.1 Global changes of anthropogenic origin 1

1.1.2 Amplified crisis in the high latitudes 1

1.2 The past is the key to the future 2

1.2.1 The Quaternary glacial and interglacial stages 2

1.2.2 The Beringia study case 3

1.3 Investigating past biodiversity 6

1.3.1 Traditional tools 6

1.3.2 Newest *seda*DNA proxies 7

1.4 Motivation and aims of the thesis 9

1.5 Structure of the thesis 9

1.6 Author's contributions 10

2 Manuscript I 12

2.1 Abstract 13

2.2 Introduction 14

2.3 Materials and Methods 16

2.3.1 Geographical settings 16

2.3.2 Fieldwork and subsampling 18

2.3.3 Core splicing and dating 18

2.3.4 Sediment-geochemical analyses 20

2.3.5 Pollen analysis 22
2.3.6 Molecular genetic preparation 22
2.3.7 Processing of *seda*DNA data 24
2.3.8 Statistical analysis and visualization 25
2.4 Results 27
2.4.1 Age model 27
2.4.2 Sediment-geochemical core composition 27
2.4.3 Pollen stratigraphy 30
2.4.4 *seda*DNA composition 32
2.4.5 Comparison between pollen and *seda*DNA 34
2.4.6 Taxa richness investigation 36
2.5 Discussion 36
2.5.1 Proxy validation 36
2.5.2 Vegetation compositional changes in response to climate inferred from pollen and *seda*DNA records 38
2.5.3 The steppe-tundra of the Late Pleistocene 39
2.5.4 The disrupted Pleistocene-Holocene transition 40
2.5.5 The boreal forest of the Holocene 41
2.5.6 Changes in vegetation richness through the Pleistocene/Holocene transition inferred from the *seda*DNA record 42
2.6 Conclusion 44
Data availability statement 45
Funding 45
References 45
3 Manuscript II 54
3.1 Abstract 55
3.2 Introduction 56

3.3 Material and Method 58
3.3.1 Site description and timeframe 58
3.3.2 Sampling, DNA extraction and PCR 60
3.3.3 Filtering and cleaning dataset 60
3.3.4 Identification of taxa – species signal 61
3.3.5 Resampling 62
3.3.6 Assessment of the species pool stability 62
3.3.7 Quantification of extinct and extirpated taxa 62
3.3.8 Characterisation of species and candidate species 63
3.4 Results 64
3.4.1 Changes in the composition and species pool at the Pleistocene - Holocene transition 65
3.4.2 Decrease in the regional plant species richness between the Pleistocene and the Holocene 66
3.4.3 Identification of loss taxa events 67
3.4.4 Characterisation of lost taxa 69
3.5 Discussion 73
3.5.1 Biotic and abiotic changes in the ecosystem - a cocktail for extinction 75
3.5.2 Identification and quantification of potential plant taxa loss 76
3.5.3 Characterisation of potential taxa loss 78
3.5.4 Limits of the method 79
3.5.5 Conclusions and perspectives 80
Funding 80
References 80
4 Manuscript III 88
4.1 Abstract 89
4.2 Introduction 90

4.3 Material & Methods 93
4.3.1 Fieldwork and subsampling 93
4.3.2 Chronology 94
4.3.3 Pollen analysis 96
4.3.4 Isolation of sedimentary ancient DNA 96
4.3.5 Metabarcoding approach 96
4.3.6 Shotgun approach 97
4.3.7 Bioinformatic processing 97
4.4 Results 101
4.4.1 General results of the three approaches: pollen, metabarcoding and shotgun sequencing 101
4.4.2 Plants (Viridiplantae) 102
4.4.3 Fungi 103
4.4.4 Mammals (Mammalia) 103
4.4.5 Birds (Aves) 105
4.4.6 Insects (Insecta) 106
4.4.7 Prokaryotes (Bacteria, Archaea) and Viruses 107
4.5 Discussion 109
4.5.1 Interglacial communities 110
4.5.2 Glacial communities 112
4.5.3 Potential and limitations of the *seda*DNA shotgun approach applied to ancient permafrost sediments 114
4.6 Conclusions 115
Data availability statement 115
Funding 115
References 115
5 Synthesis 125

5.1 Ecological changes between glacial and interglacial stages 125
5.1.1 Changes in the compositional structure 125
5.1.2 Loss of plant diversity 128
5.1.3 Potential drivers of change 130
5.2 High potential of *seda*DNA for past biodiversity reconstruction 133
5.3 Conclusions and future perspectives 135
Bibliography 136
Appendices 149
Appendix 1: Supplementary material for Manuscript I 149
Appendix 2: Supplementary material for Manuscript II 154
Appendix 3: Supplementary material for Manuscript III 159
Appendix 4: Manuscript IV 183
Eidesstattliche Erklärung 217

ACKNOWLEDGEMENTS

I was able to meet many great and supporting people and share very memorable moments over the last three and a half years that I conducted this PhD project.

I owe thanks to my supervisor Prof. Dr. Ulrike Herzschuh for giving me this opportunity to work at the Alfred Wegener institute, for guiding me, and for sharing her expertise over the years. In addition, I would like to thank my second supervisor, Dr. Kathleen Stoof-Leichsenring for her support in all laboratory issues and organizations and her role as our working group leader.

I am also very thankful for all the A43 past and present colleagues as they were all of great support, especially Thomas Böhmer, Ugur Cabuk, Viktor Dinkel, Ramesh Glückler, Barbara von Hippel, Sichao Huang, Stefan Kruse, Simeon Lisovski, Sisi Liu, Stefano Meucci, Heike Zimmermann and others I met at AWI Potsdam or at other AWI locations. Special thanks to Amedea Perfumo with whom I shared my office for the last years, she was of great moral support, and I really enjoyed our chats and valued her advice. I also want to give a special thank you to Luise Schulte; she has been of incredible help, as a colleague and friend.

For their help with DNA extraction, library builds and support in the lab I would like to thank Myron Eberle, Svetlana Karachurina, Janine Klimke, Sarah Olischläger and Theresa Wietelmann as well as Iris Eder that I had a great opportunity to supervise. I also had the chance to take part in the Batagay 2019 expedition and work more closely with colleagues with whom I had an exciting time and would like to thank. I think especially of Loeka Jongejans, Hanno Meyer, Thomas Opel, Sebastian Wetterich, and Alexander Kizyakov. I would like to thank Andrej Andreev, Boris Biskaborn, Mary Edwards, Julian Murton and all other co-authors for sharing their knowledge and their great contributions to the manuscripts. I would like to thank Sigrun Gräning for supporting me in administrative issues and wish her nice retirement. In addition, I would like to thank both Claudia Hanfland and Claudia Sprengel from POLMAR graduate school.

I am grateful to my family, my parents and my sister who are and will always support me even if I am far away. Finally, to my caring and loving partner, Antonia: my deepest gratitude. Your constant encouragement is of overwhelming comfort.

SUMMARY

Climate change of anthropogenic origin is affecting Earth's biodiversity and therefore ecosystems and their services. High latitude ecosystems are even more impacted than the rest of Northern Hemisphere because of the amplified polar warming. Still, it is challenging to predict the dynamics of high latitude ecosystems because of complex interaction between abiotic and biotic components. As the past is the key to the future, the interpretation of past ecological changes to better understand ongoing processes is possible. In the Quaternary, the Pleistocene experienced several glacial and interglacial stages that affected past ecosystems. During the last Glacial, the Pleistocene steppe-tundra was covering most of unglaciated northern hemisphere and disappeared in parallel to the megafauna's extinction at the transition to the Holocene (~11,700 years ago). The origin of the steppe-tundra decline is not well understood and knowledge on the mechanisms, which caused shifts in past communities and ecosystems, is of high priority as they are likely comparable to those affecting modern ecosystems. Lake or permafrost core sediments can be retrieved to investigate past biodiversity at transitions between glacial and interglacial stages. Siberia and Beringia were the origin of dispersal of the steppe-tundra, which make investigation this area of high priority. Until recently, macrofossils and pollen were the most common approaches. They are designed to reconstruct past composition changes but have limit and biases. Since the end of the 20th century, sedimentary ancient DNA (*seda*DNA) can also be investigated. My main objectives were, by using *seda*DNA approaches to provide scientific evidence of compositional and diversity changes in the Northern Hemisphere ecosystems at the transition between Quaternary glacial and interglacial stages.

In this thesis, I provide snapshots of entire ancient ecosystems and describe compositional changes between Quaternary glacial and interglacial stages, and confirm the vegetation composition and the spatial and temporal boundaries of the Pleistocene steppe-tundra. I identify a general loss of plant diversity with extinction events happening in parallel of megafauna' extinction. I demonstrate how loss of biotic resilience led to the collapse of a previously well-established system and discuss my results in regards to the ongoing climate change. With further work to constrain biases and limits, *seda*DNA can be used in parallel or even replace the more established macrofossils and pollen approaches as my results support the robustness and potential of *seda*DNA to answer new palaeoecological questions such as plant diversity changes, loss and provide snapshots of entire ancient biota.

ZUSAMMENFASSUNG

Der vom Menschen verursachte Klimawandel wirkt sich auf die biologische Vielfalt der Erde und damit auf die Ökosysteme und ihre Leistungen aus. Die Ökosysteme in den hohen Breitengraden sind aufgrund der verstärkten Erwärmung an den Polen noch stärker betroffen als der Rest der nördlichen Hemisphäre. Dennoch ist es schwierig, die Dynamik von Ökosystemen in den hohen Breitengraden vorherzusagen, da die Wechselwirkungen zwischen abiotischen und biotischen Komponenten sehr komplex sind. Da die Vergangenheit der Schlüssel zur Zukunft ist, ist die Interpretation vergangener ökologischer Veränderungen möglich, um laufende Prozesse besser zu verstehen. Im Quartär durchlief das Pleistozän mehrere glaziale und interglaziale Phasen, welche die Ökosysteme der Vergangenheit beeinflussten. Während des letzten Glazials bedeckte die pleistozäne Steppentundra den größten Teil der unvergletscherten nördlichen Hemisphäre und verschwand parallel zum Aussterben der Megafauna am Übergang zum Holozän (vor etwa 11 700 Jahren). Der Ursprung des Rückgangs der Steppentundra ist nicht gut erforscht, und die Kenntnis über die Mechanismen, die zu den Veränderungen in den vergangenen Lebensgemeinschaften und Ökosystemen geführt haben, ist von hoher Priorität, da sie wahrscheinlich mit denen vergleichbar sind, die sich auf moderne Ökosysteme auswirken. Durch die Entnahme von See- oder Permafrostkernsedimenten kann die vergangene Artenvielfalt an den Übergängen zwischen Eis- und Zwischeneiszeiten untersucht werden. Sibirien und Beringia waren der Ursprung der Ausbreitung der Steppentundra, weshalb die Untersuchung dieses Gebiets hohe Priorität hat. Bis vor kurzem waren Makrofossilien und Pollen die gängigsten Methoden. Sie dienen der Rekonstruktion vergangener Veränderungen in der Zusammensetzung der Bevölkerung, haben aber ihre Grenzen und Schwächen. Seit Ende des 20. Jahrhunderts kann auch sedimentäre alte DNA (*seda*DNA) untersucht werden. Mein Hauptziel war es, durch den Einsatz von *seda*DNA-Ansätzen wissenschaftliche Beweise für Veränderungen in der Zusammensetzung und Vielfalt der Ökosysteme der nördlichen Hemisphäre am Übergang zwischen den quartären Eiszeiten und Zwischeneiszeiten zu liefern.

In dieser Arbeit liefere ich Momentaufnahmen ganzer alter Ökosysteme und beschreibe die Veränderungen in der Zusammensetzung zwischen Quartärglazialen und Interglazialen und bestätige die Vegetationszusammensetzung sowie die räumlichen und zeitlichen Grenzen der pleistozänen Steppentundra. Ich stelle einen allgemeinen Verlust der Pflanzenvielfalt

fest, wobei das Aussterben der Pflanzen parallel zum Aussterben der Megafauna verlief. Ich zeige auf, wie der Verlust der biotischen Widerstandsfähigkeit zum Zusammenbruch eines zuvor gut etablierten Systems führte, und diskutiere meine Ergebnisse im Hinblick auf den laufenden Klimawandel. Mit weiteren Arbeiten zur Eingrenzung von Verzerrungen und Grenzen kann *seda*DNA parallel zu den etablierteren Makrofossilien- und Pollenansätzen verwendet werden oder diese sogar ersetzen, da meine Ergebnisse die Robustheit und das Potenzial von *seda*DNA zur Beantwortung neuer paläoökologischer Fragen wie Veränderungen der Pflanzenvielfalt und -verluste belegen und Momentaufnahmen ganzer alter Biota liefern.

1 GENERAL INTRODUCTION

1.1 A CHANGING WORLD

1.1.1 Global changes of anthropogenic origin

Ecosystems are spatially and temporarily dynamic, resilient systems of interacting biotic and abiotic components (Begon & Townsend, 2020; Gunderson & Holling, 2002; Odum & Barrett, 1971). This resilience is essential as it allows to recover from disturbances and assures their adaptability (Elmqvist et al., 2003). The biodiversity plays a considerable role in ecosystem resilience as it is a critical component of the system and trophic cascades affect ecosystem functioning (Peterson et al., 1998; Schmitz, 2013). Therefore, when key taxa are lost or new ones invade, ecosystems can drastically change (Estes & Duggins, 1995; Terborgh et al., 2001; Vitousek & Walker, 1989).

"Human influence has warmed the climate at a rate that is unprecedented in at least the last 2000 years" (IPCC 2021). Since the past decades, climate change is affecting every part of the world. Anthropogenic influence (e.g., in the form of global industrialization or deforestation) contributes to changes in weather and increases the occurrence of observed extreme events such as heat waves, heavy precipitation or droughts (IPCC 2021). Worldwide, in response to climate change, biodiversity and ecosystems are rapidly changing in sudden and irreversible ways (Malhi et al., 2020; Turner et al., 2020). When forced to follow rapid changes, resilience of the biosphere is challenged, making it vulnerable to extinction and loss of diversity, for example threatening the resilience of ecosystems (Cowie et al., 2022; Harrison, 2020). Overall, the ongoing climate change affects earth's biodiversity distribution and species interactions, altering the composition of ecological communities and therefore ecosystems, ecosystem services and human well-being (Pecl et al., 2017; Tylianakis et al., 2008; Walther, 2010).

1.1.2 Amplified crisis in the high latitudes

Amplified polar warming causes high latitude ecosystems to experience even stronger climate change than the rest of Northern Hemisphere ecosystems (Miller et al., 2010). Both abiotic and biotic components of the high latitude ecosystems are affected. Most of extreme climate change impacts that are projected to occur by 2050 elsewhere in the world have

already been observed in the Arctic (IPCC 2021). For this reason, it can be considered a flagship area for climate change, making its investigation of the highest interest.

For example, more extreme heat events and wildfires are reported in Siberia, alongside increases in carbon release, disruption of water cycles such as Arctic humidification, precipitation increase and decline in snow cover are accelerated (Box et al., 2019). Such events might lead to higher extinction rates of Arctic vascular plants, mosses, and lichens (Niitynen et al., 2018), as warming climate favours the growth of woody vegetation (Garcia Criado et al., 2020; Song et al., 2018). As a result, shrubs, and vertebrates expand northward, thus spreading into the Arctic (Davidson et al., 2020). Tundra is greening due to increase in plant biomass, with regional browning due to death of vegetation (Epstein et al., 2018; Brown et al., 2019). Finally, all these observed changes in the Arctic also threaten ecosystem services (e.g., climate regulation, Malinauskaite et al., 2019) and they can disrupt future food and nutritional security for both local and global human populations (Hicks et al., 2019).

Because of the intricate interactions and feedback processes between and within abiotic and biotic components, it is challenging to predict the dynamics of high latitude ecosystems. To improve our ability to understand recent and future climate change impacts on such ecosystems, we can work to interpret past ecological changes, in other words, the past is the key to the present and to the future (Botkin et al., 2007; Hoelzel, 2010; Lyell, 2010).

1.2 THE PAST IS THE KEY TO THE FUTURE

1.2.1 The Quaternary glacial and interglacial stages

During the Quaternary period, the Pleistocene (2,580 – 11,7 thousand years (ka) before present (BP)) experienced alternating glacial and interglacial stages (Mix & Ruddiman, 1984; Shackleton, 1967). Such cooling and warming cycles caused drastic changes in past ecosystems, affecting most biotic and abiotic systems, from global climate to local population and genetic diversity patterns (Hewitt, 2000; Jackson & Blois, 2015).

The study of palaeorecords from the Pleistocene such as lake sediments, peat bogs, ice cores or permafrost archives (Birks & Birks, 1980; Bradley, 1999) covering transitions between glacial and interglacial stages can provide information on past ecosystem

dynamics and how they adapted from one equilibrium (i.e., glacial ecosystem) to the other (i.e., interglacial ecosystem).

Records spanning several glacial and interglacial stages continuously are particularly valuable and typically originate from lake sediment or permafrost sediment sequences. Although rare, some well-known chronosequences include Lake El'gygytgyn, consisting of a core that spans the last ~3.500 ka (Andreev et al., 2014; Melles et al., 2012) and Lake Baikal the last ~3,600 ka (Demske et al., 2002; Tarasov et al., 2005; Tarasov et al., 2007). Also, several records from the Laptev Sea area cover the past 200 ka (Andreev et al., 2011 and references therein) while the Batagay megaslump provides access to permafrost formations starting from the Middle Pleistocene as old as ~650 ka (Ashastina et al., 2017; Murton et al., 2017; Murton et al., 2021; Opel et al., 2019).

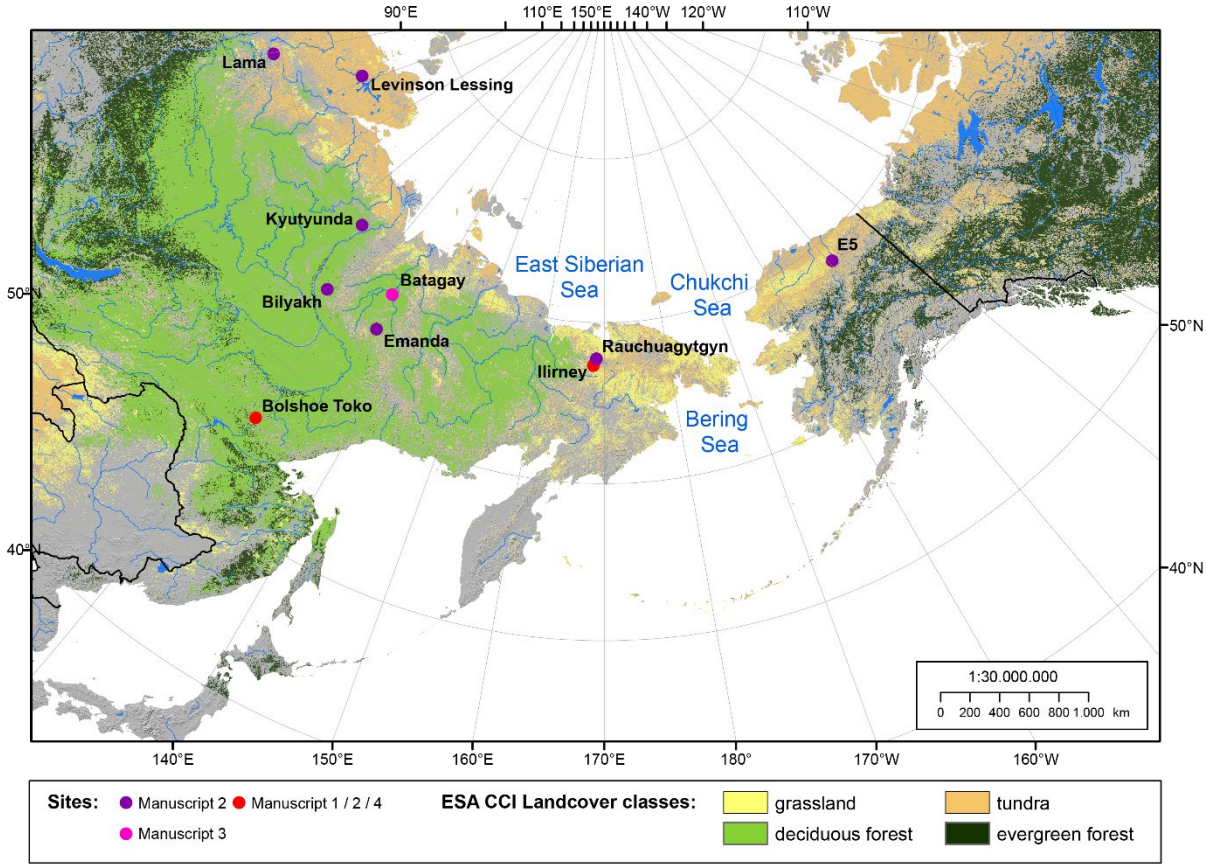


Figure 1.1: Map of the study area covering North-East Asia and Alaska. Land Cover classes are used for dominant modern (1992-2015) vegetation types. Sites investigated in the Thesis are marked with coloured circles.

1.2.2 The Beringia study case

Further records spanning shorter timescales, such as only the Holocene or the Pleistocene to Holocene transition have been retrieved from Beringia **(Figure 1.1**). Glacial stages were

characterised by the formation of mountain glaciers and vast ice sheets, causing a global sea-level drop because of massive amounts of water entrapped as ice, while during Interglacial stages the sea level rose because of ice sheets and glaciers retreats (Velichko, 1975). Such fluctuation of the sea-level affected landscapes and coastline retreats and advances. In the Northern Hemisphere, a landmass emerged during glacial stages and connected both Asia and North America: Beringia (Hultén, 1937). This connection served as a biological intercontinental land bridge (Hopkins et al., 1982; Monteath et al., 2022). It remained free of continental ice sheets during the late Quaternary, until the end of the Pleistocene (Barr & Clark, 2012; Svendsen et al., 2004). Beringia was the origin of dispersal of one of the largest northern ecosystems – tundra-steppe, also called steppe-tundra or mammoth-steppe (e.g., Guthrie, 1986; Hopkins, 1982; Kahlke, 2015; Yurtsev, 1982).

The steppe-tundra developed under the cold and dry climate of the glacial Late Pleistocene and covered most of the unglaciated Northern Hemisphere, from the Atlantic east coast to North America and across Siberia and Beringia (Guthrie, 2001; Hopkins et al., 1982; Matthews, 1982). Over its large extension, this biota shared general features despite regional variation such as a dominance of cold and dry adapted open-land taxa, also described as a mosaic landscape of herbs and grassland with woody and shrubby patches (Bezrukova et al., 2010; Bigelow et al., 2003; Blinnikov et al., 2011; Kienast et al., 2005; Kuzmina, 2015; Tarasov et al. 2013; Wang et al., 2021).

This specific vegetation structure and plant arrangement do not exist any longer and previously co-occurring plant species are now confined in either the temperate steppe or the Arctic tundra. Only the Altaian steppe-tundra could be considered a close analogue of the Pleistocene steppe-tundra, but it is considered that this past biota disappeared at the transition between the last Glacial and the Holocene (Chytrý et al., 2019; Guthrie, 2001; Zimov et al., 2012).

In addition, the Northern Hemisphere was the field of Pleistocene megafauna extinction supported by numerous findings of mammal bone remains and aDNA. This suggests that the species inhabitant of the Pleistocene steppe-tundra are nowadays either extinct or fill different ecological niches (e.g. Nehring, 1890; Tugarinov, 1929; Wang et al., 2021). Mammals of the “Ice age” such as woolly mammoth, woolly rhinoceros, cave lions or cave bears are extinct, while musk ox and reindeer inhabit the modern northern tundra, and

horse, bison or saiga antelope inhabit the southern steppes (Ashatina, 2018). Megafauna were hypothesised to be keystone species maintaining the steppe-tundra open (Bakker et al., 2016; Gill, 2014).

The latest transition between a Glacial and an Interglacial stage occurred ~11,7 ka BP at the end of the Pleistocene, marking the beginning of the Holocene (Walker et al., 2008). Climate change at this transition, between ~15 ka BP and 11.7 ka BP, was abrupt and intense. Within few thousand years, the Northern Hemisphere experienced sharp climatic oscillations with two warming events interrupted by a cooling one (Bazelmans et al., 2021). From the last "Ice-age", temperature changed rapidly from glacial to mild during the Bølling-Allerød interval (~14.5 ka BP) to then returning to glacial stage during the Younger Dryas period (~12.8 – 11.7 ka BP), just before the beginning of the warm Holocene (Steffensen et al., 2008; Liu et al., 2009; Mangerud, 2020).

In parallel to the abrupt climatic changes, modern humans (*Homo sapiens*) spread throughout the northern Hemisphere and in America. *Homo sapiens* appeared in Siberia around ~50-45 ka, were present in east Beringia ~33 ka and crossed the Bering bridge latest at 15.3 ka (Davis et al., 2018; Graf & Buvit, 2017). Because of the co-occurrence of human arrival in Beringia and megafauna extinction, Martin & Wright in 1967 suggested that *Homo sapiens* were the driver of megafauna's extinction through their population extension.

Altogether, at the end of the Pleistocene, the Northern Hemisphere experienced the decline of steppe-tundra ecosystem along with a massive extinction of megafauna, abrupt climatic changes as well as the spreading of human population. The causes of this decline are being discussed for a long time (Owen, 1861; Martin, 1958; Barnosky et al., 2004; Koch & Barnosky, 2006; Sandom et al., 2014; Stewart et al., 2021; Wang et al., 2021; Wroe et al., 2006), while many questions remain unanswered, such as the potential loss of plant diversity in parallel to the loss of the steppe-tundra. The potential co-extinction of plant taxa and megafauna has not been investigated so far and our knowledge of potential cross-kingdom interactions (e.g., trophic) is rather limited regarding ancient biota. Understanding the mechanisms which caused shifts in past communities and ecosystems is of high priority as they are likely comparable to those affected by current climate change.

1.3 INVESTIGATING PAST BIODIVERSITY

Three methods are well established to retrieve information on past biota from palaeorecords: the historical macrofossils and pollen investigations (Birks & Berglund, 2018), and the more recent investigation of ancient DNA (aDNA) and sedimentary ancient DNA (*seda*DNA) (Ogram et al., 1987; Olsen et al., 1986; Pace et al., 1986; Thomsen & Willerslev, 2015).

1.3.1 Traditional tools

Before the 21st century, only fossils and pollen (Birks & Berglund, 2018) were used to investigate past biodiversity.

Macrofossils often allow precise identification to species level of plants (Birks, 2001), invertebrates (e.g., Ashastina et al., 2018) and vertebrates (e.g., Behrensmeyer & Hill, 1988). This approach uses well preserved and identifiable macro remains of partial tissues or even entire individuals with all tissues preserved (Lopatin, 2021). The signal mostly represents the local conditions and originate directly from former individuals at the study site (Parducci et al., 2015; Pedersen et al., 2013). However, the number of identifiable macrofossils is affected by the taphonomy, and preservation that varies strongly between taxa (Gardner et al., 2016; Kienast et al., 2001; Tomescu et al., 2018). Accumulation of macrofossils might not represent the actual biodiversity as they can have been selected by animals for food or nesting (e.g., squirrel nests; Ashastina et al., 2018) or by humans (e.g., animals' bones remain; Pickering et al., 2004). Finally, the major limit of this method is that most samples being usually poor in fossils (e.g., Ashastina et al., 2018), it is little informative in chronosequences.

Pollen assemblages recovered from sediments allow reconstruction of past flora and conclusions on the presence of other taxonomic groups with the presence of other spores (e.g., coprophilous fungi to detect herbivores; Andreev et al., 2009; Andreev et al., 2011; Wicklow, 1981). Pollen grains and spores are usually well preserved as structurally and physiologically resistant (Katifori et al., 2010). The pollen signal originates from across a variable extent from extra-local to local sources providing a regional signal of past flora (Birks, 2001; van der Knaap, 1987). Because of this regional signal, this proxy can be used to investigate time series as pollen deposit at all the time whenever there are plant taxa present. Pollen grains and spores can be easily recovered from sediment records and the

methodology to identify and count it is well established (Birks & Berglund, 2018). Overall, this method is very largely accepted to investigate past flora compositional changes through time over geological timescales (e.g., up to ~3,500 ka BP; Andreev et al., 2014). However, pollen proxy has limitations such as a bias towards wind-pollinated taxa (Taberlet et al., 2018). Insect-pollinated plants are usually less represented and not all plant species produce the same amount of pollen, leading to an overrepresentation of plants with high pollen production (Birks & Berglund, 2018; Weng et al, 2006). This and the fact that taxonomic assignment is not highly specific makes investigations of past diversity changes challenging using pollen assemblages (Birks et al., 2016; Reitalu et al., 2019).

Both methods are universally used to investigate past biodiversity from palaeorecords and often complement each other, yet their biases pose a limit to what extent they can address ecological questions.

1.3.2 Newest *seda*DNA proxies

DNA (or deoxyribonucleic acid) is the molecular instruction for life found in all living and Viruses' cells. DNA holds the most basic information one can retrieve from organisms and therefore the highest potential to investigate present and past biodiversity. DNA from past organisms can be retrieve from preserved macro or microfossils but also from the environment. When a cell dies, its content is released and DNA is free in the environment (eDNA; Taberlet et al., 2012). One such example is sedimentary ancient DNA (*seda*DNA), i.e., ancient extra-cellular DNA in soil solution or bound to mineral particles from sediments, typically recovered from lake sediments or permafrost (e.g., Willerslev et al., 2003). Once the DNA is outside the cell, it undergoes degradation through biological, chemical, or physical processes (Taberlet et al., 2018). However, when the preservation conditions are ideal (e.g., frozen, and anoxic environments such as permafrost), DNA can be retrieved from sediments as old as 1.4 million years (Kirkpatrick et al., 2016). Once the DNA is released in the environment, it is vulnerable to microbial and enzymatic activities (Edwards, 2020). By binding to charged minerogenic or organic particles present in the soil, the eDNA fraction can be stabilized and resist decomposition (Morissey et al., 2015; Murchie et al., 2021; Pietramellara et al., 2009). Environmental temperature influences preservation to some extent and DNA is particularly well preserved in permafrost sediments as constantly cool temperature reduce microbial and enzymatic degradation (Jørgensen et al., 2012; Zimmermann et al., 2017). Other records yielding good

preservation are lake sediments as the water column can become thermally stratified favouring the development of anoxia at the lake bottom (Parducci et al., 2017). The signal of the newest proxy, *seda*DNA, like macrofossils, is of local origin (Alsos et al., 2018; Parducci et al., 2017).

The end of the 20th century and beginning of the 21st century saw an acceleration of technical and technological innovations. Methods such as next generation sequencing (NGS) allowed the sequencing of DNA and subsequently the analysis of high amount of data faster and cheaper, making the investigation of eDNA largely accessible (Glenn, 2011; Margulies et al., 2005; Metzker, 2010; Ogram et al., 1987; Olsen et al., 1986; Pace et al., 1986; Shendure et al., 2017; Thomsen & Willerslev, 2015). Pushed forward by this sequencing revolution, the last few years saw an increasing number of studies using *seda*DNA Quaternary reconstruction (Brown & Barnes, 2015; Capo et al., 2021; Edwards, 2020; Pedersen et al., 2016; Ruppert et al., 2019; Willerslev et al., 2014). Most of them focussed on vegetation reconstruction and identification of vascular plants, following the "metabarcoding" approach using the amplification of the P6 loop region of the chloroplast trnL (UAA) intron (Alsos et al., 2018; Boessenkool et al., 2014; Clarke et al., 2019; Crump et al., 2021; Liu et al., 2021; Niemeyer et al., 2017; Pansu et al., 2015; Parducci et al., 2012; Rijal et al., 2021; Sjögren et al., 2017; Taberlet et al., 2007). Other taxonomic groups can be investigated with the metabarcoding technique, using a short, highly variable part of the genome flanked by highly conserved regions, which can serve as primer binding sites in a PCR reaction (Ficetola et al., 2008; Giguet-Covex et al., 2014; Taberlet et al., 2007). The amplified variable fragments can then be compared to specific databases, and DNA sequences can be assigned to family, genus or even species level (Liu et al., 2020; Niemeyer et al., 2015).

Another approach to investigate *seda*DNA is to sequence the entire pool of extracted DNA, using the so-called "metagenomic shotgun sequencing" (shorten as metagenomics or shotgun sequencing). In this case, it opens to the possibility to reconstruct the entire biotic system that lived around the study site, from viruses to mammals or plants. When using this approach, it is possible to discriminate endogenous aDNA fragments from the exogenous modern ones (e.g., contaminants) by assessing the typical damage patterns occurring in the process of DNA degradation, including deamination and DNA fragmentation to authenticate ancient DNA (Ginolhac et al., 2011). Despite its high

potential, this method is not yet well established yet and only few studies used this approach (Ahmed et al., 2018; Graham et al., 2016; Pedersen et al., 2016; Wang et al., 2021). A major disadvantage of the method is that only a small fraction of reads sequenced is assigned to the databases, (e.g., 0.4% of reads assigned in Pedersen et al., 2016).

To summarise, *seda*DNA holds the highest potential for investigation of ecological time series. It also enables to answer ecological questions that could not be previously explored with other approaches such as pollen (Bálint et al., 2018; Orlando et al., 2021; Pont et al., 2019). However, the *seda*DNA research field is still in its infancy and more studies are needed to pave the way towards its full exploitation for the reconstruction of past diversity changes and the study of entire ecosystems or the investigation of potential extinction events other than megafauna.

1.4 MOTIVATION AND AIMS OF THE THESIS

The main motivation of this work was to attain a better understanding of how past ecosystems adapted to past climate changes during the Quaternary. With modern ecosystems experiencing human-driven climate change, it is of the paramount importance to gain information on the potential drivers and directions of ecosystem change. Therefore, the main goal of this thesis was to provide scientific evidence of compositional and diversity changes in the Northern Hemisphere ecosystems at the transition between Quaternary glacial and interglacial stages. *Seda*DNA techniques were used in addition to more traditional pollen analysis to be able to assess the advantages and limitations of each method in the context of palaeoenvironment reconstruction. In parallel, the potential of *seda*DNA was explored to answer new palaeoecological questions. In this thesis, I focus particularly on past plant diversity changes, potential past plant extinction and the reconstruction of potential interactions across taxonomic groups such as vegetation-megafauna.

1.5 STRUCTURE OF THE THESIS

This thesis is designed as a cumulative dissertation, consisting of a general introduction with sub-sections to give scientific background and context to the research topic (Chapter 1), four individual manuscripts (Chapters 2, 3, 4 and **Appendix 4**), and an overall

discussion of the thesis outcomes (Chapter 5). References used in Chapter 1 and 5 are given in the bibliography section.

1.6 AUTHOR'S CONTRIBUTIONS

- **Manuscript 1**:

  Authors: **Jérémy Courtin**, Andrei A. Andreev, Elena Raschke, Sarah Bala, Boris K. Biskaborn, Sisi Liu, Heike Zimmermann, Bernhard Diekmann, Kathleen R. Stoof-Leichsenring, Luidmila A. Pestryakova and Ulrike Herzschuh

  Contributions: **JC**: conceptualization, methodology, investigation, and writing – original draft preparation. AA: investigation, writing – reviewing and editing. ER: investigation. SB: investigation. BB: investigation, writing – reviewing and editing. SL: writing – reviewing and editing. HZ: methodology, writing – reviewing and editing. BD: conceptualization and funding acquisition. KS-L: writing – reviewing and editing and supervision. LP: resources. UH: conceptualization, co-writing original draft preparation, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

- **Manuscript 2**:

  Authors: **Jérémy Courtin** and Ulrike Herzschuh

  Contributions: **JC**: conceptualization, methodology, investigation, and writing – original draft preparation. UH: conceptualization, co-writing original draft preparation, supervision, project administration, and funding acquisition.

- **Manuscript 3**:

  Authors: **Jérémy Courtin**, Amedea Perfumo, Andrei A. Andreev, Thomas Opel, Kathleen R. Stoof-Leichsenring, Mary E. Edwards, Julian B. Murton and Ulrike Herzschuh

  Contributions: **JC**: designed and performed research, analysed, and interpreted the data, and wrote the first draft of the manuscript. AP: interpreted the data, contributed in detail to the microorganisms and viruses' interpretation and wrote the microbial part of the manuscript. AAA: analysed and interpreted the data and contributed in detail to the pollen analysis and the vegetation reconstruction. TO: and JBM: performed fieldwork and contributed details on regional background, cryostratigraphy and chronology of the Batagay megaslump. KRSL: supervised

palaeogenetic laboratory work and contributed to the interpretation of both metabarcoding and shotgun sequencing results. ME: reviewed the manuscript and contributed in detail to the ecological interpretation. UH designed and supervised the research and contributed in detail to the interpretation of the data and co-wrote the first draft of the manuscript. All co-authors contributed to the final discussion of results and commented on the draft of the manuscript.

- **Manuscript 4**:

  Authors: Sichao Huang, Kathleen R. Stoof-Leichsenring, Sisi Liu, **Jérémy Courtin**, Andrej A. Andreev, Luidmila. A. Pestryakova and Ulrike Herzschuh

  Contributions: SH: conceptualization (lead), formal analysis (lead), and writing – original draft and review and editing (lead). KS- L: conceptualization (supporting), formal analysis (supporting), resources (supporting), and writing – review and editing (supporting). SL, **JC**, and AA: writing – review and editing (supporting). LP: resources (supporting) and writing – review and editing (supporting). UH: conceptualization (lead), resources (lead), writing – review and editing (lead), and supervision (lead). All authors contributed to the article and approved the submitted version.

2 MANUSCRIPT I

Vegetation changes in southeastern Siberia during the Late Pleistocene and the Holocene

Authors

Jérémy Courtin[1]*, Andrei A. Andreev[1], Elena Raschke[1], Sarah Bala[1], Boris K. Biskaborn[1], Sisi Liu[1], Heike Zimmermann[1], Bernhard Diekmann[1], Kathleen R. Stoof-Leichsenring[1], Luidmila A. Pestryakova[2] and Ulrike Herzschuh[1,3,4]*

Affiliation

[1]Polar Terrestrial Environmental Systems, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany,

[2]Department of Geography and Biology, University of Yakutsk, Yakutsk, Russia,

[3]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany,

[4]Institute of Biology and Biochemistry, University of Potsdam, Potsdam, Germany

***Correspondence**

Jérémy Courtin: jeremy.courtin@awi.de; Ulrike Herzschuh, ulrike.herzschuh@awi.de.

Status

Published in *Frontiers in Ecology and Evolution*. doi: 10.3389/fevo.2021.625096

Additional information in **APPENDIX 1**.

2.1 ABSTRACT

Relationships between climate, species composition, and species richness are of particular importance for understanding how boreal ecosystems will respond to ongoing climate change. This study aims to reconstruct changes in terrestrial vegetation composition and taxa richness during the glacial Late Pleistocene and the interglacial Holocene in the sparsely studied southeastern Yakutia (Siberia) by using pollen and sedimentary ancient DNA (*seda*DNA) records. Pollen and *seda*DNA metabarcoding data using the trnL g and h markers were obtained from a sediment core from Lake Bolshoe Toko. Both proxies were used to reconstruct the vegetation composition, while metabarcoding data were also used to investigate changes in plant taxa richness. The combination of pollen and *seda*DNA approaches allows a robust estimation of regional and local past terrestrial vegetation composition around Bolshoe Toko during the last ~35,000 years. Both proxies suggest that during the Late Pleistocene, southeastern Siberia was covered by open steppe-tundra dominated by graminoids and forbs with patches of shrubs, confirming that steppe-tundra extended far south in Siberia. Both proxies show disturbance at the transition between the Late Pleistocene and the Holocene suggesting a period with scarce vegetation, changes in the hydrochemical conditions in the lake, and in sedimentation rates. Both proxies document drastic changes in vegetation composition in the early Holocene with an increased number of trees and shrubs and the appearance of new tree taxa in the lake's vicinity. The *seda*DNA method suggests that the Late Pleistocene steppe-tundra vegetation supported a higher number of terrestrial plant taxa than the forested Holocene. This could be explained, for example, by the "keystone herbivore" hypothesis, which suggests that Late Pleistocene megaherbivores were able to maintain a high plant diversity. This is discussed in the light of the data with the broadly accepted species-area hypothesis as steppe-tundra covered such an extensive area during the Late Pleistocene.

Keywords: last glacial, Holocene, Lake Bolshoe Toko, paleoenvironments, sedimentary ancient DNA, metabarcoding, trnL, pollen.

2.2 INTRODUCTION

Climate change affects species' range dynamics, eventually resulting in certain biodiversity patterns in space and time (Thomas et al., 2004; Dawson et al., 2011). Because of amplified arctic warming (Biskaborn et al., 2019b), northern boreal ecosystems experience stronger climate change than low-latitude ecosystems (Miller et al., 2010) in the same way that high elevation ecosystems are more impacted than lower elevation ones (Pepin et al., 2015). Accordingly, biodiversity in the northern ecosystems is expected to change at an exceptional pace under current global-warming processes. However, the links between climate, species composition, and species richness are largely unknown in the boreal ecosystems. Models suggest a major switch in vegetation composition in the Northern Hemisphere during the 21st century in the course of warming and boreal forests are expected to advance into the tundra zone (Pearson et al., 2013; Kruse et al., 2016). Whether this northward migration will cause an increase or decrease in plant richness is not yet predicted.

According to the hypothetical Latitudinal Diversity Gradient (LDG), diversity generally decreases with latitude (Hillebrand, 2004). For example, plant diversity in tundra is estimated to be lower than in boreal forests because of harsher habitat and climatic conditions such as permafrost making rooting systems difficult to establish, while there is less sunlight and it is often drier (Beschel, 1970; Kier et al., 2005; Mutke and Barthlott, 2005; Chytrý et al., 2007; Hofgaard et al., 2012). With co-correlation between latitude and temperature, projections based on modern spatial diversity suggest an increase of biodiversity at higher latitudes under future warming. However, the observed LDG for current distribution patterns can be explained by other processes than just the spatial temperature variability (Mittelbach et al., 2007). For example, area effect suggests that diversity decreases toward the poles because diversity and area are correlated (MacArthur and Wilson, 1967; Rosenzweig, 1995), and as area decreases, so does diversity. The LDG has been described as a robust phenomenon since the Paleozoic and has remained stable through glacial and interglacial changes (Ricklefs, 1987; Mittelbach et al., 2007; Mannion et al., 2014). Accordingly, spatial relationships between climate, vegetation type, and richness might yield misleading conclusions when used to predict temporal diversity changes. Inferences from time-series can instead be employed to overcome methodological shortcomings derived from space-for time approaches.

During the last glacial, most of northern Eurasia and North America were covered by open steppe-tundra communities (Guthrie, 2001). After the late glacial, the treeline advanced northward and uphill populating most of northern Eurasia during the Holocene (MacDonald et al., 2000; Lozhkin et al., 2007). Although the relationship between glacial-interglacial climate change and main vegetation types is generally understood, the detailed changes in plant richness are still to be explored. This lack of information is mainly caused by the fact that paleoecological tools to detect high-resolution species richness changes have only recently been developed.

Vegetation changes are mostly revealed via pollen records, but the interpretation of pollen records in terms of richness changes is challenging due to constraints on taxonomic identification to species levels or the magnitude of pollen counts (e.g., wind pollinated taxa being overrepresented in the samples) (Odgaard, 1999; Birks et al., 2015). These biases can be reduced in modern records by using the highest quality pollen data and pollen productivity estimates (PPEs) where available (Birks et al., 2015). However, when PPEs are missing, it is quite challenging to investigate diversity indices such as richness using pollen records. The relatively new sedimentary ancient DNA (*seda*DNA) metabarcoding method can be used to investigate past vegetation changes as a complementary taxonomically high-resolution proxy. Compared to pollen analysis, it typically detects a higher number of plant taxa at a higher taxonomic level and can reliably be used to investigate diversity changes, such as taxa richness (Niemeyer et al., 2017; Zimmermann et al., 2017a, b). The pollen proxy provides a robust estimation of the regional vegetation composition including a long-range catchment signal with wind-transported pollen grains (Moore et al., 1991; Willerslev et al., 2014). Compared to pollen, *seda*DNA is predominantly of local origin and thought to be mostly derived from extracellular DNA from various plant tissues (Jørgensen et al., 2012; Edwards et al., 2018). This is an advantage over pollen data, because different pollination strategies influence the amount of pollen production and pollen dispersal capacities. Wind-pollinated plants produce a large quantity of pollen, while insect-pollinated ones produce very little. Moreover, the preservation of some pollen taxa (e.g., *Larix*, *Populus*) are strongly affected by sedimentation processes. Thus, to mitigate a record biased toward taxa with high pollen production, *seda*DNA analyses provide complementary data. A metabarcoding approach is used to investigate the P6-loop of the chloroplast trnL (UAA) intron, a universal, plant-specific and short barcode marker (Taberlet et al., 2007).

Records from vast continental areas such as Siberia are often rare and mostly cover short time intervals (McKay et al., 2018) and thus do not fully cover the relevant processes involved in vegetation dynamics and related diversity changes. Many Siberian vegetation records spanning the glacial-interglacial transition originate from permafrost archives, which vary considerably in their suitability for taphonomic preservation. *seda*DNA taphonomy is still under investigation but as *seda*DNA is bounded to sediments (such as clay and silt), it could also leach, especially in permafrost sediments (Edwards, 2020). Therefore, permafrost *seda*DNA can strongly fluctuate, hampering diversity inferences. *seda*DNA from lake archives, where conditions are more stable, are rare in Siberia.

Our study explores vegetation compositional changes inferred from both pollen and *seda*DNA analyses and richness changes inferred from *seda*DNA analyses from a sediment core from Lake Bolshoe Toko located in southeastern Yakutia (Siberia). Very little information is currently available from this area. Our main objectives are to fill this knowledge gap by (1) investigating vegetation composition changes in the lake catchment and region and (2) estimating taxa richness changes during the transition between the last glacial and the Holocene interglacial. By investigating past vegetation changes during major climatic transitions, such as the transition between the late Pleistocene to the Holocene, we may understand better how vegetation composition and diversity will change under current global warming.

2.3 MATERIALS AND METHODS

2.3.1 Geographical settings

The study area is located in the Sakha Republic (southeastern Yakutia, Russia, **Figure 2.1A**). The oligotrophic freshwater Lake Bolshoe Toko (56°05’ N, 130°90’ E, 903 m above sea level) is 15.4 km long and 7.4 km wide, with a maximum water depth of about 80 m and surface area of 82.6 km2 [for details see Biskaborn et al. (2019a)]. The lake lies in a depression of tectonic and glacial origin, on the northern flank of the eastern Stanovoy Mountain Range (Imaeva et al., 2009). At its northeastern margins, the lake is bordered by moraines of three different glacials (Kornilov, 1962).

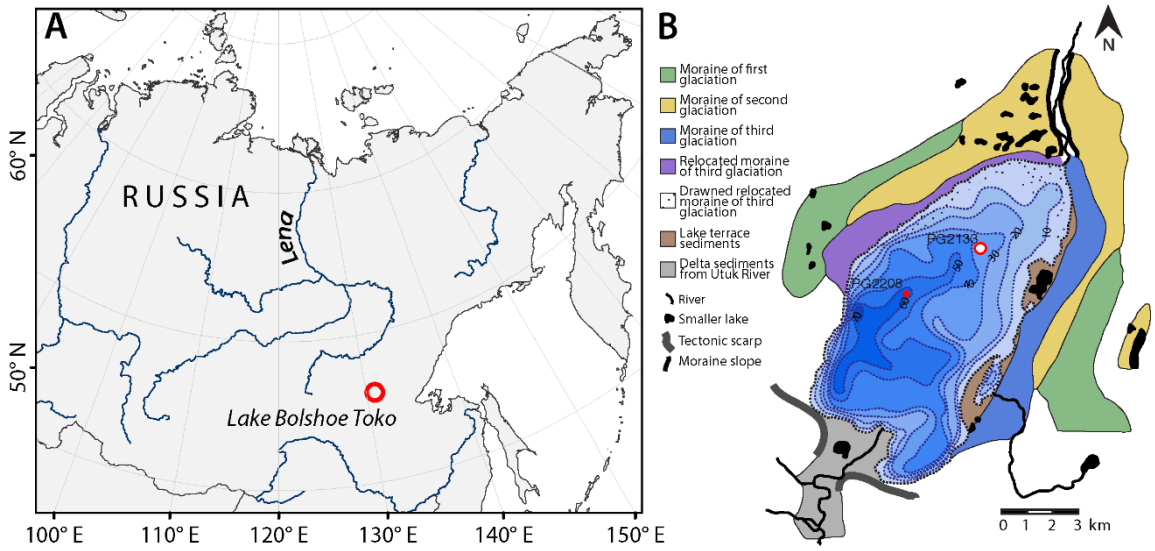


Figure 2.1 | Study area. (A) Location of Lake Bolshoe Toko in southeastern Yakutia, Siberia. **(B)** Bathymetrical map with geological information and position of the sediment core PG2133 used for vegetation analyses in the northern part of the basin. Sediment core PG2208 was used to correlate Holocene ages between cores.

According to the data from the meteorological station, "Toko", located ca. 10 km northeast of Lake Bolshoe Toko, mean annual air temperature is 11.2°C, ranging from -65°C in January to +34°C in July. Annual precipitation varies from 276 to 579 mm (Konstantinov, 2000). However, measurements taken directly at the lake suggest the influence of cold water from the Stanovoy Range that might cause slightly lower summer temperatures and thick ice during winter at the site.

Soil cover in the study area is thin and contains large amounts of gravel. The formation of soil is very slow because of deep seasonal freezing, short summers, and permafrost (Shahgedanova, 2002). The lake area belongs to the Aldan Floristic Region (Kuznetsova et al., 2010). Northern taiga dominates the study area. Forests consist of *Larix cajanderi* and *L. gmelinii* often accompanied by *Picea obovata*, *P. jezoensis*, and *Pinus sylvestris* (Konstantinov, 2000; Kuznetsova et al., 2010). The bedrock composition has an important influence on the vegetation: larch (*Larix*) and pine (*Pinus*) forest with a moss and lichen cover are especially common on the acidic metamorphic rocks, while in riparian areas small birch forests often form a belt around the mixed forest (Kuznetsova et al., 2010).

2.3.2 Fieldwork and subsampling

Fieldwork was carried out by a German-Russian expedition in spring 2013. Sediment cores were recovered using a UWITEC piston coring system operated from the ice in the northern and central part of the lake (**Figure 2.1B**). This study is based on sediment core PG2133. The ice thickness was about 1 m and the measured water depth at this sampling point (56°04.582' N, 130°54.948' E) was 26 m. The 375 cm core was then cut into 1 m long pieces, transported unfrozen in plastic tubes to the laboratories of the Alfred Wegener Institute for Polar and –Marine Research in Potsdam, Germany and stored there at 4°C until the subsampling for pollen and DNA analysis.

In December 2013 the core segments were cut into two halves. One was stored as an archive at 4°C and the other was scanned to correlate any overlaps. The opening and subsampling of the half-core for *seda*DNA analysis was performed in the climate chamber of the Helmholtz Center Potsdam - German Research Center for Geosciences (GFZ) at 10°C, where no molecular genetic studies take place to prevent contamination with modern DNA. All surfaces in the climate chamber were cleaned with DNA Exitus Plus™ (VWR, Germany) and demineralized water before working on the core. The sampling tools such as knives, scalpels, and their holders were cleaned before the taking of each sample following the recommendations of Champlot et al. (2010).

About 3 mm of each sample slice that was in touch with the plastic tube or the thin foil that covered the half-core was removed using a sterile scalpel as these parts cannot be considered sterile (Parducci et al., 2017). The samples for analysis were taken with the aid of 5 ml disposable syringes, in which the anterior cap had previously been cut off. A DNA sample was taken every two cm, giving 188 DNA samples over the 375 cm core length. In addition, 188 pollen samples were collected. The collected DNA samples were then stored at -20°C.

2.3.3 Core splicing and dating

The studied sediment core PG2133 was retrieved from the lake bottom as overlapping parts including a short core (PG2133-4) by gravity coring for undisturbed surface layers and two overlapping long-cores (PG2133-2, PG2133-3). The overlapping parts were spliced into one composite record by correlation of sediment layers using elemental (Zr) composition. We used geochemical variability in the stratigraphy validated by radiocarbon values to

correlate the upper part of this core with the parallel core from the deeper central location PG2208 (**Figure 2.1B**). The higher resolution and more reliably dated Holocene part from PG2208 provides a precise delimitation of the Pleistocene-Holocene transition in core PG2133 (**Figure 2.2**). The transfer of the age-depth relation from PG2208 onto PG2133 younger than the last reliable tie-point (8400-8440 14C yrs BP, see **Supplementary Table 1.1**) is based on the assumption that sedimentation rates at both coring sites after the Late Holocene were stable until today and minor changes would have affected both coring sites in a similar direction.

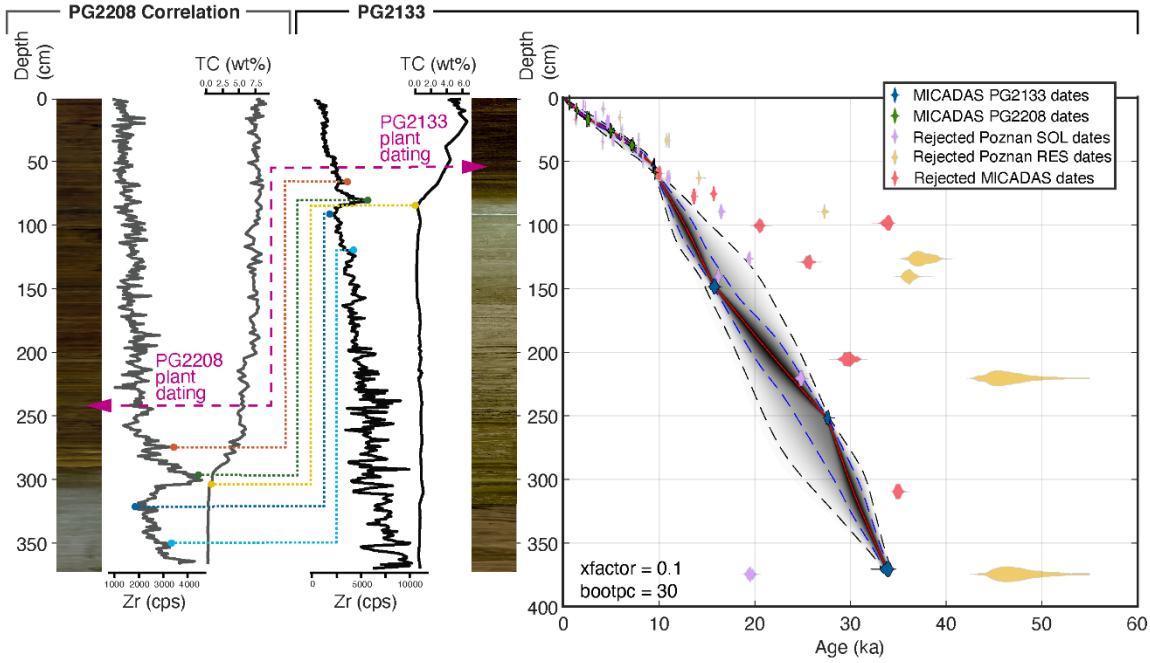


Figure 2.2 | Age-depth model based on 22 radiocarbon dated samples from sediment cores PG2133, PG2208, and two surface samples (PG2112, PG2119) from Lake Bolshoe Toko, southeastern Yakutia, Siberia. Age-depth model created with the MATLAB software package 'Undatable' (Lougheed and Obrochta, 2019) and the IntCal20 calibration curve (Reimer et al., 2020). Modeled ages are expressed as calibrated thousand years (ka) before present. The Holocene part was correlated between sediment core PG2208, retrieved from the deeper central part of the lake, and sediment core PG2133. Correlation was performed using Zr values from XRF core scanner data shown as counts per seconds (cps) and the concentrations of total carbon (wt%). The purple arrows indicate the position of plant remains in photographs of the sediment cores, dated to similar ages found at different depths of both cores. Dashed lines indicate tie points between cores used to report age information from the higher resolution Holocene part from core PG2208 onto PG2133 used in this study.

Accelerator Mass Spectrometry (AMS) radiocarbon analysis was carried out on bulk sediment material and woody plant remains at the Pozna´n Radiocarbon Laboratory, Poland, and at the MICADAS radiocarbon laboratory, AWI Bremerhaven. In total 49 samples were measured for radiocarbon including 24 samples from PG2133, 23 samples from PG2208, and two surface samples in proximity of these cores. All samples were bulk sediment samples, with the exception of two plant remains of the same age, one from each

core used for correlation between 242.5 cm (PG2208) and 53.5 cm (PG2133). The results from Pozna´n and MICADAS labs are partly inconsistent due to different pretreatment methods. Pozna´n either measured the humic acid fraction ("SOL", soluble in NaOH) and/or the humin fraction ("RES", residual, i.e., alkali-insoluble fraction). MICADAS measured following their standard acid treatment method outlined in detail in Vyse et al. (2020). Due to the large deviation between SOL and RES samples of the bulk samples from Pozna´n compared to a better down-core consistency of MICADAS samples, we used MICADAS bulk samples, and the dating of plant remains from Pozna´n for age-depth modeling. We excluded samples that are influenced by old organic carbon (Biskaborn et al., 2019a) detected by deviation from an approximate linear relationship. We used the surface samples PG2112 for reservoir correction (1203 yr) of PG2133 and PG2119 for reservoir correction (762 yr) of PG2208 (**Supplementary Table 1.1**). For modeling we used the open-source MATLAB software package 'Undatable' (Lougheed and Obrochta, 2019), version 1.3.1 with the settings: nsim = 105, bootpc = 30, xfactor = 0.1 and the IntCal20 calibration curve (Reimer et al., 2020).

2.3.4 Sediment-geochemical analyses

To gain information on the organic matter chemistry we analysed 75 samples for total carbon (TC) and nitrogen (TN), and 69 samples for total organic carbon (TOC). Freeze-dried and milled sample material was measured in tin boats using an elemental analyzer (Elementar Vario EL III) for TC and TN. For TOC, where enough material was left, material from the same sample boxes were placed into small crucibles and measured with a Vario MAX C analyzer. The per cent by weight of TOC and TN were used to calculate the TOC/TN_{atomic} ratio following Meyers and Teranes (2002). However, below 80 cm in core PG2133, TN contents were less than 0.1 wt% and could not be detected by the device. Therefore, for these samples with TN <0.1, we used 0.09 as an estimate to calculate the TOC/TNatomic ratios, marking the samples as "uncertain" in the down-core plot (**Figure 2.3**). Sediment grain-size distribution was generated using a laser to determine the grain sizes of 65 samples. We removed organic compounds by soaking samples in hydrogen peroxide (3%) for 5 weeks, while the pH was kept neutral using drops of ammonia and acetic acid 3 times per week. After washing by adding water, centrifuging, and decanting, we added sodium pyrophosphate to homogenize samples in an overhead shaker. Particle size was measured by laser diffractometry using a Malvern Mastersizer 3000. We used averaged values after a minimum of 9 measurements. Clay, silt, sand, and gravel fractions

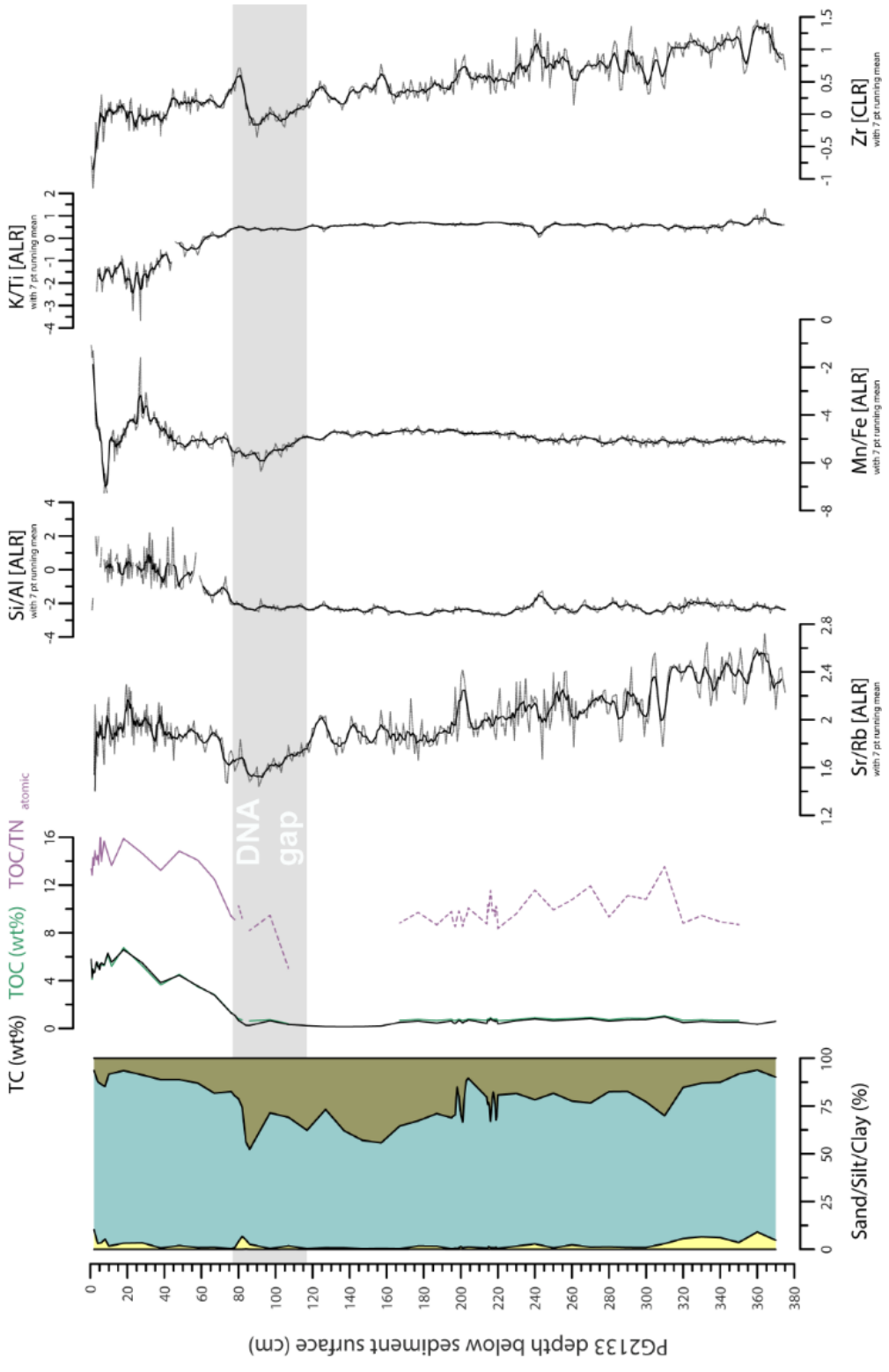


Figure 2.3 | Sedimentary and geochemical composition of sediment core PG2133 including grain-size distribution, carbon concentrations, and XRF elements. Sand, silt, and clay proportions are reported over the entire PG2133 core. Total carbon (TC), total organic carbon (TOC), and the ratio of TOC to nitrogen (TOC/TN_{atomic}) are reported to indicate input of terrestrial plant material relative to aquatic algae (Meyers, 2003; Biskaborn et al., 2012). The Strontium (Sr) over Rubidium (Rb) ratio and Zirconium (Zr) indicate relative detrital input related to coarse grain size (Biskaborn et al., 2013). The Silicon (Si) over Aluminum (Al) ratio indicates the occurrence of diatom valves (Vyse et al., 2020). The Manganese (Mn) and Iron (Fe) ratio indicate redox dynamics associated with bottom water oxygenation (Melles et al., 2012; Naeher et al., 2013), which with the Potassium (K) over Titanium (Ti) ratio, provides information on physical versus chemical weathering in the catchment (Vyse et al., 2020). Shaded area indicates a gap in the DNA data.

were calculated as the sums of the volume percentages using boundaries <2 µm, 2–63 µm, 63 µm–2 mm, and 2–63 mm, respectively. To gain information on the inorganic elemental composition of the sediments, non-destructive X-ray fluorescence was performed on one half of each core section including overlapping parts using an AVAATECH XRF core scanner at AWI in Bremerhaven. Light elements including Al, Si, K, Ti, Mn, and Fe, were measured at 10 kV using a Rh X-ray tube, 1.2 mA, 12 s count time, and no filter. Heavier elements including Rb, Sr, and Zr, were measured at 30 kV using a Rh X-ray tube, 1.5 mA, 15 s count time, and a thick Pd-Filter. To account for data compositional effects, scanning results originally expressed in counts per second (cps) were additive log-ratio (ALR) and centered logratio transformed (Weltje and Tjallingii, 2008) for ratios and single elements, respectively (**Figure 2.3**). We used original cps values to correlate between cores PG2133 and PG2208 to prevent differences in XRF device settings between cores influencing other elements included in the CLR transformation and thus bias a core-to-core correlation.

2.3.5 Pollen analysis

Sixty-seven samples of 3 g (wet weight) were selected for pollen analysis. Standard preparation including KOH, HCl treatment, boiling with HF, and acetolysis was used to concentrate palynomorphs (Faegri and Iversen, 1989). A *Lycopodium* spore tablet (batch no. 1031; n = 20,848 ± 1460) was added to each sample to calculate pollen concentration (Stockmarr, 1971). Palynomorphs were identified using a light microscope (Zeiss Axioskop 2) under 400–600 × magnification. At least 300 terrestrial plant pollen grains were counted in each sample. Pollen atlases (Sokolovskaya, 1958; Kupriyanova and Alyoshina, 1972, 1978; Moore et al., 1991; Beug, 2004; Savelieva et al., 2013) and pollen reference collections at the Arctic and Antarctic Research Institute (Saint Petersburg) and the Alfred Wegener Institute were used for taxonomic identification of pollen and spores. Non-pollen palynomorphs (NPP) were determined according to van Geel (2001); van Geel et al. (1983), and van Geel and Aptroot (2006). Freshwater algae were determined using Jankovská and Komárek (2000) and Komárek and Jankovska (2003). Only the terrestrial pollen data are presented here, however, all data are archived in DRYAD at "10.5061/dryad.34tmpg4gz".

2.3.6 Molecular genetic preparation

DNA isolation and polymerase chain reaction (PCR) setup were performed in the paleogenetic laboratory of the Alfred-Wegener- Institute Helmholtz Center for Polar and

Marine Research in Potsdam, Germany. This lab is dedicated to ancient DNA isolation and PCR setup and is located in a building devoid of any molecular genetic lab work. The lab is cleaned frequently and subjected to nightly UV-irradiation. All laboratory work was performed in a UVC/T-M-AR DNA/RNA cleaner-box (BIOSAN, Latvia). DNA isolations and PCR setups were performed on different days using dedicated sets of pipettes and equipment. Further precautions to reduce contamination included UV-irradiation of 10× buffer, BSA, MgSO4, and DEPC-treated water for 10 min in a UV crosslinker in thin-walled PCR reaction tubes (recommendations of Champlot et al., 2010).

In total, 54 samples were selected for DNA extraction ranging from 1 to 374 cm depth. Six separate extractions were performed, each with nine samples and one control (blank). Total DNA was isolated from approximately 3 g of sample material using the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories, Inc., USA) added to 1.2 ml of C1 buffer (PowerMax® Soil DNA Isolation Kit, Mo Bio Laboratories, Inc., USA), 0.4 ml of 2 mg/mL Proteinase K (protK, VWR International, Germany) and 0.5 ml of 1 M dithiothreitol (DTT, VWR International, Germany). This mixture was homogenized for 10 minutes on a vibratory mixer (VortexGenie2, Mo Bio Laboratories, USA) and incubated overnight at 56°C while rotating. All subsequent steps were performed according to the instructions of the manufacturer Qiagen, except for the last step in which 800 mL elution buffer (C6 buffer) was applied to the filter membrane, incubated for 10 minutes at room temperature, and then centrifuged for 3 minutes at 2500x g. This step was performed twice ending up with a final volume of 1.6 ml.

The PCR reactions were performed using the trnL g and h primers (Taberlet et al., 2007). Both primers were modified on the 50 end by unique 8 bp tags which varied from each other in at least five base pairs to distinguish samples after sequencing (Binladen et al., 2007) and were additionally elongated by NNN tagging to improve cluster detection on the sequencing platform (De Barba et al., 2014). The PCR reactions contained 1.25U Platinum® Taq High Fidelity DNA Polymerase (Invitrogen, USA), 1× HiFi buffer, 2 m MMgSO4, 0.25 mM mixed dNTPs, 0.8 mg Bovine Serum Albumin (VWR, Germany), 0.2 mM of each primer, and 3 uL DNA in a final volume of 25 uL. PCRs were carried out in a Professional Basic Thermocycler (Biometra, Germany) with initial denaturation at 94°C for 5 min, followed by 50 cycles of 94°C for 30 s, 50°C for 30 s, 68°C for 30 s, and a final extension at 72°C for 10 min. The extraction blank and one no template control (NTC) was

included in each PCR to identify possible contamination during extraction and PCR setup. For each extraction sample, two PCR replicates with different primers were performed. PCR success was checked with gel-electrophoresis on 2% agarose (Carl Roth GmbH & Co. KG, Germany) gels were used. The PCR products were purified using the MinElute PCR Purification Kit (Qiagen, Germany), following the manufacturer's recommendations. Elution was carried out twice with the elution buffer to a final volume of 40 uL. The DNA concentrations were estimated with the dsDNA BR Assay and the Qubit 2.0 fluorometer (Invitrogen, USA) using 1 uL of the purified amplifications. To avoid bias based on differences in DNA concentration between samples, all replicates (except for samples 20, 86, and 106 where only one sample was used) were pooled in equimolar concentrations. All extraction blanks and NTCs (9 extraction blanks and 11 PCR NTCs) were included in the sequencing run, using a standardized volume of 10 mL, even though they were negative in the PCRs. The sequencing results of extraction blanks and PCR controls are reported in the DRYAD. Fasteris SA sequencing service (Switzerland) performed the paired-end sequencing on one-tenth of a full lane of the Illumina HiSeq platform (2 × 125 bp).

2.3.7 Processing of *seda*DNA data

Filtering, sorting, and taxonomic assignments of the sequences were performed with OBITools (Boyer et al., 2015). Using the function *illuminapairedend*, forward and reverse reads were aligned to produce merged sequences. Based on exact matches to their tag-combination, they were assigned to their samples using *ngsfilter*. Sequences shorter than 10bp were excluded with *obigrep* and duplicated sequences were merged with *obiuniq* while keeping the information to which sample the sequences originally belonged. Sequences with fewer than 10 read counts were excluded from the dataset as they are probably artifacts. *Obiclean* was used to exclude sequence variants probably attributable to PCR or sequencing errors (Boyer et al., 2015). Two reference databases were used for taxonomic assignments (Epp et al., 2015). The first one, ArctBorBryo, is based on the quality-checked and curated Arctic and Boreal vascular plant and bryophyte reference libraries (Sønstebø et al., 2010; Willerslev et al., 2014; Soininen et al., 2015). The second is based on the EMBL Nucleotide Database standard sequence release 133 (Kanz et al., 2005). The sequences were assigned to taxon names based on sequence similarity to each of the reference databases using the function *ecotag*.

The database that gave the best identity was selected with a priority for ArctBorBryo database if both gave the same identity, as it is a more specific database for boreal species. Only sequences assigned with a best identity of 100% and that are present in at least two replicates or samples are used in this study. The sequence types assigned with a 100% value to an unknown family level were removed. Taxa and sequences from the extraction and PCR NTCs were checked and removed from the dataset if their relative proportion in the controls was greater than 0.5%. The PCR replicates were merged to work with 54 samples. Sequences assigned to the same taxon were merged and those sequences that appeared fewer than 10 times in a sample were removed from it. Sequences were assigned to aquatic, bryophyte, or terrestrial taxa and only the terrestrial ones were selected for this study. Samples at 78, 82, 86, and 106 cm were represented by no terrestrial plant reads after these filtering steps.

Before proceeding to the data analysis, a rarefaction analysis to compare the richness at a similar level of sampling effort was performed based on the minimum number of observed sequence counts (n = 4119 (sample at 96 cm); highest sequence count is n = 329924, from 366 cm).

2.3.8 Statistical analysis and visualization

The paleogenetic and palynological datasets are archived in DRYAD at "10.5061/dryad.34tmpg4gz".

2.3.8.1 Composition analysis

To allow visual comparison of the different records and changes of abundance through time, percentage bar plots of taxa were produced for pollen and *seda*DNA datasets using the strat.plot function from the rioja R-package (Juggins, 2019). In addition, the summed proportions of tree, shrub, or herb taxa for *seda*DNA or tree/shrub and herb taxa for pollen were plotted. A constrained hierarchical clustering approach (CONISS) (Grimm, 1987) was performed with clusters constrained by depth/ages using vegdist from the vegan R-package (Oksanen et al., 2019) and chclust from the rioja R-package. The samples with 0 read counts were excluded from this analysis. Zones were determined by the broken-stick model (MacArthur, 1957; Bennett, 1996) using bstick from the vegan R-package.

A principal component analysis (PCA) was run on the double square-root transformed relative proportions for *seda*DNA and square-root transformed pollen data using the rda function in the vegan R-package (Oksanen et al., 2019) to portray the major structure in the multivariate dataset. For the *seda*DNA record, samples with 3 or fewer taxa detected were removed before performing the PCA as those samples might not be representative of the vegetation composition. Principle component 1 and 2 (PC1 and PC2) axes scores were extracted and visualized in biplots. For better visibility, only the name of terrestrial taxa that explain most variance are printed.

To allow direct comparison of the records, pollen and *seda*DNA taxa were harmonized to the lowest common taxonomic level (e.g., Ericaceae taxa were transformed to Ericales). PCAs were rerun for each record including only those samples that are present in both datasets. To compare the fit between sample and species scores of the first PCA axes for pollen and *seda*DNA, Procrustes rotation analyses were performed (Peres-Neto and Jackson, 2001) and the significance of similarity between the datasets was tested by PROTEST (Jackson, 1995).

2.3.8.2 Richness analysis

Taxa richness is here defined as the number of different taxa counted per sample. It was measured for each depth/age from the rarefied DNA results with the ChaoRichness command from the iNEXT R-package (Hsieh et al., 2016). To test the significance of the difference in the average taxa richness between different zones determined by the broken-stick model, a non-parametric Wilcoxon rank sum test was performed.

After resampling the two most common families of the DNAZ I to 1000 reads (Saxifragaceae and Asteraceae), the same test was performed to compare the taxa richness within those families and their impact on the DNAZ I richness.

Statistical analyses and age-depth modeling were performed in R v. 3.6.1 (R Core Team, 2019). Statistics were mainly performed using the packages “vegan” version 2.5-6 (Oksanen et al., 2019), “rioja” version 0.9.21 (Juggins, 2019), “iNEXT” version 2.0.20 (Hsieh et al., 2016) and “analog” version 0.17-3 (Simpson, 2007; Simpson and Oksanen, 2016).

2.4 RESULTS

2.4.1 Age model

Modeled dates for PG2133 are expressed as calibrated years before present relative to 1950 CE (cal. yr BP). The Pleistocene-Holocene boundary detected by the main lithological transition found at ca. 80–82 cm revealed an age between 10,765 and 11,924 cal. yr BP (sigma 2 range). The maximum modeled age in the core is 33,478–34,126 cal. yr BP at 370.5 cm composite depth below sediment surface (**Figure 2.2**). Accordingly, the mean sedimentation rate during the Holocene was 0.007 cm a^{-1} and 0.01 cm a^{-1} during the Pleistocene part of the core. Differences in sedimentation rates between distant sediment cores shown in **Figure 2.2** is common in large lakes (Jenny et al., 2014). The drop in sedimentation rate after the deglaciation period is likely a result of decreased clastic input from the catchment as indicated by decreasing Zr (Biskaborn et al., 2019a). The mean sigma 2 uncertainty ranges in the Holocene and Pleistocene are estimated as 517.6 and 2378.8 yr, respectively. In this study, for simplicity, we use mean values to estimate the timing of transitions between pollen zones.

2.4.2 Sediment-geochemical core composition

The Pleistocene section below 80 cm in sediment core PG2133 reveals organic-poor clayey silt (**Figure 2.3**). Here the mean value for gravel, sand, silt, and clay is <0.01, 1.8, 73.3, and 24.9%, respectively. The TC and TOC contents are 0.6 and 0.7%, respectively, with a negligible difference resulting from measurements with different devices (TC-TOC difference over the whole core is mean -0.06 wt%, min. -0.48 wt%, and max. 0.35 wt%). The TN/TOCatomic ratios remain low at around 9.6 (note: this section is based on estimated TN values because the nitrogen values remained under the detection limit, see Materials and Methods section). Zr (both cps values, **Figure 2.2** and CLR values, **Figure 2.3**) has generally higher values in the Pleistocene section below 80 cm, with a strong peak that correlates to a small sand peak in the transition zone between 77 and 83 cm. K/Ti has uniformly high values in the Pleistocene, whereas Sr/Rb fluctuates with highest values at the bottom of the core and lowest values at the Pleistocene–Holocene transition. Si/Al is low with only minor excursions in the Pleistocene. Mn/Fe is also low but slightly decreases at 113 cm toward the top of the core.

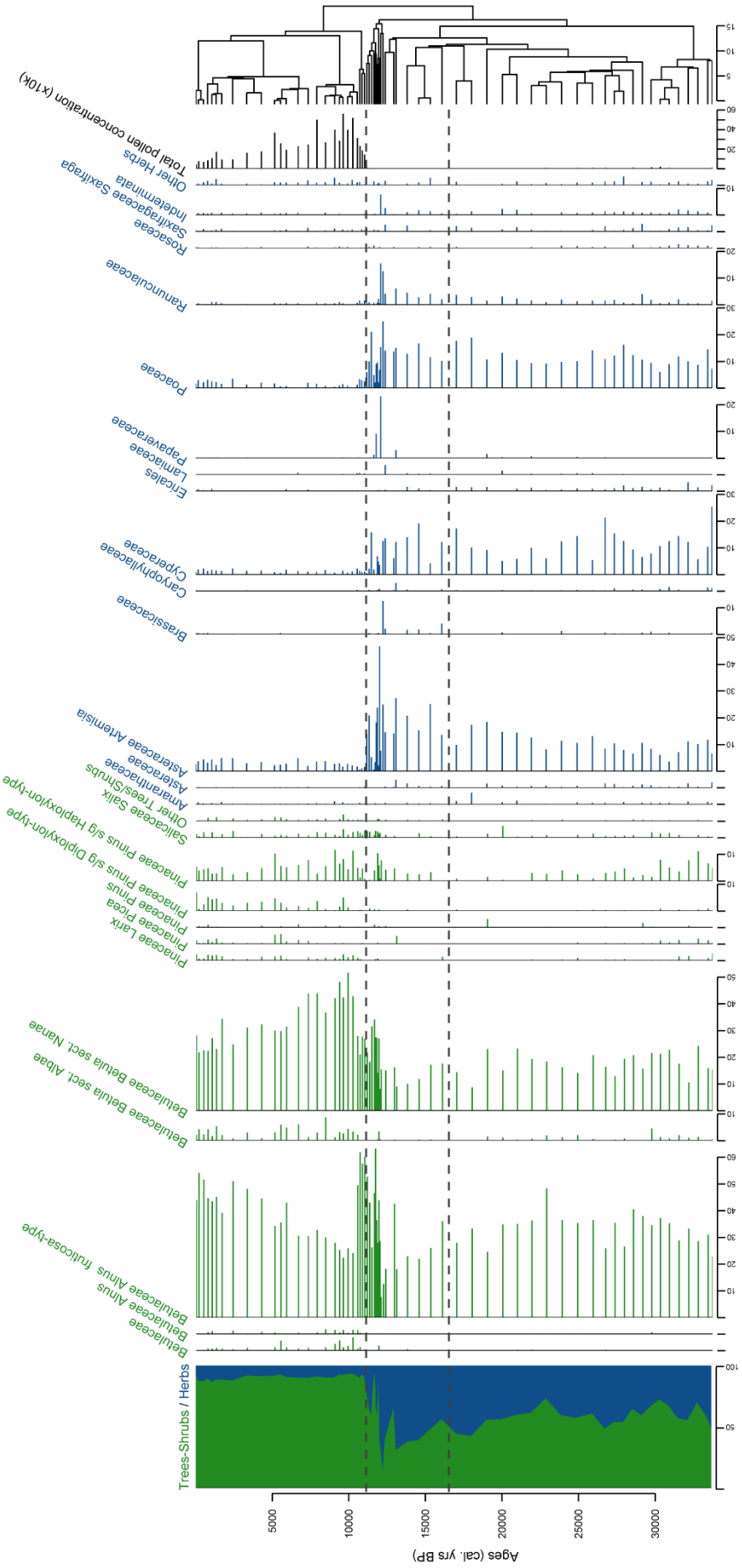


Figure 2.4 | Stratigraphic plot of terrestrial pollen with relative percentages of the taxa in each sample as horizontal bars, ratio of tree/shrub taxa vs herb taxa, and CONISS dendrogram. The taxa are sorted at genus level for trees and shrubs and at family level for herbs. Three pollen zones (PZ) are shown derived from the CONISS analysis. Samples from sediment core PG2133 are plotted against calibrated ages given in cal yr BP.

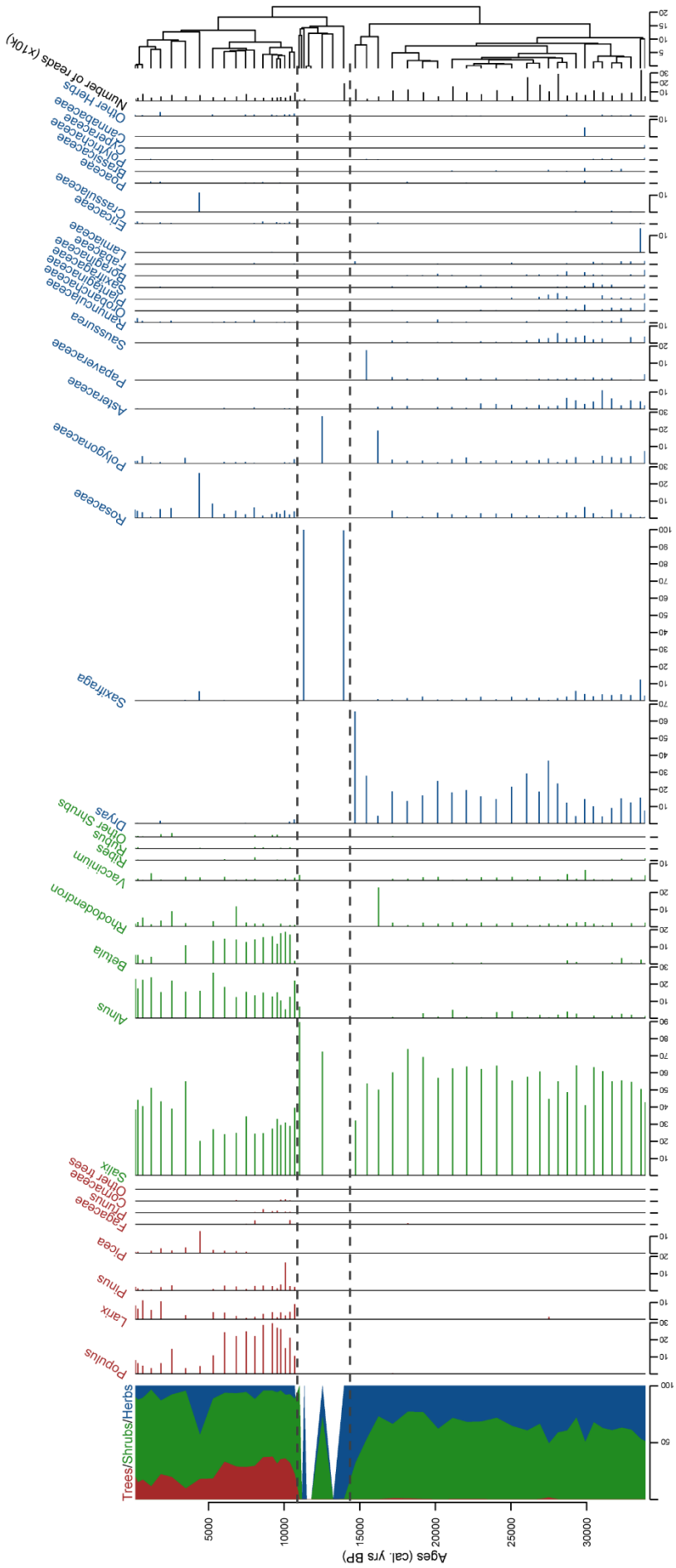


Figure 2.5 | Stratigraphic plot of terrestrial *sed*aDNA with relative percentages of the taxa in each sample as horizontal bars, ratio of tree/shrub taxa vs herb taxa, and CONISS dendrogram. The taxa are sorted at genus level for trees and shrubs and at family level for herbs. Three *sed*aDNA zones (DNAZ) are shown, derived from the CONISS analysis. Calibrated ages from the age–depth model are given in calendar years before present (cal yr BP).

In the Holocene core section, the grain-size composition is slightly less clayey, but still dominated by silt (mean values: sand 2.9, silt 84.1, clay 13.1%) with increased carbon content (mean value: TC 4.5 wt%) and TOC/TNatomic ratios (mean value: 13.7), which is within the ranges reported from Bolshoe Toko short cores covering the past few decades (Biskaborn et al., 2021). Sr/Rb, Si/Al, and Mn/Fe increase in the early Holocene, whereas K/Ti decreases. Between ca. 30 cm and the top of the core Mn/Fe decreases after a maximum, and K/Ti increases after a minimum. Close to the sediment surface the values of Sr/Rb and Mn/Fe exhibit strong fluctuations, which could reflect the effect of higher water content in the unconsolidated surface sediments.

2.4.3 Pollen stratigraphy

A total of 111 different terrestrial pollen, spores, and NPP taxa were identified in the 67 samples. Only terrestrial pollen taxa are presented on the pollen diagram (**Figure 2.4**) for a direct comparison with the *seda*DNA diagram (**Figure 2.5**). The entire pollen dataset is presented in **Supplementary Figure 1.1**. The pollen assemblages can be subdivided into 3 main pollen zones (PZ) according to CONISS analysis.

PZ I (ca. 374–96 cm,>33,800–12,474 cal yr BP is characterized by rather low pollen concentrations varying between 486 and 19,795 pollen grains per gram (gr g^{-1}) with 6,027 on average. Tree and shrub pollen average 55% (varying from 30.3 to 72.7%). The most common shrub pollen taxa are *Alnus fruticosa*-type (31.8% on average) and *Betula* sect. *Nanae* (16.8% on average). Herb pollen taxa (45% on average, varying from 27.3 to 69.3%) are mostly represented by *Artemisia* (12.3% on average), Poaceae (11.9% on average), and Cyperaceae (11.4% on average).

Pollen concentration in PZ II (ca. 94–80 cm, 12,329–11,268 cal yr BP) is even lower than in PZ I (2,479 gr g^{-1} on average, varying from 201 to 9,677 gr g^{-1}). Tree and shrub pollen taxa dominate the spectra of most samples (up to 93.6%, on average 61%), while herb taxa vary from 6.4 to 87.5% (on average 39%). *Alnus fruticosa*-type (35.3% on average) and *Betula* sect. *Nanae* (20.5% on average) are still the most dominant pollen taxa. The most represented herb pollen taxa are *Artemisia* (15.2% on average) and Poaceae (10.2%), with Cyperaceae (4.4%), Papaveraceae (3%), and Ranunculaceae (3%) also represented.

Pollen concentration drastically increases (up to 558,726 gr g^{-1}, on average 244,074 gr g^{-1}) at the beginning of PZ III (ca. 78–0 cm, 11,112 cal yr BP to present), but gradually decreases after 24 cm (4,373 cal yr BP). Tree and shrub pollen taxa dominate the spectra (on average 90%). The most common shrub pollen taxa are *Alnus fruticosa*-type (38.1% on average) and *Betula* sect. *Nanae* (33.8%). *Alnus fruticosa*-type percentages are highest between 78 and 72 cm (11,112 to 10,643 cal yr BP). *Betula* sect. *Nanae* pollen percentages increase around 72 cm (10,643 cal yr BP) and gradually decrease upward. More tree pollen taxa are also detected in this zone: *Larix* (1.3% on average), *Pinus* s/g *Diploxylon*-type (2.6%) and *P.* s/g *Haploxylon*-type (5.6%). The herb taxa are less numerous than in the lower PZs (9.9% on average, varying from 7.1 to 13.4%). They are dominated by *Artemisia* (3.2% on average), Poaceae (1.6%), and Cyperaceae (1.4%).

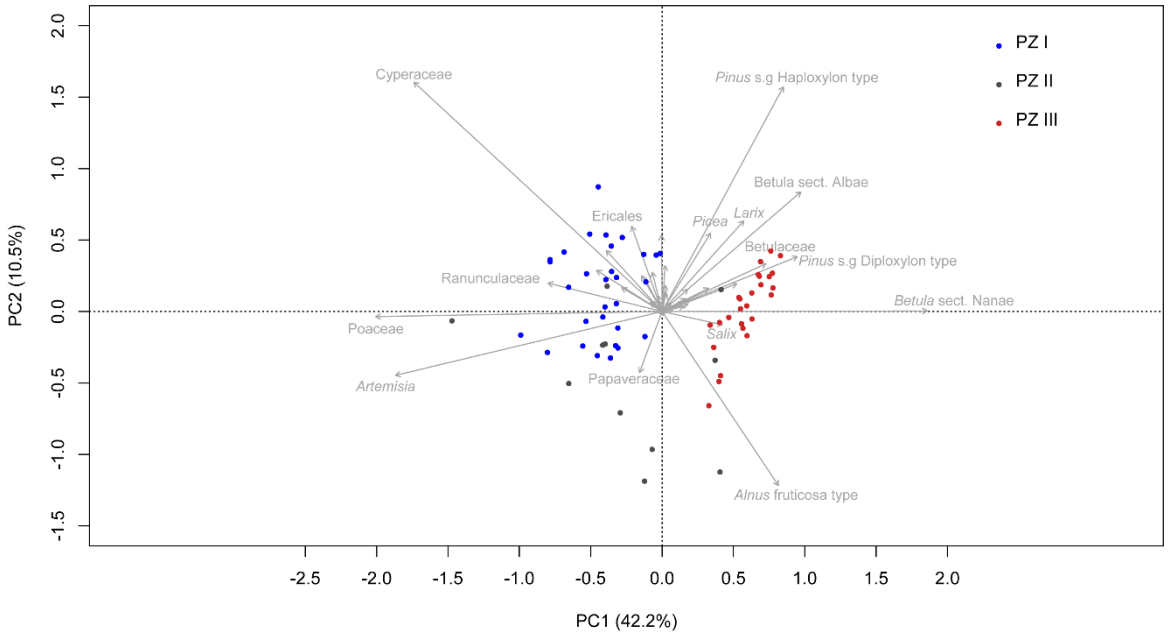


Figure 2.6 | Principal component analysis (PCA) biplot of terrestrial pollen taxa. Only the names of the 15 taxa explaining the most variance are printed. Samples are colored according to their corresponding zone. The explained variances of the two principal components (PC1 & 2) are shown in brackets.

The first two axes of the PCA (**Figure 2.6**) explain 52.7% of the variance in the dataset, supporting the zonation of the broken-stick analysis. Samples from PZ I are mostly in the left-hand quadrants, represented mainly by Cyperaceae, Poaceae, and *Artemisia*. The samples from PZ II are spread across PC1 and intermix with samples from PZ I and PZ III. Samples from PZ III are mostly located in the right-hand quadrants and are represented by taxa such as *Pinus* s/g *Haploxylon*-type, *Betula* sect. *Albae*, *Pinus* s/g *Diploxylon*-type, *Betula* sect. *Nanae*, and *Alnus fruticosa*-type as well as *Larix*, *Picea*, and *Salix*.

2.4.4 *seda*DNA composition

In total, 10,172,402 reads were assigned to 3,007 different sequence types with more than 10 counts. 5,364,355 were assigned with 100% identity to 322 sequence types. In total, 8 sequence types were present in the controls. There were three different sequence types with 274,319 (5.1% from total dataset) total reads in 3 of the 9 PCR NTCs and 6 sequence types with 904,427 (16.9% from total dataset) total reads in 6 of the 11 extraction blanks. Two sequence types assigned to Brassicaceae were removed from the dataset as all their reads were present in the controls; as was one sequence type assigned to *Taxus baccata* for which 91% of its reads were present in one NTC and the rest in only one other replicate. Five other sequence types (assigned to *Bistorta vivipara*, Salicaceae, Asteraceae, *Saxifraga cernua*, and *Pinus*) were kept for the analysis after examination. For *Bistorta vivipara*, 1 read was present in the controls, for Salicaceae, 7 reads (<0.001%) in 4 controls, and 1 read for Asteraceae and *Saxifraga cernua*.

For *Pinus*, 388,820 and 1 reads (94%) were detected in 2 extraction blanks, but is also present in 35 other PCR replicates in coherent proportions with no link to the problematic extraction blanks.

The sequence type assigned to *Convallaria majalis* with 14,702 reads was deleted from the dataset as it is a common contaminant and 9 sequence types not even assigned to the family level at best were removed for a total of 633 reads. Furthermore, 64 sequence types with a total of 9,258 reads that appear in only one PCR product were deleted. After the replicates were merged and those with the same assignment merged (54 samples), 181 taxa were detected with 4,159,565 reads in total and 4,157,784 reads when the taxa with fewer than 10 reads were deleted.

In total, 4,133,309 sequence counts were assigned to 153 terrestrial plant taxa across 50 samples: 82 to species level, 55 to genus level, and 16 only to family level at best. The most common families among the dataset that represent 90% of the reads are Salicaceae (52%), Rosaceae (13%), Betulaceae (12%), Saxifragaceae (6%), Ericaceae (4%), Pinaceae (3%), and Asteraceae (2%).

The terrestrial plant *seda*DNA dataset can be divided into three zones according to the CONISS clustering and broken stick analysis (**Figure 2.5**). Because of the high number of

taxa detected using DNA, tree and shrub taxa are plotted at the genus level and herbs at the family level. The detailed dataset is presented in **Supplementary Figure 1.2**.

DNAZ I (ca. 374–126 cm, >33,800–14,660 cal yr BP) with 26 samples approximately matches PZ I. This zone is represented mostly by shrub taxa (from 32.3 to 76.6% with 62.2% on average) with *Salix* (56.2% on average) dominant. Herb taxa also have high proportions (23 to 67.7% with 37.6% on average) with *Dryas* (18.3% on average) and other Rosaceae (2% on average) as well as Asteraceae (4.4%) represented by *Artemisia* and *Saussurea*. Tree taxa are present in DNAZ I at low percentages (0 to 1.6% with 0.2% on average).

DNAZ II (ca. 116–76 cm, 13,901–10,956 cal yr BP) is represented by 8 samples: 4 samples are represented by very few reads and 4 had no sequence types at all but were added anyway. Only a few taxa are found in the zone: *Salix* and Polygonaceae at 96 cm (12,474 cal yr BP), 2 taxa from Saxifragaceae in the samples at 116 and 80 cm (13,901 and 11,268 cal yr BP, respectively), and *Alnus*, *Salix*, and *Vaccinium* in the sample at 76 cm (10,956cal yr BP). The zone has a high proportion of herbs (up to 100% with 56.8% on average), with Polygonaceae (6.9% on average) and *Saxifraga* (49.9%) as the most abundant taxa. Shrub taxa are also found (43.2% on average, ranging from 0 to 100%) mostly represented by *Salix* (40.6% on average), but not a single tree taxon.

The uppermost zone, DNAZ III (ca. 72–0 cm, 10,643 cal yr BP to present) has 20 samples and has more tree taxa than the lower zones (24.9% on average, ranging from 9.3 to 37.5%), mostly *Populus* (15.7%), *Larix* (4.3%), *Pinus* (2.6%), and *Picea* (1.5%). Shrubs are also highly represented (38.8 to 86.1% with 65.3% on average) with 34.1% *Salix* on average, 16.8% *Alnus*, 9.9% *Betula*, and 2.5% *Rhododendron*. Some herb taxa (3.7 to 43.3% with 9.7% on average) are also found. The most represented herb taxa are Rosaceae (4.7% on average), Polygonaceae (1%), Crassulaceae (0.6%), Ranunculaceae (0.6%), and Saxifragaceae (0.6%). In DNAZ III, notable variations can be seen and this zone can be divided into two subzones. In the lower part, DNAZ IIIa (72–28 cm, 10,643–5,242 cal yr BP), trees are abundant (30.6% on average) with a high proportion of *Populus* (21.9%). At 24 cm (4,373 cal yr BP), a peak in *Picea* (13%) is observed as well as a peak in Rosaceae (26.2%), Crassulaceae (11.3%), and *Saxifraga* (4.6%). In the upper part, DNAZ IIIb (20–1 cm, 3,432–111 cal yr BP), fewer tree taxa (16.2% on average) are present, with less *Populus* (6.7%) but more *Larix* (up to 6.4% vs 3.5% in DNAZ IIIa) and more shrubs

(74.9%) such as Salicaceae (44.6%). More *Alnus* (20% vs 15% in DNAZ IIIa) and less *Betula* (4.1% vs 14.1% in DNAZ IIIa) are also noticeable.

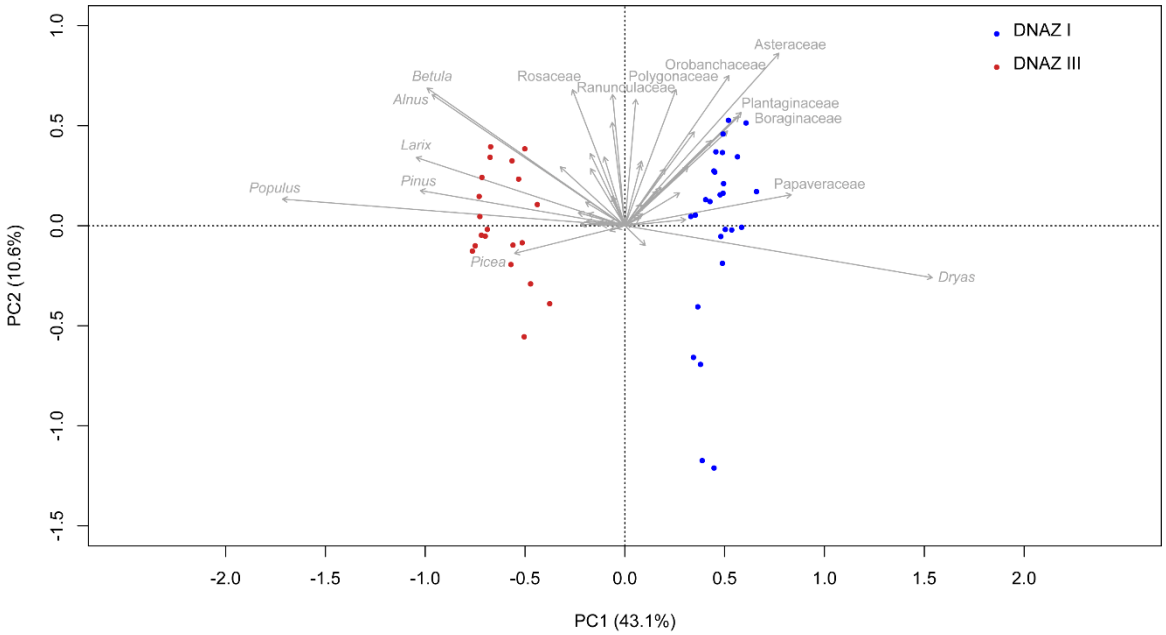


Figure 2.7 | Principal component analysis (PCA) biplot of *seda*DNA terrestrial taxa. Only the names of the 15 taxa explaining the most variance are printed. Samples are colored according to their corresponding zone. The explained variances of the two principal components (PC1 and PC2) are shown in brackets.

The first two axes of the PCA (**Figure 2.7**) explain 53.7% of the variance in the dataset, supporting the differentiation between DNAZ I and DNAZ III from the broken-stick analysis. Samples from DNAZ I are mostly in the right-hand quadrants and are mainly represented by *Dryas*, Asteraceae, Plantaginaceae, Boraginaceae, and Papaveraceae. The samples from DNAZ III are mostly located on the left-hand quadrants and are represented by taxa such as *Betula*, *Alnus*, *Larix*, *Pinus*, *Picea*, and *Populus*.

Table 2.1 | Results of Procrustes and PROTEST analyses. They indicate significant fit among all pairwise comparisons of datasets obtained by pollen and sedaDNA analyses.

Pollen vs *seda*DNA	p-value	r	m12	RMSE
Ages	0.001	0.6812	0.536	0.674201
Taxa	0.001	0.4605	0.7879	0.854638

RMSE, root mean square error

2.4.5 Comparison between pollen and *seda*DNA

Results of the Procrustes rotation analyses and PROTEST (**Table 2.1**) indicate significant concordance between the pollen and *seda*DNA PCA sample scores (m12 = 0.536, p = 0.001) as well as taxa (m12 = 0.7879, p = 0.001). Age residuals between 22,981 and 10,330

cal yr BP are mostly above average (**Figure 2.8A**) showing weak similarity for this time between both proxies. The sample dated to 33,815 cal yr BP has the lowest similarity between the *seda*DNA and pollen datasets. The sample dated to 4,373 cal yr BP is similarly weak. Taxa residuals of *Salix*, *Populus*, and *Dryas* have the highest above-average values implying very weak similarity for these taxa between the pollen and *seda*DNA datasets (**Figure 2.8B**). Furthermore, *Artemisia*, *Alnus*, Cyperaceae, Ericales, Papaveraceae, *Larix*, Polygonaceae, Rosaceae, and Saxifragaceae are also well above average and therefore dissimilar between both proxies.

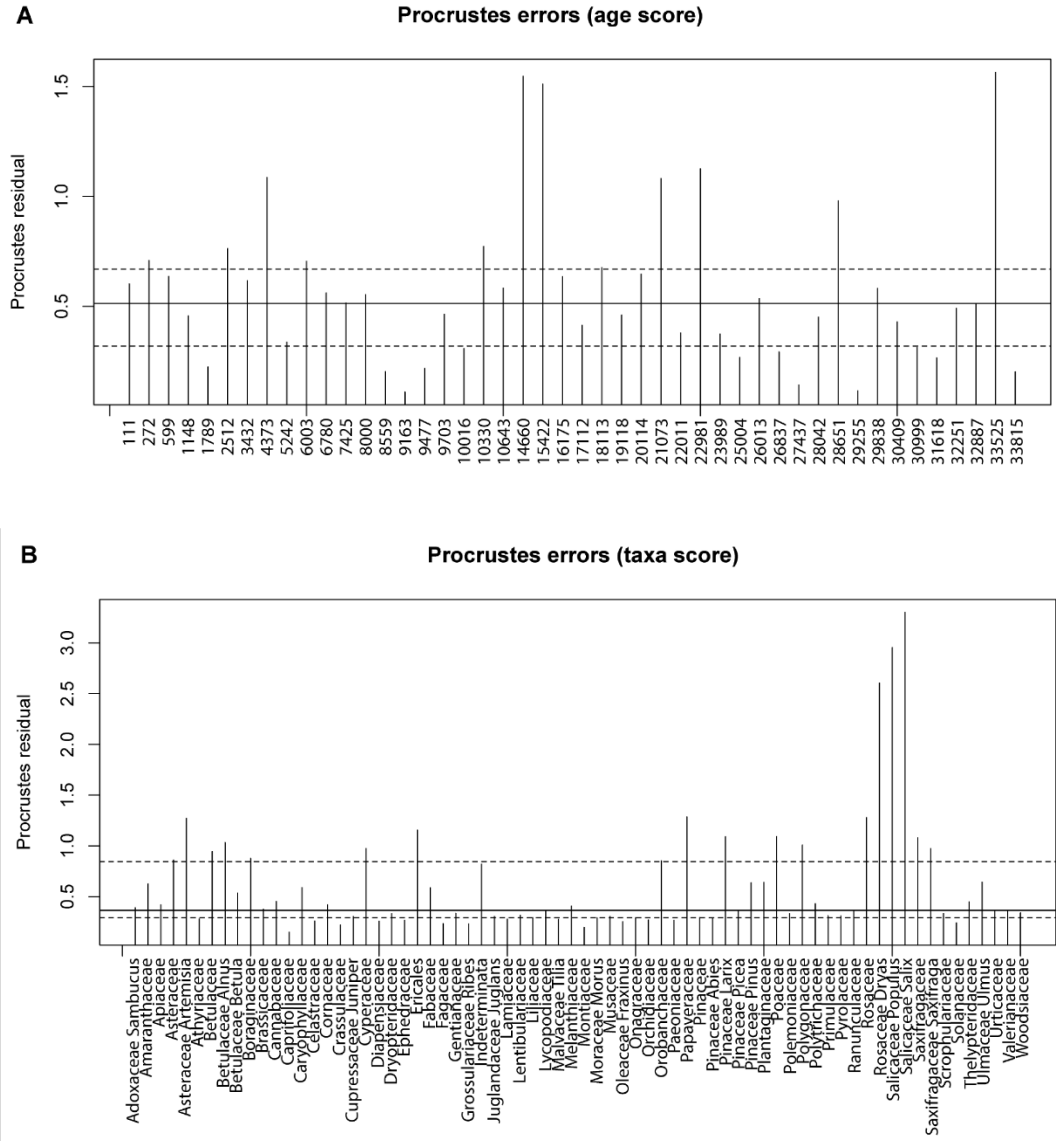


Figure 2.8 | Procrustes residuals plots. (A) Procrustes residuals plot of the comparison of principal component analysis (PCA) depth/age scores of sedaDNA and pollen data presented as residuals of depth/age scores reflecting dissimilarity between the tested datasets. Dashed lines are the first and third quartiles, while the solid line is the second quartile. **(B)** Procrustes residuals plot of the comparison of PCA taxa scores of sedaDNA and pollen data presented as residuals of taxa scores reflecting dissimilarity between the tested datasets. Dashed lines are the first and third quartiles, while the solid line is the second quartile

2.4.6 Taxa richness investigation

The test of significance of the difference in the average taxa richness focused on the difference between DNA Zone I and DNA Zone III. DNAZ II was excluded from this analysis for the same reasons as explained before. Furthermore, the three most recent samples from DNAZ I that make the transition to DNAZ II were also removed from this analysis and the visualization. DNAZ I has on average 38 taxa ranging from 10 to 54 (**Figure 2.9**), while DNAZ III has 29 taxa on average with a range of 14 to 46. Taxa richness is significantly higher in DNAZ I (Wilcoxon rank sum test with continuity correction, $W = 353$, $p = 0.001$).

The taxa richness within the two most represented families detected in DNAZ I was also investigated (**Figure 2.9**). For Asteraceae, taxa richness is significantly higher in DNAZ I (3 on average) than in DNAZ III (0.8 on average, $W = 435$, $p < 0.001$). The same is found for Saxifragaceae with an average of 4.9 taxa in DNAZ I and 2 in DNAZ III ($W = 421$, $p < 0.001$).

2.5 DISCUSSION

2.5.1 Proxy validation

While the age-depth model has wide confidence intervals for the lower part of the core, the Holocene part of the model is more precise. According to the age-depth model, the bottom part of the core is dated to a mean extrapolated age of about 35,000 cal yr BP, and the sediments with the highest herb percentages thus accumulated between ca 22,000 and 12,500 cal yr BP, during the Marine Isotope Stage (MIS) 2, known for its harsh climate and herb-dominated vegetation. Therefore, we consider that the lower part can be interpreted at a rough MIS timescale while millennial-scale variation can be inferred in the Holocene interglacial part.

In our pollen assemblages, *Alnus fruticosa*-type, *Betula* sect. *Nanae*, *Artemisia*, and Poaceae are present at high percentages (**Figure 2.3**). All these taxa are known to be high pollen producers with good dispersal abilities (Niemeyer et al., 2015) and mainly originate from the regional vegetation. As these biases are rather similar in all pollen records, pollen data from different locations are principally comparable and relative abundances of major plant taxa can be reconstructed. Accordingly, we assume that our pollen data are also useful

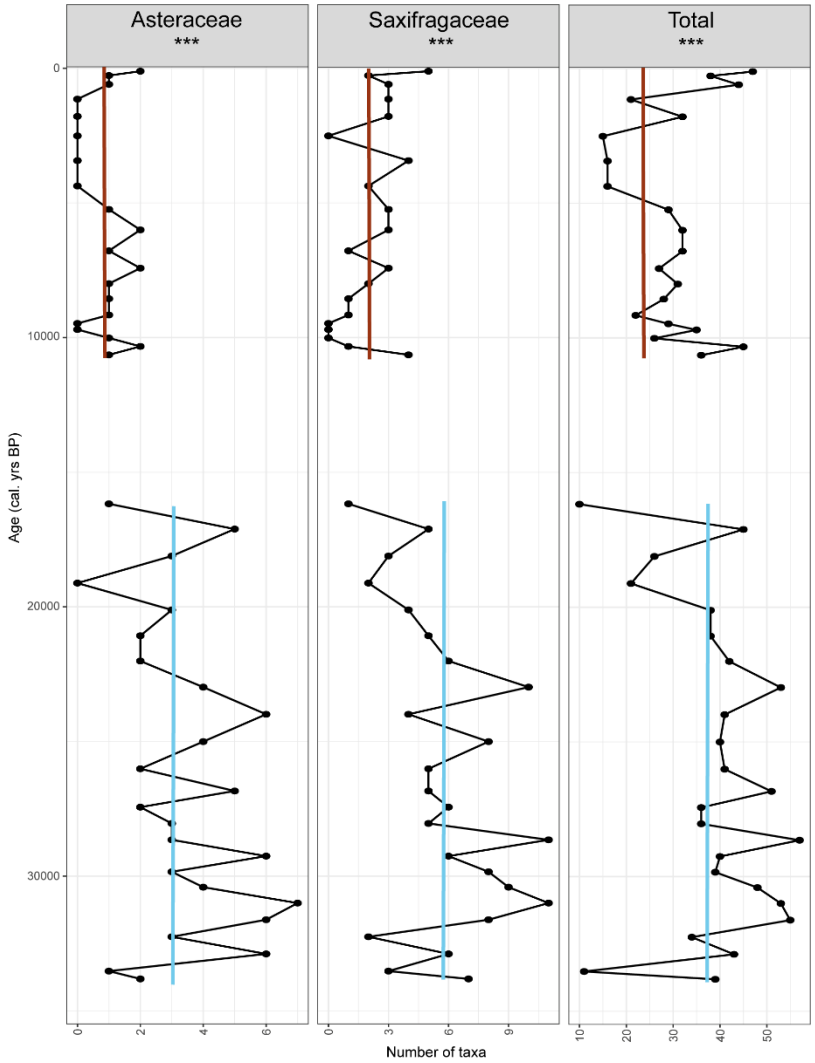


Figure 2.9 | Richness count (total and within the most represented families). Red lines indicate average values in the Holocene interglacial while blue lines indicate the average during the last glacial period.

to infer major regional-scale vegetation composition changes. On the one hand, typically diverse herb families are mostly underrepresented (e.g., Rosaceae) but on the other hand, our pollen record contains pollen likely originating from long-distance sources (e.g., *Pinus* during the glacial period). Furthermore, many pollen types can only be identified to family level (e.g., Poaceae, Cyperaceae, Brassicaceae). Therefore, we consider our pollen record as not so suitable for inferring reliable temporal plant richness changes.

SedaDNA in lake sediments originates from sources other than pollen, for example leaves, roots, and seeds (Willerslev et al., 2003). Its source signal is therefore more secure as it originates from the local environment. For instance, the *seda*DNA *Alnus* signal during the last glacial can help to infer the presence of this taxon in the direct vicinity of the lake, while no taphonomic conclusions can be drawn using just the pollen signal. Our study shows that *seda*DNA can detect twice as many taxa as palynology can (*seda*DNA: 153

terrestrial taxa; pollen: 76 taxa) and to a higher taxonomic resolution. Sixteen DNA sequence types are assigned to family level, 55 to genus and 82 to species level for the *seda*DNA data, while 23 pollen types are assigned to family, 35 to genus and 16 to species level, with one assigned only to order level. Therefore, *seda*DNA is a more suitable proxy than pollen to investigate diversity changes in vegetation (Parducci et al., 2017; Zimmermann et al., 2017a; Alsos et al., 2018).

Some biases still need to be addressed such as the common poor taxonomic resolution at the chloroplast DNA level of some families (e.g., Salicaceae, Cyperaceae, Poaceae) as using short DNA markers such as the P6 loop can be restrictive, especially for discrimination between long sequence length taxa, and might require the use of family specific markers (Sønstebø et al., 2010; Boessenkool et al., 2014; Alsos et al., 2018). However, this does not explain the very low proportion of graminoids detected in the *seda*DNA record compared to the pollen record. Poaceae and Cyperaceae are usually well represented in pollen records, which are biased toward high pollen producers, but they are also easily detected and even dominate *seda*DNA records (Zimmermann et al., 2017a,b). In our study we assume that graminoids might not have been present at high levels in the direct vicinity of the lake but that they were more frequent further afield around the lake resulting in the accumulation of graminoid pollen in the lake. The overrepresentation in the *seda*DNA records of some taxa such as Salicaceae observed in this study have already been recorded elsewhere (Niemeyer et al., 2017; Zimmermann et al., 2017a,b) and can be explained by both the fact that *Salix* densely populates the surroundings of lakes or river depressions (many leaves can end up in the lake catchment) and by high PCR amplification due to increased presence of DNA in the leaves.

2.5.2 Vegetation compositional changes in response to climate inferred from pollen and *seda*DNA records

Overall, both pollen and *seda*DNA records show similar results in terms of vegetation composition and they both reflect the "typical" vegetation, matching other existing palaeoenvironmental records from Siberia (e.g., Müller et al., 2009; Biskaborn et al., 2016). A more open vegetation composed of a mosaic of steppe-tundra and some woody taxa existed during the Late Pleistocene, became more closed boreal forest (taiga type) during the early Holocene, before reflecting the modern biome repartition with more open

vegetation under harsher (colder and wetter) conditions as seen in other records from Siberia and Beringia.

2.5.3 The steppe-tundra of the Late Pleistocene

The Late Pleistocene vegetation before ~14,500 cal yr BP is significantly different from the later periods as inferred from the pollen and *seda*DNA data. Both proxies document an open steppe-tundra landscape even though they do not capture entirely the same taxa.

The Pleistocene part of the Bolshoe Toko pollen record is characterized by a high abundance of graminoids (e.g., Poaceae, Cyperaceae) and *Artemisia*, which agrees with pollen records from northeastern Siberia (e.g., Andreev et al., 2011; Zimmermann et al., 2017a,b). Pollen shrub taxa such as *Alnus fruticosa*-type and *Betula* sect. *Nanae* dominate the major part of the Bolshoe Toko record, but are less common in the northernmost records, although they are rather common in the more southern Lake Baikal area (Shichi et al., 2009), the Verkhoyansk Mountains (Müller et al., 2009), and Chukotka (e.g., Lozhkin et al., 2007; Andreev et al., 2012) records.

The high pollen percentages of *Alnus fruticosa*-type, *Betula* sect. *Nanae*, and *Pinus* s/g *Haploxylon* and their constant presence in the DNA record point to the local presence of these shrubs in more protected habitats such as valleys around the study site. These protected habitats also supported *Rhododendron* as seen by its constant signal in the DNA record despite the extremely continental conditions during the Late Pleistocene. Therefore, the lake vicinity might have ensured sufficient moisture to provide a refugium for shrubs such as *Alnus* and *Rhododendron* during the last glacial.

Poaceae is present at high percentages in the pollen records of Bolshoe Toko. Because Poaceae is not represented in the *seda*DNA, the strong signal observed in the pollen records might come mostly from long-distance wind-transported pollen. Grasses were not dominant locally around the lake. Generally, graminoids dominate landscapes such as the mammoth steppe (Guthrie, 1990; Zimov et al., 2012). According to the *seda*DNA record, the lake vicinity during the Late Pleistocene was mostly covered by forb taxa from Asteraceae and *Dryas* with shrubby habitats hosting a high proportion of shrub willows. Our s*eda*DNA-based reconstruction is similar to other *seda*DNA reconstructions from western Beringia (Willerslev et al., 2014) and northern Siberia (Zimmermann et al., 2017a).

As our results show, the pollen record is biased toward high pollen producers. It is feasible that the steppe-tundra environment was not dominated by graminoids but by forbs, which have a high representation in the *seda*DNA record, supporting the idea that forbs may have a dominant role in full-glacial vegetation (Willerslev et al., 2014; Binney et al., 2017). Frequent occurrence of forbs in the intestines/stomach of megafaunal herbivores suggest that they supplemented their diets with high-protein forbs rather than specializing exclusively on grasses (Willerslev et al., 2014). Furthermore, because forbs may be more nutrient-rich and more easily digested than grasses, this could explain how numerous large animals were sustained (Cornelissen et al., 2004). The presence of abundant megafauna could have caused significant trampling and enhanced gap-based recruitment which would have facilitated the spread of forbs (Owen-Smith, 1987; Zimov et al., 2012).

2.5.4 The disrupted Pleistocene-Holocene transition

The *seda*DNA results of DNAZ II (116–76 cm, ~14,000–11,000 cal yr BP) have very low DNA concentrations. The pollen concentrations are also very low in these sediments. Poor *seda*DNA signals and low pollen concentrations can indicate sparse vegetation and disturbed soils in the area as otherwise trees and their root systems would have helped to limit erosion by stabilizing the ground. Stronger erosion could lead to an increased sedimentation rate related to increased detrital input as showed by the Sr/Rb ratio and Zr (**Figure 2.3**) during times where vegetation was open and scarce (Biskaborn et al., 2013). Indeed, increased water supply to the lake originating from high spring and summer meltwater supply from glaciers in the catchment caused by a warming atmosphere could carry more detrital input, increasing the *seda*DNA bound to it or diluting the signal in the sediment as vegetation was scarce. Another hypothesis would be that an increased production of algae (suggested by rising TOC and Si/Al values) could have led to an increased accumulation of organic material at the lake bottom. This could have initiated changes in the redox dynamic associated with bottom water oxygenation and hydrochemical conditions as suggested by a decrease in Mn/Fe (**Figure 2.3**). These disruptions affect both sediment deposition and hydrochemical conditions at the bottom of the lake and could lead to drastic changes and a decrease in both pollen and DNA deposition until a more stable ecological trophic system became established in the lake at the beginning of the Holocene.

In the transitional zone between the last glacial and the Holocene, *seda*DNA only provides a little information on vegetation composition and no new taxa appear: the few taxa present in the *seda*DNA record for DNAZ II were already present in DNAZ I. In PZ II (ca. 94–80cm, ~12,500–11,300 cal. yr BP), pollen assemblages likely indicate rather unstable environmental conditions at the end of the late glacial. The rapid increase in Artemisia and Poaceae with a decrease of *Betula* and *Alnus* (around 12,500 cal. yr BP) might reflect the Younger Dryas cooling. Similar trends have been observed and described for other eastern Siberian pollen records (e.g., Müller et al., 2009; Andreev et al., 2011, 2012 and references therein).

2.5.5 The boreal forest of the Holocene

Both *seda*DNA and pollen record a change in vegetation from open landscapes to forested ones at the beginning of the Holocene. Tree taxa not detected in the lower part of the core appear in the *seda*DNA record: *Populus*, *Pinus*, *Picea*, and *Larix*. These taxa are detected before the drastic increases in pollen content for *Pinus*, *Alnus fruticosa*-type, and *Betula* sect. *Nanae*.

During the early Holocene (~11,500 to 5,000 cal. yr BP), the vegetation around Bolshoe Toko was represented by summer green trees and shrubs such as *Alnus*, *Betula*, and *Populus*, which tallies with other records showing high abundance of deciduous broadleaved taxa during this warm period in central and eastern Siberia (e.g., Kremenetski et al., 1998; Binney et al., 2009 and references therein). In Yakutia today, *Populus tremula* populations are mainly located on the Lena-Aldan Plateau and more to the south, where they are common in places with sufficient moisture supply (Kuznetsova et al., 2010). The high presence of *Populus* is therefore a good indicator of a moister environment (Mann et al., 2001). Presence of *Betula* and *Salix* confirm that the early Holocene was rather moist. *Populus* is rarely detected in pollen records as its pollen grains are easily destroyed (Brubaker et al., 2005). However, establishment and high presence of *Populus* has been reported in pollen records from central and eastern Beringia (e.g., Kaufman et al., 2004 and references therein).

The upper zone (~4,500 cal. yr BP to present) represents the late Holocene. During this interval, the forest vegetation became similar to the modern vegetation, but with less *Populus* and *Betula* and more *Larix*. More shrub and herb taxa such as *Salix*,

Rhododendron, Polygonaceae, Rosaceae, Ranunculaceae, and Saxifragaceae are present in the *seda*DNA record than during the early Holocene. The onset of the late Holocene cooling is marked by an abrupt decrease in the amount of *Populus* and *Betula* alongside an increase in taxa from the Rosaceae, Crassulaceae, and Saxifragaceae at 24 cm depth (~4,400 cal. yr BP). The replacement of *Populus* forests by shrubby vegetation during the late Holocene is documented in eastern Beringia records (Kaufman et al., 2004). This event could indicate a rapid cooling and most likely represents the 4.2 ka event (Wang et al., 2016). Surprisingly, *Picea* also increased at that depth, probably reflecting the higher presence of spruce in the lake's vicinity. *Picea* may have established in the early Holocene along with other taxa such as *Populus*, and *Picea* could have survived when the occurrence of the other tree taxa decreased in the lake catchment. Indeed, when comparing growth and physiology under different soil temperatures for North American species, *Picea glauca* had a better tolerance for cold soils than *Populus tremuloides* (Landhäusser et al., 2001).

2.5.6 Changes in vegetation richness through the Pleistocene/Holocene transition inferred from the *seda*DNA record

2.5.6.1 High plant richness during the Last Glacial

The results of our vegetation analyses reveal that taxa richness during the Late Pleistocene was higher than during the Holocene (**Figures 2.5, 2.9**). This somewhat contradicts the assumption based on spatial non-forest-to-forest gradients. On average, more taxa are detected by the *seda*DNA proxy (36.7) for the Pleistocene vegetation around Bolshoe Toko than for the Holocene (28.8).

The inferred vegetation during the Late Pleistocene matches the description of the mammoth steppes: dry but diverse and covered by vast areas of grasses, forbs, and sedges (Johnson, 2009). Coprophilous fungi spores (mainly *Sporormiella*) which are usually used to detect the presence of herbivores (Lydolph et al., 2005) are found in the Pleistocene part of the studied core (datasets are archived in DRYAD at "10.5061/dryad.34tmpg4gz"). Mammoth steppe represents a combination of steppe, tundra, and some woody patches (Chytrý et al., 2019). We think that this mosaic, steppe-tundra landscape explains the higher taxa diversity dominating the study area during the Late Pleistocene. Even if we consider that during MIS 2, the observed vegetation composition was open and therefore the effective catchment of the lake was probably bigger than during the Holocene due to higher

erosion rates and more efficient transport pathways, the observed high richness as a true signal is plausible and some hypotheses can be discussed to explain it.

The "keystone herbivore" hypothesis suggests that it was the extinction of Pleistocene megafauna (primarily herbivores) that caused the steppe-tundra to disappear (Owen-Smith, 1987). This mammoth steppe supported a large number of woolly mammoths, rhinoceroses, bison, horses, and other large herbivores. It has been suggested that the Late Pleistocene megafauna played a similar role as the modern grazing megafauna in the east African savannah. Gill (2014) and Bakker et al. (2016) provide evidence that herds of large herbivores can maintain areas of open grassland even under a climate that generally supports forest or tundra. Furthermore, modern tropical grassy biomes such as savannah can compete in term of biodiversity with forests (Murphy et al., 2016). Therefore, if the extinct megafauna had been able to maintain a mosaic landscape, it is also probable that they were able to maintain high plant richness as seen in this study.

During most of the Late Pleistocene, a steppe-tundra vegetation covered the unglaciated expanses of northern Eurasia and North America (Johnson, 2009). This biome therefore spanned an area now covered by many modern biomes such as temperate and boreal forests, steppe, and tundra. Such an enormous area could have allowed more taxa to thrive during the Late Pleistocene as the number of species increases with area following one of the most successful predictions in community ecology: the species-area relationship (Holt et al., 1999; He and Legendre, 2002).

We have discussed only a couple of hypotheses here and it is possible that not just one hypothesis can explain the high richness of terrestrial plants observed during the Late Pleistocene in Yakutia. Further investigation is needed to fully understand past vegetation diversity of the Late Pleistocene steppe-tundra.

2.5.6.2 Higher richness during the Late Pleistocene within the most represented families

The diversity analysis within the most represented families in the *sedaDNA* record (Asteraceae and Saxifragaceae) indicates that families with abundant taxa during the Late Pleistocene are also represented by more taxa during that time. Indeed, if a certain family is present in high abundance, it is likely that such a family thrives under the typical

environmental conditions of the last glacial or the Holocene and would show a higher diversity.

As the last glacial landscapes were open and dominated by herbs, families such as Asteraceae or Saxifragaceae were more diverse during that period than during the Holocene. The number of taxa detected during the last glacial for these two families is surprising: up to 11 Saxifragaceae and 6 Asteraceae taxa are counted in one sample. In modern studies, fewer taxa of these two families are usually detected. It is probably a true signal rather than a biased result due to damaged old sequences as otherwise all sequences would be impacted, and more taxa would be observed for all the families during the last glacial.

The question of taxa loss in plant communities during the Late Pleistocene/Holocene transition can be addressed. Following the "keystone herbivore" hypothesis briefly discussed above, the extinction of the last glacial megaherbivores could have led to the loss of strong interactions that were limiting plant interspecific competition. An increase in competition between the terrestrial plant taxa could then have led to the loss of some diversity, for example via the extinction of some taxa that were strongly dependent on the herbivore interactions. This could potentially explain why so many plant taxa cannot be identified to species level (in this study), as they are not present in databases built from modern taxa. Investigating these lost taxa would be a step toward understanding the mass extinction that the world is facing with the ongoing global warming.

2.6 CONCLUSION

Based on the southeastern Siberian pollen and *seda*DNA records gained from the radiocarbon dated sediment core from Lake Bolshoe Toko, this study described the compositional changes in terrestrial vegetation and its species richness between the Late Pleistocene and the Holocene, thereby filling a knowledge gap in the understudied Siberian landscape. It confirmed the strength and complementarity of two palaeo-botanical proxies: the pollen record enabled us to reconstruct a regional vegetation composition, while the *seda*DNA data provided a local signal and detected more taxa at higher taxonomic levels, allowing species richness estimates. Our study supports the idea of a southward expansion of the Late Pleistocene steppe-tundra with a specific mosaic of herb and shrub patches, which were replaced by taiga during the early Holocene. Finally, the use of the *seda*DNA

record to investigate vegetation richness suggests that in southeastern Siberia, Late Pleistocene steppe-tundra supported more taxa and hence a higher alpha diversity than the boreal forests of the Holocene. This could be explained by the active effect of the megafauna shaping their environment during the Late Pleistocene and/or by the extensive area that the steppe-tundra covered at the time.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at: doi: 10.5061/dryad.34tmpg4gz.

FUNDING

This project has received funding from the Priority Project "International Continental Drilling Program" of the German Research Foundation (no. 3622/25-1) and the European Research.

REFERENCES

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., et al. (2018). Plant DNA metabarcoding of lake sediments: how does it represent the contemporary vegetation. *PLoS One* 13:e0195403. doi: 10.1371/journal.pone.0195403

Andreev, A. A., Morozova, E., Fedorov, G., Schirrmeister, L., Bobrov, A. A., Kienast, F., et al. (2012). Vegetation history of central chukotka deduced from permafrost paleoenvironmental records of the el'gygytgyn impact crater. *Clim. Past* 8, 1287–1300. doi: 10.5194/cp-8-1287-2012

Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siegert, C., et al. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. *Quat. Sci. Rev.* 30, 2182–2199. doi: 10.1016/j.quascirev.2010.12.026

Bakker, E. S., Gill, J. L., Johnson, C. N., Vera, F. W., Sandom, C. J., Asner, G. P., et al. (2016). Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 847–855. doi: 10.1073/pnas.1502545112

Bennett, K. D. (1996). Determination of the number of zones in a biostratigraphical sequence. *New Phytol.* 132, 155–170. doi: 10.1111/j.1469-8137.1996.tb04521.x Beschel, R. E. (1970). "The diversity of tundra vegetation," in Proceedings of the Northern Circumpolar Lands, (Gland: IUCN), 85–92.

Beug, H. J. (2004). Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete. München: Friedrich Pfeli.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., et al. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2:e0000197. doi: 10.1371/journal.pone.0000197

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., et al. (2017). Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. *Quat. Sci. Rev.* 157, 80–97. doi: 10.1016/j.quascirev.2016.11.022

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., et al. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quat. Sci. Rev.* 28, 2445–2464. doi: 10.1016/j.quascirev.2009.04.016

Birks, H. J. B., Felde, V. A., Bjune, A. E., Grytnes, J. A., Seppä, H., and Giesecke, T. (2015). Does pollen assemblage richness reflect floristic richness? A review of recent developments and future challenges. *Rev. Paleobot. Palynol.* 228, 1–25. doi: 10.1016/j.revpalbo.2015.12.011

Biskaborn, B. K., Herzschuh, U., Bolshiyanov, D., Savelieva, L., and Diekmann, B. (2012). Environmental variability in northeastern Siberia during the last ˜13,300 yr inferred from lake diatoms and sediment-geochemical parameters. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 32, 22–36. doi: 10.1016/j.palaeo.2012.02.003

Biskaborn, B. K., Herzschuh, U., Bolshiyanov, D. Y., Schwamborn, G., and Diekmann, B. (2013). Thermokarst processes and depositional events in a tundra lake, northeastern Siberia. *Permafrost Periglacial Process.* 24, 160–174.

Biskaborn, B. K., Narancic, B., Stoof-Leichsenring, K. R., Pestryakova, L. A., Appleby, P. G., Piliposian, G. T., et al. (2021). Effects of climate change and industrialization on Lake Bolshoe Toko, eastern Siberia. *J. Paleolimnol.* 65, 335–352.

Biskaborn, B. K., Nazarova, L., Pestryakova, L. A., Syrykh, L., Funck, K., Meyer, H., et al. (2019a). Spatial distribution of environmental indicators in surface sediments of Lake Bolshoe Toko, Yakutia, Russia. *Biogeosciences* 16, 4023–4049. doi: 10.5194/bg-16-4023-2019

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., et al. (2019b). Permafrost is warming at a global scale. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-018-08240-4

Biskaborn, B. K., Subetto, D. A., Savelieva, L. A., Vakhrameeva, P. S., Hansche, A., Herzschuh, U., et al. (2016). Late Quaternary vegetation and lake system dynamics in north-eastern Siberia: implications for seasonal climate variability. *Quat. Sci. Rev.* 147, 406–421.

Boessenkool, S., McGlynn, G., Epp, L. S., Taylor, D., Pimentel, M., Gizaw, A., et al. (2014). Use of ancient sedimentary DNA as novel conservation tool for High-Altitude tropical biodiversity. *Conserv. Biol.* 28, 446–455. doi: 10.1111/cobi.12195

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., and Coissac, E. (2015). OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* 16, 176–182. doi: 10.1111/1755-0998.12428

Brubaker, L. B., Anderson, P. M., Edwards, M. E., and Lozhkin, A. V. (2005). Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *J. Biogeogr.* 32, 833–848. doi: 10.1111/j.1365-2699.2004.01203.x

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., and Geigl, E. M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5:e0013042. doi: 10.1371/journal.pone.0013042

Chytrý, M., Danihelka, J., Ermakov, N., Hájek, M., Hájková, P., Koˇcí, M., et al. (2007). Plant species richness in continental southern Siberia: effects of pH and climate in the context of the species pool hypothesis. *Glob. Ecol. Biogeogr.* 16, 668–678. doi: 10.1111/j.1466-8238.2007.00320.x

Chytrý, M., Horsák, M., Danihelka, J., Ermakov, N., German, D. A., Hájek, M., et al. (2019). A modern analogue of the Pleistocene steppe-tundra ecosystem in southern Siberia. *Boreas* 48, 36–56. doi: 10.1111/bor.12338

Cornelissen, J. H. C., Quested, H. M., Gwynn-Jones, D., Van Logtestijn, R. S. P., De Beus, M. A. H., Kondratchuk, A., et al. (2004). Leaf digestibility and litter decomposability are related in a wide range of subarctic plant species and types. *Funct. Ecol.* 18, 779–786. doi: 10.1111/j.0269-8463.2004.00900.x

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., and Mace, G. M. (2011). Beyond predictions: biodiversity conservation in a changing climate. *Science* 332, 53–58. doi: 10.1126/science.332.6030.664-b

De Barba, M., Miquel, C., Boyer, F., Rioux, D., Coissac, E., and Taberlet, P. (2014). DNA metabarcoding multiplexing for omnivorous diet analysis and validation of data accuracy. *Mol. Ecol. Resour.* 14, 306–323.

Edwards, M. E. (2020). The maturing relationship between Quaternary paleoecology and ancient sedimentary DNA. *Quat. Res.* 96, 39–47.

Edwards, M. E., Alsos, I. G., Yoccoz, N., Coissac, E., Goslar, T., Gielly, L., et al. (2018). Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. *Holocene* 28, 2006–2016.

Epp, L. S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., et al. (2015). Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quat. Sci. Rev.* 117, 152–163. doi: 10.1016/j.quascirev.2015.03.027

Faegri, K., and Iversen, J. (1989). Textbook of Pollen Analysis, (4th edn with K Krzywinski). Chichester:Wiley.

Gill, J. L. (2014). Ecological impacts of the late Quaternary megaherbivore extinctions. *New Phytol.* 201, 1163–1169. doi: 10.1111/nph.12576

Grimm, E. C. (1987). Constrained cluster analysis by the method of incremental sum of squares. *Comput. Geosci.* 13, 13–35. doi: 10.1016/0098-3004(87)90022-7

Guthrie, R. D. (1990). Frozen Fauna of the Mammoth Steppe. Chicago: University of Chicago Press.

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. *Quat. Sci. Rev.* 20, 549–574. doi: 10.1016/S0277-3791(00)00099-8

He, F., and Legendre, P. (2002). Species diversity patterns derived from species-area models. *Ecology* 83, 1185–1198.

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *Am. Nat.* 163, 192–211.

Hofgaard, A., Harper, K. A., and Golubeva, E. (2012). The role of the circumarctic forest–tundra ecotone for Arctic biodiversity. *Biodiversity* 13, 174–181. doi: 10.1080/14888386.2012.700560

Holt, R. D., Lawton, J. H., Polis, G. A., and Martinez, N. D. (1999). Trophic rank and the species–area relationship. *Ecology* 80, 1495–1504.

Hsieh, T. C., Ma, K. H., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* 7, 1451–1456. doi: 10.1111/2041-210X.12613

Imaeva, L. P., Imaev, V. S., Koz'min, B. M., and Mackey, K. (2009). Formation dynamics of fault-block structures in the eastern segment of the Baikal-Stanovoi seismic belt. *Izvestiya Phys.* Solid Earth 45:1006.

Jackson, D. A. (1995). PROTEST: a PROcrustacean Randomization TEST of community environment concordance. *Ecoscience* 2, 297–303. doi: 10.1080/ 11956860.1995.11682297

Jankovská, V., and Komárek, J. (2000). Indicative value of Pediastrum and other coccal green algae in palaeoecology. *Folia Geobot.* 35, 59–82.

Jenny, J. P., Wilhelm, B., Arnaud, F., Sabatier, P., Covex, C. G., Melo, A., et al. (2014). A 4D sedimentological approach to reconstructing the flood frequency and intensity of the Rhône River (Lake Bourget.NWEuropean Alps). *J. Paleolimnol.* 51, 469–483. doi: 10.1007/s10933-014-9768-4

Johnson, C. N. (2009). Ecological consequences of late Quaternary extinctions of megafauna. *Proc. R. Soc. B Biol. Sci.* 276, 2509–2519. doi: 10.1098/rspb.2008.1921

Jørgensen, T., Haile, J., Möller, P. E. R., Andreev, A., Boessenkool, S., Rasmussen, M., et al. (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

Juggins, S. (2019). "rioja": Analysis of Quaternary Science Data, version 0.9-21.

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., et al. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Res.* 33, 29–33.

Kaufman, D. S., Ager, T. A., Anderson, N. J., Anderson, P. M., Andrews, J. T., Bartlein, P. J., et al. (2004). Holocene thermal maximum in the western Arctic (0-180°W). *Quat. Sci. Rev.* 23, 529–560. doi: 10.1016/j.quascirev.2003.09.007

Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., et al. (2005). Global patterns of plant diversity and floristic knowledge. J. Biogeogr. 32, 1107–1116. doi: 10.1111/j.1365-2699.2005.01272.x

Komárek, J., and Jankovska, V. (2003). *Review of the Green Algal Genus Pediastrum: Implication for Pollen analytical Research.* Berlin: Springer.

Konstantinov, A. F. (2000). *Environmental Problems of Lake Bolshoe Toko, Lakes of Cold Environments, part V: Resource Study.* Yakutsk: Resource Use, Ecology and Nature Protection Issue, 85–93.

Kornilov, B. A. (1962). *The Southeast Suburbs of Aldan Mountains.* Moscow: House of Academy of Sciences of the USSR.

Kremenetski, C. V., Sulerzhitsky, L. D., and Hantemirov, R. (1998). Holocene history of the northern range limits of some trees and shrubs in Russia. *Arctic Alpine Res.* 30, 317–333. doi: 10.2307/1552004

Kruse, S.,Wieczorek, M., Jeltsch, F., and Herzschuh, U. (2016). Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*. *Ecol. Model.* 338:101121. doi: 10.1016/j.ecolmodel.2016.08.003

Kupriyanova, L. A., and Alyoshina, L. A. (1972). Pollen and Spores of Plants from the Flora of European Part of USSR. Leningrad: Academy of Sciences USSR, Komarov Botanical Institute.

Kupriyanova, L. A., and Alyoshina, L. A. (1978). *Pollen and Spores of Plants from the Flora of European Part of USSR.* Leningrad: Academy of Sciences USSR, Komarov Botanical Institute.

Kuznetsova, L. V., Zakharova, V. I., Sosina, N. K., Nikolin, E. G., Ivanova, E. I., Sofronova, E. V., et al. (2010). "Flora of Yakutia: composition and ecological structure," in *The Far North: Plant Biodiversity and Ecology of Yakutia*, eds E. Troeva, A. Isaev, M. Cherosov, and N. Karpov (Dordrecht: Springer), 24–140.

Landhäusser, S. M., DesRochers, A., and Lieffers, V. J. (2001). A comparison of growth and physiology in *Picea glauca* and *Populus tremuloides* at different soil temperatures. *Can. J. For. Res.* 31, 1922–1929. doi: 10.1139/cjfr-31-11-1922

Lougheed, B. C., and Obrochta, S. P. (2019). A rapid, deterministic age-depth modeling routine for geological sequences with inherent depth uncertainty. *Paleoceanogr. Paleoclimatol.* 34, 122–133. doi: 10.1029/2018PA003457

Lozhkin, A. V., Anderson, P. M., Matrosova, T. V., and Minyuk, P. S. (2007). The pollen record from El'gygytgyn Lake: implications for vegetation and climate histories of northern Chukotka since the late middle Pleistocene. *J. Paleolimnol.* 37, 135–153. doi: 10.1007/s10933-006-9018-5

Lydolph, M. C., Jacobsen, J., Arctander, P., Gilbert, M. T. P., Gilichinsky, D. A., Hansen, A. J., et al. (2005). Beringian paleoecology inferred from permafrost preserved fungal DNA. *Appl. Environ. Microbiol.* 71, 1012–1017. doi: 10.1128/AEM.71.2.1012-1017.2005

MacArthur, R. H. (1957). On the relative abundance of bird species. Proc. Natl. Acad. Sci. U.S.A. 43, 293–295.

MacArthur, R. H., and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University.

MacDonald, G. M., Velichko, A. A., Kremenetski, C. V., Borisova, O. K., Goleva, A. A., Andreev, A. A., et al. (2000). Holocene treeline history and climate change across northern Eurasia. *Quat. Res.* 53, 302–311. doi: 10.1006/qres.1999.2123

Mann, D. H., Reanier, R. E., Peteet, D. M., Kunz, M. L., and Johnson, M. (2001). Environmental change and arctic paleoindians. *Arctic Anthropol.* 38, 119–138.

Mannion, P. D., Upchurch, P., Benson, R. B., and Goswami, A. (2014). The latitudinal biodiversity gradient through deep time. *Trends Ecol. Evol.* 29, 42–50. doi: 10.1016/j.tree.2013.09.012

McKay, N. P., Kaufman, D. S., Routson, C. C., Erb, M. P., and Zander, P. D. (2018). The onset and rate of Holocene neoglacial cooling in the Arctic. *Geophys. Res.Lett.* 45, 487–412. doi: 10.1029/2018GL079773

Melles, M., Brigham-Grette, J., Minyuk, P. S., Nowaczyk, N. R., Wennrich, V., DeConto, R. M., et al. (2012). 2.8 Million years of arctic climate change from Lake El'gygytgyn, NE Russia. *Science* 337:315.

Meyers, P. A. (2003). Applications of organic geochemistry to paleolimnological reconstructions: a summary of examples from the Laurentian Great Lakes. *Organ. Geochem.* 34, 261–289.

Meyers, P. A., and Teranes, J. L. (2002). "Sediment organic matter," in *Tracking Environmental Change Using Lake Sediments, Volume 2: Physical and Geochemical Methods*, eds W. M. Last and J. P. Smol (Dordrecht: Kluwer Academic Publisher), 239–269.

Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C., et al. (2010). Arctic amplification: can the past constrain the future? *Quat. Sci. Rev.* 29, 1779–1790. doi: 10.1016/j.quascirev.2010.02.008

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., et al. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* 10, 315–331. doi: 10.1111/j.1461-0248.2007.01020.x

Moore, P. D., Webb, J. A., and Collison, M. E. (1991). *Pollen Analysis.* Oxford: Blackwell Scientific Publications.

Müller, S., Tarasov, P. E., Andreev, A. A., and Diekmann, B. (2009). Late Glacial to Holocene environments in the present-day coldest region of the Northern Hemisphere inferred from a pollen record of Lake Billyakh, Verkhoyansk Mts, NE Siberia. *Clim. Past* 5, 73–84. doi: 10.5194/cp-5-73-2009

Murphy, B. P., Andersen, A. N., and Parr, C. L. (2016). The underestimated biodiversity of tropical grassy biomes. *Philos. Trans. R. Soc. B Biol. Sci.* 371:1703. doi: 10.1098/rstb.2015.0319

Mutke, J., and Barthlott, W. (2005). Patterns of vascular plant diversity at continental to global scales. *Erdkunde* 55, 521–531. doi: 10.3112/erdkunde.2007.04.01

Naeher, S., Gilli, A., North, R. P., Hamann, Y., and Schubert, C. J. (2013). Tracing bottom water oxygenation with sedimentary Mn/Fe ratios in Lake Zurich, Switzerland. *Chem. Geol.* 352, 125–133.

Niemeyer, B., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., and 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, 46–62. doi: 10.1111/ijlh.12426

Niemeyer, B., Klemm, J., Pestryakova, L. A., and Herzshuh, U. (2015). Relative pollen productivity estimates for common taxa of the northern Siberian Arctic. *Rev. Palaeobot. Palynol*. 221, 71–82. doi: 10.1016/j.revpalbo.2015.06.008

Odgaard, B. (1999). Fossil pollen as a record of past biodiversity. *J. Biogeogr.* 26, 7–17. doi: 10.1046/j.1365-2699.1999.00280.x

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). *vegan: Community Ecology Package.* version 2.5-6.

Owen-Smith, N. (1987). Pleistocene extinctions: the pivotal role of megaherbivores. *Paleobiology* 13, 351–362.

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

Pearson, R. G., Phillips, S. J., Loranty, M. M., Beck, P. S., Damoulas, T., Knight, S. J., et al. (2013). Shifts in Arctic vegetation and associated feedbacks under climate change. *Nat. Clim. Change* 3, 673–677. doi: 10.1038/nclimate1858

Pepin, N., Bradley, R. S., Diaz, H. F., Baraër, M., Caceres, E. B., Forsythe, N., et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nat. Clim. Change* 5:424.

Peres-Neto, P. R., and Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test. *Oecologia* 129, 169–178. doi: 10.1007/s004420100720

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., et al. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62, 725–757. doi: 10.1017/RDC.2020.41

Ricklefs, R. E. (1987). Community diversity: relative roles of local and regional processes. *Science* 235, 167–171.

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time.* Cambridge, MA: Cambridge University Press.

Savelieva, L. A., Raschke, E. A., and Titova, D. V. (2013). *Photographic Atlas of Plants and Pollen of the Lena River Delta.* St. Petersburg: St. Petersburg State University.

Shahgedanova, M. (2002). *The Physical Geography of Northern Eurasia.* Oxford: Oxford University Press.

Shichi, K., Takahara, H., Krivonogov, S. K., Bezrukova, E. V., Kashiwaya, K., Takehara, A., et al. (2009). Late Pleistocene and Holocene vegetation and climate records from Lake Kotokel, central Baikal region. *Quat. Int.* 205, 98–110. doi: 10.1016/j.quaint.2009.02.005

Simpson, G. L. (2007). Analogue methods in palaeoecology: using the analogue package. *J. Stat. Softw*. 22, 1–29.

Simpson, G. L., and Oksanen, J. (2016). *analogue: Analogue and Weighted averaging Methods for Palaeoecology*. R package version 0.17-3.

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., et al. (2015). Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. *PLoS One* 10:e0115335. doi: 10.1371/journal. pone.0115335

Sokolovskaya, A. P. (ed.) (1958). "Vegetation of Far North and its development'," in *Pollen of the Arctic Plants,* (Moscow: Komarov Botanical Institute), 245–292.

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., et al. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol. Ecol. Resour*. 10, 1009–1018. doi: 10.1111/j. 1755-0998.2010.02855.x

Stockmarr, J. (1971). Tablets with spores used in absolute pollen analysis. *Pollen Spores* 13, 615–621.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., et al. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35, 1–8. doi: 10.1093/nar/gkl938

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., et al. (2004). Extinction risk from climate change. *Nature* 427, 145–148. doi: 10.1038/nature02121

van Geel, B. (2001). "'Non-pollen palynomorphs," in *Tracking Environmental Change Using Lake Sediments, volume 3: Terrestrial, Algal, and Siliceous Indicators*, eds J. P. Smol, H. J. B. Birks, and W. M. Last (Dordrecht: Kluwer Academic Publishers), 99–119.

van Geel, B., and Aptroot, A. (2006). Fossil ascomycetes in Quaternary deposits. *Nova Hedwig* 82, 313–329.

van Geel, B., Hallewas, D. P., and Pals, J. P. (1983). A late Holocene deposit under the Westfriese Zeedijk near Enkhuizen (Prov. of Noord-Holland, The Netherlands): palaeoecological and archaeological aspects. *Rev. Palaeobot. Palynol.* 38, 269–335.

Vyse, S. A., Herzschuh, U., Andreev, A. A., Pestryakova, L. A., Diekmann, B., Armitage, S. J., et al. (2020). Geochemical and sedimentological responses of arctic glacial Lake Ilirney, Chukotka (far east Russia) to palaeoenvironmental change since 51.8 ka BP. *Quat. Sci. Rev*. 247:106607. doi: 10.1016/j.quascirev.2020.106607

Wang, J., Sun, L., Chen, L., Xu, L., Wang, Y., and Wang, X. (2016). The abrupt climate change near 4,400 yr BP on the cultural transition in Yuchisi, China and its global linkage. *Sci. Rep*. 6, 1–7. doi: 10.1038/srep27723

Weltje, G. J., and Tjallingii, R. (2008). Calibration of XRF core scanners for quantitative geochemical logging of sediment cores: theory and application. Earth Planet. *Sci. Lett.* 274, 423–438.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., et al. (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., et al. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791–795. doi: 10.1126/science.1084114

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schirrmeister, L., Schwamborn, G., et al. (2017a). 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* 8, 1–28. doi: 10.3390/genes8100273

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., et al. (2017b). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences* 14, 575–596. doi: 10.5194/bg-14-575-2017

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., and Chapin Iii, F. S. (2012).Mammoth steppe: a high-productivity phenomenon. *Quat. Sci. Rev.* 57, 26–45. doi: 10.1016/j.quascirev.2012.10.005

3 MANUSCRIPT II

Potential plant extinctions with Pleistocene mammoth steppe loss

Authors

Jérémy Courtin[1]* and Ulrike Herzschuh[1,2,3]*

Affiliation

[1]Polar Terrestrial Environmental Systems, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany,

[2]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany,

[3]Institute of Biology and Biochemistry, University of Potsdam, Potsdam, Germany

***Correspondence**
Jérémy Courtin: jeremy.courtin@awi.de; Ulrike Herzschuh, ulrike.herzschuh@awi.de.

Status

Original draft, in preparation for submission in a scientific journal. The data presented is preliminary.

Additional information in **APPENDIX 2**.

3.1 ABSTRACT

Human induced climate change and habitat fragmentation is responsible of the ongoing extinction that ecosystems are facing. This is happening worldwide, and every Eukaryotic taxonomic group is affected even plants in the high Latitude's ecosystems. At the transition between the Pleistocene last Glacial stage and the Holocene interglacial stage, the widespread Pleistocene steppe-tundra biota collapsed in parallel to the megafauna extinction and a general loss of plant diversity. However, no extinction of plant taxa has really been reported in parallel to those events in Siberia and Beringia mainly because of methodological limitation to detect past plant diversity. Understanding past extinction events via the investigation of Quaternary records can strengthen the current methods to forecast the effects of global warming on ecosystems. If loss of other organism groups were proportional to what has been shown for mammals, a large part of the Pleistocene steppe-tundra biota might have gone extinct. However, few example are known. The improved taxonomic resolution and high detectability of sedimentary ancient DNA provide a new tool to explore this. Here, we investigate potential plant taxa loss in the Northern Hemisphere between the late Pleistocene-Holocene transition using sedimentary ancient DNA (*seda*DNA) metabarcoding. We summarized data from 500 samples comprising nine lake sediment cores from North-East Asia and North America spanning the last 50.000 years. The patterns of past plant diversity (appearance-disappearance through time) were used to detect past taxa non-present in modern database inferring potential candidates for extinction. Our results suggest that vegetation was resilient until the Pleistocene to Holocene transition and that loss appeared in parallel to the Megafauna extinction. Finally, we characterized this vegetation loss and identified that more specialist taxa contributing less to beta diversity are more sensitive to extinction than other taxa. This work holds immense potential to reveal new insights into the evolution of the fragile boreal plant communities and the processes leading to extinction of species.

Keywords: Pleistocene, Holocene, *seda*DNA, extinction, Beringia, vegetation

3.2 INTRODUCTION

Anthropogenic impact is greatly affecting ecosystems and biodiversity (Goudie, 2018). It is now known with an overwhelming amount of evidence that human induced climate change and habitat fragmentation has an effect on species extinction worldwide (Barnosky et al., 2011; Cowie et al., 2022; Rull, 2022). Ongoing extinction rates are estimated to be up to 150-260 E/MSY (extinction per million species years) over the last 500 years much higher than background extinction rates estimated between 0.01 and 1 E/MSY (Cowie, 2022). A similar latitudinal gradient is observed for extinction and for diversity with a decreased diversity usually not coupled with species replacement (Blowes et al., 2019). Most extinction are reported in hot spots of diversity, but all ecosystems are impacted worldwide in all Eukaryota groups (Bellard et al., 2016; Dunn, 2005; Pimm 2014; Regnier et al., 2009). Since 1900, the most impacted taxonomic group has been mammals with 243 E/MSY (Cowie et al., 2022). Of course, such measures are biased toward the most studied areas and taxonomic groups as more species are investigated, reported and tracked (Cardoso et al., 2011; Cowie et al., 2017; Cowie et al., 2022; Régnier et al., 2009; Régnier et al., 2015). It is suspected that the under-studied Arthropods could be the most impacted taxonomic group by ongoing extinctions (Cowie et al., 2022; Stork, 2018). Another example is Viridiplantae as, despite being under-documented, plant species are also impacted with an extinction rate of 18-26 E/MSY, up to 500 times the background extinction rate for plants estimated between 0.05-0.35 E/MSY (De Vos et al., 2015; Humphreys et al., 2019; Lughadha et al., 2020; Pimm et al., 2020). According to the Red list, 20% of plant species are threatened which is similar to the proportion of threatened mammal species (Brummitt et al., 2015; Davies et al., 2018). In general, plant taxa the most sensitive to extinction are species at the edge of their native range, regionally rare and habitat specialists when clades with wide distributions are more likely to survive climatic shifts (Zettlemoyer et al., 2019). Furthermore, plant size has been reported to increase extinction probability when seed production and survival of seeds in the soil seed bank decreased extinction rate and there is no evidence for a phylogenetic pattern in plant extinctions (Saatkamp et al., 2018).

Before the ongoing extinction, the last well-investigated extinction event happened at the transition between the last Ice Age from the Pleistocene and the Holocene. Megafauna is the best-documented group with evidence of the disappearance of many taxa worldwide

(Galetti et al., 2017; Sandom et al., 2014; Wang et al., 2021). This extinction event happened within 3 major waves: around the last glacial maximum, at the Pleistocene-Holocene transition and during the mid-Holocene (Clark, 2009; Murchie et al., 2021; Wang et al., 2021). At the same period, the previously well-established Pleistocene steppe-tundra disappeared (Anderson & Lozhkin 2015; Courtin et al., 2021; Guthrie, 2001; Johnson, 2009; Jørgensen et al., 2012; Willerslev et al. 2014; Zimov et al., 2012). This biota was a mosaic of herbs and grasslands. It covered most of the unglaciated northern hemisphere, was species rich and stable for several glacial and interglacial stages (Bocherens, 2015; Johnson, 2009). However, it disappeared at the transition between the Pleistocene and the Holocene with no modern analogue but some relics that are reported in the Altai range (Chytrý et al., 2019). Extinction is not expected to occur only within one taxonomic group. Pleistocene megafaunas were hypothesised to be keystone species, engineering their environment and therefore interacting with many species across the phyla of life. Therefore, it would be expected that co-extinction occurred within other taxonomic groups such as plants for instance (Bakker et al., 2016; Galetti et al., 2017; Gill, 2014; Malhi et al., 2015). Surprisingly, however, only one plant taxa have been reported extinct in the area, a *Picea* species common in North America and that disappeared during the late Pleistocene (Jackson & Weng, 1999).

To investigate plant extinction or extirpation, we must be able to identify the taxonomic unit of such processes: most of the time, species. In addition, we must be able to systematically quantify species abundance, as it is most likely that species become rare before they become extinct as they first become functionally extinct before completely disappearing (Cowie et al., 2022). Most proxies that investigate past vegetation have issues to do so which could explain why loss of past plant taxa is difficult to estimate (both quantify and characterise). Macrofossils are the best for species identification, but plant remains degrade easily compared to bones and therefore this method relies on scarce material and no quantification is possible through time series (Birks, 2001; Kienast et al., 2001; Tomescu et al., 2018). The pollen proxy is limited to the identification of pollen types, and it is difficult to identify species with this method (Niemeyer et al., 2017; Zimmermann et al., 2017). Finally, ancient DNA methods can be used. The most common approach being metabarcoding with the chloroplast trnL (UAA) intron, g/h marker, best suited to investigate short *seda*DNA fragments but not optimal for species identification (Courtin et al., 2021; Taberlet et al., 2007). Metagenomic analyses are more expensive,

pipelines are not yet fully established and are overall less democratised compared to the metabarcoding approach (Bálint et al., 2018; Pedersen et al., 2016). *Seda*DNA methods rely on databases, which rely on most common and modern taxa (and therefore not extinct) and are incomplete (Breitwieser et al., 2019; Steinegger & Salzberg, 2020). Metabarcoding analyses use only the 100% assigned ASVs relying only on the identification of taxa present in the database and therefore modern, leaving on the side potentially extinct taxa which would, by definition be absent from the databases. Therefore, the methods are well established to describe compositional and diversity changes through time, but they all have a limitation to detect or quantify rare and extinct taxa, which make past plant extinction overseen in previous work (Courtin et al., 2021; Huang et al., 2021; Rijal et al., 2021).

In this study, we aim to identify and quantify potential plant taxa loss that happened during the loss of the extended steppe-tundra in parallel to the megafauna extinction. We will do so by (1) improving the detection of rare taxa in the trnL g/h datasets of multiple cores from the previous Pleistocene steppe-tundra area and covering the Pleistocene / Holocene transition (2) identifying critical potential plants extirpation and extinction events (3) and characterising the potentially extinct plant taxa.

3.3 Material and Method

3.3.1 Site description and timeframe

From 9 different sites, we selected 10 lake sediment cores spanning the Pleistocene to Holocene transition. The sites span over Siberia, Beringia and Alaska and allow the investigation of vegetation changes over an area previously covered by the Pleistocene steppe-tundra (**Figure 3.1**). Age models for the sediment records were adopted from previous studies indicated in **Table 3.1**. The cores cover the last 55,000 years for the oldest and the last 28,000 years for the most recent. For this reason, we investigated only the last 28,000 years and split the record in 14 time slices of 2,000 years. We assigned the time slices to 3 different groups. The 5 oldest (between ~27,000 calibrated years Before Present (cal. yrs BP) and ~19,000 cal. yrs BP) are assigned to the Pleistocene, the following 4 (between ~17,000 cal. yrs BP and ~11,000 cal. yrs BP) to the transition between the Pleistocene and the Holocene and the 5 most recent ones (between ~9,000 cal. yrs BP and ~1,000 cal. yrs BP) to the Holocene.

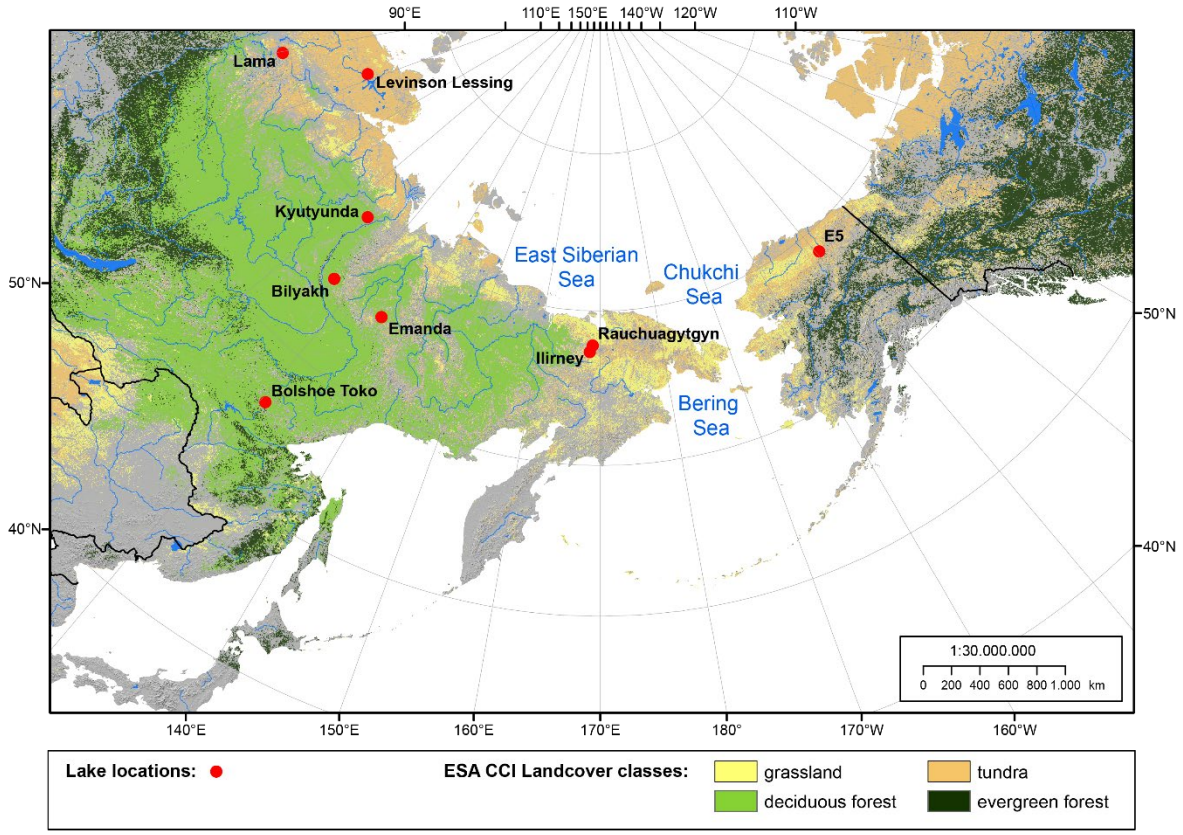


Figure 3.1: Location of investigated lakes in Siberia, Beringia and Alaska. Lake Bilyakh (Müller et al., 2010); Lake Bolshoe Toko (Courtin et al., 2021); Lake E5 (Vachula et al., 2019); Lake Emanda (Baumer et al., 2021); Lake Ilirney (2 cores – EN18208, Vyse et al., 2020; 16-KP-01-L02, Andreev et al., 2021); Lake Kyutyunda (Biskaborn et al., 2016); Lake Lama (von Hippel et al., 2021); Lake Levinson Lessing (Lenz et al., 2022); Lake Rauchuagytgyn (Andreev et al., 2021). Land Cover classes are used for dominant modern (1992-2015) vegetation types.

Table 3.1: Summarised information on the 10 lake sediment cores investigated.

Core	Site - Lake	Latitude	Longitude	Collection Date	Length (cm)	Number of samples used	Age range (cal. yrs BP)	Age model - references
PG1755	Bilyakh	65°17'N	126°47'E	2005	936	62	50000	Müller et al., 2010
PG2133	Bolshoe Toko	56°15'N	130°30'E	2013	380	58	34000	Courtin et al. 2021
E5-1A	E5	68°38.5N	149°27.5'W	2014	500	59	35000	Vachula et al., 2019
Co1412	Emanda	65°17'N	135°45'E	2017	610	45	57000	Baumer et al., 2021
EN18208	Ilirney	67°21'N	168°19'E	2018	1055	88	52000	Vyse et al., 2020
16-KP-01-L02	Ilirney	67°21'N	168°19'E	2016	235	59	28000	Andreev et al., 2021
PG2023	Kyutyunda	69°38'N	123°38'E	2010	700	9	39000	Biskaborn et al., 2016
PG1341	Lama	69°32'N	90°12'E	1997	1885	15	23000	von Hippel et al., 2021
Co1401	Levinson Lessing	74°27.9'N	98°39.97'E	2017	4600	86	62000	Lenz et al., 2022
EN18218	Rauchuagytgyn	67°47.3'N	168°44.3'E	2016	653	66	29000	Andreev et al., 2021

3.3.2 Sampling, DNA extraction and PCR

Core sampling, DNA extraction and PCR amplifications were done as described in Courtin et al. (2021) and Zimmermann et al. (2017). Each sequencing run was performed by Fasteris SA sequencing service (Switzerland) with the paired-end sequencing on the Illumina HiSeq platform (2 × 125 bp).

3.3.3 Filtering and cleaning dataset

Each metabarcoding datasets were analysed individually using OBITOOLS (version 3; Boyer et al., 2015). Overlapping paired reads were merged using the command *alignpairedend*. The unmerged reads were removed with *grep -a* mode: alignment. The command *ngsfilter* assigned each sequence to its corresponding sample according to the tag combination. Reads were de-replicated with the command *unique* and cleaned from PCR and sequencing errors using *clean -r 0.05* (-r defines the maximum ratio allowed between the counts of sequence variants). Taxonomic assignments of the amplicon sequence variants (ASVs) were performed using *ecotag* with the use of two databases. One, the ArctBorBryo is based on the curated arctic and boreal vascular plant and bryophyte reference database published by Sønstebø et al. 2010, Soininen et al. 2015 and Willerslev et al. 2014. The second database based on the EMBL Nucleotide Database standard sequence release 143 (ftp://ftp.ebi.ac.uk/pub/databases/ena/sequence/release/).

For each project, we merged both assigned datasets. This gave us the best assignment for each ASVs based on both databases. We prioritised ArctborBryo so if for one ASV, we have similar confidence threshold, and we would use the name given by the curated ArctborBryo in comparison to the un-curated EMBL143. If not, we always use the name given by the best assignment confidence. Once the best assignment was chosen for each ASV, we only kept sequences assigned with a minimum confidence of 90%, We excluded exotic assignments, and kept only assignments to at least family at best. Then, we investigated extraction blanks and NTCs by investigating each PCR batches. We excluded sequences from the dataset if for each project, they had more than 2% of their reads count in the extraction blanks and NTCs for each batch and if they appear more than 2% of their total reads count in the blanks of the entire dataset. Once the blanks were clean for each project, we collapsed reads of different PCR replicates for each sample. More details are present in the **Supplementary Table 2.1**. The last step was to merge all ASVs from different projects into one dataset and keeping only samples with a minimum of 1000 reads

and ASVs appearing in at least 10 samples with a minimum read count of 100 reads per ASV. From this final dataset, we annotated the ASVs assigned at 100%: 100% ASVs and the ASVs not assigned at 100% (from 90%): candidate ASVs. We will use those names throughout the entire study.

3.3.4 Identification of taxa – species signal

We assume that each unique ASV does not necessarily carry a unique species signal. We also do not expect species to exist on their own and therefore have unique occurrence patterns in the dataset. They are expected to exist in interaction with other species and form interspecies communities. This means that through time and space (characteristics held by our samples), we expect to find patterns of correlation of occurrence (co-occurrence) between ASVs. A spearman pairwise correlation test with holm adjustment was performed with the corr.test R function (psych package) on our four-time squared root transform reads count data frame with the spearman method and the holm adjustment for non-normally distributed data. Only keeping positive correlation scores superior to 0.4, we built a network with the graph_from_data_frame function of the igraph R package followed by the cluster_label_prop function to detect community structure in the newly built network and underline ASVs co-occurrence.

Finally, we worked with all ASVs as part of a community of a minimum of 5 ASVs. We collapsed ASVs according to their assignments. As we assumed that a community can be made only of unique taxa but that the same taxa can overlap between communities, the 100% ASVs and candidates ASVs reads with similar assignments were collapsed under unique taxa. The candidates ASVs reads were collapsed with the 100% ASVs if the candidate ASV best assignment was to family or genus and was comprise in the best assignment of the 100% ASV (e.g., candidate ASV assigned at best to Asteraceae or *Artemisia*, is collapsed with the 100% ASV assigned to Asteraceae *Artemisia vulgaris*).

With this method, we reduced the number of candidates ASVs and kept only unique assignments with specific co-occurrence patterns and then assumed that each unique assignment represented a taxon signal.

3.3.5 Resampling

As each of the time slice have different total number of reads assigned, different number of total samples and span different number of sites, we performed a three steps resampling script to normalise each time slice (**Supplementary Table 2.2**). The resampling was performed 100 times and metrics such as number of reads per taxa per time slice were measured at each iteration. The results presented are the mean values of those metrics with the standard deviation measured from the 100 resampling iterations.

3.3.6 Assessment of the species pool stability

To investigate vegetation composition changes, the taxa were categorised within 4 plant types: herb, grass, shrub or tree. The relative abundance of reads assigned to each plant type at every time slice were plotted as percentage bar plots using the ggplot R-package.

The average taxa richness and its standard deviation per time slice is plotted with the trendline using R. We then tested for the significance of the average richness differences between time slices' zones with a Wilcoxon rank sum test with continuity correction using the wilcox.test function in R.

Finally, the taxa replacement rates were estimated by measuring the Podani-family decompositions of the Jaccard dissimilarity coefficients and their quantitative forms from taxa abundance at each time slice from ~25,000 cal. yrs BP until ~1,000 cal. yrs BP using the beta.div.comp function of the adespatial R-package. The average Podani indices for each time slices' zone are then compared to identify potential changes in taxa stability.

3.3.7 Quantification of extinct and extirpated taxa

We estimated the number of taxa extirpated from the study area by accounting for the number of taxa absent from the modern time slice (~1,000 cal. yrs BP). The extirpation rate was estimated following the measure of the extinction rate and using the number of extinction or extirpation per million species years' unit (E/MSY) as described in Pimm et al., 2006. We also measured the number of taxa absent, disappearing and reappearing at each time slice for all taxa, the 100% and the candidate taxa (**Supplementary Figure 2.2**). We checked among the taxa absent at the modern time slice, their last appearance and noticed that no taxa absent at ~1000 cal. yrs BP disappeared before ~21,000 cal. yrs BP.

To look for extinction events, we needed to account for reappearance rates. Therefore, for every taxon disappearing at every time slice, we reported the number of time slice they need to reappear in the dataset. From the three time slices for which no extinction has been reported (100% of the taxa absent reappear later in the dataset): ~27,000, ~25,000 and ~23,000 cal. yrs BP, we estimated the number of the time slices needed for taxa to reappear as percentages and used this rate as the expected reappearance rate. We compared this expected rate to the observed rate at each time slice. After visual inspection of both rates, we could estimate, if the expected reappearance rate was higher than the observed potential time slices with more extinction than expected and therefore, potential extinction events. We measured the expected number of extinctions by getting the relative expected rates compared to the observed number of taxa absent at each time slice and compared it the observed number of extinctions. We could then measure the "net" taxa loss at each time slice. The extinction rates at each time slice were measured are described in Pimm et al., 2006 over a period of the time slice: 2,000 years.

3.3.8 Characterisation of species and candidate species

We assume that for taxa to be extinct, it needs to be absent from the modern time slice (~1,000 cal. yrs BP) and not present in the database (not assigned at 100%). Therefore, we assume that the taxa that hold the highest probability to be extinct are candidates' taxa absent from the ~1000 cal. yrs BP time slice.

To characterise the taxa potentially extinct from the region, we compare each of their characteristics to 100% taxa with similar total number of reads (+/-10%) and number of samples occurrence (+/-10%). We call this subset of the 100% ASVs, the *similar*100% ASVs.

For each taxon, we estimate if they are more specialist or generalist based on the number of taxa present in their communities (from the communities built using the cluster_label_prop function). We consider that taxa present in communities with less taxa are more specialist than taxa present in communities with more taxa, which are therefore considered more specialist. We compared the average number of taxa per community for the potential candidates and the *similar*100% with a wilcoxon statistic test (wilcox.test function in R). The results are plotted as a box plot.

To estimate the contribution of the taxa to functionality of the system, we assessed their contribution to beta diversity by measuring their species contribution to beta diversity (SCBD) scores with the beta.div function from the adespatial R package on squared root relative abundance data. We compared the average SCBD score of the candidates absent and present at ~1,000 cal. yrs BP to *similar*100% and 100% taxa with Wilcoxon test (wilcox.test function in R). The results are plotted as box plot.

Because the plant types are not present in a similar abundance between the potential candidate taxa and 100% assigned taxa, we compare the relative proportion changes of each plant type for candidate taxa absent or present from ~1000 cal. years BP to the average number of *similar*100% and 100% taxa. Significative differences between groups were tested by pairwise comparison with bonferonni correction using the pairwise.t.test function with the p.adjust.method="bonferroni " option in R. The results are plotted as a box plot.

The same procedure as for plant types was performed to compare families. Significative differences between groups were tested by pairwise comparison with bonferonni correction using the pairwise.t.test function with the p.adjust.method="bonferroni " option in R. The results are plotted as a box plot.

Table 3.2 | Number of ASVs passing the different filtering steps.

Filtering step	Type of ASV	Total	Percentage
1. > 10 samples	Total	21977	
	100%	556	2.5%
	candidates	21424	97.5%
2. >100 reads	Total	5302	
	100%	475	8.9%
	candidates	4827	91.1%
3. Community > 5 ASVs	Total	4957	
	100%	409	8.3%
	candidates	4548	91.7%
4. Different assignments	Total	474	
	100%	340	72%
	candidates	134	28%

3.4 RESULTS

A total of 474 potential taxa were identified from the 21,977 ASVs present in a minimum of 10 samples (**Table 3.2**). After filtering for a minimum number of 100 reads and only

ASVs part of communities, 74% (409 out of 556) of the ASVs assigned with 100% confidence were kept which allowed the identification of 340 potential taxa assigned at 100% whereas, 21% (4,548 out of 21,424) of the candidate ASVs (not assigned at 100%) were kept, to identify 134 potential taxa not assigned at 100%.

Using ASVs taxonomically assigned with a minimum of 90% confidence and the filtering pipeline explained in the methods, it allows us to identify 28% more taxa than we would if we would investigate only the 100% assigned ASVs.

3.4.1 Changes in the composition and species pool at the Pleistocene - Holocene transition

3.4.1.1 Changes in the species pool at the Pleistocene-Holocene transition

Increases in the replacement rate would suggest changes and instability of the species pool (**Figure 3.2**).

The maximum replacement rate is reached at ~11,000 cal. yrs BP with a Podani index of 0.327 +- 0.0371 when the minimum of 0.202 +- 0.0449 is reached at ~23,000 cal. yrs BP. During the Pleistocene, between ~27,000 and ~19,000 cal. yrs BP, the average Podani index is 0.237. This is lower than the average Podani index of 0.284 at the transition between the Pleistocene and the Holocene (~17,000 to ~11,000 cal. yrs BP). During the Holocene (~9,000 to ~1,000 cal. yrs BP), the average Podani index is lower than at the transition but higher than during the Pleistocene at an average of 0.263 similar to the one observed during the Pleistocene.

The Pleistocene has the smallest replacement rate suggesting that the species pool was the most stable at this period. At the Pleistocene to the Holocene transition, the higher replacement rate suggests a more unstable and disturbed transition in comparison to both the Pleistocene and the Holocene. Therefore, going from one stable Pleistocene pool to an unstable transition and again a more stable Holocene suggests an overall shift in the species pool between the Pleistocene and the Holocene.

3.4.1.2 Vegetation compositional shift between the Pleistocene and the Holocene

During the Pleistocene (~25,000 to ~19,000 cal. yrs BP), more than 50% of the reads are assigned to either herb or grass taxa (**Figure 3.2**). We observe a peak of grass just before

the Last glacial Maximum (~23,000 and ~21,000 cal. yrs BP). At the transition between the Pleistocene and the Holocene (between ~17,000 and ~11,000 cal. yrs BP), there is a shift with more and more reads assigned to shrubs and trees over herbs and grasses. During the Holocene (~9,000 to ~1,000 cal. yrs BP), 75% of all reads are assigned to shrubs or trees and we observe at the beginning of the Holocene, there is a peak of reads assigned to trees.

Overall, we observe a drastic vegetation compositional change at the transition between the Pleistocene (more herbs and grasses) and the Holocene (more shrubs and trees).

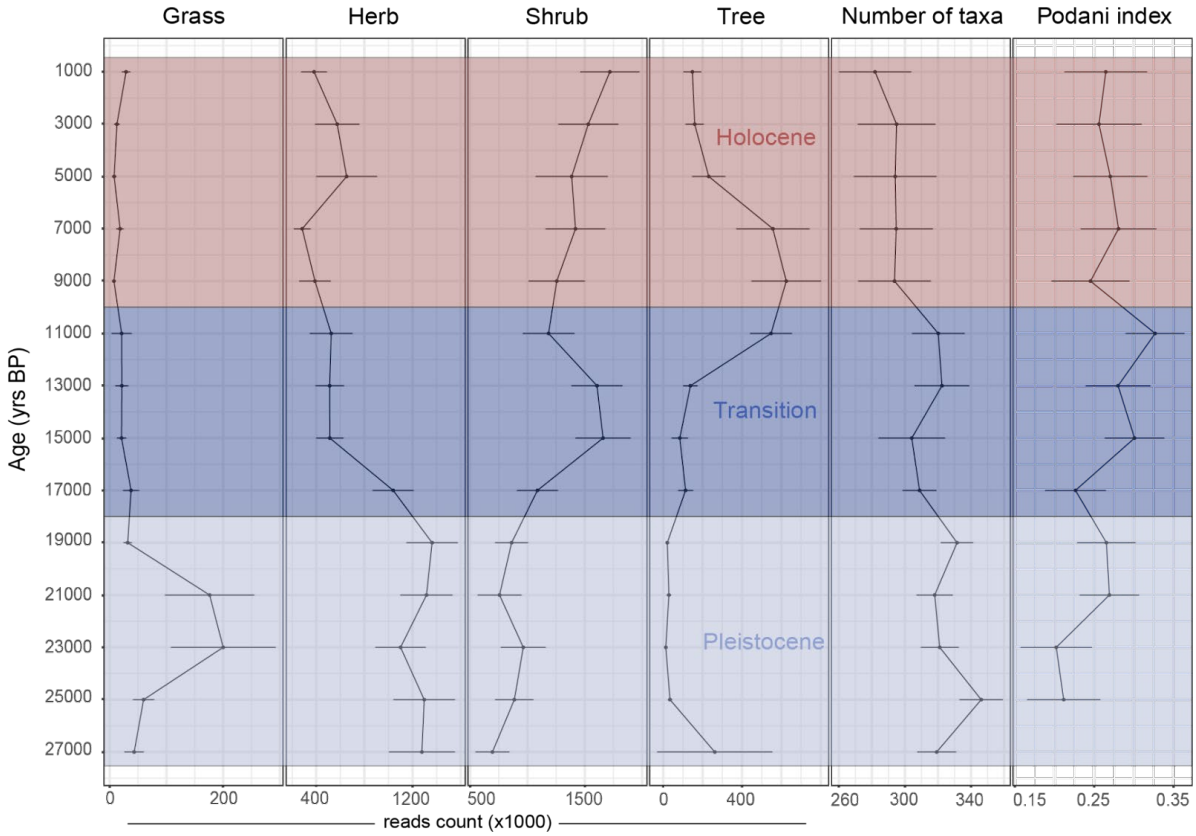


Figure 3.2 | Stratigraphic plot of terrestrial *seda*DNA. The number of reads counts per plant types assigned at each time slice. The number of plant taxa identified per time slice is reported as well as the podani index score. For each measured metrics, the average value is plotted as well as the standard deviation inferred from the 100 resampling iterations.

3.4.2 Decrease in the regional plant species richness between the Pleistocene and the Holocene

On average, 311 (+-7.67) taxa are present in each time slice. During the Pleistocene, between ~27,000 and ~19,000 cal. yrs BP, an average of 327 (+/-7.10) taxa are present per time slice; which is more (wilcox.test, p-value < 2.2e-16) than the average number of 314 (+-10.1) taxa present at the transition between the Pleistocene and the Holocene between ~17,000 and ~11,000 cal. yrs BP; which is more (wilcox.test, p-value < 2.2e-16) than the

average number of 292 (+-10.9) taxa present at during the Holocene between ~9000 and ~1,000 cal. yrs BP (**Figure 3.2**).

There is a regional loss of taxa richness between the Pleistocene and the Holocene.

Most potential taxa are present in all time slices with an average of 276.62 (+/-14.9) taxa (58% of all species assigned). On average, 147.22 (+/-8.8) 100% taxa are present in all time slices (73% of all 100% taxa) and on average 70.66 (+/-5.21) candidate taxa are present in all time slices (52.73% of all candidate taxa).

This suggests that there is a core species pool in the area made of 58% of the potential plant taxa that compose the vegetation in the area over the last 28,000 years. This core species pool is very stable through time and only 40% of all detected potential taxa are impacted by extirpation and potential extinction.

Overall, we see a shift from the stable vegetation species pool of the Pleistocene (grass and herbs dominated) to the stable vegetation species pool of the Holocene (dominated by shrubs and trees). This transition from one state to the other happened with an increased replacement rate during the so-called Pleistocene to Holocene transition illustrating instability in the species pool during that time and replacement of one steppe-grassland dominated vegetation to a more forested one. This compositional shift is accompanied by a general decrease in species richness through time between the Pleistocene and the Holocene. The changes observed concern only 40% of all potential taxa detected and still 60% of the species pool is stable though time in the area.

3.4.3 Identification of loss taxa events

3.4.3.1 Extirpation rate

None of the taxa absent from the modern time slice disappear from the dataset before ~21,000 cal. yrs BP meaning that all taxa that extirpated from the dataset before ~21,000 cal. yrs BP reappeared later in the area. In the dataset, taxa can reappear up to 10 time slices after their disappearance (20,000 years) (**Supplementary Table 2.3**).

The calculated extirpation rate over the last 28,000 years is 4.64 E/MSY in the study area.

3.4.3.2 Extinction events

Out of the 121 potential taxa absent from the modern time slice, 41 are candidates. This means that the extirpated and maximum potential extinction rates is of 3.1 over the last 28,000 years in the study area.

Because the more time a taxon has to reappear, the more likely it is to reappear (**Supplementary Table 2.3**), to account for the real extinction rate and distinguish between the extirpation and extinction; we need to measure the reappearance rate

We measured the expected reappearance rates from the time slices from ~27,000 to ~23,000 cal. yrs BP, when all taxa absent reappeared and they still have 10 time slices to potentially reappear in the dataset (meaning that they are only extirpated and not extinct) and compared it to the observed reappearance rate (**Figure 3.3**; **Supplementary Table 2.3**).

We observe at 3 time slices potential events where the observed reappearance rate is lower than the expected one by 5% or more: ~17,000, ~11,000 and ~5,000 cal. yrs BP. We could therefore expect potential extinction events to have occurred during those 3 time slices with a difference of 2.4 taxa at ~17,000 cal. yrs BP, 4.4 at ~11,000 cal. yrs BP and 4.1 at ~5,000 cal. yrs BP between the observed loss taxa and the expected. This would suggest a total of 10 to 11 plant taxa that have gone extinct since ~28,000 cal. years BP.

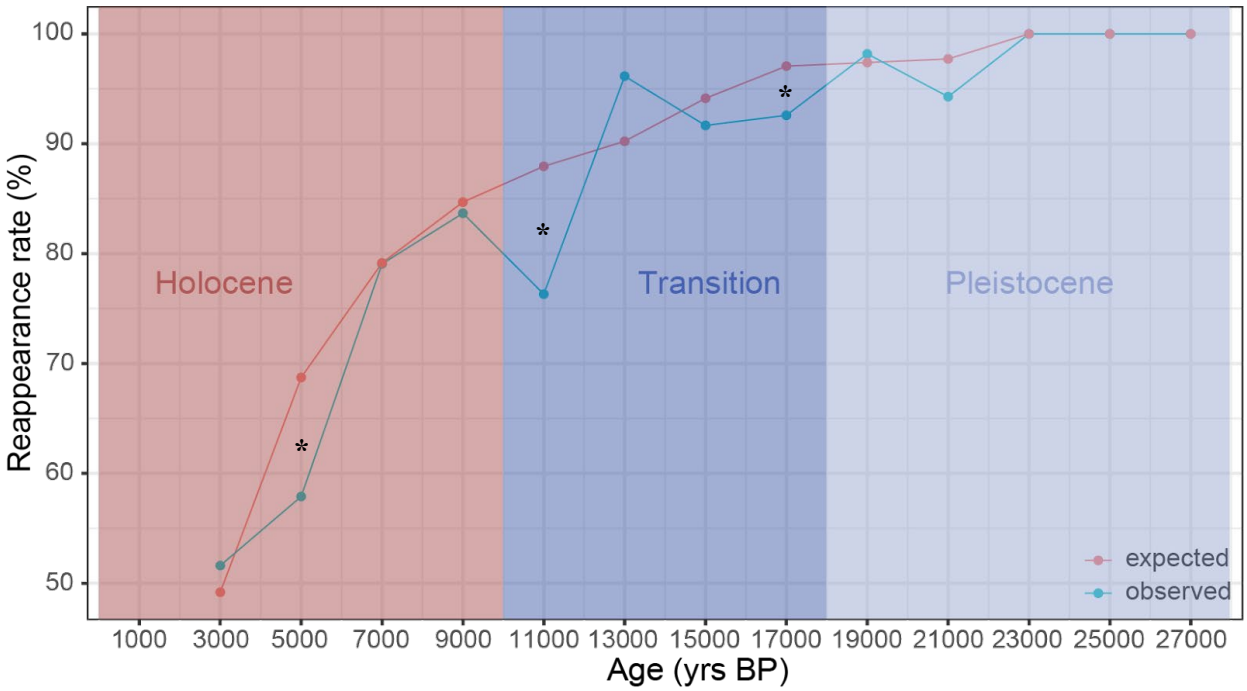


Figure 3.3: Expected vs observed reappearance rates per time slice. *: time slices when the difference between the expected and the observed reappearance rate is above 5%.

This translates to an estimated extinction rate of 2.53 E/MSY between ~19,000 and ~17,000 cal. yrs BP, 4.64 E/MSY between ~13,000 and ~11,000 cal. yrs BP and 4.32 E/MSY between ~7,000 and ~5,000 cal. yrs BP. On average, the balance between observed and expected loss is 0.9 taxa lost per time slice between ~21,000 and ~3,000 cal. yrs BP. This translates to an average extinction rate of 0.95 E/MSY per time slice between 21,000 and ~3,000 cal. yrs BP.

3.4.4 Characterisation of lost taxa

Extirpated taxa are absent from the modern time slice (~1,000 cal. yrs BP). The extirpated taxa that hold the highest probability of being extinct are the taxa absent from the modern time slice of our dataset and absent from the database. Furthermore, all the potentially extinct taxa are absent from the modern time slice. In addition, the probability that an extirpated taxon is extinct increases with the duration that they are extirpated. We characterise the lost taxa as candidates absent from the modern time slice and compare different metrics of 100% assigned taxa present and absent at ~1,000 cal. yrs BP with or without similar abundance than candidate taxa.

We identified 142 potential taxa assigned at 100% with a total read count and an occurrence in sample similar to the maximum values (+- 10%) of the 41 potential candidate taxa absent at ~1,000 cal. yrs BP. This represents 42% of all 100% taxa. The 142 *similar*100% and 134 candidates' potential taxa are considered to be the rare fraction of our identified potential taxa as they are detected with fewer abundance and in fewer samples than the other potential taxa. This means that 58% (276 out of 474) of all detected taxa are rare.

3.4.4.1 Specialist vs Generalist

Taxa absent from the modern time slice are more specialists. Overall, candidates are the most specialist of all; even the candidates present at ~1,000 cal. yrs BP are more specialist than the other groups (**Figure 3.4A**).

With our method, we established communities for each taxon. To estimate if a taxon is more specialist or more generalist, we count the number of taxa present in its community.

The taxa absent at ~1,000 cal. yrs BP are more specialist when compared to the taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.933e-06). They are part of significantly smaller communities.

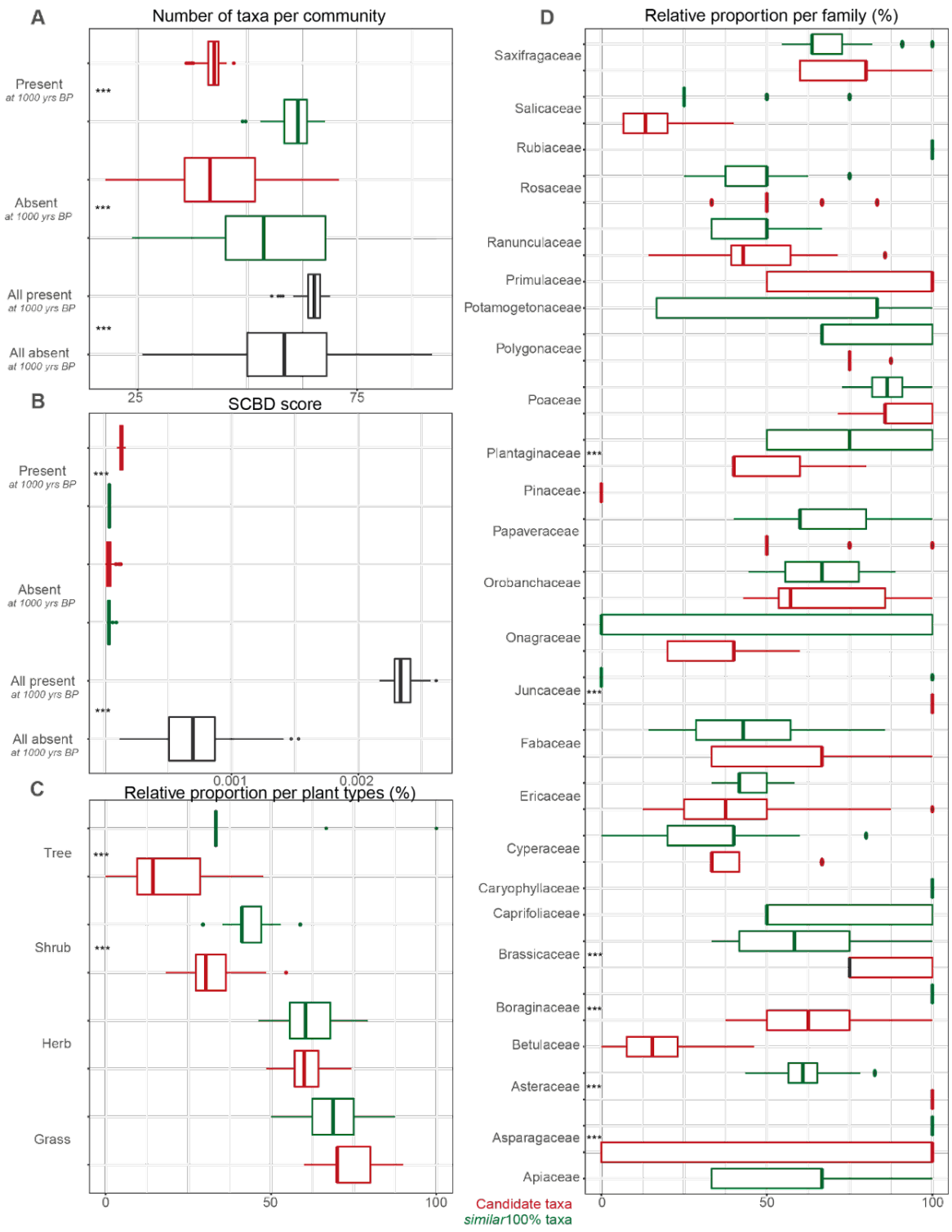


Figure 3.4: Characterisation of potentially lost plant taxa. Comparison with box plots of the distribution of candidate taxa to the *similar*100% taxa absent and present from the present time slice for (**A**), specialist or generalist based on the community size; (**B**) contribution to beta diversity based on the species contribution to beta diversity score (SCBD); (**C**) relative proportion per plant types absent at from the modern time slice; (**D**) relative proportion per plant families absent at from the modern time slice. In red are plotted the candidate taxa and in green the *similar*100% taxa. In black are plotted all taxa. When no taxa are represented in a group (e.g., family), the group is not plotted. *** indicates the significance of the statistic test described in the methods part.

The candidates absent at ~1,000 cal. yrs BP have the smallest number of taxa in their communities. They are the "most specialists". They are more specialist than taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.2e-16), especially *similar*100% (wilcoxon.test; p-value < 2.2e-16) but they are not more specialist than candidates present at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 0.6147). They are also more specialist than *similar*100% taxa absent at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.2e-16).

The *similar*100% taxa absent at ~1,000 cal. yrs BP have are more specialist than taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 3.076e-06), especially *similar*100% (wilcoxon.test; p-value = 0.004269) as they are more generalist than candidates present at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 3.617e-08). They are also more generalist than candidate taxa absent at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 1.364e-13).

3.4.4.2 Contribution to beta-diversity

Taxa, absent the modern time slice at ~1,000 cal. yrs BP, contribute less to the beta diversity (**Figure 3.4B**). s*imilar*100% and candidates taxa absent at ~1,000 cal. yrs BP have a similar SCBD score, on average lower than the SCBD scores of the taxa present at ~1,000 cal. yrs BP.

The potential taxa absent at ~1,000 cal. yrs BP have a lower SCBD than the taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.2e-16).

The candidate taxa absent at ~1,000 cal. yrs BP have a lower SCBD compared to the taxa present at ~1,000 cal. yr BP (wilcoxon.test; p-value < 2.2e-16) both to candidates (wilcoxon.test; p-value < 2.2e-16) and to *similar*100% (wilcoxon.test; p-value = 0.01383). They do not have a different SCBD score when compared to *similar*100% absent at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 0.1974).

The *similar*100% taxa absent at ~1,000 cal. yrs BP have the lowest SCBD score. They have a lower SCBD compared to the taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.2e-16) both to the candidates (wilcoxon.test; p-value < 2.2e-16) and to the *similar*100% (wilcoxon.test; p-value = 0.004681) present at ~1,000 cal. yrs BP. They do not have a different SCBD score when compared to the candidates absent at ~1,000 cal. yrs BP (wilcoxon.test; p-value = 0.1974).

The candidates present at ~1,000 cal. yrs BP have a higher SCBD score than the *similar*100% taxa present at ~1,000 cal. yrs BP (wilcoxon.test; p-value < 2.2e-16).

3.4.4.3 Plants type

Herbs represent the majority of our taxa with 320 assigned taxa (67.5% of all taxa) followed by shrubs with 75 taxa (15.8%), grasses with 46 (9.7%) and trees with 33 taxa (7%) (**Supplementary Table 2.4**).

The most impacted plant groups (absent from ~1,000 cal. yrs BP) are herbs, maybe grasses (*similar*100%) and maybe shrubs (candidates). Trees are not really impacted (**Figure 3.4C**).

Because the plant types are not present in a similar abundance between the potential candidate taxa and 100% assigned taxa, we compare the relative proportion changes of each plant type.

We observe no significant difference between the proportion of grass candidates and the proportion of grass *similar*100% taxa absent from the modern time slice (Bonferroni correction test; p-value = 0.35). Similarly, there is no difference between the proportion of herb candidates and the proportion of herb *similar*100% taxa absent from the modern time slice (Bonferroni correction test; p-value = 1). For shrubs, the proportion of *similar*100% taxa absent at ~1,000 cal. yrs BP is higher than the proportion of candidates.

Overall, it seems that there is no plant type more likely to go extinct, as the proportion of candidate grasses, herbs, shrubs and trees absent from the modern time slice is never significantly higher than the proportion observed for the *similar*100% taxa.

3.4.4.4 Families

From the 18 families represented by both candidates and *similar*100% absent from the modern time slice, only 6 have significative difference in relative proportion of candidate taxa when compared to the proportion of *similar*100% taxa absent at ~1000 cal. yrs BP: Juncaceae (Bonferroni correction test; p-value = 4.35E-113), Asparagaceae (Bonferroni correction test; p-value = 1.23E-40), Asteraceae (Bonferroni correction test; p-value = 4.76e-26), Boraginaceae (Bonferroni correction test; p-value = 1.25E-22), Plantaginaceae

(Bonferroni correction test; p-value = 1.34E-11) and Brassicaceae (Bonferroni correction test; p-value = 1.25E-06) (**Figure 3.4D**).

According to the Bonferroni correction test, the proportion of candidate taxa from Asteraceae, Juncaceae and Brassicaceae absent from ~1,000 cal. yrs BP is higher than the proportion of *similar*100% taxa. This suggests that taxa from those families would be more likely to have undergone extinction in the study area over the last 28,000 years. For Asparagaceae, Boraginaceae and Plantaginaceae, the proportion of candidate taxa absent from ~1,000 cal. yrs BP is lower than the proportion of *similar*100% taxa. They are therefore not significantly impacted by potential extinction over the covered timeframe.

3.4.4.5 Name of extirpated and potentially extinct candidate taxa

Among the 121 taxa extirpated, 41 are candidates and hold potentially extinct taxa (**Table 3.3**). We previously identified 3 extinction events. Candidates' taxa absent at ~1,000 cal. yrs BP and that disappeared in one of the extinction events could hold the highest probability to be extinct. Therefore, the taxa that have the highest probability to have gone extinct could be closely related to taxa from the following list: *Rheum*, *Ranunculus glacialis* and *Micranthes melaleuca* that disappeared at ~17,000 cal. yrs BP; *Platyspermation crassifoium*, *Pedicularis groenlandica*, *Puccinellia* and *Saxifraga hiculus* that disappeared at ~11,000 cal. yrs BP; *Dryas*, Pooideae, *Aira praecox*, *Pedicularis groenlandica*, Poeae, Loliinae, *Draba* and *Koenigia* sect. *Aconogonon* that disappeared at ~ 5,000 cal. yrs BP.

3.5 DISCUSSION

This study is the first to compile lake sediment trnL gh data on a sub-continental scale with samples spanning the Pleistocene to Holocene transition providing a sub-continental vegetation composition and richness changes over the last 28,000 years. Our results point toward an overall resilient vegetation species spool in the study area over the last 28,000 years with 60% of the detected taxa present in all time slices. Resilience in the plant species pool can be expected in the area as it covers a large area nowadays covering diverse biota (Hillebrand, 2004; Johnson, 2009). In addition, the Pleistocene was known to have been species rich so potentially taxa from the steppe-tundra can have adapted to more specialised niche and structures modern biota (Courtin et al., 2021; Huang et al., 2021; Zimov; 2012). Still, 40% of the species pool is sensitive to changes and a shift in the vegetation composition is observed between the Pleistocene and the Holocene. Our results are in

accordance with the literature describing a more open vegetation during the Pleistocene dominated by herbaceous taxa such as graminoids and forbs in contrast to the more forested Holocene (Anderson & Lozhkin 2015; Courtin et al., 2021; Guthrie, 2001; Johnson, 2009; Jørgensen et al., 2012; Willerslev et al., 2014; Zimov et al., 2012). In addition, we observe a steady decrease of taxonomic richness through time over our entire record starting at the Pleistocene to Holocene transition. Such loss of plant diversity between the Pleistocene and the Holocene have been reported in the study area (Courtin et al., 2021; Huang et al., 2021).

Table 3.3: List of extirpated candidate taxa. Taxa with the highest probability of extinction are marked in bold.

Plant type	Family	Best assignment	Age of disappearance (cal. yrs BP)	Plant type	Family	Best assignment	Age of disappearance (cal. yrs BP)
herb	Plantaginaceae	*Lagotis glauca*	21000	**grass**	**Poaceae**	**Loliinae**	**5000**
herb	**Polygonaceae**	***Rheum***	**17000**	**grass**	**Poaceae**	**Poeae**	**5000**
herb	**Ranunculaceae**	***Ranunculus glacialis***	**17000**	**grass**	**Poaceae**	**Pooideae**	**5000**
herb	**Saxifragaceae**	***Micranthes melaleuca***	**17000**	**herb**	**Polygonaceae**	***Koenigia* sect. *Aconogonon***	**5000**
herb	Asteraceae	*Saussurea*	15000	**shrub**	**Rosaceae**	***Dryas***	**5000**
herb	Boraginaceae	*Myosotis alpestris*	15000	herb	Boraginaceae	Boraginaceae	3000
herb	Primulaceae	*Douglasia*	13000	herb	Boraginaceae	*Myosotis alpestris*	3000
shrub	**Alseuosmiaceae**	***Platyspermation crassifolium***	**11000**	herb	Brassicaceae	*Parrya*	3000
herb	**Orobanchaceae**	***Pedicularis groenlandica***	**11000**	herb	Asteraceae	Asteroideae	1000
grass	**Poaceae**	***Puccinellia***	**11000**	grass	Cyperaceae	*Carex*	1000
herb	**Saxifragaceae**	***Saxifraga hirculus***	**11000**	shrub	Ericaceae	*Cassiope tetragona*	1000
herb	Papaveraceae	*Papaver*	9000	shrub	Humiriaceae	*Humiria*	1000
herb	Plantaginaceae	*Callitriche hermaphroditica*	9000	shrub	Ixonanthaceae	*Cyrillopsis paraensis*	1000
herb	Rosaceae	Rosoideae	9000	herb	Juncaceae	*Luzula*	1000
herb	Polygonaceae	*Bistorta plumosa*	7000	herb	Onagraceae	*Epilobium rigidum*	1000
herb	Saxifragaceae	*Saxifraga serpyllifolia*	7000	herb	Orobanchaceae	*Pedicularis lanata*	1000
herb	**Brassicaceae**	***Draba***	**5000**	herb	Papaveraceae	*Meconopsis integrifolia*	1000
herb	**Brassicaceae**	***Draba***	**5000**	herb	Polygonaceae	*Bistorta vivipara*	1000
herb	**Orobanchaceae**	***Pedicularis groenlandica***	**5000**	herb	Polygonaceae	*Bistorta vivipara*	1000
grass	**Poaceae**	***Aira praecox***	**5000**	herb	Polygonaceae	*Oxyria digyna*	1000
grass	**Poaceae**	**Loliinae**	**5000**	shrub	Salicaceae	Saliceae	1000

3.5.1 Biotic and abiotic changes in the ecosystem - a cocktail for extinction

Both compositional shift and diversity decrease happen during the Pleistocene to Holocene transition between the last glacial maximum (LGM) and the early Holocene as illustrated with an increased taxonomic replacement rate at the time, which contrasts with the relative lower replacement rate of the Pleistocene. This lower replacement rate during the Pleistocene suggests a resilient and stable species pool between ~28,000 and 19,000 years BP. This could be explained by the age and range spanned by this past biota (Johnson 2009). Indeed, according to the species-area hypothesis, the larger the area is covered by a biota, the more diverse it can be (Connor & McCoy, 1979; He & Legendre, 2002; Holt et al., 1999). In addition, the Pleistocene steppe-tundra can be compared to old-growth grasslands, which are ancient ecosystems with species-diverse herbaceous plant communities, supported by fires, megafaunal herbivores that limit tree growth (Buisson et al., 2018; Veldman et al., 2015). Such old-growth grasslands not only are resilient but also are dependent on endogenous disturbances that are part of their evolutionary history and internal dynamics to maintain plant diversity and vegetation structure as illustrated in modern tropical and treeless old-growth grasslands (Buisson et al., 2018). Higher taxonomic richness leads to higher temporal stability of the ecosystem, which allowed persistence of the steppe-tundra for thousands of years under similar climatic conditions (Johnson, 2009; Pennekamp et al., 2018).

At the transition from the Pleistocene to the Holocene, starting from the post-LGM (~17,000 yrs BP) until the early Holocene high climate variability within a short period of time was observed in the study area starting with an extreme cold event, a warming during the Bølling-Allerød interstadial, another cold event with the Younger Dryas and a warming at the onset of the Holocene (Rasmussen et al., 2006). Such rapid changes in the climate regime might have impacted the resilience of the previously stable steppe-tundra as large and species rich areas are more stable to endogenous disturbances through time (e.g., frequent fires and native megafaunal herbivory) but more sensitive to exogenous disturbance such as climatic changes (Buisson et al., 2018; Pennekamp et al., 2018). During the Holocene, the previously widely spread steppe-tundra has disappeared and the area it spanned is now fragmented in different biota (Binney et al., 2009). Each of them established recently (only few thousand years ago) and did not reach their final, stable state which could explain why the replacement rate is decreasing through the Holocene but did not

reach a plateau as stable as the Pleistocene as the last major disturbance is still recent (Chesson & Huntly, 1997; Wang et al., 2020).

Therefore, at the Pleistocene – Holocene transition, in our study area, a shift in climate regime, plant composition, habitat loss and decrease in species richness were observed. This is the recipe of changes the plant biota have to face today worldwide, usually leading to extinction, even in the Arctic (Cowie et al., 2022).

3.5.2 Identification and quantification of potential plant taxa loss

Despite the core, resilient species pool of 60% of the detected plant taxa present at all times in the study area, 121 potential plant taxa are extirpated which represents 26% of all identified potential plant taxa. Among them, 41 are assigned from candidate ASVs. This means that the maximum extinction rate over the last 28,000 years in the study area would be equivalent to the extirpation rates of the candidate taxa: 3.1 extinction per million species years (E/MSY) measured following Pimm et al 2006. This rate being the maximal extinction rate is, even if higher than the background extinction rate of plants by 9-62 times, much lower than the estimated ongoing extinction rates of 18-26 E/MSY (De Vos, et al., 2015; Humphreys et al., 2019). Here we use this method to estimate past extinction and be able to compare both fossil and modern and cross kingdom rates even if we acknowledge that comparison of fossil-to-modern extinction rate are not optimal, background rates can vary from one taxon to the next and that comparison of modern short-term rates with fossil long-term rates does not take into account interval-rate effect (Barnosky et al., 2011 and references therein). One explanation of the lower extinction at the transition between the Pleistocene and the Holocene compared to the modern one could be that modern extinction can be explained by a decrease of richness not coupled with an increased species replacement rate but here, we observe both a decrease of richness and an increase of replacement rate pointing toward a shift of composition and therefore an adaptational response of species to change (Blowes, 2019). The other difference is that the modern extinction rates are measured taking into account the worldwide diversity with higher extinction rates in the tropics and lower ones towards the poles (Humphreys et al., 2019) when in this study, we only inventoried the extinction species at a sub-continental level.

Extirpation is observed at every time slice. Most of the time, extirpated taxa re-appear at a later time slice (**Supplementary Figure 3.1**). As the longer a taxon is extirpated from the

area, the more likely it is for it to have undergone extinction, after accounting for reappearance rates, we identified 3 potential extinction events at ~17,000, 11,000 and 5,000 cal. yrs BP with extinction rates of respectively 2.53 E/MSY, 4.64 E/MSY and 4.32 E/MSY. At those times, the observed extinction rates were 7 to 90 times above background plant extinction rates estimated between 0.05 to 0.35 E/MSY (De Vos, et al., 2015; Humphreys et al., 2019). Such waves of extinction can happen in species-rich communities as, after extinctions, they are more likely to suffer secondary extinctions for example with negative diversity-stability relationship (Ives & Carpenter 2007).

This three-stage pattern almost perfectly fit the megafaunal extinctions and extirpations after 40,000 cal. yrs BP in the Holarctic. The first wave occurring during the Last Glacial Maximum (LGM, 26,500–19,000 cal. yrs BP), the second wave occurred between 15,000–10,000 cal. yrs BP, when other species until at least ~6,000 years ago (Clark 2009, Murchie et al., 2021). Therefore, plant and megafauna extinction events seem to have happened in parallel but not with the same amplitude as megafaunal extinction rates were higher than plant extinction rates (Johnson, 2009; Meltzer, 2020). For instance, if we would measure the extinction rate of megafaunal genera over the last 50,000 years, 97 genera went extinct from 150 total, which resulted in 12.93 E/MGY (Barnosky et al., 2004). But, extinction rates of species and genera should be broadly similar, we can therefore compare it to our maximum plant extinction rate of 3.1 E/MSY which would be 4 times lower than the megafauna one (Barnosky et al., 2011; Pimm et al., 2014; Stuart, 2015). In addition, two pulses of community turnover of smaller mammals were observed around 14,500 cal. yrs BP and between 11,000 to 7,500 cal. yrs BP with evidence for extirpations (Blois et al., 2010).

When compared to modern extinction rates, mammalian ones are 10 times higher than plant ones over the last 500 years but with mammals being the most studied taxonomic group over that same time (Cowie et al., 2022). When looking at the IUCN red list, a similar proportion of 20% of modern mammals and plant species are threatened (Brummitt et al., 2015; Davies, 2018). Our results could suggest that plants are less resilient now than they were at the transition between the Pleistocene and the Holocene when more megafauna (only a fragment of the mammalian community) went extinct in comparison to the plant community.

Not all glacial-interglacial transitions are alike. During the last ~2.5 million years of the Pleistocene species witnessed and survived many glacial-interglacial transitions (Lisiecki & Raymo 2005, Railsback et al., 2015). Among them, 19 out of the 38 large land mammal's genera from North America that went extinct at the Pleistocene-Holocene transition were present throughout the entire Pleistocene (Meltzer, 2020). The Pleistocene to Holocene transition was unique in that sense as it was the stage of observed plant taxa loss, which seems to happen in parallel to the shift of plant composition change and the loss of the Pleistocene steppe-tundra biota as well as the megafauna's extinction.

3.5.3 Characterisation of potential taxa loss

In addition to the identification of plant taxa loss, we characterised the taxa most likely to have undergone extinction. From our results, they are rare, specialists contributing less than average to the functionality of the ecosystem, generally fitting modern observations of plant extinction (Zettlemoyer et al., 2019).

There is no bias on the plant type with no plant type significatively more impacted by extinction than another one and this is similar to other studies showing that extinction is not group specific. In addition, there is no real bias on taxonomic groups even if the three plant families Asteraceae, Juncaceae and Brassicaceae might be more sensitive to plant extinction between the Pleistocene and the Holocene than the other families detected in this study. Asteraceae and Brassicaceae are also among the families with the most taxa detected in this study (**Supplementary Table 2.4**). It has already been identified that most abundant families are also the most sensitive for loss of richness and extirpation of taxa between the Pleistocene and the Holocene (Courtin et al., 2021; Huang et al., 2021)

This shows that the loss of the steppe-tundra biota is likely not due to extinction of key plant species as the observed likely lost taxa were not abundant nor key contributors to the functionality of their past ecosystem. It seems more likely that the lost taxa were rare and unable to adapt fast enough to their changing environment. One explanation could be that with fast changes happening in the structure of the vegetation due to the loss of keystone, plant assemblages' changes and each taxa had to respond individually. This led to the creation of new assemblages "recycling" plants taxa previously present in sufficient abundance and more generalist that could respond and adapt faster than more rare taxa with more specialised and unique interactions.

3.5.4 Limits of the method

Databases are made from modern extant species that by definition are not extinct. Usually using the 100% match to such a database, studies investigating make the assumption that (1) the databases are complete and (2) the species pool of past vegetation exactly mirror the modern species pool. It fully oversees potentially rare and/or extinct taxa that would be absent from databases. With the investigation of all ASVs assigned with a minimum of 90% confidence to the databases, we were able detect more potential plant taxa than we would have, using the widespread stringent method investigating only ASVs assigned at 100% confidence. This allowed us to explore potentially extinct and/or rare taxa. This first improvement of the method is a step forward as taxa are expected to become rare before extinct (Zettlemoyer et al., 2019). Therefore, with the aim to detect potential extinction, here, we identify and quantify taxonomic signals the best way possible. This is the first study to use amplicon sequence variants (ASVs) using the trnL (gh) marker not assigned at 100% to one of the databases to investigate potential past plant taxa loss. With this new approach, we could identify more potentially rare taxa than in previous studies.

Still, we faced some limits with our methods that would need further improvement in future work. For instance, via the presence of candidates (absent from the database) but present in the modern time slice we point to the fact that databases are incomplete, especially from rare taxa (Breitwieser et al., 2019). To identify past extinction, we must be able to identify species and therefore have the most complete databases possible.

Furthermore, taxa assigned at 100% but not to species level hold the potential of containing more than one species signal. Among which are potentially extinct or absent from the database taxa. It is therefore very challenging to disentangle the different signals within ASVs, which retain information on more than one taxon. This point toward the limit of the genetic marker, which performs the best to work with ancient, short DNA fragments but is not optimal to discriminate species (Taberlet et al., 2007). Because the detection of species is so difficult, disentangling the past from the modern signal is challenging.

Indeed, ASVs assigned at 100% to family or genus level at best can hold the signal of more than one species as it only shows that the marker cannot distinguish between different species. As different species share the same sequence, there is also a probability that it was also the case for extinct species. Therefore, in the case of 100% assignment to family of

genus level at best, it holds the information on a mixture of species extant and potentially extinct. This would need to be accounted for in future studies and one solution would be to use another marker or set markers for better species identification. The other best alternative would be to use metagenomics, which holds the potential for best species identification (Murchie et al., 2021; Pedersen et al., 2016; Sangwan et al., 2016).

3.5.5 Conclusions and perspectives

With this study, using the trnL gh marker on *seda*DNA, we were able to detect more taxa than other studies, more rare and potentially extinct by using the information held on ASVs assigned with a 90% confidence and more against the databases. In addition, it is the first study to reconstruct past vegetation changes over the last 28,000 years compiling 10 lake sediment records from 9 sites covering the previous Pleistocene steppe-tundra area. In addition to the standard plant composition and richness in through time, this is the first study to investigate potential plant extinction using *seda*DNA at the transition between the Pleistocene and the Holocene. We were able to quantify an above background plant extinction rate that happened in three waves in parallel to both the Pleistocene steppe-tundra loss and the megafauna extinction showing the potential importance in past trophic interaction and potential co-extinction events in parallel to the loss of an entire ecosystem. This kind of study could be improved by refining even more the species detection and quantification using a metagenomic approach for example, but it paves the way to understand interactions between fast ecosystem loss with plant compositional shift and cascades of extinction across taxonomic kingdoms even in the Arctic.

FUNDING

This project has received funding from the Priority Project "International Continental Drilling Program" of the German Research Foundation and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772852) and the Initiative and Networking fund of the Helmholtz Association.

REFERENCES

Anderson, P. M., & Lozhkin, A. V. (2015). Late quaternary vegetation of Chukotka (Northeast Russia), implications for glacial and Holocene environments of Beringia. *Quaternary Science Reviews*, *107*, 112-128. https://doi.org/10.1016/j.quascirev.2014.10.016

Andreev, A. A., Raschke, E., Biskaborn, B. K., Vyse, S. A., Courtin, J., Böhmer, T., ... & Herzschuh, U. (2021). Late Pleistocene to Holocene vegetation and climate changes in northwestern Chukotka (Far East Russia) deduced from lakes Ilirney and Rauchuagytgyn pollen records. Boreas, 50(3), 652-670. https://doi.org/10.1111/bor.12521

Bakker, E. S., Gill, J. L., Johnson, C. N., Vera, F. W., Sandom, C. J., Asner, G. P., & Svenning, J. C. (2016). Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy of Sciences*, *113*(4), 847-855. https://doi.org/10.1073/pnas.1502545112

Bálint, M., Pfenninger, M., Grossart, H. P., Taberlet, P., Vellend, M., Leibold, M. A., ... & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, *33*(12), 945-957. https://doi.org/10.1016/j.tree.2018.09.003

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L., & Shabel, A. B. (2004). Assessing the causes of late Pleistocene extinctions on the continents. *science*, 306(5693), 70-75. https://doi.org/10.1126/science.1101476

Barnosky, A., Matzke, N., Tomiya, S. et al. (2011) Has the Earth's sixth mass extinction already arrived? *Nature* 471, 51–57. https://doi.org/10.1038/nature09678

Baumer, M. M., Wagner, B., Meyer, H., Leicher, N., Lenz, M., Fedorov, G., ... & Melles, M. (2021). Climatic and environmental changes in the Yana Highlands of north-eastern Siberia over the last c. 57 000 years, derived from a sediment core from Lake Emanda. *Boreas*, 50(1), 114-133. https://doi.org/10.1111/bor.12476

Bellard, C., Cassey, P., & Blackburn, T. M. (2016). Alien species as a driver of recent extinctions. *Biology letters*, 12(2), 20150623. https://doi.org/10.1098/rsbl.2015.0623

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., ... & Vazhenina, L. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, *28*(23-24), 2445-2464. https://doi.org/10.1016/j.quascirev.2009.04.016

Birks, H. H. (2001). Plant macrofossils. In *Tracking environmental change using lake sediments* (pp. 49-74). Springer, Dordrecht.

Biskaborn, B. K., Subetto, D. A., Savelieva, L. A., Vakhrameeva, P. S., Hansche, A., Herzschuh, U., ... & Diekmann, B. (2016). Late Quaternary vegetation and lake system dynamics in north-eastern Siberia: Implications for seasonal climate variability. *Quaternary Science Reviews*, 147, 406-421. https://doi.org/10.1016/j.quascirev.2015.08.014

Blois, J. L., McGuire, J. L., & Hadly, E. A. (2010). Small mammal diversity loss in response to late-Pleistocene climatic change. *Nature*, 465(7299), 771-774. https://doi.org/10.1038/nature09077

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., ... & Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, 366(6463), 339-345. https://doi.org/10.1126/science.aaw1620

Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U., & Kowalczyk, R. (2015). European bison as a refugee species? Evidence from isotopic data on Early Holocene bison and other large herbivores in northern Europe. *PloS one*, *10*(2), e0115090. https://doi.org/10.1371/journal.pone.0115090

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

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in bioinformatics*, *20*(4), 1125-1136. https://doi.org/10.1093/bib/bbx120

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., ... & Nic Lughadha, E. M. (2015). Green plants in the red: A baseline global assessment for the IUCN sampled Red List Index for plants. *PloS one*, 10(8), e0135152. https://doi.org/10.1371/journal.pone.0135152

Buisson, E., Le Stradic, S., Silveira, F. A., Durigan, G., Overbeck, G. E., Fidelis, A., ... & Veldman, J. W. (2019). Resilience and restoration of tropical and subtropical grasslands, savannas, and grassy woodlands. *Biological Reviews*, *94*(2), 590-609. https://doi.org/10.1111/brv.12470

Cardoso, P., Erwin, T. L., Borges, P. A., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological conservation*, 144(11), 2647-2655. https://doi.org/10.1016/j.biocon.2011.07.024

Chesson, P., & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *The American Naturalist*, *150*(5), 519-553.

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. https://doi.org/10.1111/bor.12338

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... & McCabe, A. M. (2009). The last glacial maximum. *science*, *325*(5941), 710-714. https://doi.org/10.1126/science.1172873

Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist*, *113*(6), 791-833.

Courtin, J., Andreev, A. A., Raschke, E., Bala, S., Biskaborn, B. K., Liu, S., ... & Herzschuh, U. (2021). Vegetation changes in southeastern Siberia during the Late Pleistocene and the Holocene. *Frontiers in Ecology and Evolution*, *9*, 625096. https://doi.org/10.3389/fevo.2021.625096

Cowie, R. H., Regnier, C., Fontaine, B., & Bouchet, P. (2017). Measuring the sixth extinction: what do mollusks tell us. *The Nautilus*, 131(1), 3-41.

Cowie, R. H., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation?. *Biological Reviews.* 97: 640-663. https://doi.org/10.1111/brv.12816

Davies, T., Cowley, A., Bennie, J., Leyshon, C., Inger, R., Carter, H., ... & Gaston, K. (2018). Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment. *PloS one*, 13(9), e0203694. https://doi.org/10.1371/journal.pone.0203694

De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R., & Pimm, S. L. (2015). Estimating the normal background rate of species extinction. *Conservation biology*, 29(2), 452-462. https://doi.org/10.1111/cobi.12380

Dunn, R. R. (2005). Modern insect extinctions, the neglected majority. *Conservation biology*, 19(4), 1030-1036. https://doi.org/10.1111/j.1523-1739.2005.00078.x

Earth Resources Observation And Science (EROS) Center (2017) "Global 30 Arc-Second Elevation (GTOPO30)." U.S. Geological Survey. doi: 10.5066/F7DF6PQS.

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Galetti, M., Brocardo, C. R., Begotti, R. A., Hortenci, L., Rocha-Mendes, F., Bernardo, C. S. S., ... & Siqueira, T. (2017). Defaunation and biomass collapse of mammals in the largest Atlantic forest remnant. *Animal Conservation*, 20(3), 270-281. https://doi.org/10.1111/acv.12311

Gill, J. L. (2014). Ecological impacts of the late Quaternary megaherbivore extinctions. *New Phytologist*, 201(4), 1163-1169. https://doi.org/10.1111/nph.12576

Goudie, A. S. (2018). *Human impact on the natural environment*. John Wiley & Sons.

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. https://doi.org/10.1016/S0277-3791(00)00099-8

He, F., & Legendre, P. (2002). Species diversity patterns derived from species–area models. *Ecology*, *83*(5), 1185-1198. https://doi.org/10.1890/0012-9658(2002)083[1185:SDPDFS]2.0.CO;2

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, *163*(2), 192-211.

Holt, R. D., Lawton, J. H., Polis, G. A., & Martinez, N. D. (1999). Trophic rank and the species–area relationship. *Ecology*, *80*(5), 1495-1504. https://doi.org/10.1890/0012-9658(1999)080[1495:TRATSA]2.0.CO;2

Huang, S., Stoof-Leichsenring, K., Liu, S., Courtin, J., Andreev, A. A., Pestryakova, L. A., & Herzschuh, U. (2021). Plant sedimentary ancient DNA from Far East Russia covering the last 28,000 years reveals different assembly rules in cold and warm climates. *Frontiers in Ecology and Evolution*, 873. https://doi.org/10.3389/fevo.2021.763747

Humphreys, A. M., Govaerts, R., Ficinski, S. Z., Nic Lughadha, E., & Vorontsova, M. S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature ecology & evolution*, 3(7), 1043-1047. https://doi.org/10.1038/s41559-019-0906-2

Ives, A. R., & Carpenter, S. R. (2007). Stability and diversity of ecosystems. *science*, *317*(5834), 58-62. https://doi.org/10.1126/science.1133258

Jackson, S. T., & Weng, C. (1999). Late Quaternary extinction of a tree species in eastern North America. *Proceedings of the National Academy of Sciences*, 96(24), 13847-13852. https://doi.org/10.1073/pnas.96.24.13847

Johnson, C. N. (2009). Ecological consequences of Late Quaternary extinctions of megafauna. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1667), 2509-2519. https://doi.org/10.1098/rspb.2008.1921

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. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Kienast, F., Siegert, C., Dereviagin, A., & Mai, D. H. (2001). Climatic implications of Late Quaternary plant macrofossil assemblages from the Taymyr Peninsula, Siberia. *Global and Planetary Change*, *31*(1-4), 265-281. https://doi.org/10.1016/S0921-8181(01)00124-2

Lenz, M., Lenz, M. M., Andreev, A. Scheidt, S., Gromig, R., lebas, E., … & Wagner, B. (2022). Climate and environmental history of Lake Levinson-Lessing, Taymyr Peninsula, during the last 62 kyr. Journal of Quaternary Science, 37(5), 836-850. https://doi.org/10.1002/jqs.3384

Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*, 20(1). https://doi.org/10.1029/2004PA001071

Lughadha, E. N., Bachman, S. P., Leão, T. C., Forest, F., Halley, J. M., Moat, J., ... & Walker, B. E. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet*, 2(5), 389-408. https://doi.org/10.1002/ppp3.10146

Malhi, Y., Doughty, C. E., Galetti, M., Smith, F. A., Svenning, J. C., & Terborgh, J. W. (2015). Megafauna and ecosystem function from the Pleistocene to the Anthropocene. *Proceedings of the National Academy of Sciences*, 113(4), 838-846. https://doi.org/10.1073/pnas.1502540113

Meltzer, D. J. (2020). Overkill, glacial history, and the extinction of North America's Ice Age megafauna. *Proceedings of the National Academy of Sciences*, 117(46), 28555-28563. https://doi.org/10.1073/pnas.2015032117

Müller, S., Tarasov, P. E., Andreev, A. A., Tütken, T., Gartz, S., & Diekmann, B. (2010). Late Quaternary vegetation and environments in the Verkhoyansk Mountains region (NE Asia) reconstructed from a 50-kyr fossil pollen record from Lake Billyakh. *Quaternary Science Reviews*, 29(17-18), 2071-2086. https://doi.org/10.1016/j.quascirev.2010.04.024

Murchie, T. J., Monteath, A. J., Mahony, M. E., Long, G. S., Cocker, S., Sadoway, T., ... & Poinar, H. N. (2021). Collapse of the mammoth-steppe in central Yukon as revealed by ancient environmental DNA. *Nature Communications*, *12*(1), 1-18. https://doi.org/10.1038/s41467-021-27439-6

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

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., ... & 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

Pennekamp, F., Pontarp, M., Tabi, A., Altermatt, F., Alther, R., Choffat, Y., ... & Petchey, O. L. (2018). Biodiversity increases and decreases ecosystem stability. *Nature*, *563*(7729), 109-112. https://doi.org/10.1038/s41586-018-0627-8

Pimm, S., Raven, P., Peterson, A., Şekercioğlu, Ç. H., & Ehrlich, P. R. (2006). Human impacts on the rates of recent, present, and future bird extinctions. *Proceedings of the National Academy of Sciences*, *103*(29), 10941-10946. https://doi.org/10.1073/pnas.0604181103

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *science*, 344(6187), 1246752. https://doi.org/10.1126/science.1246752

Pimm, S. L. (2020). What we need to know to prevent a mass extinction of plant species. *Plants, People, Planet*, 3(1), 7-15. https://doi.org/10.1002/ppp3.10160

Railsback, L. B., Gibbard, P. L., Head, M. J., Voarintsoa, N. R. G., & Toucanne, S. (2015). An optimized scheme of lettered marine isotope substages for the last 1.0 million years, and the climatostratigraphic nature of isotope stages and substages. *Quaternary Science Reviews*, 111, 94-106. https://doi.org/10.1016/j.quascirev.2015.01.012

Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., ... & Ruth, U. (2006). A new Greenland ice core chronology for the last glacial termination. *Journal of Geophysical Research: Atmospheres*, *111*(D6). https://doi.org/10.1029/2005JD006079

Régnier, C., Fontaine, B., & Bouchet, P. (2009). Not knowing, not recording, not listing: numerous unnoticed mollusk extinctions. *Conservation Biology*, 23(5), 1214-1221. https://doi.org/10.1111/j.1523-1739.2009.01245.x

Régnier, C., Achaz, G., Lambert, A., Cowie, R. H., Bouchet, P., & Fontaine, B. (2015). Mass extinction in poorly known taxa. *Proceedings of the National Academy of Sciences*, 112(25), 7761-7766. https://doi.org/10.1073/pnas.1502350112

Rijal, D. P., Heintzman, P. D., Lammers, Y., Yoccoz, N. G., Lorberau, K. E., Pitelkova, I., ... & Alsos, I. G. (2021). Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. *Science Advances*, *7*(31), eabf9557. https://doi.org/10.1126/sciadv.abf9557

Rull, V. (2022). Inductive prediction in biology: Are long-term ecological and evolutionary processes predictable? *EMBO reports*, e54846. https://doi.org/10.15252/embr.202254846

Saatkamp, A., Cochrane, A., Commander, L., Guja, L. K., Jimenez-Alfaro, B., Larson, J., ... & Walck, J. L. (2019). A research agenda for seed-trait functional ecology. *New Phytologist*, *221*(4), 1764-1775. https://doi.org/10.1111/nph.15502

Sandom, C., Faurby, S., Sandel, B., & Svenning, J. C. (2014). Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1787), 20133254. https://doi.org/10.1098/rspb.2013.3254

Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome,* 4(1), 1-11. https://doi.org/10.1186/s40168-016-0154-5

Soininen, J., Bartels, P. I. A., Heino, J., Luoto, M., & Hillebrand, H. (2015). Toward more integrated ecosystem research in aquatic and terrestrial environments. *BioScience*, *65*(2), 174-182. https://doi.org/10.1093/biosci/biu216

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

Stuart, A. J. (2015). Late Quaternary megafaunal extinctions on the continents: a short review. *Geological Journal*, *50*(3), 338-363. https://doi.org/10.1002/gj.2633

Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome biology*, *21*(1), 1-12. https://doi.org/10.1186/s13059-020-02023-1

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences*, 112(24), 7519-7523. https://doi.org/10.1073/pnas.1502408112

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

Tomescu, A. M., Bomfleur, B., Bippus, A. C., & Savoretti, A. (2018). Why are bryophytes so rare in the fossil record? A spotlight on taphonomy and fossil preservation. In *Transformative paleobotany* (pp. 375-416). Academic Press. https://doi.org/10.1016/B978-0-12-813012-4.00016-4

Vachula, R. S., Huang, Y., Longo, W. M., Dee, S. G., Daniels, W. C., & Russell, J. M. (2019). Evidence of Ice Age humans in eastern Beringia suggests early migration to North America. *Quaternary Science Reviews*, 205, 35-44. https://doi.org/10.1016/j.quascirev.2018.12.003

Veldman, J. W., Buisson, E., Durigan, G., Fernandes, G. W., Le Stradic, S., Mahy, G., ... & Bond, W. J. (2015). Toward an old-growth concept for grasslands, savannas, and woodlands. *Frontiers in Ecology and the Environment*, *13*(3), 154-162. https://doi.org/10.1890/140270

von Hippel, B., Stoof-Leichsenring, K. R., Schulte, L., Seeber, P., Epp, L. S., Biskaborn, B. K., ... & Herzschuh, U. (2021). Long-term fungus–plant co-variation from multi-site sedimentary

ancient DNA metabarcoding in Siberia. *bioRxiv*. https://doi.org/10.1101/2021.11.05.465756

Vyse, S. A., Herzschuh, U., Andreev, A. A., Pestryakova, L. A., Diekmann, B., Armitage, S. J., & Biskaborn, B. K. (2020). Geochemical and sedimentological responses of arctic glacial Lake Ilirney, chukotka (far east Russia) to palaeoenvironmental change since∼ 51.8 ka BP. Quaternary Science Reviews, 247, 106607. https://doi.org/10.1016/j.quascirev.2020.106607

Wang, Y., Shipley, B. R., Lauer, D. A., Pineau, R. M., & McGuire, J. L. (2020). Plant biomes demonstrate that landscape resilience today is the lowest it has been since end-Pleistocene megafaunal extinctions. *Global Change Biology*, 26(10), 5914-5927. https://doi.org/10.1111/gcb.15299

Wang, Y., Pedersen, M. W., Alsos, I. G., De Sanctis, B., Racimo, F., Prohaska, A., ... & Willerslev, E. (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nature*, *600*(7887), 86-92. https://doi.org/10.1038/s41586-021-04016-x

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

Zettlemoyer, M. A., Schultheis, E. H., & Lau, J. A. (2019). Phenology in a warming world: differences between native and non-native plant species. *Ecology letters*, *22*(8), 1253-1263. https://doi.org/10.1111/ele.13290

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., ... & Herzschuh, U. (2017). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences*, *14*(3), 575-596. https://doi.org/10.5194/bg-14-575-2017

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. https://doi.org/10.1016/j.quascirev.2012.10.005

4 MANUSCRIPT III

Pleistocene glacial and interglacial ecosystems inferred from ancient DNA analyses of permafrost sediments from Batagay megaslump, East Siberia

Authors

Jérémy Courtin[1*], Amedea Perfumo[1], Andrei A. Andreev[1], Thomas Opel[1,2], Kathleen R. Stoof-Leichsenring[1], Mary E. Edwards[3,4], Julian B. Murton[5], Ulrike Herzschuh[1,6,7*]

Affiliation

[1]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Potsdam, Germany

[2]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, PALICE Helmholtz Young Investigator Group, Bremerhaven, Germany

[3]Geography and Environmental Science, University of Southampton, Southampton, UK

[4]Alaska Quaternary Centre, University of Alaska, Fairbanks. USA

[5]Department of Geography, University of Sussex, Brighton, UK

[6]Institute of Environmental Science and Geography, University of Potsdam, Potsdam-Golm, Germany

[7]Institute of Biology and Biochemistry, University of Potsdam, Potsdam-Golm, Germany

***Correspondence**

Jérémy Courtin: jeremy.courtin@awi.de; Ulrike Herzschuh, ulrike.herzschuh@awi.de.

Status

Published in *Environmental DNA.* doi: 10.1002/edn3.336

Additional information in **APPENDIX 3**.

4.1 ABSTRACT

Pronounced glacial and interglacial climate cycles characterised northern ecosystems during the Pleistocene. Our understanding of the resultant community transformations and past ecological interactions strongly depends on the taxa found in fossil assemblages. Here, we present a shotgun metagenomic analysis of sedimentary ancient DNA (*seda*DNA) to infer past ecosystem-wide biotic composition (from viruses to megaherbivores) from the Middle and Late Pleistocene at the Batagay megaslump, east Siberia. The shotgun DNA records of past vegetation composition largely agree with pollen and plant metabarcoding data from the same samples. Interglacial ecosystems at Batagay attributed to Marine Isotope Stage (MIS) 17 and MIS 7 were characterised by forested vegetation (*Pinus*, *Betula*, *Alnus*) and open grassland. The microbial and fungal communities indicate strong activity related to soil decomposition, especially during MIS17. The local landscape likely featured more open, herb-dominated areas, and the vegetation mosaic supported birds and small omnivorous mammals. Parts of the area were intermittently/partially flooded as suggested by the presence of water-dependant taxa. During MIS 3, the sampled ecosystems are identified as cold-temperate, periodically flooded grassland. Diverse megafauna (*Mammuthus*, *Equus*, *Coelodonta*) coexisted with small mammals (rodents). The MIS 2 ecosystems existed under harsher conditions, as suggested by the presence of cold-adapted herbaceous taxa. Typical Pleistocene megafauna still inhabited the area. The new approach, in which shotgun sequencing is supported by metabarcoding and pollen data, enables the investigation of community composition changes across a broad range of taxonomic groups and inferences about trophic interactions and aspects of soil microbial ecology.

Keywords: Batagay megaslump, metabarcoding, Pleistocene, pollen, sedaDNA, shotgun sequencing.

4.2 INTRODUCTION

Ongoing climate change affects biota at the ecosystem level (Pecl et al., 2017; Tylianakis, et al., 2008; Walther, 2010). Amplified arctic warming causes northern boreal and arctic ecosystems to experience more extreme climate change than other ecosystems in the Northern Hemisphere (Miller et al., 2010) fostering the northward extension of shrubs and trees into tundra ecosystems (Myers-Smith et al., 2019; Tape et al., 2006). Due to the complexity of biotic and abiotic interactions, predictions of species distribution patterns at ecosystem scale are difficult to establish (HilleRisLambers et al., 2013; Walther et al., 2002).

Understanding past ecological changes enhances our ability to understand recent and future warming effects on ecosystems (Botkin et al., 2007). The Pleistocene Epoch (~2,600–11.7 thousand years (kyr) before present, was characterised by alternating glacial and interglacial stages (Mix & Ruddiman, 1984; Shackleton, 1967), which involved major changes in past climate, ecosystems and genetic diversity patterns (Hewitt, 2000; Jackson & Blois, 2015). Continuous environmental records spanning several glacial-interglacial cycles are rare in boreal and arctic regions of Eurasia, being primarily represented by two deep sediment cores from Lake Baikal and Lake El'gygytgyn (Andreev et al., 2014; Brigham-Grette et al., 2013; Demske et al., 2002; Melles et al., 2012; Tarasov et al., 2005). Other Eurasian palaeoecological records older than MIS 3 are discontinuous, being derived from permafrost sediment sequences, but some date to the MIS 7 (e.g., Andreev et al., 2011, Wetterich et al., 2019 and references therein). The Batagay megaslump in east Siberia provides access to ecological information since the Middle Pleistocene (Ashastina et al., 2017; Murton et al., 2017, 2021; Opel et al., 2019).

A particular challenge is to retrieve information on past biota from such environmental records. Beyond the palaeofaunas that have been well documented from bone assemblages; plant macrofossil, pollen analysis (e.g., Andreev et al., 2011, 2012) and insect remains (e.g., Kuzmina et al., 2011) are the traditional methods for retrieving information on terrestrial ecological changes from permafrost sediments. Recently, the investigation of sedimentary ancient DNA (*seda*DNA) has become a promising approach (e.g., Willerslev et al., 2003, 2014). In particular several plant metabarcoding studies have focussed on vascular plants (Courtin et al., 2021; Epp et al., 2015; Liu et al. 2020; Sønstebø et al., 2010; Willerslev et al 2014; Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schirrmeister, et al., 2017;

Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schwamborn, et al., 2017). Other groups have also been investigated, such as fungi (Bellemain et al., 2013; Epp et al., 2012; Lydolph et al., 2005), bacteria (Wagner et al., 2007; Willerslev et al., 2004), invertebrates (Epp et al., 2012), birds (Epp et al., 2012), and mammals (e.g. Arnold et al., 2011; Haile et al., 2009; Wang et al., 2021; Willerslev et al., 2003, 2014). Recent advances in computational analysis and the development of new bioinformatic tools (such as Kraken2) and pipelines (Hübler et al., 2019; Piro et al., 2020; Wood et al., 2019), pave the way for shotgun sequencing of environmental DNA (Bálint et al., 2018), which allows the investigation of a broad range of organisms, from non-living viruses to mammals and plants within one single experiment (Ahmed et al., 2018; Murchie et al., 2021; Pedersen et al., 2016; Wang et al., 2021). Applied to *seda*DNA, it has the potential to provide, at least partially, an ecosystem view and snapshots of past glacial and interglacial environments.

Macrofossils mostly originate from plants growing at the study site and represent local conditions. However, the number of identifiable macrofossils is affected by taphonomy and preservation, which can vary strongly among taxa (e.g., Foote & Raup, 1996; Kienast et al., 2001; Tomescu et al., 2018). Similar issues affect vertebrates (Behrensmeyer, 1988; Gardner et al., 2016). Pollen records reflect regional ecological changes well; but information is constrained by different pollen productivities and transportability (Birkset al., 2012; van der Knaap, 1987). Moreover, the taxonomic resolution of pollen is limited, typically in trees and shrubs to genus/sub-genus, but more variably from species to family in herbaceous taxa (Beug, 1961; Moore et al., 1991). In a rigorous comparison, pollen and *seda*DNA (metabarcoding) from a lake-sediment core (Polar Urals; Clarke et al., 2019) yielded a similar floristic richness, 114 and 137 vascular plant taxa, respectively, but each method featured a different configuration of taxa. How *seda*DNA represents vegetation at a sampling site requires further work as both taphonomy, for which investigation methods are still in their infancy, and geomorphological processes, which are site dependant, impact the scale of the *seda*DNA signal (Giguet-Covex et al., 2019; Alsos et al., 2018). However, there is increasing evidence that DNA from lacustrine sediments reflects the catchment vegetation, while DNA from terrestrial sediments (e.g., palaeosols) provides a highly local signal (Edwards, 2020). *Seda*DNA is particularly well preserved in permafrost sediments as persistent sub-zero temperatures reduce microbial and enzymatic degradation (Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schwamborn, et al., 2017) and because *seda*DNA binds strongly to minerogenic sediments (Edwards, 2020).

Here we present data showing ecosystem-wide changes between Pleistocene glacial and interglacial intervals in east Siberia inferred from *seda*DNA shotgun sequencing, plant metabarcoding and pollen analysis. Five sediment samples from the Batagay megaslump are analysed. The dating of samples is imprecise: the oldest dates from the period relating to MIS 17 to MIS 16 (~650 kyr), another from MIS 16 to MIS 7 (~600 to ~200 kyr), the third to MIS 3 (~50 kyr), and the two youngest samples date to MIS 2 (~27 and ~23 kyr).

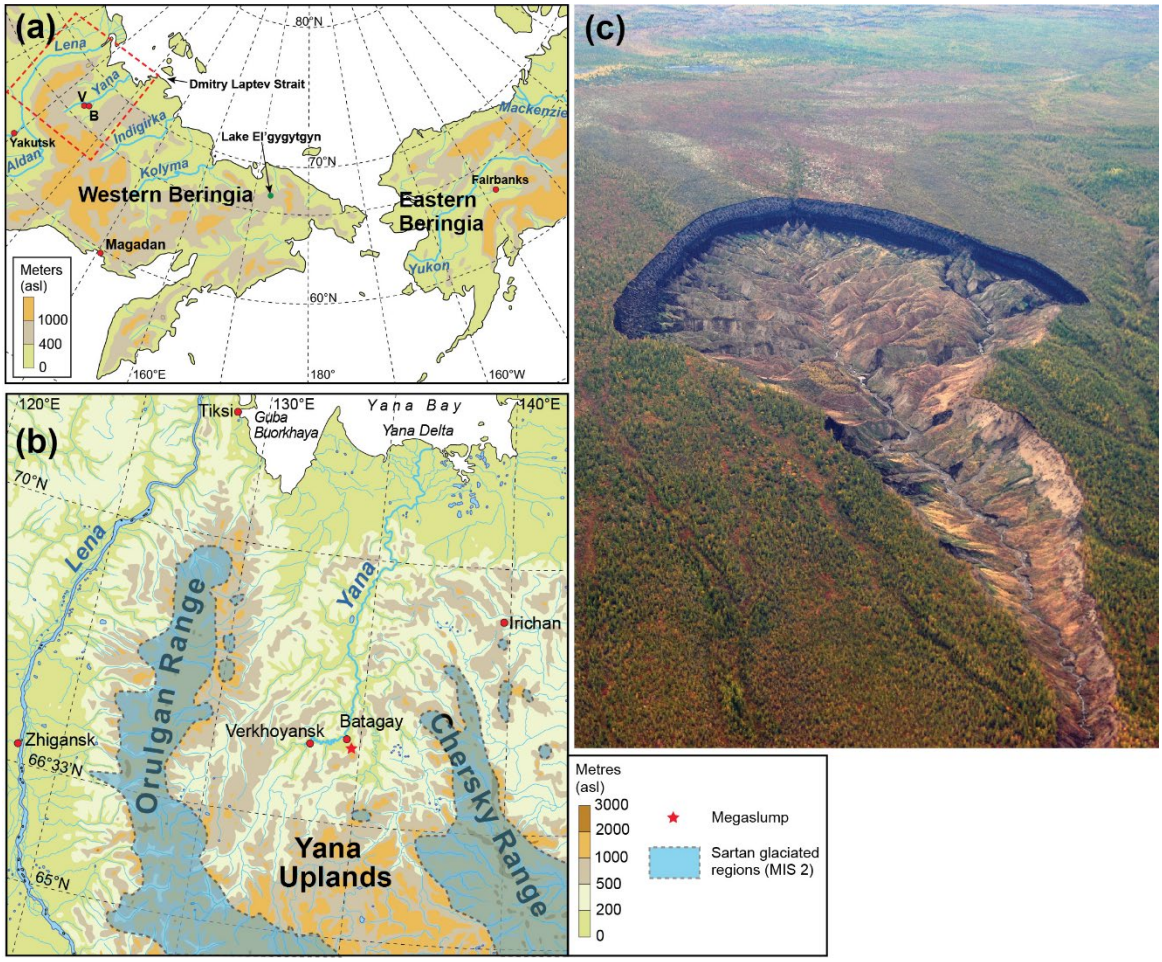


Figure 4.1 | Map of the study site and picture of the Batagay megaslump. A. Location of the study area in Beringia. **B.** Location FFI of the Batagay megaslump in the Yana Uplands. **C.** Aerial picture of the Batagay megaslump. Adapted from Murton et al., 2017.

Site description

The Batagay megaslump (67.58° N, 134.77° E) is located on a hillslope underlain by permafrost just of the Yana River in northern Yakutia, near the town of Batagay (**Figure 4.1**). It is the world's largest known retrogressive thaw slump. The slump formed during recent decades and was ~1.8 km long and >0.8 km wide in 2019 (Kunitsky et al., 2013). More than 60 m deep, it exposes a discontinuous sequence of Middle Pleistocene to

Holocene permafrost deposits ranging in age from >MIS 6 (Ashastina et al., 2017) and even >MIS 17 to 16 (Murton et al., 2021) to MIS 1.

The climate in the region of the Yana Highlands is continental subarctic (defined by Köppen et al., 2011) and characterised by low precipitation and the largest seasonal temperature range globally. The mean air temperature between 1988 and 2017 was -40°C in winter (December to February) and 13.7°C in summer (June to August). The mean annual precipitation was 203 mm. Both air temperature and precipitation have increased at Batagay since the mid-20th century.

The Batagay megaslump is located within the northern taiga vegetation zone of the upper Yana floral district (Isaev et al., 2010). The modern vegetation is dominated by larch (*Larix gmelinii*) taiga. Siberian dwarf pine (*Pinus pumila*) is also common around the study site. The local flora includes shrubs such as *Alnus*, *Betula*, and *Salix* and dwarf shrubs such as *Ledum*, *Vaccinium*, *Arctous* and *Empetrum*. The ground is moist and covered with a thick layer of lichens and mosses. Only a few grass species and herbs such as *Polygonum*, *Corydalis*, and *Pedicularis* occur (Ashastina, 2018). The local fauna is typical for Yakutian taiga comprising small mammals such as arctic hare (*Lepus arcticus*) and Siberian chipmunk (*Eutamias sibiricus*), which are prey for birds such as the northern goshawk (*Accipiter gentilis*) and the golden eagle (*Aquila chrysaetos*). Other predators include brown bear (*Ursus arctos*), lynx (*Lynx lynx*), and smaller species such as mink and wolverine (Mustelidae family). Wolves (*Canis lupus*) predate bigger prey such as reindeer (*Rangifer tarandus*) and elk (*Alces alces*). Insect-eating birds come to the taiga to breed, whereas seedeaters or omnivorous birds such as finches (Fringillidae), sparrows (Passeridae), and crows (Corvidae) are present all year round.

4.3 Material & Methods

4.3.1 Fieldwork and subsampling

During fieldwork in late July and early August 2017, we studied several cryostratigraphic units of the Batagay megaslump and took permafrost sediment and ground-ice samples for cryostratigraphic, sedimentological, geochemical, stable-isotope, chronological and palaeoecological analyses. Before sampling, the exposure was cleaned and described. Ten samples were taken for *seda*DNA analysis using a battery-driven hand-drill or hammer and

axe cleaned with bleach between each sample. Samples were placed in two whirl pack bags, stored for several hours in a plastic container placed directly on the permafrost table, covered by a thick moss layer, and then moved to a freezer. Samples were kept frozen during transportation to the DNA laboratory in Potsdam.

Five samples were selected for this study based on the cryostratigraphic and chronological data (**Table 4.1**, **Supplementary Figure 3.1**). Sub-sampling took place in a climate chamber, in a different building than that housing the molecular genetics laboratories. The temperature was set to -10°C to prevent thawing of the core material. The sub-sampling procedure was carried out following the protocol described by Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schwamborn et al., 2017 using the same equipment, treatments and facilities. Ten subsamples were collected: two per sample for both *seda*DNA and palynological analyses and stored at -20°C, and +4°C, respectively.

4.3.2 Chronology

The ancient permafrost exposed at the Batagay megaslump is the oldest known in northern Eurasia (Murton et al., 2021). The pilot chronology of the exposed permafrost sediments is based on four dating techniques: radiocarbon dating of organic material, optically stimulated luminescence (OSL) dating of quartz grains, post-infrared infrared (pIRIR) luminescence dating of K-feldspar grains, and 36Cl dating of ice wedges. Estimated ages for the sampled cryostratigraphic units are provided in **Table 4.1** (for details see Murton et al., 2021). The lower ice complex, hosting the sample B17-D3, developed at least 650 kyr ago, likely at some time within MIS 17 or 16. The lower sand unit, discordantly overlying the lower ice complex, was formed at some time between MIS 16 and MIS 6 but is likely nearer in age to the latter (Ashastina et al., 2017). Sample B17-D5 was taken from the lowermost part of this unit and may be partially reworked, and thus it may represent a transition, or a remnant of older interglacial deposits rather than the environmental conditions associated with deposition of the main body of the lower sand. The upper ice complex overlies a wood-rich layer attributed to the Last Interglaciation (Ashastina et al., 2017). It formed mainly during MIS 3 and includes sample B17-D10. The two uppermost samples (B17-D7 and B17-D6) from the upper sand unit date to 22.4 and 27.3 kyr (MIS 2) (**Supplementary Figure 3.1**). Sample names within the manuscript consist of the sample ID and its associated MIS. In this way, we provide age context to the reader even if we constrain the age information later on.

Table 4.1 | Stratigraphic units of the permafrost sequence exposed in the Batagay megaslump.

Unit and approximate thickness (m)	Description	Interpretation	Preliminary age (Marine Isotope Stage, MIS) Dating method	*sedaDNA* sample number, depth below ground surface and estimated age of each sample
7. Near-surface sand (1–3)	Ice-cemented sand	Colluvium	MIS 1 (Holocene) Radiocarbon dating	
6. Upper sand (20–30)	Ice-cemented fine sand with narrow syngenetic ice wedges and composite wedges. Unit thickens downslope	Aeolian sand sheet with reworking by slopewash	MIS 3–2 Radiocarbon dating OSL dating	B17–D–6 (2.15 m) - ~27kyr BP B17–D–7 (2.3 m) - ~23kyr BP
5. Upper ice complex (20–25)	Yedoma containing large syngenetic ice wedges	Growth of large syngenetic ice wedges	MIS 4–3 Radiocarbon dating OSL dating	B17–D–10 (25.8 m) - ~50kyr BP
4. Woody debris (≤3)	Discontinuous lenses of woody debris cut down into lower sand unit.	Forest bed above erosional surface (disconformity)	MIS 5e (Last Interglacial) Palaeoecological interpretation	
3. Lower sand (≤20)	Ice-cemented fine sand with narrow syngenetic ice wedges and composite wedges	Aeolian sand sheet with forest bed near top	MIS 6 or older IRSL dating	B17–D–5 (~47.6 m) - ~200kyr BP
2. Lower ice complex (3–7)	Contains ice wedges, angular clasts of slate, in situ tree stumps and woody debris. Thaw unconformity along top.	Growth of large syngenetic ice wedges; forest bed near top; disconformity along top of unit	MIS 17–16 pIRIR luminescence dating ^{36}Cl dating	B17–D–3 (~49.5 m) - ~650kyr BP
1. Diamicton (≥0.5)	Contains abundant clasts of slate and overlies slate bedrock	Colluvium, locally derived from bedrock	MIS 17–16 or older (at least 650 ka)	

IRSL: infrared stimulated luminescence, pIRIR luminescence: post-infrared infrared luminescence
Modified from Murton et al. (2021).

4.3.3 Pollen analysis

For each sample, 2.85-4.58 g of wet sediment was taken for pollen preparation. A standard preparation method including KOH, HCl, HF, and acetolysis was used (Faegri et al., 1989). Palynomorphs were identified using a light microscope (Zeiss Axioskop 2) under 400-600x magnification. Pollen atlases (Beug, 2004; Kupriyanova & Alyoshina, 1972, 1978) were used for identification of pollen and spores. Non-pollen palynomorphs (NPP) were determined according to van Geel et al., (1983) and van Geel & Aptroot (2006). Freshwater algae were determined using Jankovská & Komárek (2000) and Komárek & Jankovská (2003). At least 300 pollen grains and spores were counted in each sample except for B17-D3_MIS17-16, where only 178 palynomorphs were counted. The complete pollen record is available in PANGAEA.

4.3.4 Isolation of sedimentary ancient DNA

DNA isolation, polymerase chain reaction (PCR) setup and library preparations were performed in the palaeogenetic laboratory of the Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research in Potsdam, Germany. This lab is dedicated to ancient DNA isolation and PCR setup (see **supplementary material**). Precautions to reduce contamination were taken following the recommendations of Champlot et al. (2010).

All samples, 4.74-10.65 g were prepared for DNA isolation within one extraction procedure including one extraction blank. Total DNA was isolated using the DNeasy PowerMax Soil kit (Qiagen, Hilden, Germany; see supplementary material for detailed procedure). The extracts with a concentration lower than 3 ng µL-1 were concentrated with Genejet PCR purification kit (Thermo Fisher, Germany) and samples were diluted to a final concentration of 3 ng µL-1.

4.3.5 Metabarcoding approach

The PCR reactions were performed with modified trnL g and h primers (Taberlet et al., 2007; see supplementary material) using the following cycle: initial denaturation at 94°C for 5 min, followed by 50 cycles of 94°C for 30 s, 50°C for 30 s, 68°C for 30 s and a final extension at 72°C for 10 min. To monitor for potential contamination, one no template control (NTC) was included in each PCR. For each extraction, three PCR replicates with different tag combinations were performed. The PCR products were purified using the

MinElute PCR Purification Kit (Qiagen, Germany), following the manufacturer's recommendations (see supplementary material). All replicates were pooled in equimolar concentrations. The amplified extraction blank and PCR NTCs were included in the sequencing run, using a standardised volume of 5 µL. Fasteris SA sequencing service (Switzerland) performed the paired-end sequencing on one-tenth of a full lane of the Illumina HiSeq 2500 platform (2×125 bp).

4.3.6 Shotgun approach

The DNA libraries were prepared following the single-stranded DNA library preparation protocol of Gansauge et al. (2017) except that the ligation of the second adapter (CL53/CL73) took place in a rotating incubator (Schulte et al., 2021; detailed procedure in supplementary material). The five libraries were pooled in equimolar ratios to a final pool of 10 nM with the blanks accounting for a molarity of 20% compared to the samples. The sequencing of the pool was performed twice using a modified forward sequencing primer CL72 as described in Gansauge & Meyer (2013) by Fasteris SA sequencing service (Switzerland). After the first sequencing on one lane of an Illumina HiSeq 2500 platform (2 x 125 bp), we decided to increase the sequencing depth—especially for the deepest samples—and sequenced the pool a second time on an Illumina Novaseq SP platform (2 x 100 bp).

4.3.7 Bioinformatic processing

4.3.7.1 Metabarcoding

Filtering, sorting, and taxonomic assignments of the metabarcoding sequences were performed with the OBITools package (Boyer et al., 2015). With the *illuminapairedend* function, we merged forward and reverse reads, and demultiplexing and sample sorting were performed with the *ngsfilter* function. With the *obigrep* command, sequences shorter than 10bp and less than 10 read count were excluded. Duplicated sequences were merged with *obiuniq* and cleaning of potential PCR or sequencing errors was performed with the *obiclean* function. Two reference databases were used for taxonomic assignments. The first one, ArctBorBryo, is based on the quality-checked and curated Arctic and Boreal vascular plant and bryophyte reference libraries (Soininen et al., 2015; Sønstebø et al., 2010; Willerslev et al., 2014). The second is based on the EMBL Nucleotide Database standard sequence release 138 (Kanz et al., 2005; http://www.ebi.acuk/embl/).

Final taxonomic assignments were determined by selecting the best identity given by embl or the ArctBorBryo database; if both assignments showed the same sequence identity, the taxonomic name was selected from ArctBorBryo database because of its specificity to arctic and boreal vegetation. Only sequences assigned with a best identity of 100% and present in at least two replicates of a sample or two different samples were used for this study. The sequences assigned to a taxonomic level higher than family were removed from the dataset. Sequences were also removed if more than 0.2% of their total read counts were present in the extraction blanks and PCR NTCs (**Supplementary Table 3.1**). Sequence data of the PCR replicates were merged to represent samples and sequences assigned to the same taxon were merged.

Before proceeding to the data analysis, a rarefaction analysis was performed based on the minimum number of sequence counts (n=84,348, for sample B17-D3_MIS17-16) to normalise the total counts of each sample (https://github.com/StefanKruse/R_Rarefaction). The complete metabarcoding record, before and after rarefaction is available in **Supplementary Table 3.1**.

4.3.7.2 Shotgun

Demultiplexed FASTQ files obtained from the two sequencing runs were quality checked, trimmed, and paired end merged using fastp (Chen et al., 2018). Statistics of both sequencing runs are reported in the **Supplementary Table 3.2**. We then merged both sequencing run files and proceeded to taxonomic classification using kraken2 (v. 2.1.2, Wood et al., 2019) against two different databases.

The first database used was the non-redundant nucleotide database (nt) from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz; downloaded in May 2020), and the NCBI taxonomy (retrieved via the kraken2-build command). It was used with a confidence threshold of 0.05, firstly, to differentiate Eukaryota from Prokaryota and Viruses and, secondly, for taxonomic assignment of Prokaryota (Bacteria and Archaea) and Viruses.

The sequences previously unassigned and assigned to Eukaryota against the nt database were extracted and taxonomically reassigned using kraken2 with default confidence parameters against a second, customised, database of curated full chloroplast and mitochondrion genomes from the NCBI Reference Sequence Database (Refseq, https://www.ncbi.nlm.nih.gov/refseq/, 10,010 sequences downloaded in December 2020).

Using the nt database to taxonomically assign our reads to Prokaryota and Viruses and the curated Refseq database for Eukaryota, we created different subsets for working taxonomic groups. We included bacterial families, (34,521,704 reads assigned to a family), Archaea (317,626 reads), and Viruses (42,112 reads). For the Eukaryota, we investigated families from Viridiplantae (435,733 reads), Fungi (348,619 reads), and selected terrestrial classes representing Metazoa with at least 0.5% of total metazoan reads (6,887 total reads): Insecta (1,493 reads), Mammalia (294 reads), and Aves (29 reads). Only families that represented at least 0.5% of the total reads assigned to each subset were kept. The only exception was for the Viridiplantae, where we kept all families that represented at least 0.2% of total reads assigned, the rationale being to capture as many as possible of the families detected via the other methods. An overview of the read counts and relative proportion of the selected families is summarised in **Table 4.2**.

If the abundance of reads detected in the blanks for an assigned family exceeded 1% of its total reads, this family was removed from the dataset. Furthermore, Primate reads were removed from the dataset as probable human contaminants. Finally, as shotgun sequencing almost inevitably yields "unlikely" taxa, only the families likely to be present in the study area were kept; therefore, we excluded metazoan aquatic classes and aquatic families of Mammalia. Detailed information about reads present in the blanks and the families excluded are reported in **Supplementary Table 3.2**.

Resampling to the sample with the minimum read counts was performed with the rarefy function of the "vegan" R package (v. 2.5-7 Oksanen et al. 2020) using R v. 3.6.1 (R Core Team, 2019) for all the subsets except the metazoan ones. Because few reads were assigned to the selected metazoan classes, differences in families detected per sample are reported instead of relative abundance, as was done in other phyla. Detailed information about lower taxonomic-level taxa detected from selected families is reported in **Supplementary Table 3.2**.

4.3.7.3 Damage pattern analysis

When performing data analysis on a shotgun sequenced dataset, we were particularly careful that the investigated signal comes from the ancient DNA of past organisms and not from modern contamination. Higher nucleotide mis-incorporation (cytosine deamination or C to T substitution) rates or DNA damage patterns can be observed on the 5'-overhang

Table 4.2 | Summary of reads assigned with kraken2 to main taxonomic groups.

	Total samples	B17-D3 MIS17-16	B17-D5_MIS16-6	B17-D10_MIS3	B17-D6_MIS2	B17-D7_MIS2	Total blanks	extraction blank	library blank
Total reads	771139326	166860854	156036581	138057439	155656957	154527495	1265153	404162	860991
custom database assigned	1033627	38384	373315	247415	175676	198837	1238	407	831
Viridiplantae	415593 (40.21%)	840 (2.19%)	4466 (1.2%)	156242 (63.15%)	121110 (68.94%)	132935 (66.86%)	765 (61.79%)	271	494
Fungi	317836 (30.75%)	17259 (44.96%)	281344 (75.36%)	10827 (4.38%)	3155 (1.80%)	5251 (2.64%)	56 (4.52%)	7	49
Mammalia	243 (0.02%)	6 (0.02%)	55 (0.01%)	70 (0.03%)	35 (0.02%)	77 (0.04%)	0 (0%)	0 (0%)	0 (0%)
Aves	17 (0%)	1 (0%)	2 (0%)	1 (0%)	8 (0%)	5 (0%)	0 (0%)	0 (0%)	0 (0%)
Insecta	1122 (0.11%)	73 (0.19%)	385 (0.1%)	374 (0.15%)	153 (0.09%)	137 (0.07%)	0 (0%)	0 (0%)	0 (0%)
NCBI nt database assigned	94940445	14963330	21842470	19540932	18773725	19819988	147770	62056	85714
Bacteria	28014036 (29.51%)	5114286 (29.51%)	6183906 (28.31%)	7416218 (37.95%)	4560976 (24.29%)	4738650 (23.91%)	39999 (27.07%)	11238 (18.11%)	28761 (33.55%)
Archaea	303556 (0.32%)	35333 (0.32%)	22004 (0.1%)	60196 (0.31%)	130211 (0.69%)	55812 (0.28%)	570 (0.39%)	213 (0.34%)	357 (0.42%)
Virus	37976 (0.04%)	4282 (0.04%)	5271 (0.02%)	6913 (0.04%)	13561 (0.07%)	7949 (0.04%)	74 (0.05%)	30 (0.05%)	44 (0.05%

and statistically estimated. Here we used the mapDamage 2.0 tool using a Bayesian framework to investigate the presence of such damage by comparing our sequenced reads to several reference genomes from each of our investigated taxonomic groups: Viridiplantae, Bacteria, Archaea, Virus, Metazoa and Fungi (Jónsson et al. 2013). We used entire genomes as reference sequences for the Bacteria, Archaea and Viruses; chloroplast entire genomes for the Viridiplantae and mitochondrion entire genomes for the Metazoa and Fungi. Detailed methodological information and examples of damage pattern plots are reported in **Supplementary Figure 3.2**.

4.3.7.4 Composition analysis

From the rarefied shotgun datasets, each studied taxonomic group is plotted as positive mean centred. This transformation compares the distance of each sample's read counts to the average read counts and sets the minimum value to zero, adding this value to the other samples to remove noise and facilitate the inspection of biodiversity compositional patterns and trends. Using these intermediate plots, a final selection of target families was undertaken, and below we describe only the most represented families likely to appear in the study area (**Supplementary Figure 3.3**).

Bubble charts were created to visualise relative abundance extracted from shotgun data of selected families recovered within larger taxonomic groups including Viruses, Bacteria, Archaea, Fungi, and Viridiplantae as well as read counts of selected families of Mammalia, Insecta, and Aves. Along with the plant shotgun data, we present relative abundance resulting from the pollen and metabarcoding analyses. We also compared the relative abundance of fungal spores identified in the palynological record with fungal shotgun data.

4.4 RESULTS

4.4.1 General results of the three approaches: pollen, metabarcoding and shotgun sequencing

In total, 1,520 pollen and non-pollen palynomorphs were counted in the five permafrost samples. *seda*DNA analyses yielded 650,985 reads assigned with a 100% match to plant families with metabarcoding and 559,922 reads retrieved by the metagenomic approach (415,593 reads assigned to the selected families). The oldest sample (B17-D3_MIS17-16) had the lowest read count (84,348 reads for the metabarcoding and 840 reads for shotgun

dataset) and the lowest pollen concentration (460 grains per gram). Fungi were detected, both among the non-pollen palynomorphs and in the shotgun data. In total, 317,836 reads were assigned to selected fungus families with the shotgun method, 88.5% of which (281,344 reads) came from the B17-D5_MIS16-6 sample. B17-D7_MIS2 had the lowest fungi read count with 3,155. For the three selected terrestrial classes of Metazoa, in total 1,122 reads were assigned to selected insect (Insecta) families, 243 reads to mammal (Mammalia) families, and 17 to bird (Aves) families. Finally, 28,355,568 reads were assigned to selected families from microbial communities and viruses; mostly Bacteria (28,014,036), whereas reads of Archaea (303,556 reads) and Viruses (37,976 reads) were 2-3 orders of magnitude lower.

The data indicated that two samples (B17-D3_MIS17-16 and B17-D5_MIS16-6) represent interglacial ecosystems and the three other samples (B17-D10_MIS3, B17-D7_MIS2, and B17-D6_MIS2) represent glacial ecosystems.

Finally, the damage pattern investigation confirmed the ancient origin of our shotgun *seda*DNA signal. Indeed, we identified damage patterns for all investigated subsets except Metazoa as not enough reads were aligning to the reference genomes and no damage pattern investigation was possible. A detailed description of the tested signal is available in the **Supplementary Figure 3.2**.

4.4.2 Plants (Viridiplantae)

All the palynological, *seda*DNA metabarcoding, and shotgun proxies similarly showed higher relative proportions of trees and woodland taxa in the interglacial samples (B17-D3_MIS17-16 and B17-D5_MIS16-6) compared to the glacial samples (B17-D10_MIS3, B17-D7_MIS2 and B17-D6_MIS2; **Figure 4.2A**).

The interglacial pollen spectra were characterised by the presence of Pinaceae (*Larix*, *Picea*, *Pinus*) in the pollen record (23.2% in B17-D3_MIS17-16, 18.2% in B17-D5_MIS16-6). With metabarcoding, *Cedrus* (Pinaceae) reached 6% in sample B17-D3_MIS17-16. Other woody taxa such as Betulaceae (*Alnus*, *Betula*) in the pollen record and sequences assigned to Menispermaceae in the shotgun data were also abundant in both interglacial spectra. The three methods also detected numerous herbs including Poaceae and other families typical of steppe communities, such as Asteraceae (*Artemisia*). Rosaceae

was also detected with the metabarcoding approach and Solanaceae with the shotgun approach in B17-D3_MIS17-16 and B17-D5_MIS7.

The three methods used revealed that the glacial spectra had a high proportion of non-woody taxa, in particular, a high relative abundance of Asteraceae and Poaceae. B17-D10_MIS3 was particularly rich in Plantaginaceae (*Plantago*) and Boraginaceae (*Eritrichium*) detected both using metabarcoding (13.4% of Plantaginaceae, 6.7% of Boraginaceae) and shotgun (13.5% of Plantaginaceae, 2.5% of Boraginaceae). Both B17-D7_MIS2 and B17-D6_MIS2 samples were characterised by a high abundance of Cyperaceae, Caryophyllaceae and Fabaceae (including *Astragalus* and *Oxytropis*).

4.4.3 Fungi

Both palynological and *seda*DNA shotgun proxies provided information on fungi. Among the non-pollen palynomorphs *Glomus* (Glomeraceae) spores were numerous in all samples (**Figure 4.2A**).

B17-D5_MIS16-6 had the highest number of reads assigned to fungi, mainly to the saprophytic and commonly psychrophilic Pseudeurotiaceae family (*Pseudogymnoascus* genus, 72.5% of all fungal reads for this sample).

The palynomorph assemblage contained spores of coprophilous fungi from Sordariaceae (*Sordaria*, 10.6%) in B17-D10_MIS3, where both Glomerellaceae (12%) and Aspergillaceae (11.1%) were abundant in the shotgun dataset. Furthermore, high relative abundance of lichens (Parmeliaceae), saprophylic Dermateaceae, and the plant-pathogen Pleosporaceae were detected in the three glacial spectra by the *seda*DNA shotgun approach.

4.4.4 Mammals (Mammalia)

Overall, this class had few reads assigned to family level, especially for the deepest and oldest sample (B17-D3_MIS17-16): six reads assigned to selected families. Selected families were detected even at higher confidence thresholds (--confidence 0.5 with kraken2), to taxa such as *Rangifer* (Cervidae), *Bison* (Bovidae), *Equus* (Equidae), *Prionodon* (Viverridae), and even *Mammuthus* (Elephantidae) but almost no exotic terrestrial taxa were found (**Supplementary Table 3.2**). In addition, among the few metazoans selected families detected, three (out of 23) were representative of the extinct

Figure 4.2 | Bubble chart illustrating compositional changes between the five studied samples for each family in relation to their respective groups. A. Relative vegetation changes per sample and method (P - Pollen, M - Metabarcoding, S - Shotgun). **B**. Relative metazoan changes per sample. A selection of families from three metazoan classes are represented here. **C**. Relative microorganism changes per sample.Pleistocene megafauna: Elephanthidae (mammoths), Rhinocerotidae (woolly rhinoceros) and Megalonychidae (giant ground sloths).

Overall, the B17-D10_MIS3, B17-D7_MIS2, B17-D6_MIS2 and B17-D3_MIS17-16 samples provided evidence mostly of larger herbivores and predators, whereas the B17-D5_MIS16-6 provided evidence mostly of smaller and more forest-adapted mammals in general (**Figure 4.2B**).

In the oldest sample (B17-D3_MIS17-16), very few reads were assigned to Mammalia, but we detected the presence of larger mammals such as the giant ground sloth family (Megalonychidae) and bovids (Bovidae). Two families of bats from Pteropodidae and Vespertilionidae also occurred in the oldest sample.

In B17-D5_MIS16-6, reads of smaller forest carnivores from Viverridae (45 reads) were quite abundant given to the small number of reads assigned overall (55 reads assigned to all families). Other families detected in this sample included the squirrel family (Sciuridae), the wild boar family (Suidae), and three families of bats (Pteropodidae, Vespertilionidae, and Molossidae). Among the larger mammals, only the mammoth family (Elephantidae) was detected.

B17-D10_MIS3's mammalian community was represented mainly by larger species such as mammoths (Elephantidae, 25 reads assigned), horses (Equidae, 20 reads assigned), and ruminants (Bovidae and Cervidae, 6 and 5 reads, respectively). Camelids (Camelidae, 3 reads assigned) were detected only in this sample. Predators from the Canidae were also detected in this sample. In addition to this community of large mammals, families of smaller mammals were detected such as rodents (Spalacidae, Octodontidae, Echimyidae) and insectivores in the hedgehog family (Erinaceidae).

Evidence of larger mammals, including horses (Equidae) and cervids (Cervidae), occurred again in the two samples dated MIS2. Predators such the Canidae and even the Felidae were detected. Smaller mammals occur in B17-D6_MIS2 but not in B17-D7_MIS2 including squirrels (Sciuridae), insectivores (Erinaceidae), and rodents (Cricetidae and Octodontidae), as well as the wild boar family (Suidae), bat families (Phyllostomidae and Pteropodidae), and the giant ground sloth family (Megalonychidae). In B17-D7_MIS2, however, there was even greater megafaunal representation: mammoths (Elephantidae), bovids (Bovidae) and the woolly rhinoceros' family (Rhinocerotideae). Some smaller mammals occurred in B17-D7_MIS2, including rodents (Echimyidae), moles (Spalacidae), and viverrids (Viverridae).

4.4.5 Birds (Aves)

Only very few reads were assigned to this class and no representative of the selected families was present in all five samples (**Figure 4.2B**).

In the two lowermost samples, woodland families were detected, with reads assigned to tits (Paridae), Asian barbets (Megalaimidae) and warblers (Sylviidae) detected in B17-D3_MIS17-16, and to the arboreal cuckoo family (Neomorphidae) detected only in B17-D5_MIS16-6. The ground-dwelling bird family of pheasants (Phasianidae) was also detected in B17-D5_MIS16-6.

In B17-D10_MIS3 the woodland tit family of Paridae was detected. We also detected sandpipers (Scolopacidae) which, like Paridae, need an availability of invertebrates as a food source in addition to water; a need which is shared by herons (Ardeidae) that were also detected. Ground-dwelling rails (Rallidae) were detected too.

In B17-D7_MIS2 and B17-D6_MIS2, the avian community was similar, characterised by ground-dwelling birds from Phasianidae. Birds of prey such as falcons (Falconidae) in B17-D7_MIS2 and eagles (Accipitridae) in B17-D6_MIS2 were also detected. In addition, some bird families that need water habitat such as sandpipers (Scolopacidae) or cranes (Gruidae) were detected in B17-D6_MIS2. Lastly, families that typically need proximity to woodlands were detected in both samples, such as Paridae in B17-D7_MIS2 or doves and pigeons (Columbidae) in B17-D6_MIS2.

4.4.6 Insects (Insecta)

Reads were assigned to Insecta in every sample even at higher thresholds (--confidence 0.5). For the three most recent samples (B17-D7_MIS2, B17-D6_MIS2, and B17-D10_MIS3), between 137 and 374 reads were assigned to selected Insecta families, while only 73 were assigned to the oldest sample, B17-D3_MIS17-16, and 385 to B17-D5_MIS16-6. Culicidae, the mosquito family, and Tettidoniidae, bush crickets, were the only two families (out of the 31 selected) to be detected in all samples. In terms of insect communities, a clear difference was observed between the three most recent and the two most ancient ones (**Figure 4.2B**).

In the two oldest samples (B17-D3_MIS17-16 and B17-D5_MIS16-6), more parasitoids (Trichogrammatidae) and mosquitoes (Culicidae) were detected than in the three other samples. In B17-D5_MIS16-6, we also observed a high abundance of skippers (Hesperiidae), butterflies (Nymphalidae) and moths (Yponomeutidae). Several families that were only present in this sample indicate the local presence of dead wood, for example, flat bark beetles (Cucujidae) and termites (Rhinotermitidae), while the abundance of

mayflies (Siphlonuridae) suggests a wetter environment. This sample is also the only one in which we detected skin beetles (Dermestidae).

In the three uppermost samples (B17-D7_MIS2, B17-D6_MIS2 and B17-D10_MIS3), several beetle families were detected, including lady beetles (Coccinellidae) and rove beetles (Straphylinidae) (**Supplementary Table 3.2**). Several families indicative of herb and shrub habitats were present, such as grasshoppers (Acrididae), grasshopper parasites (Meloidae), bush crickets (Tettigoniidae), leaf hoppers (Cicadellidae), and weevils (Curculionidae). Moths and butterflies were represented by three families: Nymphalidae, Lycaenidae, and Papilionidae. Some scavengers (Sarcophagidae) occurred in the three uppermost samples, alongside ants (Formicidae) and termites (Termitidae). Wasps were only detected in B17-D10_MIS3 but were abundant (251 reads) in this sample.

4.4.7 Prokaryotes (Bacteria, Archaea) and Viruses

We identified a core community consisting of Actinobacteria (especially the soil-associated families Nocardioidaceae, Streptomycetaceae, Microbacteriaceae and Mycobacteriaceae) and Haloarchaea (**Figure 4.2C**). Beside this shared community, we detected several distinctive patterns along the depositional sequence that likely reflect changes to the microbial community in response to environmental conditions.

In the most ancient sample (B17-D3_MIS17-16), Actinobacteria (Micrococacceae; 8.5% of all selected families of bacteria), particularly the genera *Arthrobacter* and *Pseudoarthrobacter*, (**Supplementary Table 3.2**) and Sphingomonadaceae (4.5%), had a higher relative abundance when compared to younger samples. Halophilic archaea were also well represented in the oldest samples, especially by Natrialbaceae (23.7% of all selected archaea families), Haloarculaceae (18.3%), Haloferacaceae (8.5%), Halorubraceae (13.7%) and Halobacteriaceae (9.9%), which shared a similar age-related distribution.

In B17-D5_MIS16-6, all major changes in community composition reflected an increase of taxa related to the degradation and cycling of organic matter in soil. These taxa included the biopolymer-degrading bacteria *Clostridium* (Clostridiaceae; 3.5%), *Nocardioides* and *Pseudonocardia* (Nocardioidaceae; 24.2%) and the fermenting *Propionibacterium* (Propionibacteriaceae; 2.2%)

In B17-D10_MIS3, Flavobacteriaceae were particularly abundant (47.5% of the entire selected bacteria families) together with Oxalobacteraceae (2.6%). Overall, well-known cold-adapted bacteria, such as *Polaribacter* and *Chryseobacterium* (Flavobacteriaceae), *Cryobacterium* (Microbacteriaceae; 5.5%) and *Polaromonas* (Comamonadaceae; 2.1%) were found in variety in high numbers. Other minor taxa typically associated with algae (e.g. *Tenacibaculum*, *Formosa*, *Cellulophaga*, *Lacinutrix*, Flavobacteriaceae family) also showed a distinct increase.

The two most recent samples (B17-D6_MIS2, B17-D7_MIS2) had a particularly high degree of similarity in microbial community composition and only marginal differences with regards to bacteria assigned to Cellulomonadaceae (4.5% in B17-D6_MIS2) and Conexibacteriaceae (4.4% in B17-D7_MIS2 and 3.5% in B17-D6_MIS2). Instead, ammonia-oxidizing archaea (AOA) were highly dominant, in particular Nitrososphaeraceae (70.3% in B17-D6_MIS2), with the genus *Candidatus Nitrosocosmicus*, and Nitrosopumilaceae (10.8% in B17-D6_MIS2), including the genera *Nitrosopumilus*, *Nitrososphaera* and *Candidatus Nitrosotalea*.

Sequences affiliated with methanogenic archaea represented 5-13% of the archaeal community. In line with their known distribution and abundance in arctic soil ecosystems, Methanosarcinaceae and Methanobacteriaceae were the most abundant families. The two families followed an inverse distribution trend, with Methanobacteriaceae being dominant in the youngest samples (B17-D10_MIS3, B17-D6_MIS2, B17-D7_MIS2) and Methanosarcinaceae in the oldest ones (B17-D3_MIS17-16, B17-D5_MIS16-6).

Viral communities throughout the permafrost sequence were dominated by bacteriophages of the Siphoviridae, Mimiviridae and Myoviridae families, which together accounted for 56-76% of the entire viral community. Overall, there was a high representation of giant viruses (e.g., *Mimivirus*, *Pandoravirus*, *Klosneuvirus* and *Tupanvirus*) that generally infect protozoa (**Supplementary Table 3.2**). In some samples, virus presence correlated well with that of plants or animals acting as either host or vector, most notably, *Mastrevirus* (Geminiviridae, 15.4% in B17-D7_MIS2), which infects monocotyledon plants via grasshopper transmission, was characteristic in B17-D7_MIS2.

4.5 Discussion

The Batagay megaslump data date back discontinuously to MIS 17-16 (~650 kyr) or older. Hence, this study presents one of the few records from northern Eurasia that reach back to the Middle Pleistocene (Andreev et al., 2011; Ashastina et al., 2018; Tarasov et al., 2005, Wetterich et al., 2019). It is also one of the first to successfully investigate *sedaDNA* as old as MIS 17 (Parducci, 2019) and identified ancient DNA damage patterns for several representatives of investigated phyla supporting the signal's authenticity (**Supplementary Figure 3.2**). Furthermore, we provide unique snapshots of Pleistocene glacial and interglacial biota that include all biological kingdoms. Previous shotgun sequencing investigations have been limited to fewer kingdoms (e.g. Murchie et al., 2021; Pedersen, 2016). Our results indicate distinct ecosystem-wide differences between forested (interglacial) and non-forested (glacial) intervals (**Figure 4.3**).

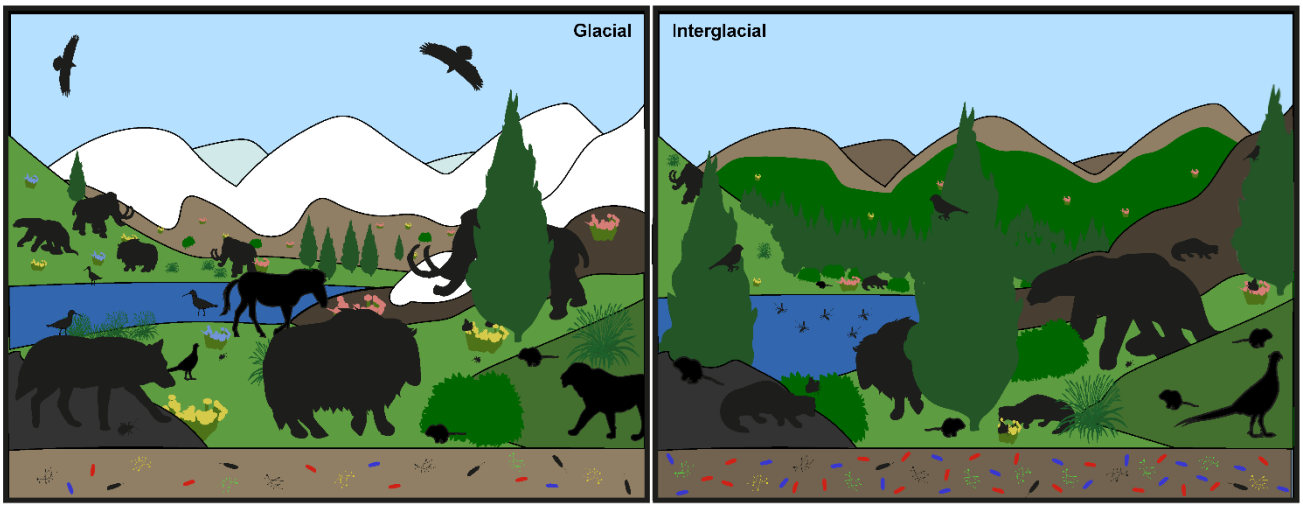


Figure 4.3 | Schematic comparison of Pleistocene glacial and interglacial landscapes and their respective biodiversity. Using the silhouettes highlighted in **Figure 4.2**, we made a schematic comparison ecosystems reconstructed from our datasets. Left: glacial, right: interglacial.

Our results for the samples B17-D3 and B17-D5 clearly point to interglacial ecosystems. Hence, our palaeoecological information contributes to constraining the ages of the respective samples. In the case of B17-D3, pilot dating results suggest an age of MIS 17-16 (Murton et al., 2021) (**Table 4.1**). The reconstructed interglacial ecosystem rules out the glacial MIS 16 and hence, we attribute this sample to MIS 17. Similarly, we can exclude a glacial environment for sample B17-D5. Based on the existing dating results of the lower sand unit suggesting a MIS 7 to MIS 6 as the age of this unit (Ashastina et al., 2017) we attribute this sample tentatively to MIS 7 even though an earlier interglaciation such as MIS 9 and up to MIS 15 cannot be excluded yet. The transition between the lower ice complex (MIS 17 sample) and the lower sand unit (MIS 7 sample) is marked by an erosional surface.

Such erosional features can be caused during warm stages (i.e., interglaciations) by intense permafrost degradation with widespread thermokarst (e.g., Kienast et al., 2011; Reyes et al., 2010). This suggests that other interglacial and glacial deposits between MIS 17 and MIS 7 may not have been preserved. Further dating is needed to constrain the chronology of the sediments exposed in the Batagay megaslump.

4.5.1 Interglacial communities

The two lowermost samples, attributed to MIS 17 and MIS 7, reflect forest vegetation with *Larix*. Understory shrubs (*Salix*, *Alnus*, *Betula*) and typical temperate families represented by herbs and low shrubs (e.g., Rosaceae, Orchidaceae, Poaceae, Cyperaceae, Ericaceae) were abundant. This assemblage aligns with data from the other kingdoms. These two samples should not be considered representative of the entire MIS 17 or MIS 7 stages, but they do provide snapshots of those two past interglacials, which featured an organic-rich soil and were probably intermittently flooded and/or characterized by wetlands.

Our record reveals that the MIS 17 forests were probably rather open, as indicated by the presence of grassland taxa such as Asteraceae and Poaceae, plus bovids and giant sloths. The presence of *Cedrus* DNA in this sample is surprising as *Cedrus* distribution today is only as far North as the Tibetan Plateau and the Himalayas, and it disappeared from southwestern China after the Late Pliocene, ~2.5 Myr ago (Su et al., 2013). Similar unlikely taxa were found in other records (e.g. *Juglans* in Jørgensen et al., 2012), but these finds could well be explained by reworking of older sediments or a misleading taxonomic identification, for instance due to incomplete databases or the use of loose classification thresholds. Alternatively, *Cedrus* can be a rare contaminant already detected in other studies using the same marker (Voldstad et al., 2020). The occurrences of Menispermaceae and Solanaceae are similarly unexpected, though in the case of these two families, representative species are reported from north-central Asia, closer to the study site than the modern *Cedrus* distribution.

During MIS 7, fewer megafauna and more smaller mammals were detected. The data are in accordance with an inference of dense forest vegetation, as this favours smaller taxa while megafauna co-occur with open vegetation (Bakker et al., 2016; Smith et al. 2016). This confirms the hypothesis that large herbivores play a role in maintaining forests as a mosaic landscape composed of closed canopy woodland and open areas (Vera, 2000). Our

data indicate that forest-dwelling taxa including viverrids (Viverridae) were common during MIS 7, as well as smaller taxa from the squirrel family (Sciuridae). The high proportion of insects in this sample includes forest-specific families such as flat bark beetles (Cucujidae) and termites (Rhinotermitidae) supporting the inference of forest conditions. Furthermore, the presence of parasitic insects such as Trichogrammatidae suggest a high host density, that is, other insects or mammals (Abrams, 2000).

In the MIS 7 sample, the microbial community is rather distinctive, owing to the marked increase of taxa involved in the degradation and cycling of organic matter in soil, which include the fungus *Pseudogymnoascus sp.* and bacteria such as Nocardioidaceae and *Clostridia*. Members of *Pseudogymnoascus* are associated with decaying roots or plants, especially in boreal forests (Sigleret al., 2000). They can form mycorrhizal associations with arctic shrubs, which aligns with the increased presence of shrubs and woody taxa in MIS 7 (Semenova et al., 2015). Similarly, Nocardiodaceae of the genera *Nocardioides*, *Corinebacterium* and *Pseudonocardia* are aerobic saprophytic bacteria, considered to be consumers of organic material. Many, especially *Nocardioides*, secrete extracellular enzymes, and degrade and metabolise a wide range of natural organic compounds, thus being important players in the turnover of organic matter in ecosystems, including cold ones (Babalola et al., 2009; Perez-Mon et al., 2020). They have a copiotrophic lifestyle, implying that they are fast-growing and metabolically versatile, and are consequently fast responders to increased nutrient availability (Perez-Mon et al., 2020). *Clostridium* is a dominant bacterium of the permafrost subsurface environment and related DNA sequences have been detected in ancient samples up to ~1 Myr (Burkert et al., 2019; Liang et al., 2019). It can hydrolyse biopolymers such as cellulose, lignocellulose and lignin, in deep sediments under anoxic conditions and produce alcohols, organic acids (e.g. lactate, acetate, butyrate) and hydrogen as metabolic by-products (Ueno et al., 2016). These compounds - especially lactate - can then be fermented by *Propionibacterium spp*. to propionate, acetate, and carbon dioxide, which can be assimilated by methanogenic archaea (Methanosarcinaceae) with a final synthesis of methane (Drake et al., 2009). Thus, these data suggest high organic matter turnover going on during MIS 7, both at the surface and subsurface. In permafrost and cold ecosystems generally, this is directly related to an increase in dissolved organic carbon mobilised through soil horizons, which is a consequence of relatively high ambient temperature (Bracho et al., 2016; Selvam et al.,

2017). This may well mean that climate and environmental conditions were mild during this period.

4.5.2 Glacial communities

Our data clearly indicate that herb communities (mostly Asteraceae, Poaceae, Cyperaceae) dominated the vegetation in the Batagay area under glacial conditions during MIS 3 and MIS 2. This is in line with the palaeo-vegetation reconstruction based on plant macrofossils and pollen previously reported for this site (Ashastina et al., 2018). It also confirms the open steppe-tundra landscape that has been described by other *seda*DNA and pollen records from north-eastern Siberia (Andreev et al., 2011; Binney et al., 2017; Lozhkin et al., 2007; Willerslev et al., 2014; Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schirrmeister, et al., 2017; Zimmermann, Raschke, Epp, Stoof-Leichsenring, Schwamborn, et al., 2017; Wetterich et al., 2011; Zimov et al., 2012).

This open steppe-tundra landscape (Dale Guthrie, 2001; Zimov et al., 2012) spanned much of unglaciated northern Eurasia and Beringia during the Late Pleistocene (Johnson, 2009). Coprophilous fungi found in the pollen assemblage, such as *Sordaria*, indirectly point to the presence of herbivorous megafauna during the Late Pleistocene glacial interval. The presence of the latter is confirmed by the *seda*DNA mammal record, which shows traces of megaherbivores such as mammoths (Elephantidae), woolly rhinoceroses (Rhinoceritadeae), camels (Camelidae), horses (*Equus*), bovids (Bovidae), and deer and reindeer (Cervidae). This availability of prey supported large predators, for example taxa within Canidae and Felidae. Our shotgun data support the macrofossil remains (*Equus sp*., *Mammuthus primigenius*, *Bison priscus*, *Panthera leo spelaea*) found in the Yana River valley region (Novgorodov et al., 2013) and agree well with previous ancient DNA data (Seersholm et al., 2020).

The data can represent trophic interactions (Krebs et al., 2014). For example, smaller mammals such as hedgehogs (Erinaceidae) and rodents are detected in the MIS 3 sample. They feed on vegetative material, seeds and arthropods and are in turn consumed by predators. The increased proportion of Plantaginaceae (*Plantago*) during MIS 3 is correlated with the presence of beetles and predators such as wasps (Vespidae), grasshoppers (Acrididae) and parasites (Meloidae), small insectivores - mammals (Erinaceidae) and birds (Scolopacidae) - and their predators in the Canidae. This increased

proportion of *Plantago*, and the related trophic links suggest that the MIS 3 interstadial was milder than the MIS 2 stadial but the high abundance of psychrophilic bacteria, a typical inhabitant of glacial environments (Boetius et al., 2015), provides further evidence for a glacial climate. The presence of sand pipers (Scolopacidae), herons (Ardeidae), mayflies (Siphlonuridae) and mosquitoes (Culicidae) indicate the availability of water in the area, which points to at least an intermittent but substantial moisture supply. This is in accordance with an increased presence of *Flavobacterium spp*. and other minor taxa (*Tenacibaculum*, *Formosa*, *Cellulophaga*, *Lacinutrix*), which are considered as "first responders" to phytoplankton blooms due to their activity in degrading algal cells and detritus (Williams et al., 2013; Zeder et al., 2009). This may relate to spring bloom of snow algae or algae in permafrost polygonal or thaw ponds during the cold MIS 3.

The reconstructed glacial mammoth steppe ecosystems lack modern analogues. It is hypothesised that the extinct megaherbivores played a critical role in Pleistocene ecosystems. They directly affected the abiotic component of the system by functioning as "nutrient pumps" (Gross, 2016), allowing distribution of nutrients in the ecosystem. The community structure of the mammoth steppes can be compared to that of the modern savannah (Zimov et al., 2012). Indeed, the megafauna taxa, by their trampling and sapling consumption, may, during milder climatic intervals, have suppressed colonization by trees, keeping the open vegetation largely intact (Guthrie, 2001; Bakker et al., 2016), with the exception, perhaps of woodland patches in habitats with a favourable mesoclimate (Edwards et al 2014, Chytrý et al 2019). In our record, low abundance of tree taxa and relatively more abundant megafauna such as Equidae, Elephantidae, Rhinocerotidae, and Megalonychidae and bigger predators such as Canidae and Felidae were detected during Pleistocene glacial time, in contrast to the Pleistocene interglacial time, when megafauna were less abundant and forest taxa more abundant. Modern vegetation in the Altai region (Chytrý et al., 2019) is the most similar analogue yet identified for steppe-tundra of the Pleistocene, being characterized by forest patches within grassland and shrub areas. Such a mosaic may explain the presence of sequences related to small- and medium-sized birds such as rails (Rallidae) or tits (Paridae), typical inhabitants of wooded areas, and of cellulose-degrading bacteria (Cellulomonadaceae and Conexibacteriaceae families) during MIS 2 and MIS 3.

4.5.3 Potential and limitations of the *seda*DNA shotgun approach applied to ancient permafrost sediments

Using the *seda*DNA shotgun approach, we were able to identify taxa from multiple kingdoms and infer trophic structure and biotic interactions at the ecosystem level. Information about vegetation was in some cases confirmed by the more established methods of metabarcoding and pollen. However, we faced some challenges, in particular the handling of taxonomic assignment to optimise the ratio between false negative and false positives (Guillera-Arroita et al., 2017). Highly curated databases are often used to limit the biases of false positives during taxonomic assignments via filtering out unlikely taxa from the database (Pedersen, 2016). In our opinion, such methods are too sensitive to false negatives, particularly with respect to groups where we have only limited knowledge on past biota, and which are poorly presented in the databases. Therefore, we used non-curated databases and worked with likely families to allow the reconstruction of communities to be as broad as possible, limiting the false negatives but reducing the impact of false positives with a taxonomic assignment at higher taxonomic levels. Furthermore, we opted for a more stringent taxonomic assignment using kraken2 v.2.1.2, which requires 2 hit groups (set of overlapping k-mers with the same minimizer) in comparison to the previous version kraken2 v.2.0.7-beta requiring only 1 hit group. While this leads to fewer reads assigned especially for very short DNA fragments (i.e., as those originating from old sediment up to MIS17), it minimizes the rate of false positive thus increasing the overall confidence that the taxonomic assignments correspond to real signals. This would also provide a stronger support to the reads assigned for example to Metazoa, which are known to be underrepresented in *seda*DNA metagenomics when compared to other taxonomic groups as reported also by others (Perdersen et al. 2016).

To overcome the technical challenges of the metagenomics analyses of *seda*DNA and further reduce taxonomic assignment biases such as false positive rates, one could improve the quality of metagenomic reference databases that are incomplete, contaminated, biased toward modern taxa, domain specific and have a skewed taxonomic representation (Breitwieser et al., 2019; Steinegger & Salzberg, 2020). The use of other tools is also advised, such as alignment-based classification with Malt or de-novo assembly of contigs and investigation of metagenome-assembled genomes (MAGs) (Herbig et al., 2016; Hübler et al., 2019; Sangwan et al., 2016) but they are still computationally very intense and will need further optimisation for use with large databases.

4.6 CONCLUSIONS

We reconstructed the palaeoecology associated with permafrost sediments from the Batagay Megaslump in eastern Siberia using *seda*DNA shotgun sequencing, metabarcoding and pollen analyses. We inferred ecosystem-wide information on glacial and interglacial biotic assemblages including Bacteria, Archaea, Virus, Fungi, Viridiplantae, Mammalia, Aves and Insecta. All methods indicate that samples at Batagay attributed to interglacial ecosystems in MIS 17 and MIS 7 were characterised by forested vegetation. The local landscape from the MIS 17 snapshot likely featured more open, herb-dominated areas, probably maintained by large herbivores. Around the study site, during MIS 3, diverse megafauna coexisted with smaller mammals on cold grassland. The MIS 2 ecosystems existed under harsher conditions, as suggested by the presence of cold-adapted taxa, while typical Pleistocene megafauna still inhabited the area.

DATA AVAILABILITY STATEMENT

Pollen datasets are uploaded to PANGAEA under doi: 10.1594/PANGAEA.945257

DNA sequences after shotgun sequencing are uploaded to the European Nucleotide Archive under Project Number PRJEB43506 (Courtin et al., 2022).

DNA metabarcoding sequencing and DNA datasets after metabarcoding are uploaded to DRYAD under doi: 10.5061/dryad.xpnvx0kj1.

FUNDING

This project has received funding from the Priority Project "International Continental Drilling Program" of the German Research Foundation and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772852) and the Initiative and Networking fund of the Helmholtz Association. TO acknowledges funding by the Leverhulme Trust Research Project Grant RPG-2020-334.

REFERENCES

Abrams, P. A. (2000). The Evolution of Predator-Prey Interactions: Theory and Evidence. *Annual Review of Ecology and Systematics*, 31(1), 79–105. doi: 10.1146/annurev.ecolsys.31.1.79

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., … Wohlfarth, B. (2018). Archaeal community changes in Lateglacial lake sediments: Evidence from ancient DNA. *Quaternary Science Reviews*, 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

Andreev, A. A., Morozova, E., Fedorov, G., Schirrmeister, L., Bobrov, A. A., Kienast, F., & Schwamborn, G. (2012). Vegetation history of central Chukotka deduced from permafrost paleoenvironmental records of the El'gygytgyn Impact Crater. *Climate of the Past*, 8(4), 1287–1300. doi: 10.5194/cp-8-1287-2012

Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siegert, C., … Hubberten, H. W. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. *Quaternary Science Reviews*, 30(17–18), 2182–2199. doi: 10.1016/j.quascirev.2010.12.026

Andreev, A.A., Tarasov, P.E., Wennrich, V., Raschke, E., Herzschuh, U., Nowaczyk, N.R., … Melles, M. (2014). Late Pliocene and Early Pleistocene environments of the north-eastern Russian Arctic inferred from the Lake El'gygytgyn pollen record. *Climate of the Past*, 10, 1-23. doi: 10.5194/cp-10-1-2014

Arnold, L. J., Robert, R. G., Macphee, R. D. E., Haile, J. S., Brock, F., Möller, P., … Willerslev, E. (2011). Paper II - Dirt, dates and DNA: OSL and radiocarbon chronologies of perennially frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. *Boreas*, 40(3), 417–445. doi: 10.1111/j.1502-3885.2010.00181.x

Ashastina, K. (2018). *Palaeo-environments at the Batagay site in West Beringia during the late Quaternary* (Doctoral dissertation, Friedrich-Schiller-Universität Jena).

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. doi: 10.1016/J.QUASCIREV.2018.07.032

Ashastina, K., Schirrmeister, L., Fuchs, M., & Kienast, F. (2017). Palaeoclimate characteristics in interior Siberia of MIS 6-2: First insights from the Batagay permafrost mega-thaw slump in the Yana Highlands. *Climate of the Past*, 13(7), 795–818. doi: 10.5194/cp-13-795-2017

Babalola, O. O., Kirby, B. M., Le Roes-Hill, M., Cook, A. E., Cary, S. C., Burton, S. G., & Cowan, D. A. (2009). Phylogenetic analysis of actinobacterial populations associated with Antarctic Dry Valley mineral soils. *Environmental Microbiology*, 11(3), 566–576. doi: 10.1111/j.1462-2920.2008.01809.x

Bakker, E. S., Gill, J. L., Johnson, C. N., Vera, F. W. M., Sandom, C. J., Asner, G. P., & Svenning, J.-C. (2016). Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy of Sciences of the United States of America*, 113(4), 847–855. doi: 10.1073/pnas.1502545112

Bálint, M., Pfenninger, M., Grossart, H. P., Taberlet, P., Vellend, M., Leibold, M. A., … Bowler, D. (2018). Environmental DNA Time Series in Ecology. *Trends in Ecology and Evolution*, 33(12), 945–957. doi: 10.1016/j.tree.2018.09.003

Behrensmeyer, A. K. (1988). Vertebrate preservation in fluvial channels. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 63(1–3), 183–199. doi: 10.1016/0031-0182(88)90096-X

Bellemain, E., Davey, M. L., Kauserud, H., Epp, L. S., Boessenkool, S., Coissac, E., … Haile, J. (2013). Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. *Environmental Microbiology*, 15(7), 1176–1189. doi: 10.1111/1462-2920.12175

Beug, H.-J. (1961). *Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete: Mit 17 Abbildungen und 8 Tafeln*. G. Fischer.

Beug, H.-J. (2004). *Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete*, *Verlag Dr. Friedrich Pfeil*, München, 542 pp., 2004

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., … Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157, 80–97. doi: 10.1016/j.quascirev.2016.11.022

Birks, H. J. B., Lotter, A. F., Juggins, S., & Smol, J. P. (2001). *Tracking environmental change using lake sediments*. doi: 10.1007/978-94-007-2745-8

Boetius, A., Anesio, A. M., Deming, J. W., Mikucki, J. A., & Rapp, J. Z. (2015). Microbial ecology of the cryosphere: Sea ice and glacial habitats. *Nature Reviews Microbiology*, Vol. 13, pp. 677–690. Nature Publishing Group. doi: 10.1038/nrmicro3522

Botkin, D. B., Saxe, H., Araújo, M. B., Betts, R., Bradshaw, R. H. W., Cedhagen, T., … Stockwell, D. R. B. (2007). Forecasting the Effects of Global Warming on Biodiversity. *BioScience*, 57(3), 227–236. doi: 10.1641/B570306

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2015). OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, (16), 176–182. doi: 10.1111/1755-0998.12428

Bracho, R., Natali, S., Pegoraro, E., Crummer, K. G., Schädel, C., Celis, G., … Schuur, E. A. G. (2016). Temperature sensitivity of organic matter decomposition of permafrost-region soils during laboratory incubations. *Soil Biology and Biochemistry*, 97, 1–14. doi: 10.1016/j.soilbio.2016.02.008

Breitwieser, F. P., Lu, J. & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4), 1125–1136. doi: 10.1093/bib/bbx120

Brigham-Grette, J., Melles, M., Minyuk, P. S., Andreev, A. A., Tarasov, P. E., DeConto, R. M., … Herzschuh, U. (2013). Pliocene warmth, polar amplification, and stepped pleistocene cooling recorded in NE Arctic Russia. *Science*, 340, 1421–1427. doi: 10.1126/science.1233137

Burkert, A., Douglas, T. A., Waldrop, M. P., & Mackelprang, R. (2019). Changes in the active, dead, and dormant microbial community structure across a pleistocene permafrost chronosequence. *Applied and Environmental Microbiology*, 85(7). doi: 10.1128/AEM.02646-18

Champlot, S., Berthelot, C., Pruvost, M., Andrew Bennett, E., Grange, T., & Geigl, E. M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE*, 5(9). doi: 10.1371/journal.pone.0013042

Chen, S., Zhou, Y., Chen, Y., Gu, J., (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics* 34(17), i884–i890. doi: 10.1093/bioinformatics/bty560

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. doi: 10.1111/bor.12338

Clarke, C. L., Edwards, M. E., Gielly, L., Ehrich, D., Hughes, P. D. M., Morozova, H., … Alsos, I. G. (2019). Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Scientific reports*, 9, 19613. doi: 10.1038/s41598-019-55989-9

Courtin, J., Andreev, A. A., Raschke, E., Bala, S., Biskaborn, B. K., Liu, S., … Herzschuh, U. (2021). Vegetation changes in southeastern Siberia during the Late Pleistocene and the Holocene. *Frontiers in Ecology and Evolution* 9:625096. doi: 10.3389/fevo.2021.625096

Demske, D., Mohr, B., & Oberhänsli, H. (2002). Late Pliocene vegetation and climate of the Lake Baikal region, southern East Siberia, reconstructed from palynological data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 184(1–2), 107–129. doi: 10.1016/S0031-0182(02)00251-1

Drake, H. L., Horn, M. A. & Wüst, P. K. (2009). Intermediary ecosystem metabolism as a main driver of methanogenesis in acidic wetland soil. *Environmental Microbiology Reports*, 1: 307-318. doi: 10.1111/j.1758-2229.2009.00050.x

Edwards, M. E. (2020). The maturing relationship between Quaternary paleoecology and ancient sedimentary DNA. *Quaternary Research*, 96, 39–47. doi: 10.1017/qua.2020.52

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(0318), 152–163. doi: 10.1016/j.quascirev.2015.03.027

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

Faegri, K., Iversen, J., Kaland, P. E., & Krzywinski, K. (1989). *Textbook of pollen analysis*, 4th Edn. Hoboken, NJ: John Wiley & Sons Ltd.

Foote, M., & Raup, D. M. (1996). Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, 22(2), 121–140.

Gansauge, M. T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., … Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*, 45(10), 79. doi: 10.1093/nar/gkx033

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

Gardner, E. E., Walker, S. E., & Gardner, L. I. (2016). Palaeoclimate, environmental factors, and bird body size: A multivariable analysis of avian fossil preservation. *Earth-Science Reviews*, 162, 177–197. doi: 10.1016/j.earscirev.2016.07.001

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, … 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, 14676. doi: 10.1038/s41598-019-50339-1

Gross, M. (2016). Megafauna moves nutrients uphill. *Current Biology*, 26(1), R1–R5. doi: 10.1016/j.cub.2015.12.028

Guillera-Arroita, G., Lahoz-Monfort, J. J., Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, 8(9), 1081–1091. doi: 10.1111/2041-210X.12743

Guthrie, R. (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, 549-574. doi: 10.1016/S0277-3791(00)00099-8

Haile, J., Froese, D. G., MacPhee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. V., … Willerslev, E. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), 22352–22357. doi: 10.1073/pnas.0912510106

Hausmann, B., Knorr, K. H., Schreck, K., Tringe, S. G., Glavina Del Rio, T., Loy, A., & Pester, M. (2016). Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *ISME Journa*l, 10(10), 2365–2375. doi: 10.1038/ismej.2016.42

Herbig, A., Maixner, F., Bos, K., Zink, A., Krause, J., & Huson, D. (2016). MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *BioRxiv*, 050559. doi: 10.1101/050559

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907–913. doi: 10.1038/35016000

HilleRisLambers, J., Harsch, M. A., Ettinger, A. K., Ford, K. R., & Theobald, E. J. (2013). How will biotic interactions influence climate change-induced range shifts? *Annals of the New York Academy of Sciences*, 1297(1), 112-125. doi: 10.1111/nyas.12182

Hoelzel, A. R. (2010). Looking backwards to look forwards: Conservation genetics in a changing world. *Conservation Genetics*, 11, 655–660. doi: 10.1007/s10592-010-0045-4

Hübler, R., Key, F. M., Warinner, C., Bos, K. I., Krause, J., & Herbig, A. (2019). HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biology*, 20(1), 1-13. doi: 10.1186/s13059-019-1903-0

Isaev, A. P., Protopopov, A. V., Protopopova, V. V., Egorova, A. A., Timofeyev, P. A., Nikolaev, A. N., … Sleptsova, N. P. (2010). *Vegetation of Yakutia: Elements of Ecology and Plant Sociology*. Springer, Dordrecht. doi: 10.1007/978-90-481-3774-9_3

Jackson, S. T., & Blois, J. L. (2015). Community ecology in a changing environment: Perspectives from the Quaternary. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 4915–4921. doi: 10.1073/pnas.1403664111

Jankovská, V., Komárek, J. (2000). Indicative value of Pediastrum and other coccal green algae in palaeoecology. *Folia Geobot*, 35, 59–82. doi: 10.1007/BF02803087

Johnson, C. N. (2009). Ecological consequences of late quaternary extinctions of megafauna. *Proceedings of the Royal Society B: Biological Sciences*, 276(1667), 2509–2519. doi: 10.1098/rspb.2008.1921

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, 29(13), 1682–1684. doi: 10.1093/bioinformatics/btt193

Jørgensen, T., Haile, J., Möller, P., 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. doi: 10.1111/j.1365-294X.2011.05287.x

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., ... & Apweiler, R. (2005). The EMBL nucleotide sequence database. *Nucleic acids research*, 33(suppl_1), D29-D33.

Kienast, F., Siegert, C., Dereviagin, A., & Mai, D. H. (2001). Climatic implications of Late Quaternary plant macrofossil assemblages from the Taymyr Peninsula, Siberia. *Global and Planetary Change*, 31(1–4), 265–281. doi: 10.1016/S0921-8181(01)00124-2

Kienast, F., Wetterich, S., Kuzmina, S., Schirrmeister, L., Andreev, A. A., Tarasov, P., … Kunitsky, V. V. (2011). Paleontological records indicate the occurrence of open woodlands in a dry inland climate at the present-day Arctic coast in western Beringia during the Last Interglacial. *Quaternary Science Reviews*, 30(17–18), 2134–2159. doi: 10.1016/j.quascirev.2010.11.024

Komárek, J., Jankovská, V. (2001) *Review ofthe Green Algal Genus Pediastrum; Implication for Pollenanalytical Research*; Bibliotheca Phycologica: Berlin, Germany; Schweizerbart Science Publishers: Stuttgart, Germany, 108.

Ḱoppen, W., Volken, E., & Brönnimann, S. (2011). The thermal zones of the Earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world. *Meteorologische Zeitschrift*, 20(3), 351–360. doi: 10.1127/0941-2948/2011/105

Krebs, C. J., Boonstra, R., Boutin, S., Sinclair, A. R. E., Smith, J. N. M., Gilbert, B. S., … Turkington, R. (2014). Trophic dynamics of the boreal forests of the Kluane Region. *Arctic*, 67, 71–81. doi: 10.14430/arctic4350

Kunitsky, V. V., Syromyatnikov, I. I., Schirrmeister, L., Skachov, Y. B., Grosse, G., Wetterich, S., & Grigoriev, M. N. (2013). Ice-rich permafrost and thermal denudation in the Batagay area (Yana Upland, East Siberia), *Kriosfera Zemli (Earth Cryosphere)*, 17, 56–58.

Kupriyanova, L. A. & Alyoshina L. A. (1972). *Pollen and spores of plants from the flora of European part of USSR*. I. Academy of Sciences of the USSR, the Komarov Botanical Institute.

Kupriyanova, L. A. and Alyoshina L. A. (1978). *Pollen and spores of plants from the flora of European part of USSR*, Academy of Sciences of the USSR, the Komarov Botanical Institute.

Liang, R., Lau, M., Vishnivetskay, T., Lloyd, K. G., Wang, W., Wiggins, J., … Onstott, T. C. (2019). Predominance of anaerobic, spore-forming bacteria in metabolically active microbial communities from ancient Siberian permafrost. *Applied and Environmental Microbiology*, 85(15), 560–579. doi: 10.1128/AEM.00560-19

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., & Herzschuh, U. (2020). Holocene vegetation and plant diversity changes in the north-eastern Siberian treeline region from pollen and sedimentary ancient DNA. *Frontiers in Ecology and Evolution*, 8, 304. doi: 10.3389/fevo.2020.560243

Lozhkin, A. V., & Anderson, P. M. (2013). Vegetation responses to interglacial warming in the Arctic: Examples from Lake El'gygytgyn, Far East Russian Arctic. *Climate of the Past*, 9(3), 1211–1219. doi: 10.5194/cp-9-1211-2013

Lozhkin, Anatoly V., Anderson, P. M., Matrosova, T. V., & Minyuk, P. S. (2007). The pollen record from El'gygytgyn Lake: Implications for vegetation and climate histories of northern Chukotka since the late middle Pleistocene. *Journal of Paleolimnology*, 37(1), 135–153. doi: 10.1007/s10933-006-9018-5

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. doi: 10.1128/AEM.71.2.1012-1017.2005

Melles, M., Brigham-Grette, J., Minyuk, P. S., Nowaczyk, N. R., Wennrich, V., DeConto, R. M., … Wagner, B. (2012). 2.8 Million years of Arctic climate change from lake El'gygytgyn, NE Russia. *Science*, 315, 315–320. doi: 10.1016/j.actbio.2010.11.011

Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C., & White, J. W. C. (2010). Arctic amplification: can the past constrain the future? *Quaternary Science Reviews*, 29(15–16), 1779–1790. doi: 10.1016/J.QUASCIREV.2010.02.008

Mix, A. C., & Ruddiman, W. F. (1984). Oxygen-isotope analyses and Pleistocene ice volumes. *Quaternary Research*, 21(1), 1–20. doi: 10.1016/0033-5894(84)90085-1

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

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., … Poinar, H. (2021). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quaternary Research*, 99, 305–328. doi: 10.1017/qua.2020.59

Murton, J. B., Edwards, M. E., Lozhkin, A. V., Anderson, P. M., Savvinov, G. N., Bakulina, N., … Zanina, O. G. (2017). Preliminary paleoenvironmental analysis of permafrost deposits at Batagaika megaslump, Yana Uplands, northeast Siberia. *Quaternary Research*, 87(2), 314–330. doi: 10.1017/qua.2016.15

Murton, J. B., Opel, T., Toms, P., Blinov, A., Fuchs, M., Wood, J., … Wetterich, S. (2021). A multi-method pilot dating study of ancient permafrost, Batagay megaslump, east Siberia. *Quaternary Research*, 1-22. doi: 10.1017/qua.2021.27

Myers-Smith, I. H., Grabowski, M. M., Thomas, H. J. D., Angers-Blondin, S., Daskalova, G. N., Bjorkman, A. D., … Eckert, C. D. (2019). Eighteen years of ecological monitoring reveals multiple lines of evidence for tundra vegetation change. *Ecological Monographs*, 89(2), e01351. doi: 10.1002/ecm.1351

Novgorodov, G. P., Grigorev, S. E., Cheprasov, M. Y. (2013). Prospective location of the mammoth fauna in the river basin Yana. *International journal of applied and fundamental research*, 8, pp. 255- 259.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2020). *vegan: Community Ecology Package*. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Opel, T., Murton, J. B., Wetterich, S., Meyer, H., Ashastina, K., Günther, F., … Schirrmeister, L. (2019). Past climate and continentality inferred from ice wedges at Batagay megaslump in the Northern Hemisphere's most continental region, Yana Highlands, interior Yakutia. *Climate of the Past*, 15(4), 1443–1461. doi: 10.5194/cp-15-1443-2019

Parducci, L. (2019). Quaternary DNA: a multidisciplinary research field. *Quaternary*, 2(4), 37. doi: 10.3390/quat2040037

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355(6332), eaai9214. doi: 10.1126/science.aai9214

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., … Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, 537(7618), 45–49. doi: 10.1038/nature19085

Pedersen, M. W. (2016). *Paleo-Environmental Reconstruction Using Ancient DNA*. https://www.forskningsdatabasen.dk/en/catalog/2349594394

Perez-Mon, C., Frey, B., & Frossard, A. (2020). Functional and structural responses of arctic and alpine soil prokaryotic and fungal communities under freeze-thaw cycles of different frequencies. *Frontiers in Microbiology*, 11, 982. doi: 10.3389/fmicb.2020.00982

Piro, V. C., Dadi, T. H., Seiler, E., Reinert, K., & Renard, B. Y. (2020). Ganon: precise metagenomics classification against large and up-to-date sets of reference sequences. *Bioinformatics*, 36(Supplement_1), i12–i20. doi: 10.1093/bioinformatics/btaa458

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Reyes, A. V., Froese, D. G., & Jensen, B. J. L. (2010). Permafrost response to Last Interglacial warming: Field evidence from non-glaciated Yukon and Alaska. *Quaternary Science Reviews*, 29(23–24), 3256–3274. doi: 10.1016/j.quascirev.2010.07.013

Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(8). doi: 10.1186/s40168-016-0154-5

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2021). Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Molecular Ecology Resources*. doi: 10.1111/1755-0998.13311

Seersholm, F. V., Werndly, D. J., Grealy, A., Johnson, T., Keenan Early, E. M., Lundelius, E. L., … Bunce, M. (2020). Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nature Communications*, 11(1), 1–10. doi: 10.1038/s41467-020-16502-3

Selvam, B. P., Lapierre, J. F., Guillemette, F., Voigt, C., Lamprecht, R. E., Biasi, C., … Berggren, M. (2017). Degradation potentials of dissolved organic carbon (DOC) from thawed permafrost peat. *Scientific Reports*, 7(1), 1–9. doi: 10.1038/srep45811

Semenova, T. A., Morgado, L. N., Welker, J. M., Walker, M. D., Smets, E., & Geml, J. (2015). Long-term experimental warming alters community composition of ascomycetes in Alaskan moist and dry arctic tundra. *Molecular Ecology*, 24(2), 424–437. doi: 10.1111/mec.13045

Shackleton, N. (1967). Oxygen isotope analyses and Pleistocene temperatures re-assessed. *Nature*, 215(5096), 15–17. doi: 10.1038/215015a0

Sigler, L., Lumley, T. C., & Currah, R. S. (2000). New species and records of saprophytic ascomycetes (Myxotrichaceae) from decaying logs in the boreal forest. *Mycoscience*, 41(5), 495–502. doi: 10.1007/bf02461670

Smith, F. A., Doughty, C. E., Malhi, Y., Svenning, J. C., & Terborgh, J. (2016). Megafauna in the Earth system. *Ecography*, 39(2), 99–108. doi: 10.1111/ecog.02156

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), 1–18. doi: 10.1371/journal.pone.0115335

Sonstebo, 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. doi: 10.1111/j.1755-0998.2010.02855.x

Steinegger, M. & Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, 21(115). doi: 10.1186/s13059-020-02023-1

Su, T., Liu, Y. S. (C.), Jacques, F. M. B., Huang, Y. J., Xing, Y. W., & Zhou, Z. K. (2013). The intensification of the East Asian winter monsoon contributed to the disappearance of Cedrus (Pinaceae) in southwestern China. *Quaternary Research*, 80(2), 316–325. doi: 10.1016/j.yqres.2013.07.001

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), 1–8. doi: 10.1093/nar/gkl938

Tape, K. D., Sturm, M., & Racine, C. (2006). The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. *Global Change Biology*, 12(4), 686–702. doi: 10.1111/j.1365-2486.2006.01128.x

Tarasov, P., Granoszewski, A. W., Bezrukova, A. E., Brewer, S., Nita, A. M., Abzaeva, A. A., & Oberha¨nsli, A. H. O. (2005). Quantitative reconstruction of the Last Interglacial vegetation and climate based on the pollen record from Lake Baikal, Russia. *Climate Dynamics*, 25, 625–637. doi: 10.1007/s00382-005-0045-0

Tomescu, A. M. F., Bomfleur, B., Bippus, A. C., & Savoretti, A. (2018). Why are bryophytes so rare in the fossil record? A spotlight on taphonomy and fossil preservation. In *Transformative Paleobotany* (pp. 375–416). Elsevier. doi: 10.1016/b978-0-12-813012-4.00016-4

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12), 1351–1363. doi: 10.1111/j.1461-0248.2008.01250.x

Ueno, A., Shimizu, S., Tamamura, S., Okuyama, H., Naganuma, T., & Kaneko, K. (2016). Anaerobic decomposition of humic substances by Clostridium from the deep subsurface. *Scientific Reports*, 6(1), 1–9. doi: 10.1038/srep18990

van der Knaap, W. O. (1987). Long-distance transported pollen and spores on Spitsbergen and Jan Mayen. *Pollen et Spores*, 29(4) 449-453.

Van Geel, B., Aptroot, A. (2006). Fossil ascomycetes in Quaternary deposits. *Nova Hedwig*. 2006, 82, 313–329. doi: 10.1127/0029-5035/2006/0082-0313

Van Geel, B., Hallewas, D. P., & Pals, J. P. (1983). A late holocene deposit under the Westfriese Zeedijk near Enkhuizen (Prov. of Noord-Holland, The Netherlands): Palaeoecological and archaeological aspects, *Review of Palaeobotany and Palynology*, 38(3-4), 269-335. doi: 10.1016/0034-6667(83)90026-X

Vera, F. W. M. (2000). *Grazing Ecology and Forest History*. CABI, Wallingford.

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. doi: 10.1016/j.quascirev.2020.106207

Wagner, M., Smidt, H., Loy, A., & Zhou, J. (2007). Unravelling microbial communities with DNA-microarrays: challenges and future directions a decade of DNA microarray research in microbial ecology. *Microbial Ecology*, 53(3), 498–506. doi: 10.1007/s00248-006-9197-7

Walther, G., Post, E., Convey, P., Menzel, A., Parmesank, C., Beebee, T. J. C., … Bairlein, F. (2002). Ecological response to recent climate change. *Nature*, 416, 389–395. doi: 10.1038/416389a

Walther, G. R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2019–2024. doi: 10.1098/rstb.2010.0021

Wetterich, S., Rudaya, N., Kuznetsov, V., Maksimov, F., Opel, T., Meyer, H., … Schirrmeister, L. (2019). Ice Complex formation on Bol'shoy Lyakhovsky Island (New Siberian Archipelago, East Siberian Arctic) since about 200 ka. *Quaternary Research*, 92(2), 530-548. doi: 10.1017/qua.2019.6

Wetterich, S., Rudaya, N., Tumskoy, V., Andreev, A. A., Opel, T., Schirrmeister, L., & Meyer, H. (2011). Last Glacial Maximum records in permafrost of the East Siberian Arctic. *Quaternary Science Reviews*, 30(21-22), 3139-3151. doi: 10.1016/j.quascirev.2011.07.020

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. doi: 10.1038/nature12921

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

Willerslev, E., Hansen, A. J., Rønn, R., Brand, T. B., Barnes, I., Wiuf, C., … Cooper, A. (2004). Long-term persistence of bacterial DNA. *Current Biology*, 14(1), R9–R10.

Williams, T. J., Wilkins, D., Long, E., Evans, F., DeMaere, M. Z., Raftery, M. J., & Cavicchioli, R. (2013). The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environmental Microbiology*, 15(5), 1302–1317. doi: 10.1111/1462-2920.12017

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 1–13. doi: 10.1186/s13059-019-1891-0

Zeder, M., Peter, S., Shabarova, T., & Pernthaler, J. (2009). A small population of planktonic Flavobacteria with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake. *Environmental Microbiology*, 11(10), 2676–2686. doi: 10.1111/j.1462-2920.2009.01994.x

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*, 8(10), 1–28. doi: 10.3390/genes8100273

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., … Herzschuh, U. (2017). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences*, 14(3), 575–596. doi: 10.5194/bg-14-575-2017

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin, I. S. (2012). Mammoth steppe: A high-productivity phenomenon. *Quaternary Science Reviews*, 57, 26–45. doi: 10.1016/j.quascirev.2012.10.005

5 SYNTHESIS

The main goal for this thesis was to acquire evidence of compositional and diversity changes in high latitude ecosystems at the transition between Quaternary glacial and interglacial stages with the help of new methods based on *seda*DNA proxies. Here, I synthetize and discuss the findings of my work in this context.

5.1 ECOLOGICAL CHANGES BETWEEN GLACIAL AND INTERGLACIAL STAGES

5.1.1 Changes in the compositional structure

In this work, I investigated the Siberian and Beringian Late Pleistocene glacial stage in addition to the Holocene and the Marine Isotope Stage (MIS) 7 and MIS 17 interglacial stages' taxonomic assemblages and I identified distinct communities' composition between glacial and interglacial stages.

During the last glacial stage, Siberia and Beringia were dominated by an open landscape composed of a mixture of forbs and grasses with some shrubs and few trees (**Manuscript I**, **II**, **III** and **IV**). In this thesis, I highlighted the overall signal comparability of different methods used for vegetation composition reconstruction between the *seda*DNA metabarcoding, metagenomic and the pollen approaches (**Manuscript I** and **III**). I could confirm that the Pleistocene steppe-tundra was covering most of unglaciated Siberia and Beringia, expanding as far south as the 56th parallel north as inferred from the Bolshoe Toko lake site (**Manuscript I**; Bocherens, 2015). From each of the investigated study sites, I could identify a local and regional vegetation signal matching the previously described steppe-tundra biota with the *seda*DNA (**Manuscript I**, **II**, **III** and **IV**) and pollen proxies (**Manuscript I** and **IV**; Guthrie, 2001). Whereas the overall signal was consistent between methods and sites, some unique characteristics - method and site dependent - were noted, possibly suggesting a more complex mosaic landscape than previously described.

To highlight differences between the methods, I show that *seda*DNA approaches detected more forbs and less graminoids than the pollen approach (**Manuscript I** and **III**). This emphasises the role of forbs in the Pleistocene steppe-tundra that was previously thought to be dominated by grasses (Blinnikov et al., 2011; Guthrie, 2013; Zimov et al., 2012). It

has been discussed that forbs were actually dominant during the Last Glacial especially in areas with herbivorous megafauna, as it was also shown based on the analysis of their intestine and stomach content **(Manuscript I** and **IV**, Binney et al., 2017; Willerslev et al., 2014). The dominance of graminoids in the steppe-tundra was previously overestimated by pollen proxies because biases towards wind pollinated and high pollen productivity taxa (Parducci et al., 2017). On the other hand, some plant taxa are overrepresented in the *seda*DNA datasets. For example, Salicaceae dominate *seda*DNA records (**Manuscript I** and **IV**). One explanation is that Salicaceae (e.g., *Salix*) have small and leathery leaves that are usually well preserved in glacial sediments (Birks, 2014). Because of this, Salicaceae signal may result in a stronger PCR amplification (**Manuscript I**). The overrepresentation of Salicaceae has been observed in many others *seda*DNA metabarcoding studies. Besides, Salicaceae were present locally in dense population around lakes and rivers (Niemeyer et al., 2017). This second hypothesis is supported by lower abundance of Salicaceae when *seda*DNA was recovered from very continental permafrost sediments (**Manuscript III**, Zimmermann et al., 2017). This underlines the local origin of the *seda*DNA signal in comparison to the pollen signal (Capo et al., 2021; Parducci et al., 2017). Nevertheless, it is possible to compile *seda*DNA datasets from different records and locations to provide a regional and even subcontinental signal. From already analysed records, *seda*DNA datasets handled with similar laboratory approaches can be reanalysed using updated taxonomic databases or methodologies with high reproducibility (**Manuscript II**).

The Pleistocene steppe-tundra biodiversity was also dynamic with differences regionally and locally through time during the ~ 130,000 yrs covered by the late Pleistocene. For example, grasses were generally abundant during the Last Glacial Maximum (LGM) in the *seda*DNA datasets and less after the LGM (**Manuscript II**, Willerslev et al., 2014). This general trend is observed in northeast Siberia (**Manuscript IV**) but not in southeast Siberia, (**Manuscript I**) pointing to the local specificity and the difficulty to infer general conclusions from sparse sites with unique histories (Giguet-Covex et al., 2019; Jia et al., 2021; Lozhkin & Anderson, 2011; Wang et al., 2021). Sites can be a refugium of some taxa that will be found only in sparse locations (Edwards et al., 2014; Herzschuh et al., 2016; Herzschuh, 2020). It is therefore of high importance to use multiple sites to infer regional and subcontinental conclusions on past biodiversity (**Manuscript II**). The Pleistocene steppe-tundra system also changed through the Late Pleistocene. For example, based only on Batagay, I observed distinct features between the MIS 3 and MIS 2 steppe-tundra

(**Manuscript III**). The MIS 3 was likely milder and moister allowing plants from Plantaginaceae and Boraginaceae to thrive more than during the MIS 2. Evidence of milder and moister MIS 3 are confirmed by the presence of taxa from other kingdoms that need some availability of water such as sand pipers, herons, mayflies or mosquitoes. In addition, the microbial soil community investigated with the *seda*DNA approach allowed to finally confirm moister MIS 3 at Batagay with the higher abundance of taxa degrading algal cells and detritus in the sample in comparison to the MIS 2. From other works, it was already recognized that MIS 3 was milder and moister than the MIS 2 at least at different sites in northeast Siberia and Beringia (Lozhin & Anderson, 2011; Reuther et al., 2019; Zanina et al., 2011).

In this work, I was also able to provide snapshots of past interglacial biota likely spanning the period between MIS 17 and MIS 7 (**Manuscript III**; Murton et al., 2021). From the Batagay megaslump, I could infer that the system was, at least locally, forested with an undercover of shrubs. Locations providing access to such old records are extremely rare and so far none has been investigated with *seda*DNA analyses, making **Manuscript III** the first study to study such ancient environmental DNA originating from permafrost sediment. Therefore, comparison of the observed signal at the Batagay megaslump with other locations is challenging. The lake El'gygytgyn in northeast Siberia gives access to ~3.6 million years of continuous sediments (Layer, 2000). Nowadays, the vegetation around this site is more tundra-like, differing from the larch deciduous forest at Batagay (**Figure 1.1**, Isaev et al., 2010). Still, similar differences between MIS17 and MIS7 can be observed both at El'gygytgyn and at Batagay. At ~200 ka BP (MIS 7), the region around El'gygytgyn was covered by shrubs from steppes and tree taxa from cold deciduous forest when at MIS 17, fewer tree taxa were detected with the pollen proxy (Lozhkin et al., 2017; Zhao et al., 2019). Comparable results are observed from the Batagay record, suggesting that the forest cover was probably denser during MIS 7 than during the MIS 17 in Siberia. With the methods explored in the **Manuscript III**, I could confirm this not only with the direct signal of pollen and plant *seda*DNA but also from other taxonomic groups. For example, exclusively at MIS 7, I could identify traces of taxa specific to forested environments from diverse taxonomic groups such as Mammalia, Aves or Insecta. In addition, smaller mammals are abundant and fewer megafaunas are detected at MIS 7 in comparison to MIS 17, which is in line with the hypothesis that large herbivores can maintain forests as a mosaic landscape composed of closed canopy woodland and more open areas (Vera, 2000).

The main difference between the past investigated interglacial stages and the Holocene systems is that at the transition between the Pleistocene and the Holocene, the megafauna went extinct, while we still detected some traces of megafauna during previous interglacial stages (**Manuscript III**). At a sub-continental scale, the shift in vegetation composition between the Pleistocene and the Holocene started already during the Last Glacial Maximum around ~ 19–17 cal. ka BP with a decrease in abundance of herbaceous taxa and an increase in shrubs (**Manuscript II** and **IV**). Between ~17-11 cal. ka BP, I observed a period of transition and instability in the plant assemblages which led to the collapse, at the onset of the Holocene of the steppe-tundra. This unique biota disappeared and got replaced generally by open woodlands of shrubs and trees (**Manuscript II**) in Siberia and Beringia, especially by boreal forests in south-eastern Siberia (**Manuscript I**), dwarf-shrub tundra in north-eastern Siberia (**Manuscript IV**) but also temperate deciduous forests, temperate grasslands, and more diverse biota (Binney et al., 2017). There is no modern analogue of the Pleistocene steppe-tundra but some relics in the Altai Mountain range (Chytrý et al., 2019). The vegetation during the early Holocene was, generally, more densely forested by deciduous broadleaved taxa (e.g., *Betula*, *Populus*) under warmer and moister conditions (**Manuscript I** and **II**; Binney et al., 2009 and references therein). After ~5,000 cal. yrs BP, the forest vegetation became similar to the modern vegetation, with less *Populus* and *Betula*, but more deciduous *Larix* (**Manuscript I**, **II** and **IV**; Kaufman et al., 2004).

Overall, in this thesis, I could describe compositional changes between Quaternary glacial and interglacial stages, confirming the steppe-tundra floral composition and the spatial and temporal boundaries of the Pleistocene steppe-tundra collapse with the support of *seda*DNA in addition to pollen proxies. I also demonstrated the robustness of signal from the newest method and with this, I was able to investigate the entire biotic structure of past ecosystems with the shotgun sequencing method. This method allows to identify potential co-occurrences, cross kingdom and trophic interactions and can provide snapshots of entire ecosystems as illustrated in this work.

5.1.2 Loss of plant diversity

Additionally, I was able to identify a general loss of plant taxa richness and diversity coupled with potential plant extinction events between the Pleistocene and the Holocene in Siberia and Beringia at the study sites (**Manuscript I**, **II** and **IV**). Using *seda*DNA, the

investigation and estimations of past alpha and beta diversity is possible (Alsos et a., 2018; Boessenkool et al., 2014; Clarke et al., 2019; Liu et al., 2020, Rijal et al., 2021).

According to the results presented in this work, the Siberian and Beringian Last Glacial steppe-tundra was taxonomically rich and diverse until the Last Glacial Maximum. I observed a general decrease of floral alpha diversity (taxonomic richness, Shannon and Simpson diversity) through time between the last Glacial and the Holocene (**Manuscript I**, **II** and **IV**). These results generally differ from other works based on *seda*DNA extracted from Northern European records, which observed the opposite trend: a general increase in floristic diversity and richness towards the Holocene (Clarke et al., 2019; Rijal et al., 2021). In addition, they are in contrast to the rule of modern diversity distribution that follows a latitude – diversity gradient (LDG), according to which diversity generally decreases with latitude (Hillebrand, 2004; Willig et al., 2003). The direct comparison with modern distribution is biased as the Pleistocene steppe-tundra biota was unique and has no real modern analogue (Chytrý et al., 2019). The LDG can be explained by the species area effect which suggests that diversity and distribution area are correlated and when area decreases towards the poles, species diversity declines as well (Hillebrand, 2004; Fine, 2015). The Pleistocene steppe-tundra covered most of unglaciated Northern Hemisphere, providing a massive potential distribution area for species that could explain the high diversity of the Pleistocene steppe-tundra communities (Johnson, 2009; Guthrie, 2001). Climate stability is another factor that explains the geographic variation in species diversity (Fine, 2015). The Pleistocene Last Glacial was relatively climatically stable for ~ 120 ka after the last interglacial stage (Sandel et al., 2017). In comparison, the Holocene is relatively young with ~12 ka, which could also explain why the diversity during the last Glacial was higher than during the Holocene. Finally, modern grassy biomes such as savannah are comparable in term of biodiversity to tropical forests (Murphy et al., 2016). Open grassland, especially in the context of a mosaic landscape, can sustain a highly taxonomically diverse plant assemblage (**Manuscript I**).

The loss of diversity affected especially the most abundant Pleistocene steppe-tundra plant families (e.g., Asteraceae, Brassicaceae, Cyperaceae, **Manuscript I** and **IV**) and was not accompanied by increased species replacement rates, making plant extinction likely to occur (**Manuscript II**). Indeed, in parallel to this general loss of floral diversity, I could identify potential events in Siberia and Beringia with plant extinction rates higher than the

background extinction rate (**Manuscript II**). The taxa most likely to have undergone extinction are rare ones, specialists from abundant families that contribute less to the functionality of the system (**Manuscript II**). This can be explained by the close relatedness between species that composed the diverse Pleistocene steppe-tundra (**Manuscript IV**). Under increased stress, competition between closely related taxa that would share similar traits would lead to the exclusion of taxa sharing similar niches but less flexible to changes because rarer or specialists (Zettlemoyer et al., 2019).

In this work, I identified three extinction events of plant taxa at the transition between the Pleistocene and the Holocene in Siberia and Beringia at ~17, 11 and 5 ka BP (**Manuscript II**). The described extinction events started during the LGM and continued until the mid-Holocene and occurred in parallel to the megafauna extinction with a first wave during the LGM, a second wave between 15–10 ka BP, while other species survived until at least ~6,000 years ago (Clark, 2009, Murchie et al., 2021). This underlies the close interaction and feedback processes between the megafauna and the steppe-tundra plant species assemblage.

Therefore, while there was a shift of plant composition at the transition between the Pleistocene and the Holocene in Siberia and Beringia, it was coupled to a general loss of plant diversity and not the entire species pool rearranged in the tundra and the steppes as previously discussed (Chytrý et al., 2019; Nehring, 1890). Extinction is happening in parallel of at least both megafauna and plant species.

5.1.3 Potential drivers of change

To understand what could have led to the collapse of the Pleistocene steppe-tundra biota, several hypotheses have been made and are discussed in the literature. Some argue that climate change alone could have driven such drastic changes, others that human impact and megafaunal extinction are the main drivers as the steppe-tundra faced several glacial and interglacial stages during the Pleistocene without collapsing.

At the transition between glacial and interglacial stages, there was a change in the climate regime between one stable glacial to one stable interglacial and vice versa (Hewitt, 2000; Jackson & Blois, 2015). Those stable stages lasted thousands of years during the Pleistocene (Gornitz, 2021). Each of the transitions led to shifts in biodiversity composition (**Manuscript III**; Jackson & Blois, 2015). Such shifts in composition stress and challenge

the resilience of ecosystems (Hewitt, 2000; Jackson & Blois, 2015). It can lead to the collapse of stable systems with changes in the species pool and potentially extinction of species as observed in Siberia and Beringia at the transition between the Pleistocene and the Holocene (**Manuscript II**).

At the same time, between the Pleistocene and the Holocene, megafauna's species went extinct. The "keystone herbivore" hypothesis suggests that the extinction of Pleistocene megaherbivores triggered the steppe-tundra to disappear (Owen-Smith, 1987). The Late Pleistocene megafauna likely played a similar role as the modern grazing megafauna in the modern African savannahs as herds of large herbivores can maintain areas of open grassland (Bakker et al., 2016; Gill, 2014). In addition, the extinct megaherbivores directly affected the abiotic component of the system by functioning as nutrient pumps, allowing distribution of nutrients in the ecosystem (Gross, 2016) and their trampling and sapling consumption likely suppressed colonization by trees, keeping the open vegetation largely intact (Bakker et al., 2016; Guthrie, 2001).

Finally, the "overkill hypothesis" suggests that modern human were the likely driver of megafauna's extinctions worldwide (Martin & Wright, 1967). Late Quaternary extinctions of megafauna have been more severe in areas where modern humans were the first hominin to arrive, introducing a new, effective predator into regions with megafauna subject to human hunting (Sandom et al., 2014). Humans arrived in Siberia around 40 ka BP and in Beringia around 30 ka BP (Graf & Buvit, 2017) and it has been discussed that with slow population growth and low hunting activity, they could have been responsible for the extinction of megafauna in North America (Alroy, 2001).

When looking at ecosystem changes in general, it is unlikely that they are driven by single factors (e.g., volcanic event or meteorite impact; Chapin et al., 1998; Dutch, 1999; Petersen et al., 2016). It is more common that drastic and abrupt changes originate from multiple stress factors, initiating a feedback-loop (Norby & Luo, 2004). Since the industrial revolution, modern humans have increased stress compared to the past, being directly responsible of both indirect (e.g., climate warming) and direct effects (e.g., fragmentation of habitat) on the ecosystem. This leads to unprecedented changes in climate and structure of ecosystems. However, in the past, human impact was limited to hunting or deforestation. Direct human impact on ecosystems with reduced resilience could, however, make a previously stable system collapse. For example, one potential hypothesis for the collapse

of the Pleistocene steppe-tundra could be that with warming of the high latitudes, the steppe-tundra was naturally invaded with shrubs and trees and transformed into open forests. This reduced the distribution range of megafaunal population only to the remaining open grasslands and steppes like observed during previous interglacial stages (**Manuscript III**). With a reduced range and potentially smaller populations, megafauna was then more sensitive to hunting by human, thus inducing megafaunal extinction. Such a collapse of megafauna resulted in less active landscape maintenance, facilitation of tree invasion in the remaining open areas, denser forests and the start of the feedback-loop that led to the collapse of both megafauna population and the steppe-tundra system.

From palaeoreconstruction of biodiversity, we can support prediction on future ecosystem changes in high latitudes (Boessenkool et al., 2014; Orlando et al., 2021). Generally, it can provide temporal ecological data needed for robust prediction in temporal ecology. Besides, time series records allow to test for biodiversity models to improve future predictions (Bálint et al., 2018). For example, studies predicting dynamics at the ecotone between taiga and tundra in high latitudes use models that feed on data retrieved from palaeorecords (Kruse & Herzschuh, 2022; Kruse et al., 2016). Tundra greening in High Arctic has been identified during the Last Interglacial stages which can be compared to the greening observed in modern tundra (Crump et al., 2021). Moreover, inferences on past ecological interactions between fauna and flora are used to establish conservation projects of the tundra (Zimov et al., 2012).

In this work, I demonstrated how loss of biotic resilience in the past led to the collapse of a previously well-established system with shifts in community composition, loss of species diversity and triggered extinction cascades. Such drastic changes occured in high latitude ecosystems, yet they had repercussion worldwide. Lastly, similar cascade of events are observed in high latitude systems at the transition between past glacial and interglacial stages could happen elsewhere in the world. If ecosystems collapse, they will be replaced by new ones, providing new ecosystem services and therefore affecting human well-being. As soon as the threshold of resilience is passed, feedback-loops across every part of an ecosystem trigger drastic and relatively fast global changes that would require direct effort to counter.

5.2 HIGH POTENTIAL OF *SEDA*DNA FOR PAST BIODIVERSITY RECONSTRUCTION

In this work, I demonstrated the high value and potential of *seda*DNA as a proxy to reconstruct not only past community composition but also to unravel past diversity in high latitude ecosystems. In general, the use of multiple proxies in concert is advised to provide the strongest and most robust results from palaeorecords (Edwards, 2020). This general recommendation was followed in **Manuscript I** and **III** but there is an increasing amount of evidence that *seda*DNA alone can provide, in general, a similar signal to pollen and macrofossils, while giving access to potential of more information that both conventional proxies could not provide.

The *seda*DNA approach is capable to explore very old records from permafrost or marine sediments (**Manuscript III**; Kirkpatrick et al., 2016). From such records, the investigation of past alpha and beta diversity is now possible, while it was previously a limited estimation from pollen records (**Manuscript I**, **II** and **IV**; Clarke et al., 2019; Rijal et al., 2021). Still, this approach is constrained by biases such as taxonomic assignment or by the amplified region (e.g., trnL gh P6-loop) and the database used for taxonomic assignment (Rijal., et al., 2021). In addition, because the *seda*DNA can provide a higher taxonomic resolution than pollen, and its preservation is more consistent than the macrofossils, it becomes easier to investigate potential loss of species in the past (**Manuscript II;** Alsos et al., 2018; Parducci et al., 2017). Finally, the investigation of full ecosystems from one location at different timesteps is now possible with the *seda*DNA metagenomics approach (Pedersen et al., 2016). In this work, I showed that it can highlight co-occurrences of taxa across kingdom and potential trophic interactions. Previously, megafauna presence could only be hypothesized through the direct presence of fossils or the indirect pollen proxies with the presence of coprophilous fungi such as *Sordaria*, while now it is possible to directly identify them to genus and even species level as well as species from other taxonomic groups (**Manuscript III**). Depending on the research question, *seda*DNA alone will be able to take over the other proxies. It provides generally similar composition reconstruction but has higher potential to answer new palaeoecological questions such as species richness dynamics, metacommunity impacts, evolutionary model, cross-kingdom interactions (Bálint et al., 2018; Pedersen et al., 2016)

Despite its high potential, the *seda*DNA approach is still a relatively young method and has limitations that need to be overcome in the future. Understanding preservation conditions and taphonomy will be essential to further refine the origin of the *seda*DNA signal (Capo et al., 2021). Optimisation of *seda*DNA retrieval during sampling, subsampling and DNA extraction need to be adjusted both to increase DNA yields and to minimise contamination (Capo et al., 2021). With technical and methodological innovations, it becomes more and more possible for palaeoscientists to retrieve and investigate *seda*DNA to its full potential. For example, we can now estimate if the origin of the DNA signal is endogenous or exogenous (e.g., modern contaminant) with the estimation of DNA degradation by assessing the typical deamination and DNA fragmentation to authenticate ancient DNA (Ginolhac et al., 2011). New approaches to further improve detection of rare or under-represented taxa are developed such as capture sequencing (e.g., Schulte et al., 2021). In addition, de-novo assembly of contigs and investigation of metagenome-assembled genomes (MAGs; Sangwan et al., 2016) that can increase the robustness of the taxonomic assignment are being developed. In addition, work on curation databases which are currently incomplete, contaminated or biased towards modern or domain specific sequences, is needed (Breitwieser et al., 2019; Steinegger & Salzberg, 2020).

Finally, there are also some major limits that will not be possible to overcome. One can think of the limit in number of records providing access to well preserved and continuous sediment that would allow investigation of palaeoecological questions. This limit is constraint to *seda*DNA but also common to macrofossils and pollen approaches. However, DNA degradation is inherent to the *seda*DNA approach. DNA degrades with time and it is supposed that even under very good preservation conditions, enzymatic activities are slowed down but not stopped. Therefore, with time, DNA gets more and more fragmented and damaged until the point that it is not possible to retrieve useful information from *seda*DNA anymore.

Until the establishment of further and more robust methodologies to investigate *seda*DNA datasets, I can still recommend the use of *seda*DNA approach in concert to the established methods. However, one could imagine in the future that *seda*DNA approaches will become even more accessible with established pipelines and have the potential to be the reference method to tackle palaeoecological questions. In this thesis, I highlighted the investigation of past biodiversity changes (**Manuscript I** and **IV**), past extinction (**Manuscript II**),

snapshots of past communities (**Manuscript III**), but also the investigation of past food webs, population genetics, and functional diversity is possible.

5.3 CONCLUSIONS AND FUTURE PERSPECTIVES

In this work, I showed that using *seda*DNA proxies, it is possible to obtain comparable signals to the established pollen and macrofossils approaches, supporting the *seda*DNA general robustness. In addition, *seda*DNA based methods allowed to highlight and confirm that during the Pleistocene, Siberia and Beringia were covered by a dynamic, diverse and species rich steppe-tundra biota, supporting a variety of fauna and megafauna. This system collapsed at the onset of the warmer Holocene with several extinction events of plant species in parallel to megafauna's extinction events. The collapse of this steppe-tundra led to the modern distribution of flora in Siberian and Beringian high latitudes, with e.g., boreal forests, steppes and tundra, being generally less diverse than the Pleistocene steppe-tundra. With the use of *seda*DNA to investigate new palaecological questions, I could explore and discuss the origin of the collapse of the Pleistocene steppe-tundra and megafauna's extinction, giving clues and ground data for prediction of future ecological changes in high-latitude ecosystems that could resonate worldwide. With the ongoing great methodological progresses and accelerated technical advances, future work using *seda*DNA could explore the functionality of past biota and extract unique and potential extinct genes that could translate into novel proteins with a high potential for biotechnology development. Another approach could use MAGs to deepen the understanding of past and extinct genomes that might help understand the origin or mechanisms of extinction and identify differences between extinct and extant genomes. Future work could also make use of higher resolution records from previous Pleistocene interglacial stages to investigate the dynamics of past interglacial ecosystems in more detail while putting it in parallel to the advanced knowledge we now have from the Holocene ecosystem.

BIBLIOGRAPHY

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., ... & Wohlfarth, B. (2018). Archaeal community changes in Late glacial lake sediments: Evidence from ancient DNA. *Quaternary Science Reviews*, 181, 19-29. https://doi.org/10.1016/j.quascirev.2017.11.037

Alroy, J. (2001). A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. *Science*, 292(5523), 1893-1896. https://doi.org/10.1126/science.1059342

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

Andreev, A. A., Grosse, G., Schirrmeister, L., Kuznetsova, T. V., Kuzmina, S. A., Bobrov, A. A., ... & Kunitsky, V. V. (2009). Weichselian and Holocene palaeoenvironmental history of the Bol'shoy Lyakhovsky Island, New Siberian Archipelago, Arctic Siberia. *Boreas*, 38(1), 72-110. https://doi.org/10.1111/j.1502-3885.2008.00039.x

Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siegert, C., ... & Hubberten, H. W. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. *Quaternary Science Reviews*, 30(17-18), 2182-2199. https://doi.org/10.1111/j.1502-3885.2011.00203.x

Andreev, A. A., Tarasov, P. E., Wennrich, V., Raschke, E., Herzschuh, U., Nowaczyk, N. R., ... & Melles, M. (2014). Late Pliocene and Early Pleistocene vegetation history of northeastern Russian Arctic inferred from the Lake El'gygytgyn pollen record. *Climate of the Past*, 10(3), 1017-1039. https://doi.org/10.5194/cp-10-1017-2014

Ashastina, K., Schirrmeister, L., Fuchs, M., & Kienast, F. (2017). Palaeoclimate characteristics in interior Siberia of MIS 6–2: first insights from the Batagay permafrost mega-thaw slump in the Yana Highlands. *Climate of the Past*, 13(7), 795-818. https://doi.org/10.5194/cp-13-795-2017

Ashastina, K. (2018). *Palaeo-environments at the Batagay site in West Beringia during the late Quaternary*. Dissertation, Friedrich-Schiller-Universität Jena.

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. https://doi.org/10.1016/j.quascirev.2018.07.032

Bakker, E. S., Gill, J. L., Johnson, C. N., Vera, F. W., Sandom, C. J., Asner, G. P., & Svenning, J. C. (2016). Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy of Sciences*, 113(4), 847-855. https://doi.org/10.1073/pnas.1502545112

Bálint, M., Pfenninger, M., Grossart, H. P., Taberlet, P., Vellend, M., Leibold, M. A., ... & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, 33(12), 945-957. https://doi.org/10.1016/j.tree.2018.09.003

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L., & Shabel, A. B. (2004). Assessing the causes of late Pleistocene extinctions on the continents. *science*, 306(5693), 70-75. https://doi.org/10.1126/science.1101476

Barr, I. D., & Clark, C. D. (2012). Late Quaternary glaciations in Far NE Russia; combining moraines, topography and chronology to assess regional and global glaciation synchrony. *Quaternary Science Review*s, 53, 72-87. https://doi.org/10.1016/j.quascirev.2012.08.004

Bazelmans, J., van Balen, R., Bos, J., Brinkkemper, O., Colenberg, J., Doeve, P., ... & van der Woude, J. (2021). Environmental changes in the late Allerød and early Younger Dryas in the Netherlands: a multiproxy high-resolution record from a site with two *Pinus sylvestris* populations. *Quaternary Science Reviews*, 272, 107199. https://doi.org/10.1016/j.quascirev.2021.107199

Begon, M., & Townsend, C. R. (2020). *Ecology: from individuals to ecosystems.* John Wiley & Sons.

Behrensmeyer, A. K., & Hill, A. P. (Eds.). (1988). *Fossils in the making: vertebrate taphonomy and paleoecology* (No. 69). University of Chicago Press.

Bezrukova, E. V., Tarasov, P. E., Solovieva, N., Krivonogov, S. K., & Riedel, F. (2010). Last glacial–interglacial vegetation and environmental dynamics in southern Siberia: Chronology, forcing and feedbacks. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 296(1-2), 185-198. https://doi.org/10.1016/j.palaeo.2010.07.020

Bigelow, N. H., Brubaker, L. B., Edwards, M. E., Harrison, S. P., Prentice, I. C., Anderson, P. M., ... & Volkova, V. S. (2003). Climate change and Arctic ecosystems: 1. Vegetation changes north of 55° N between the last glacial maximum, mid-Holocene, and present. *Journal of Geophysical Research: Atmospheres*, 108(D19). https://doi.org/10.1029/2002JD002559

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., ... & Vazhenina, L. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, 28(23-24), 2445-2464. https://doi.org/10.1016/j.quascirev.2009.04.016

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., ... & Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, 157, 80-97. https://doi.org/10.1016/j.quascirev.2016.11.022

Birks, H. J. B., & Birks, H. H. (1980). *Quaternary palaeoecology*. Edward Arnold: London.

Birks, H. H. (2001). Plant macrofossils. In *Tracking environmental change using lake sediments* (pp. 49-74). Springer, Dordrecht.

Birks, H. J. B. (2014). Challenges in the presentation and analysis of plant-macrofossil stratigraphical data. *Vegetation History and Archaeobotany*, 23(3), 309-330. https://doi.org/10.1007/s00334-013-0430-2

Birks, H. J. B., Felde, V. A., Bjune, A. E., Grytnes, J. A., Seppä, H., & Giesecke, T. (2016). Does pollen-assemblage richness reflect floristic richness? A review of recent developments and future challenges. *Review of Palaeobotany and Palynology*, 228, 1-25. https://doi.org/10.1016/j.revpalbo.2015.12.011

Birks, H. J. B., & Berglund, B. E. (2018). One hundred years of Quaternary pollen analysis 1916–2016. *Vegetation History and Archaeobotany*, 27(2), 271-309. https://doi.org/10.1007/s00334-017-0630-2

Blinnikov, M. S., Gaglioti, B. V., Walker, D. A., Wooller, M. J., & Zazula, G. D. (2011). Pleistocene graminoid-dominated ecosystems in the Arctic. *Quaternary Science Reviews*, 30(21-22), 2906-2929. https://doi.org/10.1016/j.quascirev.2011.07.002

Bocherens, H., Hofman-Kamińska, E., Drucker, D. G., Schmölcke, U., & Kowalczyk, R. (2015). European bison as a refugee species? Evidence from isotopic data on Early Holocene bison and other large herbivores in northern Europe. *PloS one*, 10(2), e0115090. https://doi.org/10.1371/journal.pone.0115090

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. https://doi.org/10.1111/cobi.12195

Botkin, D. B., Saxe, H., Araujo, M. B., Betts, R., Bradshaw, R. H., Cedhagen, T., ... & Stockwell, D. R. (2007). Forecasting the effects of global warming on biodiversity. *Bioscience*, 57(3), 227-236. https://doi.org/10.1641/B570306

Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F. J. W., ... & Olsen, M. S. (2019). Key indicators of Arctic climate change: 1971–2017. *Environmental Research Letters*, 14(4), 045010. https://doi.org/10.1088/1748-9326/aafc1b

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

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in bioinformatics*, 20(4), 1125-1136. https://doi.org/10.1093/bib/bbx120

Brown, T. A., & Barnes, I. M. (2015). The current and future applications of ancient DNA in Quaternary science. *Journal of Quaternary Science*, 30(2), 144-153. https://doi.org/10.1002/jqs.2770

Brown, D. R., Brinkman, T. J., Verbyla, D. L., Brown, C. L., Cold, H. S., & Hollingsworth, T. N. (2018). Changing river ice seasonality and impacts on interior Alaskan communities. *Weather, Climate, and Society*, 10(4), 625-640. https://doi.org/10.1175/WCAS-D-17-0101.1

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., Vuillemin, A., ... & 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

Chapin, F. S., Sala, O. E., Burke, I. C., Grime, J. P., Hooper, D. U., Lauenroth, W. K., Lombard, A., Mooney, H. A., Mosier, A. R., Naeem, S., Pacala, S. W., Roy, J., Steffen, W. L., & Tilman, D. (1998). Ecosystem Consequences of Changing Biodiversity. *BioScience*, 48(1), 45–52. https://doi.org/10.2307/1313227

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. https://doi.org/10.1111/bor.12338

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... & McCabe, A. M. (2009). The last glacial maximum. *science*, 325(5941), 710-714. https://doi.org/10.1126/science.1172873

Clarke, C. L., Edwards, M. E., Brown, A. G., Gielly, L., Lammers, Y., Heintzman, P. D., ... & Alsos, I. G. (2019). Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (sedaDNA) and pollen. *Boreas*, 48(2), 299-316. https://doi.org/10.1111/bor.12357

Cowie, R. H., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews*. 97: 640-663. https://doi.org/10.1111/brv.12816

Crump, S. E., Fréchette, B., Power, M., Cutler, S., de Wet, G., Raynolds, M. K., ... & Miller, G. H. (2021). Ancient plant DNA reveals High Arctic greening during the Last Interglacial. *Proceedings of the National Academy of Sciences*, 118(13), e2019069118. https://doi.org/10.1073/pnas.2019069118

Davidson, S. C., Bohrer, G., Gurarie, E., LaPoint, S., Mahoney, P. J., Boelman, N. T., ... & Hebblewhite, M. (2020). Ecological insights from three decades of animal movement tracking across a changing Arctic. *Science*, 370(6517), 712-715. https://doi.org/10.1126/science.abb7080

Davis, L. G., Madsen, D. B., Becerra-Valdivia, L,. Highman, T., Sisson, D. A., Skinner, S. M., ... & Buvit, I. (2018). Late Upper Paleolithic occupation at Cooper's Ferry, Idaho, USA, ~16,000 years ago. *Science*, 365(6456), 891-897. https://doi.org/10.1126/science.aax9830

Demske, D., Mohr, B., & Oberhänsli, H. (2002). Late Pliocene vegetation and climate of the Lake Baikal region, southern East Siberia, reconstructed from palynological data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 184(1-2), 107-129. https://doi.org/10.1016/S0031-0182(02)00251-1

Dutch, S. (1999). Volcanoes, impacts on ecosystems. In: *Environmental Geology. Encyclopedia of Earth Science*. Springer, Dordrecht. https://doi.org/10.1007/1-4020-4494-1_347

Edwards, M. E., Armbruster, W. S., & Elias, S. E. (2014). Constraints on post-glacial boreal tree expansion out of far-northern refugia. *Global Ecology and Biogeography*, 23(11), 1198-1208. https://doi.org/10.1111/geb.12213

Edwards, M. E. (2020). The maturing relationship between Quaternary paleoecology and ancient sedimentary DNA. *Quaternary Research*, 96, 39-47. https://doi.org/10.1017/qua.2020.52

Elmqvist, T., Folke, C., Nyström, M., Peterson, G., Bengtsson, J., Walker, B., & Norberg, J. (2003). Response diversity, ecosystem change, and resilience. *Frontiers in Ecology and the Environment*, 1(9), 488-494. https://doi.org/10.1890/1540-9295(2003)001[0488:RDECAR]2.0.CO;2

Epstein, H., Bhatt, U. S., Raynolds, M., Walker, D., Forbes, B., Phoenix, G., ... & Jia, G. (2018). *Tundra greenness*. Arctic Report Card 2018.

Earth Resources Observation And Science (EROS) Center (2017) "Global 30 Arc-Second Elevation (GTOPO30)." U.S. Geological Survey. doi: 10.5066/F7DF6PQS.

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Estes, J. A., & Duggins, D. O. (1995). Sea otters and kelp forests in Alaska: generality and variation in a community ecological paradigm. *Ecological monographs*, 65(1), 75-100. https://doi.org/10.2307/2937159

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology letters*, 4(4), 423-425. https://doi.org/10.1098/rsbl.2008.0118

Fine, P. V. (2015). Ecological and evolutionary drivers of geographic variation in species diversity. Annual Review of Ecology, *Evolution, and Systematics*, 46, 369-392.

García Criado, M., Myers-Smith, I. H., Bjorkman, A. D., Lehmann, C. E., & Stevens, N. (2020). Woody plant encroachment intensifies under climate change across tundra and savanna biomes. *Global Ecology and Biogeography*, 29(5), 925-943. https://doi.org/10.1111/geb.13072

Gardner, E. E., Walker, S. E., & Gardner, L. I. (2016). Palaeoclimate, environmental factors, and bird body size: A multivariable analysis of avian fossil preservation. *Earth-Science Review*s, 162, 177–197. https://doi.org/10.1016/j.earscirev.2016.07.001

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

Gill, J. L. (2014). Ecological impacts of the late Quaternary megaherbivore extinctions. *New Phytologist,* 201(4), 1163-1169. https://doi.org/10.1111/nph.12576

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, 27(15), 2153-2155. https://doi.org/10.1093/bioinformatics/btr347

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular ecology resources*, 11(5), 759-769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Gornitz, V. (2021) Timescales of Climate Change. In, *Encyclopedia of Geology* (Second Edition), Pages 318-327, Academic Press, https://doi.org/10.1016/B978-0-08-102908-4.00107-7

Graf, K. E., & Buvit, I. (2017). Human dispersal from Siberia to Beringia: Assessing a Beringian standstill in light of the archaeological evidence. *Current Anthropology*, 58(S17), S583-S603. https://doi.org/10.1086/693388

Graham, R. W., Belmecheri, S., Choy, K., Culleton, B. J., Davies, L. J., Froese, D., ... & Wooller, M. J. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proceedings of the National Academy of Sciences*, 113(33), 9310-9314. https://doi.org/10.1073/pnas.1604903113

Gross, M. (2016). Megafauna moves nutrients uphill. *Current Biology*, 26(1), R1-R5. https://doi.org/10.1016/j.cub.2015.12.028

Guthrie, R. D. (1982). Mammals of the mammoth steppe as paleoenvironmental indicators. In *Paleoecology of Beringia* (pp. 307-326). Academic Press. https://doi.org/10.1016/B978-0-12-355860-2.50030-2

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. https://doi.org/10.1016/S0277-3791(00)00099-8

Guthrie, R. D. (2013). *Frozen fauna of the mammoth steppe: the story of Blue Babe*. University of Chicago Press.

Harrison, S. (2020). Plant community diversity will decline more than increase under climatic warming. *Philosophical Transactions of the Royal Society B*, 375(1794), 20190106. https://doi.org/10.1098/rstb.2019.0106

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), 907-913. https://doi.org/10.1038/35016000

Herzschuh, U., Birks, H. J. B., Laepple, T., Andreev, A., Melles, M., & Brigham-Grette, J. (2016). Glacial legacies on interglacial vegetation at the Pliocene-Pleistocene transition in NE Asia. *Nature Communications,* 7(1), 1-11. https://doi.org/10.1038/ncomms11967

Herzschuh, U. (2020). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, 29(2), 198-206. https://doi.org/10.1111/geb.13018

Hicks, C. C., Cohen, P. J., Graham, N. A., Nash, K. L., Allison, E. H., D'Lima, C., ... & MacNeil, M. A. (2019). Harnessing global fisheries to tackle micronutrient deficiencies. *Nature*, 574(7776), 95-98. https://doi.org/10.1038/s41586-019-1592-6

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163(2), 192-211. https://doi.org/10.1086/381004

Hoelzel, A. R. (2010). Looking backwards to look forwards: conservation genetics in a changing world. *Conservation Genetics*, 11(2), 655-660. https://doi.org/10.1007/s10592-010-0045-4

Holling, C. S., & Gunderson, L. H. (2002). Resilience and adaptive cycles. In: *Panarchy: Understanding Transformations in Human and Natural Systems*, 25-62.

Hopkins, D. M. (1982). Aspects of the paleogeography of Beringia during the late Pleistocene. *Paleoecology of Beringia*, 3-28. https://doi.org/10.1016/B978-0-12-355860-2.50008-9

Hultén, E. (1937). *Outline of the history of arctic and boreal biota during the Quaternary period*. Lehre J Cramer, New York.

IPCC, 2021: Summary for Policymakers. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 3−32.

Isaev, A. P., Protopopov, A. V., Protopopova, V. V., Egorova, A. A., Timofeyev, P. A., Nikolaev, A. N., … & Sleptsova, N. P. (2010). *Vegetation of Yakutia: Elements of Ecology and Plant Sociology*. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-3774-9_3

Jackson, S. T., & Blois, J. L. (2015). Community ecology in a changing environment: Perspectives from the Quaternary. *Proceedings of the National Academy of Sciences*, 112(16), 4915-4921. https://doi.org/10.1073/pnas.1403664111

Jia, W., Liu, X., Stoof-Leichsenring, K. R., Liu, S., Li, K., & Herzschuh, U. (2021). Preservation of sedimentary plant DNA is related to lake water chemistry. *Environmental DNA*, 4(2), 425-439. https://doi.org/10.1002/edn3.259

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. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Kahlke, R. D. (2015). The maximum geographic extension of Late Pleistocene *Mammuthus primigenius* (Proboscidea, Mammalia) and its limiting factors. Quaternary International, 379, 147-154. https://doi.org/10.1016/j.quaint.2015.03.023

Katifori, E., Alben, S., Cerda, E., Nelson, D. R., & Dumais, J. (2010). Foldable structures and the natural design of pollen grains. *Proceedings of the National Academy of Sciences*, 107(17), 7635-7639. https://doi.org/10.1073/pnas.0911223107

Kaufman, D. S., Ager, T. A., Anderson, N. J., Anderson, P. M., Andrews, J. T., Bartlein, P. J., ... & Wolfe, B. B. (2004). Holocene thermal maximum in the western Arctic (0–180 W). *Quaternary Science Reviews*, 23(5-6), 529-560. https://doi.org/10.1016/j.quascirev.2003.09.007

Kienast, F., Siegert, C., Dereviagin, A., & Mai, D. H. (2001). Climatic implications of Late Quaternary plant macrofossil assemblages from the Taymyr Peninsula, Siberia. *Global and Planetary Change*, 31(1-4), 265-281. https://doi.org/10.1016/S0921-8181(01)00124-2

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. https://doi.org/10.1016/j.yqres.2005.01.003

Kirkpatrick, J. B., Walsh, E. A., & D'Hondt, S. (2016). Fossil DNA persistence and decay in marine sediment over hundred-thousand-year to million-year time scales. *Geology*, 44(8), 615-618. https://doi.org/10.1130/G37933.1

Koch, P. L., & Barnosky, A. D. (2006). Late Quaternary extinctions: state of the debate. Annual *Review of Ecology, Evolution, and Systematics*, 37. https://doi.org/10.1146/annurev.ecolsys.34.011802.132415

Kruse, S., Wieczorek, M., Jeltsch, F., & Herzschuh, U. (2016). Treeline dynamics in Siberia under changing climates as inferred from an individual-based model for *Larix*. *Ecological Modelling*, 338, 101-121. https://doi.org/10.1016/j.ecolmodel.2016.08.003

Kruse, S., & Herzschuh, U. (2022). Regional opportunities for tundra conservation in the next 1000 years. *eLife*, 11, e75163. https://doi.org/10.7554/eLife.75163

Kuzmina, S. A. (2015). Quaternary insects and environment of northeastern Asia. P*aleontological Journal*, 49(7), 679-867. https://doi.org/10.1134/S0031030115070011

Layer, P. W. (2000). Argon-40/argon-39 age of the El'gygytgyn impact event, Chukotka, Russia. *Meteoritics & Planetary Science*, 35(3), 591-599. https://doi.org/10.1111/j.1945-5100.2000.tb01439.x

Liu, J., Saito, Y., Wang, H., Zhou, L., & Yang, Z. (2009). Stratigraphic development during the Late Pleistocene and Holocene offshore of the Yellow River delta, Bohai Sea. *Journal of Asian Earth Sciences*, 36(4-5), 318-331. https://doi.org/10.1016/j.jseaes.2009.06.007

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., & Herzschuh, U. (2020). Holocene vegetation and plant diversity changes in the north-eastern Siberian treeline region from pollen and sedimentary ancient DNA. *Frontiers in Ecology and Evolution*, 304. https://doi.org/10.3389/fevo.2020.560243

Liu, S., Kruse, S., Scherler, D., Ree, R. H., Zimmermann, H. H., Stoof-Leichsenring, K. R., ... & Herzschuh, U. (2021). Sedimentary ancient DNA reveals a threat of warming-induced alpine habitat loss to Tibetan Plateau plant diversity. *Nature communications*, 12(1), 1-9. https://doi.org/10.1038/s41467-021-22986-4

Lopatin, A. V. (2021). Yuka the Mammoth, a Frozen Mummy of a Young Female Woolly Mammoth from Oyogos. *Paleontological Journal*, 55(11), 1270-1274. https://doi.org/10.1134/S0031030121110046

Lozhkin, A. V., & Anderson, P. M. (2011). Forest or no forest: implications of the vegetation record for climatic stability in Western Beringia during Oxygen Isotope Stage 3. *Quaternary Science Reviews*, 30(17-18), 2160-2181. https://doi.org/10.1016/j.quascirev.2010.12.022

Lozhkin, A. V., Minyuk, P. S., Anderson, P. M., Nedorubova, E. Y., & Korzun, J. V. (2017). Variability in landscape and lake system responses to glacial and interglacial climates during the Middle Pleistocene based on palynological and geochemical data from Lake El'gygytgyn, Eastern Arctic. *Review of Palaeobotany and Palynology*, 246, 1-13. https://doi.org/10.1016/j.revpalbo.2017.06.004

Lyell, C. (2010). Principles of Geology, Volume 2. In *Principles of Geology, Volume 2*. University of Chicago Press. https://doi.org/10.7208/9780226498027

Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., & Knowlton, N. (2020). Climate change and ecosystems: Threats, opportunities and solutions. *Philosophical Transactions of the Royal Society B*, 375(1794), 20190104. https://doi.org/10.1098/rstb.2019.0104

Mangerud, J. (2020). The discovery of the Younger Dryas, and comments on the current meaning and usage of the term. *Boreas*, 50(1), 1-5. https://doi.org/10.1111/bor.12481

Margulies, L., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., … & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376–380. https://doi.org/10.1038/nature03959

Martin, P. S. (1958). Pleistocene ecology and biogeography of North America. In C. L. Hubbs (Ed.), *Zoogeography* (pp. 375–420). Washington, DC: American Association for the Advancement of Science.

Martin, P. S., & Wright, H. E. (1967). *Pleistocene extinctions; the search for a cause*. Yale University Press

Matthews Jr, J. V. (1982). East Beringia during Late Wisconsin time: a review of the biotic evidence. *Paleoecology of Beringia*, 127-150. https://doi.org/10.1016/B978-0-12-355860-2.50015-6

Melles, M., Brigham-Grette, J., Minyuk, P. S., Nowaczyk, N. R., Wennrich, V., DeConto, R. M., ... & Wagner, B. (2012). 2.8 million years of Arctic climate change from Lake El'gygytgyn, NE Russia. *science*, 337(6092), 315-320. https://doi.org/10.1126/science.1222135

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1), 31-46. https://doi.org/10.1038/nrg2626

Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C., & White, J. W. (2010). Arctic amplification: can the past constrain the future? *Quaternary Science Reviews,* 29(15-16), 1779-1790. https://doi.org/10.1016/j.quascirev.2010.02.008

Mix, A. C., & Ruddiman, W. F. (1984). Oxygen-Isotope Analyses and Pleistocene Ice Volumes 1. *Quaternary Research*, 21(1), 1-20. https://doi.org/10.1016/0033-5894(84)90085-1

Monteath, A. J., Gaglioti, B. V., Edwards, M. E., & Froese, D. (2022). Late Pleistocene shrub expansion preceded megafauna turnover and extinctions in eastern Beringia. *Proceedings of the National Academy of Sciences*, 118(52), e2107977118. https://doi.org/10.1073/pnas.2107977118

Morrissey, E. M., McHugh, T. A., Preteska, L., Hayer, M., Dijkstra, P., Hungate, B. A., & Schwartz, E. (2015). Dynamics of extracellular DNA decomposition and bacterial community composition in soil. *Soil Biology and Biochemistry*, 86, 42-49. https://doi.org/10.1016/j.soilbio.2015.03.020

Murchie, T. J., Monteath, A. J., Mahony, M. E., Long, G. S., Cocker, S., Sadoway, T., ... & Poinar, H. N. (2021). Collapse of the mammoth-steppe in central Yukon as revealed by ancient environmental DNA. *Nature Communications*, 12(1), 1-18. https://doi.org/10.1038/s41467-021-27439-6

Murphy, B. P., Andersen, A. N., & Parr, C. L. (2016). The underestimated biodiversity of tropical grassy biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1703), 20150319. https://doi.org/10.1098/rstb.2015.0319

Murton, J. B., Edwards, M. E., Lozhkin, A. V., Anderson, P. M., Savvinov, G. N., Bakulina, N., ... & Zanina, O. G. (2017). Preliminary paleoenvironmental analysis of permafrost deposits at Batagaika megaslump, Yana Uplands, northeast Siberia. *Quaternary Research*, 87(2), 314-330. https://doi.org/10.1017/qua.2016.15

Murton, J. B., Opel, T., Toms, P., Blinov, A., Fuchs, M., Wood, J., ... & Wetterich, S. (2021). A multimethod dating study of ancient permafrost, Batagay megaslump, east Siberia. *Quaternary Research*, 105, 1-22. https://doi.org/10.1017/qua.2021.27

Nehring, A. (1890). *Ueber Tundren und Steppen der jetzt-und vorzeit: mit besonderer Berücksichtigung ihrer Fauna*. F. Dümmler. In german.

Niemeyer, B., Klemm, J., Pestryakova, L. A., & Herzschuh, U. (2015). Relative pollen productivity estimates for common taxa of the northern Siberian Arctic. *Review of Palaeobotany and Palynology*, 221, 71-82. https://doi.org/10.1016/j.revpalbo.2015.06.008

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

Niittynen, P., Heikkinen, R. K., & Luoto, M. (2018). Snow cover is a neglected driver of Arctic biodiversity loss. *Nature Climate Change*, 8(11), 997-1001. https://doi.org/10.1038/s41558-018-0311-x

Norby, R. J., & Luo, Y. (2004). Evaluating ecosystem responses to rising atmospheric CO2 and global warming in a multi-factor world. *New phytologist*, 162(2), 281-293. https://doi.org/10.1111/j.1469-8137.2004.01047.x

Odum, E. P., & Barrett, G. W. (1971). *Fundamentals of ecology* (Vol. 3, p. 5). Philadelphia: Saunders.

Ogram, A., Sayler, G. S., & Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *Journal of microbiological methods*, 7(2-3), 57-66. https://doi.org/10.1016/0167-7012(87)90025-X

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annual review of microbiology*, 40, 337-365. https://doi.org/10.1146/annurev.mi.40.100186.002005

Opel, T., Murton, J. B., Wetterich, S., Meyer, H., Ashastina, K., Günther, F., ... & Schirrmeister, L. (2019). Past climate and continentality inferred from ice wedges at Batagay megaslump in the Northern Hemisphere's most continental region, Yana Highlands, interior Yakutia. *Climate of the Past*, 15(4), 1443-1461. https://doi.org/10.5194/cp-15-1443-2019

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., ... & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, 1(1), 1-26. https://doi.org/10.1038/s43586-020-00011-0

Owen, R. (1861) *Palaeontology, or a systematic study of extinct animals and their geological relations*. Adam and Charles Black.

Owen-Smith, N. (1987). Pleistocene extinctions: the pivotal role of megaherbivores. *Paleobiology*, 13(3), 351-362.

Pace, N. R., Stahl, D. A., Lane, D. J., & Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. In *Advances in microbial ecology* (pp. 1-55). Springer, Boston, MA.

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., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., ... & Willerslev, E. (2012). Glacial survival of boreal trees in northern Scandinavia. *science*, 335(6072), 1083-1086. https://doi.org/10.1126/science.1216043

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. https://doi.org/10.1098/rstb.2013.0382

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

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., ... & Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355(6332), eaai9214. https://doi.org/10.1126/science.aai9214

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. https://doi.org/10.1016/j.quascirev.2013.06.006

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., ... & 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

Petersen, S. V., Dutton, A., & Lohmann, K. C. (2016). End-Cretaceous extinction in Antarctica linked to both Deccan volcanism and meteorite impact via climate change. *Nature communications*, 7(1), 1-9. https://doi.org/10.1038/ncomms12079

Peterson, G., Allen, C. R., & Holling, C. S. (1998). Ecological Resilience, Biodiversity, and Scale. *Ecosystems* 1, 6–18. https://doi.org/10.1007/s100219900002

Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., ... & Ecker, J. R. (2013). Patterns of population epigenomic diversity. *Nature* 495, 193–198. https://doi.org/10.1038/nature11968

Pickering, T. R., Clarke, R. J., & Moggi-Cecchi, J. (2004). Role of carnivores in the accumulation of the Sterkfontein Member 4 hominid assemblage: a taphonomic reassessment of the complete hominid fossil sample (1936–1999). A*merican Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 125(1), 1-15. https://doi.org/10.1002/ajpa.10278

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, 45(3), 219-235. https://doi.org/10.1007/s00374-008-0345-8

Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., & Salse, J. (2019). Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biology*, 20(1), 1-17. https://doi.org/10.1186/s13059-019-1627-1

Reitalu, T., Bjune, A. E., Blaus, A., Giesecke, T., Helm, A., Matthias, I., ... & Birks, H. J. B. (2019). Patterns of modern pollen and plant richness across northern Europe. *Journal of Ecology*, 107(4), 1662-1677. https://doi.org/10.1111/1365-2745.13134

Reuther, J. D., Rogers, J., Druckenmiller, P., Bundtzen, T. K., Wallace, K., Bowman, R., ... & Cherkinsky, A. (2019). Late Quaternary (≥ MIS 3 to MIS 1) stratigraphic transitions in a highland Beringian landscape along the Kuskokwim River, Alaska. *Quaternary Research*, 93, 139-154. https://doi.org/10.1017/qua.2019.51

Rijal, D. P., Heintzman, P. D., Lammers, Y., Yoccoz, N. G., Lorberau, K. E., Pitelkova, I., ... & Alsos, I. G. (2021). Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. *Science Advances*, 7(31), eabf9557. https://doi.org/10.1126/sciadv.abf9557

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, 17, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Sandel, B., Monnet, A. C., Govaerts, R., & Vorontsova, M. (2017). Late Quaternary climate stability and the origins and future of global grass endemism. *Annals of Botany*, 119(2), 279-288. https://doi.org/10.1093/aob/mcw178

Sandom, C., Faurby, S., Sandel, B., & Svenning, J. C. (2014). Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787), 20133254. https://doi.org/10.1098/rspb.2013.3254

Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1), 1-11. https://doi.org/10.1186/s40168-016-0154-5

Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., ... & Ecker, J. R. (2013). Patterns of population epigenomic diversity. *Nature*, 495(7440), 193-198. https://doi.org/10.1038/nature11968

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2021). Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Molecular Ecology Resources*, 21(3), 801-815. https://doi.org/10.1111/1755-0998.13311

Shackleton, N. (1967). Oxygen isotope analyses and Pleistocene temperatures re-assessed. *Nature*, 215(5096), 15-17. https://doi.org/10.1038/215015a0

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, 550(7676), 345-353. https://doi.org/10.1038/nature24286

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. https://doi.org/10.1111/nph.14199

Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, 560(7720), 639-643. https://doi.org/10.1038/s41586-018-0411-9

Steffensen, J. P., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Fischer, H., ... & White, J. W. (2008). High-resolution Greenland ice core data show abrupt climate change happens in few years. *science*, 321(5889), 680-684. https://doi.org/10.1126/science.1157707

Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome biology*, 21(1), 1-12. https://doi.org/10.1186/s13059-020-02023-1

Stewart, M., Carleton, W. C., & Groucutt, H. S. (2021). Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. *Nature communications*, 12(1), 1-15. https://doi.org/10.1038/s41467-021-21201-8

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., ... & Stein, R. (2004). Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, 23(11-13), 1229-1271. https://doi.org/10.1016/j.quascirev.2003.12.008

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

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). *Environmental dna*. Molecular ecology, 21(8), 1789-1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

Tarasov, P., Granoszewski, W., Bezrukova, E., Brewer, S., Nita, M., Abzaeva, A., & Oberhänsli, H. (2005). Quantitative reconstruction of the last interglacial vegetation and climate based on the pollen record from Lake Baikal, Russia. *Climate Dynamics*, 25(6), 625-637. https://doi.org/10.1007/s00382-005-0045-0

Tarasov, P., Bezrukova, E., Karabanov, E., Nakagawa, T., Wagner, M., Kulagina, N., ... & Riedel, F. (2007). Vegetation and climate dynamics during the Holocene and Eemian interglacials derived from Lake Baikal pollen records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 252(3-4), 440-457. https://doi.org/10.1016/j.palaeo.2007.05.002

Tarasov, P. E., Andreev, A. A., Anderson, P. M., Lozhkin, A. V., Leipe, C., Haltia, E., ... & Melles, M. (2013). A pollen-based biome reconstruction over the last 3.562 million years in the Far East Russian Arctic–new insights into climate–vegetation relationships at the regional scale. *Climate of the Past*, 9(6), 2759-2775. https://doi.org/10.5194/cp-9-2759-2013

Terborgh, J., Lopez, L., Nuñez, P., Rao, M., Shahabuddin, G., Orihuela, G., ... & Balbas, L. (2001). Ecological meltdown in predator-free forest fragments. *Science*, 294(5548), 1923-1926. https://doi.org/10.1126/science.106439

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

Tomescu, A. M., Bomfleur, B., Bippus, A. C., & Savoretti, A. (2018). Why are bryophytes so rare in the fossil record? A spotlight on taphonomy and fossil preservation. In *Transformative paleobotany* (pp. 375-416). Academic Press. https://doi.org/10.1016/B978-0-12-813012-4.00016-4

Tugarinov, A. Y. (1929). On the origin of the Arctic fauna. *Priroda*, 7, 653-680.

Turner, M. G., Calder, W. J., Cumming, G. S., Hughes, T. P., Jentsch, A., LaDeau, S. L., ... & Carpenter, S. R. (2020). Climate change, ecosystems and abrupt change: science priorities. *Philosophical Transactions of the Royal Society B*, 375(1794), 20190105. https://doi.org/10.1098/rstb.2019.0105

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology letters*, 11(12), 1351-1363. https://doi.org/10.1111/j.1461-0248.2008.01250.x

van der Knaap, W. O. (1987). Five short pollen diagrams of soils from Jan Mayen, Norway: a testimony of a dynamic landscape. *Polar research*, 5(2), 193-206. https://doi.org/10.3402/polar.v5i2.6876

Velichko, A. A. (1975). Paragenesis of a cryogenic (periglacial) zone. *Biuletyn Peryglacjalny*, 24, 89-110.

Vera, F. W. M. (2000). *Grazing ecology and forest history*. Cabi.

Vitousek, P. M., & Walker, L. R. (1989). Biological invasion by *Myrica faya* in Hawai'i: plant demography, nitrogen fixation, ecosystem effects. *Ecological monographs*, 59(3), 247-265. https://doi.org/10.2307/1942601

Walker, M., Johnsen, S., Rasmussen, S. O., Steffensen, J. P., Popp, T., Gibbard, P., ... & Schwander, J. (2008). The global stratotype section and point (GSSP) for the base of the Holocene series/epoch (Quaternary system/period) in the NGRIP ice core. *Episodes Journal of International Geoscience*, 31(2), 264-267. https://doi.org/10.18814/epiiugs/2008/v31i2/016

Walther, G. R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549), 2019-2024. https://doi.org/10.1098/rstb.2010.0021

Wang, Y., Pedersen, M. W., Alsos, I. G., De Sanctis, B., Racimo, F., Prohaska, A., ... & Willerslev, E. (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. *Nature*, 600(7887), 86-92. https://doi.org/10.1038/s41586-021-04016-x

Weng, C., Hooghiemstra, H., & Duivenvoorden, J. F. (2006). Challenges in estimating past plant diversity from fossil pollen data: statistical assessment, problems, and possible solutions. *Diversity and distributions*, 12(3), 310-318. https://doi.org/10.1111/j.1366-9516.2006.00230.x

Wicklow, D. T., & Yocom, D. H. (1981). Fungal species numbers and decomposition of rabbit faeces. *Transactions of the British Mycological Society*, 76(1), 29-32. https://doi.org/10.1016/S0007-1536(81)80005-8

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., ... & 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

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

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual review of ecology, evolution, and systematics*, 273-309.

Wroe, S., Field, J., & Grayson, D. K. (2006). Megafaunal extinction: climate, humans and assumptions. *Trends in Ecology & Evolution*. https://doi.org/10.1016/j.tree.2005.11.012

Yurtsev, B. A. (1982). Relics of the xerophyte vegetation of Beringia in northeastern Asia. In *Paleoecology of Beringia* (pp. 157-177). Academic Press. https://doi.org/10.1016/B978-0-12-355860-2.50018-1

Zanina, O. G., Gubin, S. V., Kuzmina, S. A., Maximovich, S. V., & Lopatina, D. A. (2011). Late-Pleistocene (MIS 3-2) palaeoenvironments as recorded by sediments, palaeosols, and ground-squirrel nests at Duvanny Yar, Kolyma lowland, northeast Siberia. *Quaternary Science Reviews*, 30(17-18), 2107-2123. https://doi.org/10.1016/j.quascirev.2011.01.021

Zettlemoyer, M. A., McKenna, D. D., & Lau, J. A. (2019). Species characteristics affect local extinctions. *American Journal of Botany*, 106(4), 547-559. https://doi.org/10.1002/ajb2.1266

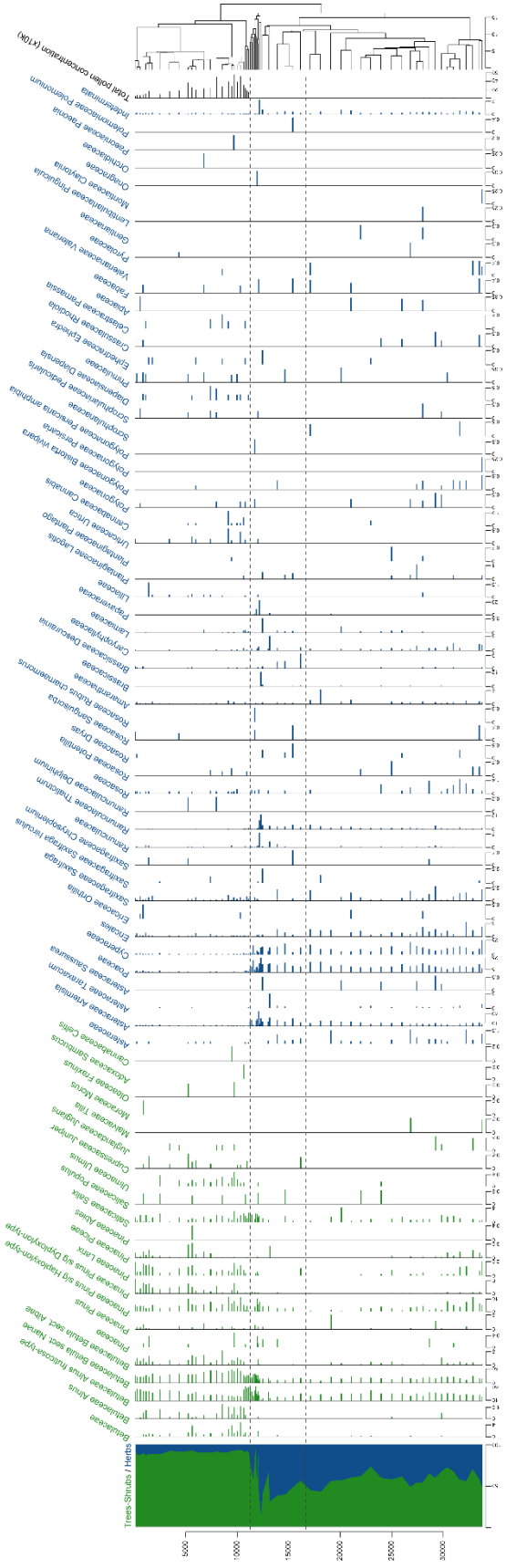
Zhao, W., Andreev, A. A., Tarasov, P. E., Wennrich, V., & Melles, M. (2019). Vegetation and climate during the penultimate interglacial of the northeastern Russian Arctic: the Lake El'gygytgyn pollen record. *Boreas*, 48(2), 507-515. https://doi.org/10.1111/bor.12373

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., ... & Herzschuh, U. (2017). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences*, 14(3), 575-596. https://doi.org/10.5194/bg-14-575-2017

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. https://doi.org/10.1016/j.quascirev.2012.10.005

APPENDICES

APPENDIX 1: SUPPLEMENTARY MATERIAL FOR MANUSCRIPT I



Supplemetary Figure 1.1 | Detailed stratigraphic plot of terrestrial pollen with relative percentages of the taxa in each sample as horizontal bars, ratio of tree/shrub taxa vs herb taxa, and CONISS dendrogram. Taxa are sorted at for trees and shrubs and at family level for herbs. Three pollen zones (PZ) are shown derived from the CONISS analysis. Samples from sediment core PG2133 are plotted against calibrated ages given in cal yr BP.

Supplemetary Figure 1.2 | Detailed stratigraphic plot of terrestrial sedaDNA with relative percentages of the taxa in each sample as horizontal bars, ratio of tree/shrub taxa vs herb taxa, and CONISS dendrogram. The taxa are sorted for trees and shrubs and herbs. Three sedaDNA zones (DNAZ) are shown, derived from the CONISS analysis. Calibrated ages from the age–depth model are given in calendar years before present (cal yr BP).

Supplementary Table 1.1 | Uncalibrated results of radiocarbon dating from the MICADAS (AWI) and Poznan laboratories including sample depth and its range within each sub-core and the composite depth below sediment surface given as mid-point (centred) cm. Material and dated fractions are shown, as well as the error provided by the laboratory.

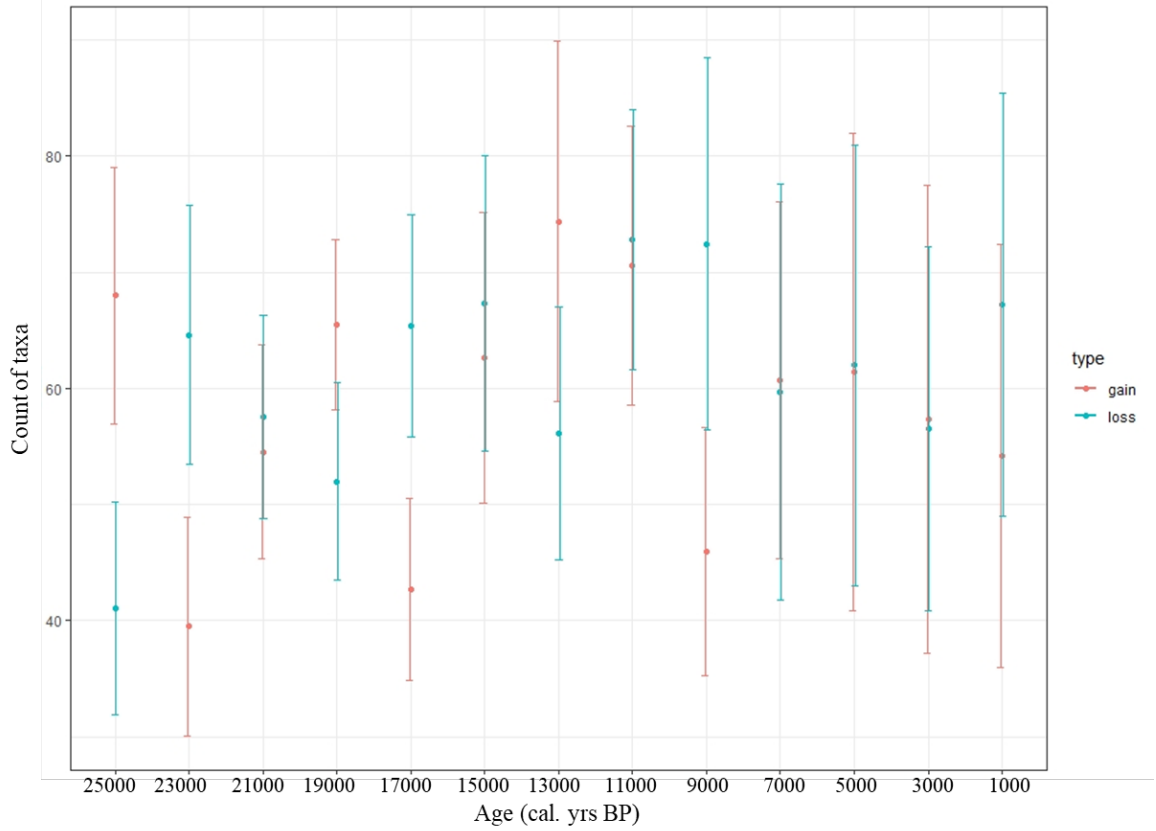
Core sample ID and range in cm	Composite depth (mid-point cm)	Lab	Lab-ID	Material	Lab comment (µg C)	Uncalibrated Age (14C yr BP)	Uncalibrated Age Error (+/- yr)	Method	Reservoir Value (yr)	Reservoir Error (+/- yr)	Comments
PG2112 0-1cm	0.5	AWI MICADAS	AWI-5388.1.1	bulk sediment	987	1140	45	A	0	45	Surface sample used as reservoir estimate
PG2119 0-1cm	0.5	AWI MICADAS	AWI-5387.1.1	bulk sediment	142	699	66	A	0	66	Surface sample used as reservoir estimate
PG2133-2_6-7_(9) SOL	9.5	Poznan	Poz-102011	bulk sediment		3765	35	SOL	0	0	
PG2133-4 17-17.5cm	17.25	AWI MICADAS	AWI-5377.1.1	bulk sediment	998	2578	53	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2133-2_16-17_(19) SOL	19.5	Poznan	Poz-101853	bulk sediment		4110	35	SOL	0	0	
PG2208-1 20-20.5cm	20.25	AWI MICADAS	AWI-5389.1.1	bulk sediment	994	1336	47	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-2_31-32 RES	33.5	Poznan	Poz-69756	bulk sediment		9510	50	RES	0	0	
PG2133-2_31-32 SOL	33.5	Poznan	Poz-69749	bulk sediment		7010	40	SOL	0	0	
PG2208-1 37-37,5 SOL	37.25	Poznan	Poz-101854	bulk sediment		2130	35	SOL	0	0	
PG2133-2 46-47cm	49.5	AWI MICADAS	AWI-5633.1.1	bulk sediment	999	8696	35	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2133-2 51-52	53.5	Poznan	Poz-92873	plant remain		8400	50	RES	0	0	
PG2208-2 55-56cm	55.5	AWI MICADAS	AWI-5390.1.1	bulk sediment	987	2234	51	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-2 56-57cm	59.5	AWI MICADAS	AWI-5379.1.1	bulk sediment	995	10068	121	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2133-2 85-86cm	77.5	AWI MICADAS	AWI-5380.1.1	bulk sediment	439	12893	60	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2133-2 85-86cm	77.5	AWI MICADAS	AWI-5380.1.1_duplicate1	bulk sediment	439	13007	56	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-2_85-86 RES	85.5	Poznan	Poz-69760-	bulk sediment	300	5150	50	RES	0	0	

Core sample ID and range in cm	Composite depth (mid-point cm)	Lab	Lab-ID	Material	Lab comment (µg C)	Uncalibrated Age (14C yr BP)	Uncalibrated Age Error (+/- yr)	Method	Reservoir Value (yr)	Reservoir Error (+/- yr)	Comments
PG2208-2_85-86 SOL	85.5	Poznan	Poz-69753	bulk sediment		3125	35	SOL	0	0	
PG2133-2_87-88 RES	89.5	Poznan	Poz-69757	bulk sediment		22950	160	RES	0	0	
PG2133-2_87-88 SOL	89.5	Poznan	Poz-69751	bulk sediment		13630	80	SOL	0	0	
PG2208-2 91-92cm	91.5	AWI MICADAS	AWI-5391.1.1	bulk sediment	990	3192	56	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-2 106-107cm	98.5	AWI MICADAS	AWI-5381.1.1	bulk sediment	339	30539	263	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-2_137-138 SOL	137.5	Poznan	Poz-101855	bulk sediment		5250	40	SOL	0	0	
PG2133-2_138-139 RES	140.5	Poznan	Poz-69758	bulk sediment		31740	430	RES	0	0	
PG2133-2_138-139 SOL	140.5	Poznan	Poz-69752	bulk sediment	300	13430	100	SOL	0	0	
PG2208-2 141-142cm	141.5	AWI MICADAS	AWI-5392.1.1	bulk sediment	983	5089	68	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-2 156-157cm	148.5	AWI MICADAS	AWI-5382.1.1	bulk sediment	129	14333	156	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-2_175-176 SOL	175.5	Poznan	Poz-105351	bulk sediment	300	4480	40	SOL	0	0	
PG2208-2_188-189 RES	188.5	Poznan	Poz-69761-	bulk sediment	400	5420	50	RES	0	0	
PG2208-2_188-189 SOL	188.5	Poznan	Poz-69755-	bulk sediment		3710	40	SOL	0	0	
PG2208-2_189-190 SOL	189.5	Poznan	Poz-102012	bulk sediment		3725	30	SOL	0	0	
PG2208-2 203-204cm	203.5	AWI MICADAS	AWI-5393.1.1	bulk sediment	984	7024	85	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-2 213-214cm	205.5	AWI MICADAS	AWI-5383.1.1	bulk sediment	128	26717	539	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-2_215-216 SOL	215.5	Poznan	Poz-105352	bulk sediment	200	6940	50	SOL	0	0	
PG2133-3 70-71 RES	220.5	Poznan	Poz-92874	bulk sediment		43000	2000	RES	0	0	
PG2133-3 70-71 SOL	220.5	Poznan	Poz-92911	bulk sediment	400	20590	130	SOL	0	0	
PG2208-2 242-243	242.5	Poznan	Poz-92872	plant remains	600	8440	50	RES	0	0	

Core sample ID and range in cm	Composite depth (mid-point cm)	Lab	Lab-ID	Material	Lab comment (µg C)	Uncalibrated Age (14C yr BP)	Uncalibrated Age Error (+/- yr)	Method	Reservoir Value (yr)	Reservoir Error (+/- yr)	Comments
PG2133-3 100-101cm	251.5	AWI MICADAS	AWI-5384.1.1	bulk sediment	263	24612	178	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-3 57-58cm	257.5	AWI MICADAS	AWI-5394.1.1	bulk sediment	982	9304	36	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2208-3 67-68 RES	267.5	Poznan	Poz-92871-	bulk sediment		12220	60	RES	0	0	
PG2208-3 67-68 SOL	267.5	Poznan	Poz-92910	bulk sediment		9600	50	SOL	0	0	
PG2208-3 89-90cm	289.5	AWI MICADAS	AWI-5395.1.1	bulk sediment	547	13846	38	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-3 158-159cm	309.5	AWI MICADAS	AWI-5385.1.1	bulk sediment	337	31778	293	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2208-3 132-133cm	332.5	AWI MICADAS	AWI-5396.1.1	bulk sediment	132	17729	220	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2208-3 162-163 RES	362.5	Poznan	Poz-92870-	bulk sediment	300	32700	700	RES	0	0	
PG2208-3 162-163 SOL	362.5	Poznan	Poz-93076	bulk sediment	350	16080	110	SOL	0	0	
PG2208-3 165-166cm	365.5	AWI MICADAS	AWI-5397.1.1	bulk sediment	132	22081	314	A	762	66	*Reservoir ages are from surface sample PG2119 (ID: 5387.1.1) plus 63 years for expedition start
PG2133-3 219-220cm	370.5	AWI MICADAS	AWI-5386.1.1	bulk sediment	268	30487	276	A	1203	45	*Reservoir ages are from surface sample PG2112 (ID: 5388.1.1) plus 63 years for expedition start
PG2133-3_215-216 RES	374.5	Poznan	Poz-69759	bulk sediment		44000	2000	RES	0	0	
PG2133-3_215-216 SOL	374.5	Poznan	Poz-69650	bulk sediment	80	16160	220	SOL	0	0	

APPENDIX 2: SUPPLEMENTARY MATERIAL FOR MANUSCRIPT II

Supplementary Figure 2.1 | Number of taxa gain and loss at each time slice. The average values of the 100 iterations of resampling are plotted. In red are plotted the average gain and in blue, the average loss. The error bars are the minimum and maximum values.



Supplemetary Table 1: Raw information after OBITOOLs and number of reads and ASVs excluded because exotic or present in the blanks. Overall, it represents a large amount of data and here is only a compilation of some metrics of ASVs and reads handled.

Core	Site - Lake	Number of reads assigned 100%	Number of ASVs assigned 100%	Number of reads assigned > 90%	Number ASVs assigned > 90%	Number of ASVs excluded and %
PG1755	Bilyakh	17697877	488	20348066	156656	12066 (7.7%)
PG2133	Bolshoe Toko	4822610	350	5262619	22305	353 (1.6%)
E5-1A	E5	20509643	406	23521019	202984	4711 (2.3%)
Co1412	Emanda	8210695	273	9397885	3067	21 (0.7%)
EN18208	Ilirney	14341508	414	16904128	109159	1775 (1.6%)
16-KP-01-L02	Ilirney	33569990	546	38280155	234852	70 (0.03%)
PG2023	Kyutyunda	4241206	624	6541937	123410	1110 (0.8%)
PG1341	Lama	8563021	567	10620146	96243	736 (0.7%)
Co1401	Levinson Lessing	19932963	678	22200122	120322	483 (0.4%)
EN18218	Rauchuagytgyn	15780883	424	18446852	133310	901 (0.7%)

Supplementary Table 2.2 | Metrics for each time slice before resampling. The "28000+" time slice aggregates the metrics for the samples dated older than 28000 cal. yrs BP. It is not part of the resampling. In blue are highlighted the minimum values and in red the maximum values for each metric.

Time slice (cal. yrs BP)	Number of cores covered	Number of samples	Number of reads assigned	Number of taxa
28000+	9	126	21,298,728	443
27000	9	39	**10,524,112**	393
25000	**10**	25	5,116,414	395
23000	9	25	4,317,628	361
21000	**10**	23	4,427,615	371
19000	9	20	**3,463,489**	359
17000	9	**19**	3,977,927	**339**
15000	**10**	29	5,622,699	376
13000	9	25	5,023,605	376
11000	**8**	39	7,228,455	405
9000	9	35	6,635,600	375
7000	9	33	6,521,288	372
5000	**10**	35	6,864,532	377
3000	**8**	**40**	7,686,373	377
1000	**8**	34	5,235,825	353

Supplementary Table 2.3 | Summary of number of taxa extirpating and the calculated expected and observed reappearance rate at each time slice. The time slices with the difference between expected and observed reappearance rate superior to 5% are highlighted in bold.

Time slice (cal. yrs BP)	27000	25000	23000	21000	19000	**17000**	15000	13000	**11000**	9000	7000	**5000**	3000	1000
Potential number of time slice to recover	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Total number of taxa extirpated	81	30	52	35	55	54	36	52	38	49	43	38	31	52
Number of taxa that do not reappear	0	0	0	2	1	4	3	2	9	8	9	16	15	52
Number of taxa that reappear after 10 time slices	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of taxa that reappear after 9 time slices	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of taxa that reappear after 8 time slices	6	3	0	0	0	0	0	0	0	0	0	0	0	0
Number of taxa that reappear after 7 time slices	11	0	0	0	1	0	0	0	0	0	0	0	0	0
Number of taxa that reappear after 6 time slices	5	0	2	0	0	0	1	1	0	0	0	0	0	0
Number of taxa that reappear after 5 time slices	4	3	2	0	1	0	1	0	2	0	0	0	0	0
Number of taxa that reappear after 4 time slices	7	1	3	2	3	1	0	2	4	3	0	0	0	0
Number of taxa that reappear after 3 time slices	9	3	6	5	6	3	0	1	5	3	2	0	0	0
Number of taxa that reappear after 2 time slices	5	7	10	3	23	12	12	8	10	7	5	4	0	0
Number of taxa that reappear after 1 time slice	32	13	29	23	20	34	19	38	8	28	27	18	16	0
Observed proportion of taxa that reappear	100%	100%	100%	94%	98%	93%	92%	96%	76%	84%	79%	58%	52%	0%
Expected reappearance rate (calculated from time slice with still 10 or more time slices to recover)				98%	97%	98%	94%	90%	88%	85%	79%	69%	49%	
Difference between expected and observed reappearance rates				3%	-1%	**5%**	2%	-6%	**12%**	1%	0%	**11%**	-2%	

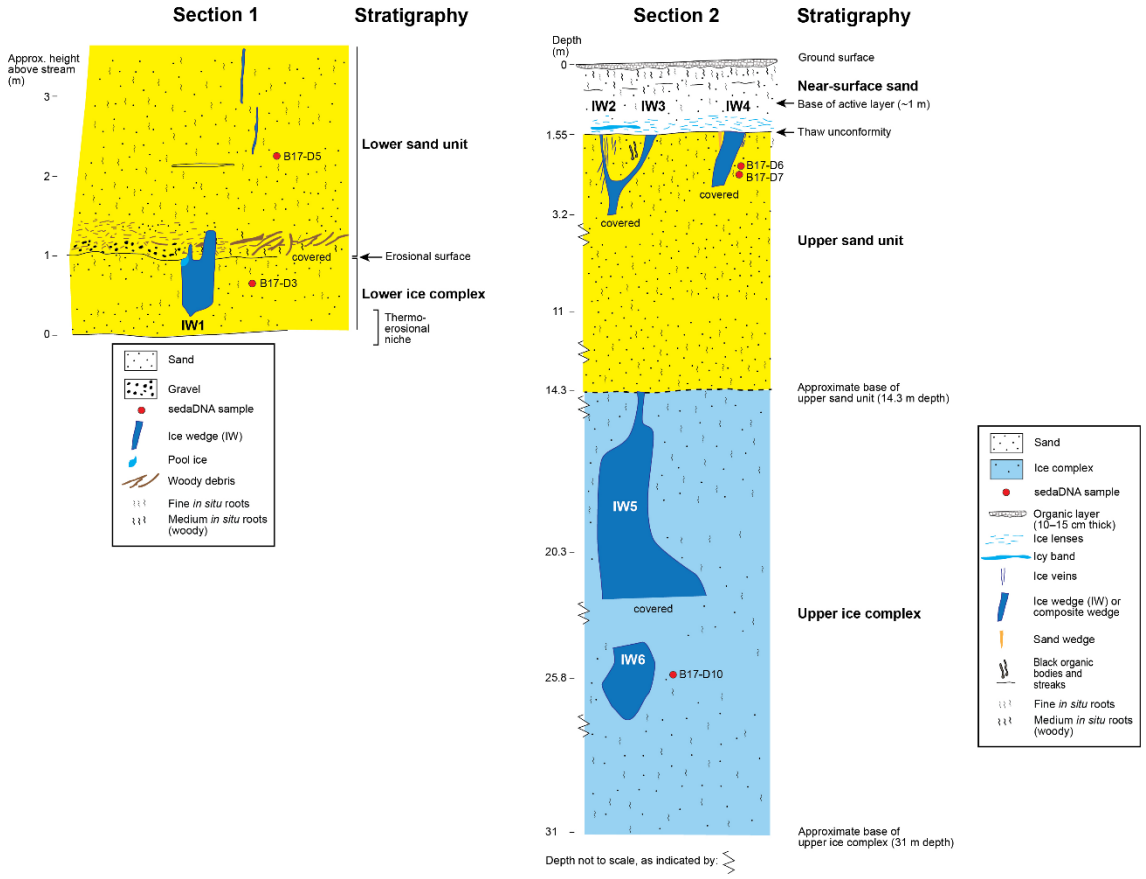
Supplementary Table 2.4 | Number of taxa identified per family.

Family	Number of taxa	Family	Number of taxa	Family	Number of taxa
Asteraceae	59	Onagraceae	7	Campanulaceae	1
Ericaceae	37	Apiaceae	6	Cardiopteridaceae	1
Poaceae	33	Pinaceae	5	Ceratophyllaceae	1
Orobanchaceae	28	Primulaceae	5	Cordiaceae	1
Ranunculaceae	28	Juncaceae	4	Cornaceae	1
Rosaceae	27	Asparagaceae	3	Crassulaceae	1
Saxifragaceae	25	Caprifoliaceae	3	Cupressaceae	1
Brassicaceae	22	Rubiaceae	3	Icacinaceae	1
Polygonaceae	21	Adoxaceae	2	Isoetaceae	1
Salicaceae	21	Euphorbiaceae	2	Ixonanthaceae	1
Betulaceae	16	Grossulariaceae	2	Juglandaceae	1
Fabaceae	15	Haloragaceae	2	Melanthiaceae	1
Cyperaceae	13	Humiriaceae	2	Nymphaeaceae	1
Boraginaceae	12	Lecythidaceae	2	Solanaceae	1
Papaveraceae	12	Myrothamnaceae	2	Theaceae	1
Plantaginaceae	12	Polemoniaceae	2	Ticodendraceae	1
Potamogetonaceae	12	Alseuosmiaceae	1	Tofieldiaceae	1
Caryophyllaceae	9	Araliaceae	1	Typhaceae	1

APPENDIX 3: SUPPLEMENTARY MATERIAL FOR MANUSCRIPT III

Here are attached the most important information to follow manuscript III. Because the supplementary tables 3.2 and 3.3 were too big to be completely added in this Thesis, further information and higher quality pictures can be retrieved on the supplementary information published with the article: https://doi.org/10.1002/edn3.336.

Supplementary Figure 3.1 | Schematic and descriptive representation of the stratigraphic features of the 2 sections of the Batagay mega slump from which the permafrost samples originate. The position of the samples recovered are indicated with a red dot.



Supplementary Figures 3.2 | Mapdamage2 plots – Detailed method explanation

For each investigated subset, we selected from investigated families a reference genome from NCBI and downloaded it as fasta file. We extracted from the fastq files all reads assigned to the family of interest and aligned them to the downloaded reference genome using bwa aln (version 0.7.17; with modified parameters –n 0.001, -l 1000 and –o 2), picard (version 2.24.1) and samtools (version 1.10). We then used mapdamage2.0 to estimate damage patterns at the end of the aligned DNA fragments for each sample. A sufficient number of reads must align to the reference genome to allow mapdamage2.0 to provide a good damage patterns estimation. For this reason, we used the most represented families from each subset and different reference genomes and share here some good examples of the results obtained. We were not able to provide evidence of damage patterns for Mammalia, Insecta nor Aves as too few reads were assigned to families and aligned to the reference genome selected. We still share as an example our trial vs to estimate an extinct taxon: Mammuthus, which cannot be modern.

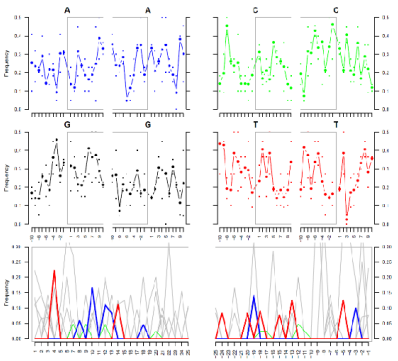
Overall, we observed damage patterns for every investigated groups and selected families for damage pattern investigation. We identified that a minimum of ~1000 reads must align to the reference genome for it to provide a good enough alignment and a robust result to conclude on the ancient or modern origin of the signal. This is a limit of this modern reference based method to investigate damage pattern as only well represented signal can be statistically confirm from ancient origin. However, in this study, whenever the alignment was possible, the sequences investigated showed damage patterns suggesting an ancient origin and not a modern contamination. From this observation and confirmation of the ancient signal from all investigated families, we assume that the metazoan signal is also from ancient origin even if not possible to prove using this technic.

In the following pages, we present in detail a selection of investigated families and a description of the observed results supporting the robustness of our signal.

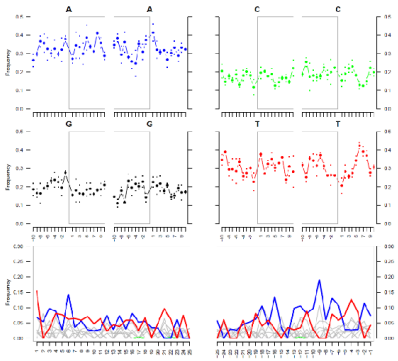
Mapdamage2 plots – Bacteria – Clostridiaceae signal

Reference used: *Clostridium estertheticum* genome - NZ_CP015756.1 (NCBI)

B17-D7



B17-D6

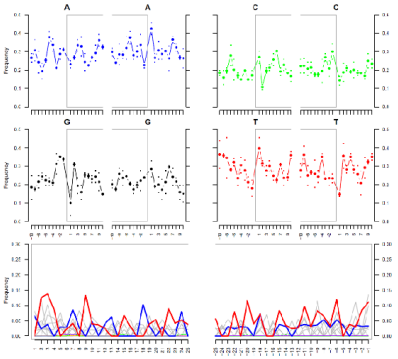


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	3906	12962	6936	204047	1976
Reads aligned	42	192	142	41675	67

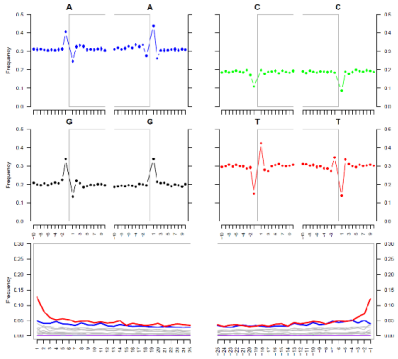
Alignment was correct for B17-D5. Alignment not perfect and bad for B17-D7, B17-D6, B17-D10 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded and/or the selected reference genome was not optimal.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference, mainly for B17-D5. The signal caught in this study for this family is probably of ancient origin.

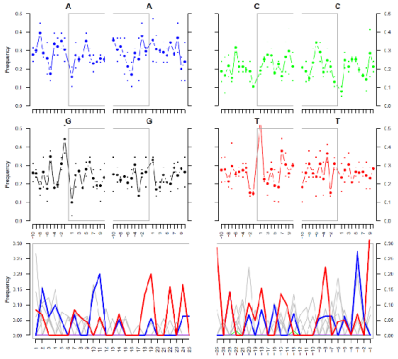
B17-D10



B17-D5



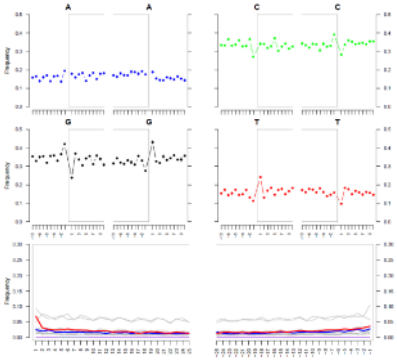
B17-D3



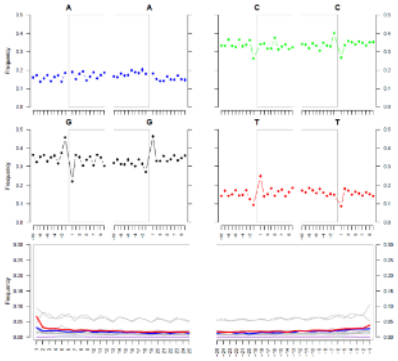
Mapdamage2 plots – Bacteria – Nocardioidaceae signal

Reference used: *Nocardioides mesophilus* genome - NZ_CP060713.1 (NCBI)

B17-D7



B17-D6

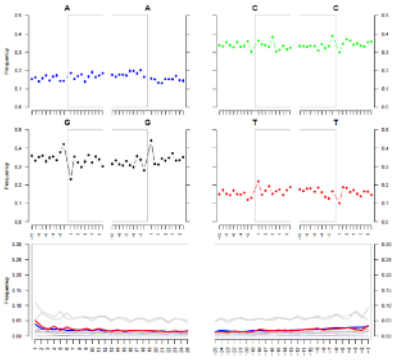


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	486086	584127	302082	1472137	653927
Reads aligned	25441	34606	16072	53615	35909

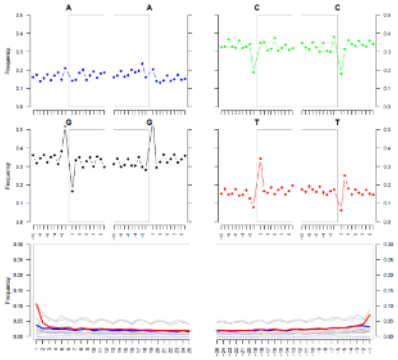
Alignment was correct for all samples.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

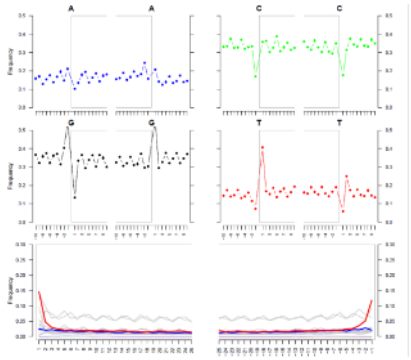
B17-D10



B17-D5



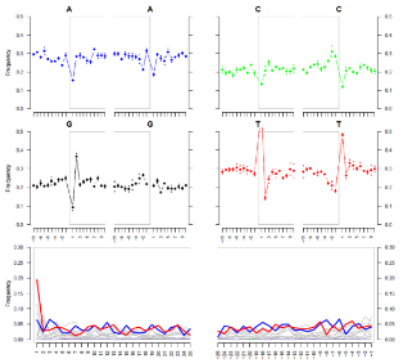
B17-D3



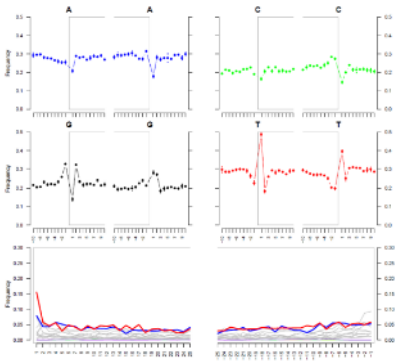
Mapdamage2 plots – Archaea – Nitrososphaeraceae signal

Reference used: *candidatus Nitrisocosmicus hydrocola* genome - NZ_CP017922 (NCBI)

B17-D7



B17-D6

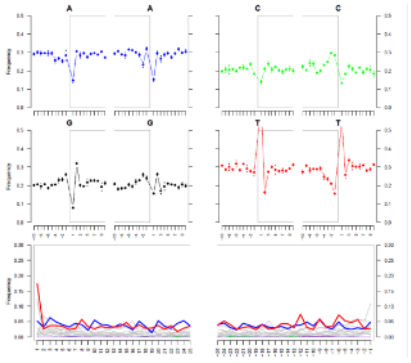


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	22779	90029	37402	2778	6110
Reads aligned	1184	4155	1861	137	297

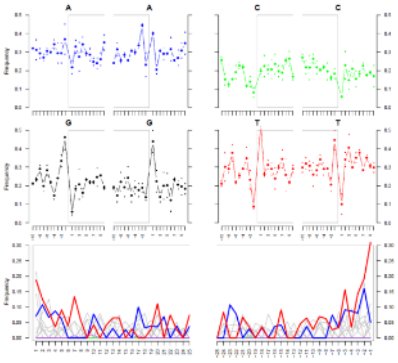
Alignment was correct for B17-D7, B17-D6 and B17-D10. Alignment not perfect and bad for B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded and/or the selected reference genome was not optimal.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference, mainly for B17-D5 and B17-D3. The signal caught in this study for this family is probably of ancient origin.

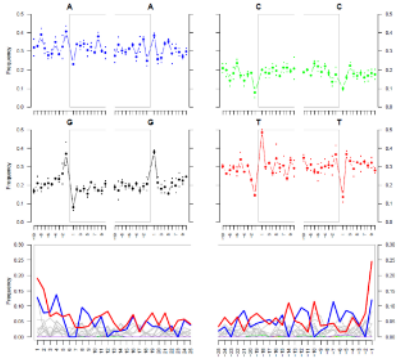
B17-D10



B17-D5



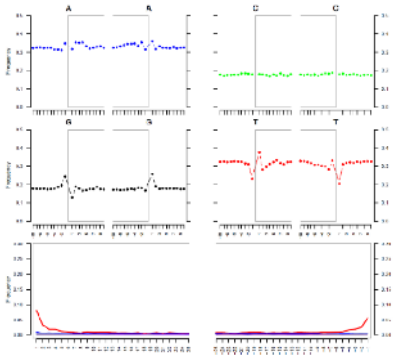
B17-D3



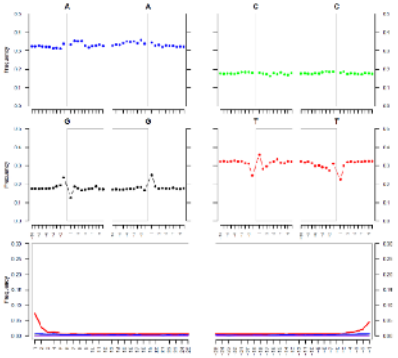
Mapdamage2 plots – Viridiplantae – Asteraceae signal

Reference used: *Artemisia gmelinii* chloroplast - NC_031399.1 (NCBI)

B17-D7



B17-D6

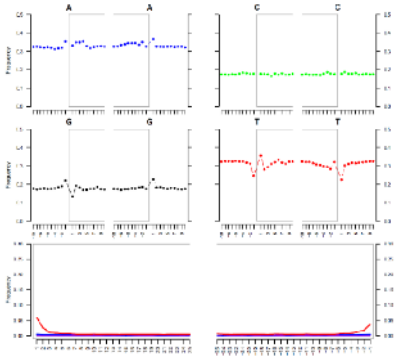


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	57332	63761	74640	237	408
Reads aligned	50222	59267	71362	222	355

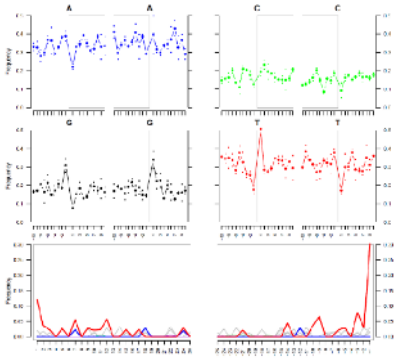
Alignment was correct for B17-D7, B17-D6 and B17-D10. Alignment not perfect for B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

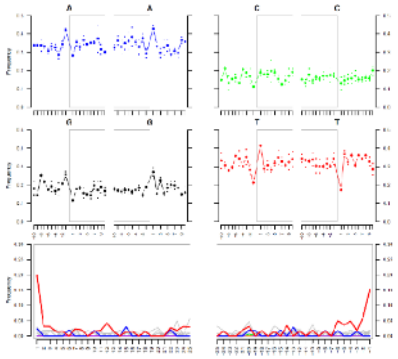
B17-D10



B17-D5



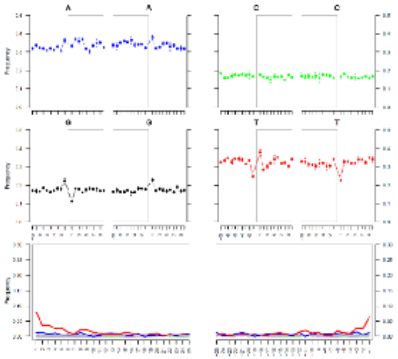
B17-D3



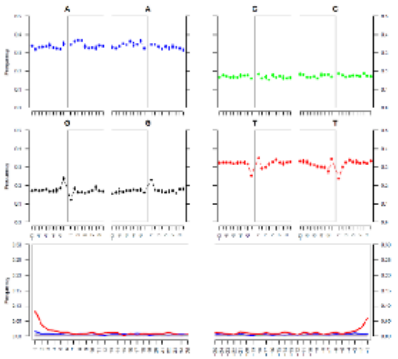
Mapdamage2 plots – Viridiplantae – Cyperaceae signal

Reference used: *Carex neurocarpa* chloroplast - NC_036037 (NCBI)

B17-D7



B17-D6

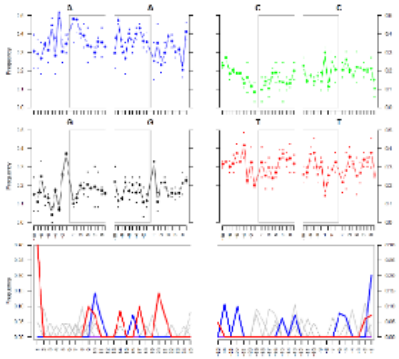


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	2276	7129	107	7	6
Reads aligned	2084	6599	70	6	3

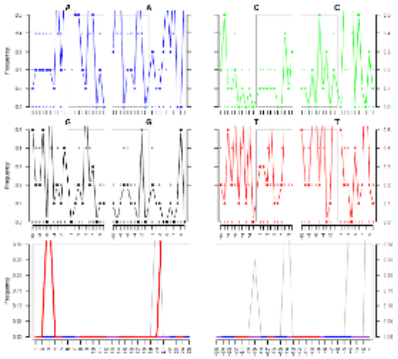
Alignment was correct for B17-D7, B17-D6 and B17-D10. Alignment not perfect and bad for B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

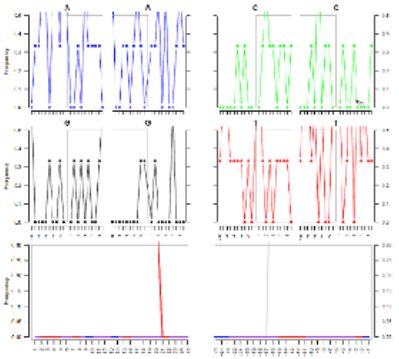
B17-D10



B17-D5



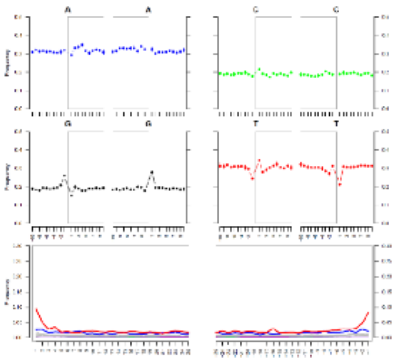
B17-D3



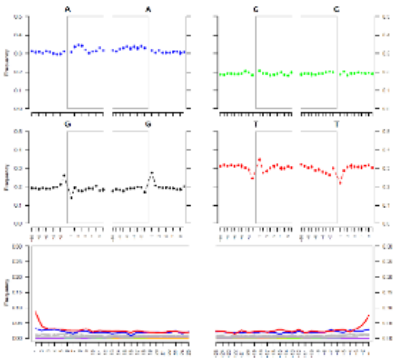
Mapdamage2 plots – Viridiplantae – Poaceae signal

Reference used: *Triticum monococcum* chloroplast - NC_021760.1 (NCBI)

B17-D7



B17-D6

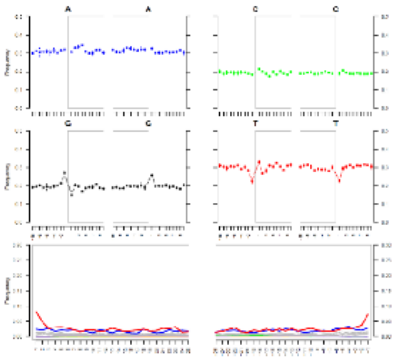


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	18440	15973	9354	42	58
Reads aligned	13391	11370	6662	33	35

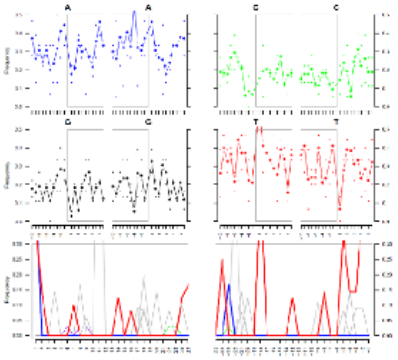
Alignment was correct for B17-D7, B17-D6 and B17-D10. Alignment not perfect and bad for B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

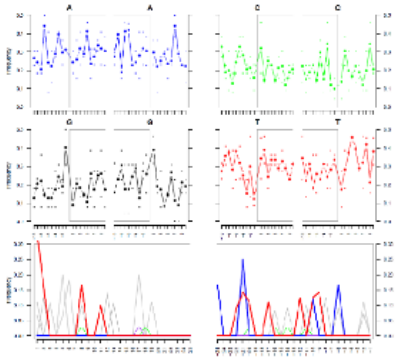
B17-D10



B17-D5



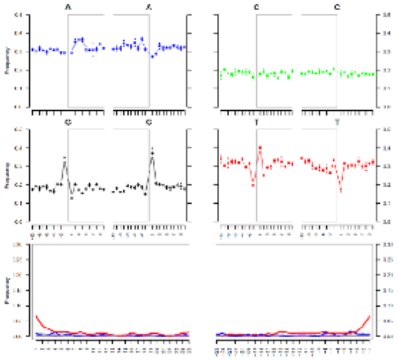
B17-D3



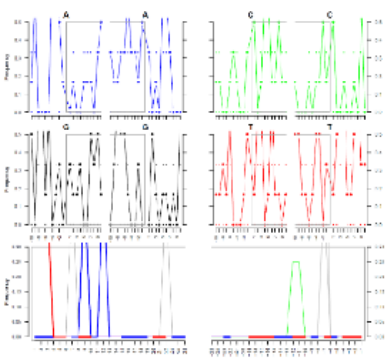
Mapdamage2 plots – Viridiplantae – Plantaginaceae signal

Reference used: *Plantago depressa* chloroplast - NC_041161.1 (NCBI)

B17-D7



B17-D6

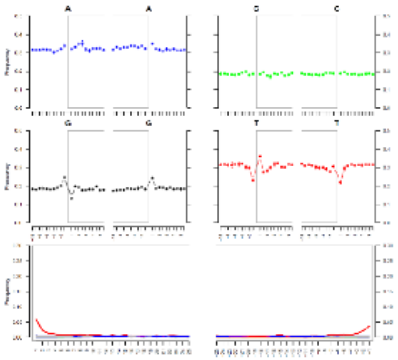


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	2269	12	18573	46	34
Reads aligned	2143	4	17459	41	33

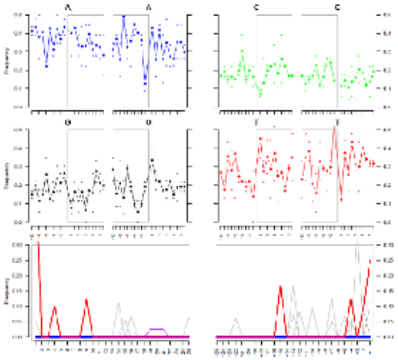
Alignment was correct for B17-D7 and B17-D10. Alignment not perfect and bad for B17-D6, B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

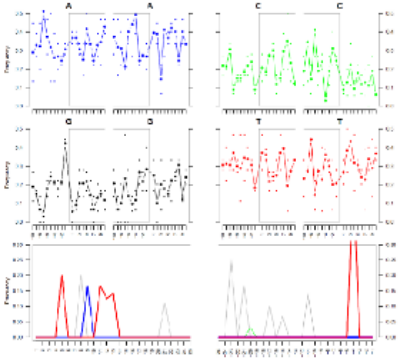
B17-D10



B17-D5



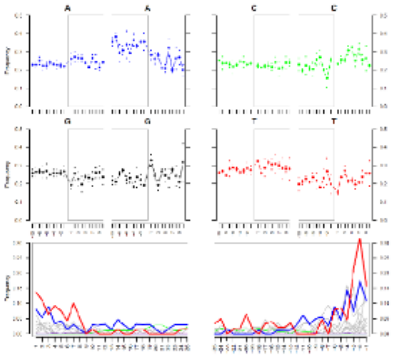
B17-D3



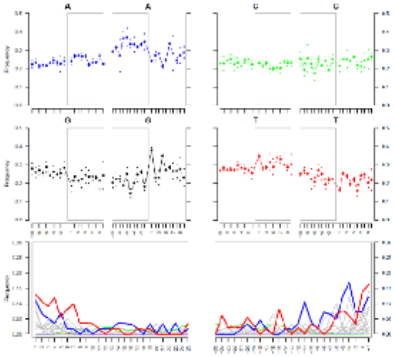
Mapdamage2 plots – Viridiplantae – Solanaceae signal

Reference used: *Nicotiana otophora* chloroplast - NC_032724.1 (NCBI)

B17-D7



B17-D6

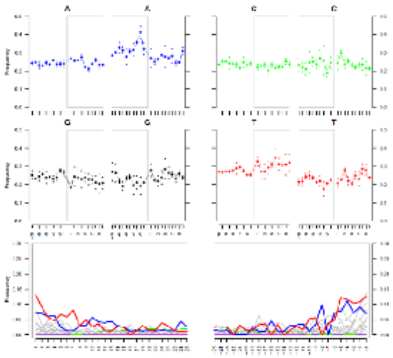


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	2355	1817	2247	1943	132
Reads aligned	276	235	313	80	12

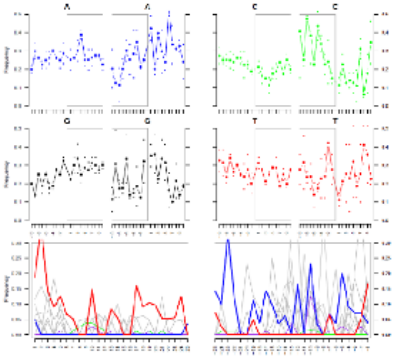
Alignment was not perfect for B17-D7, B17-D6 and B17-D10 and bad for B17-D5 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded and/or the selected reference genome was not optimal.

C to T and G to A substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference. The signal caught in this study for this family is probably of ancient origin.

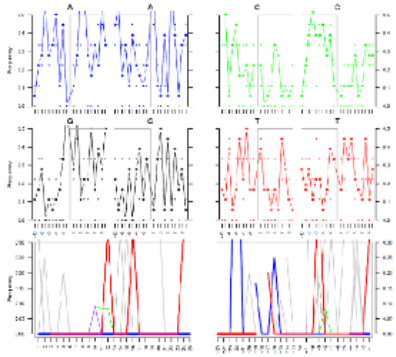
B17-D10



B17-D5



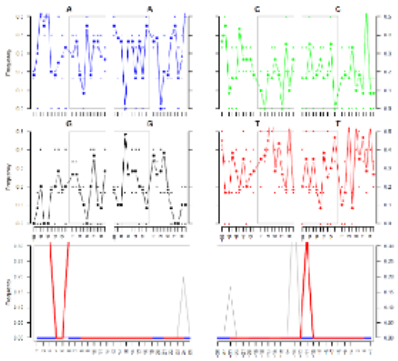
B17-D3



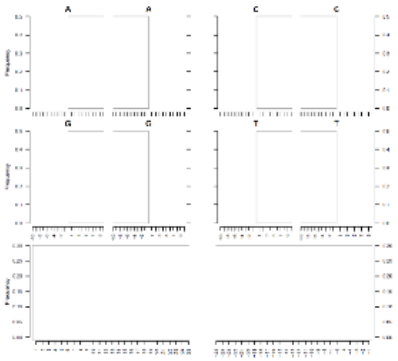
Mapdamage2 plots – Mammalia – Elephantidae signal

Reference used: *Mammuthus primmigenius* mitochondrion - NC_007596.2 (NCBI)

B17-D7



B17-D6

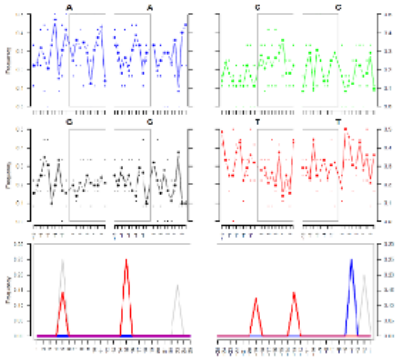


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	11	0	21	2	0
Reads aligned	11	0	21	2	0

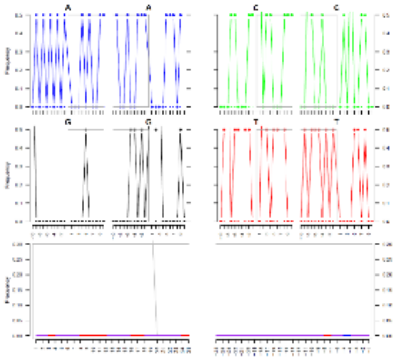
Alignment was bad for all samples as fewer DNA reads were assigned to this family in the samples.

We cannot conclude here on the presence of damage patterns. However, all reads extracted from the samples assigned to Elephantidae mapped to the reference genome chosen.

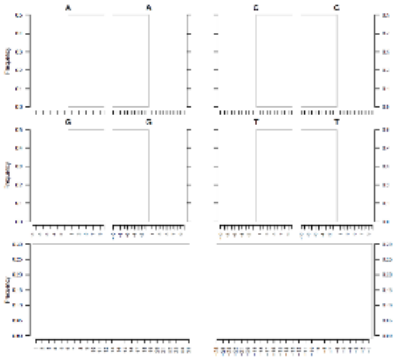
B17-D10



B17-D5



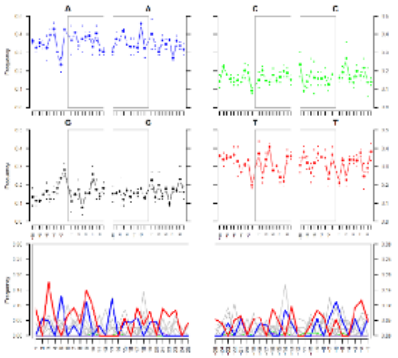
B17-D3



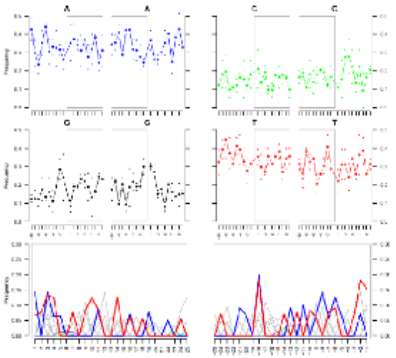
Mapdamage2 plots – Fungi – Pseudeurotiaceae signal

Reference used: *Pseudogymnoascus destructans* mitochondrion - NC_007596.2 (NCBI)

B17-D7



B17-D6

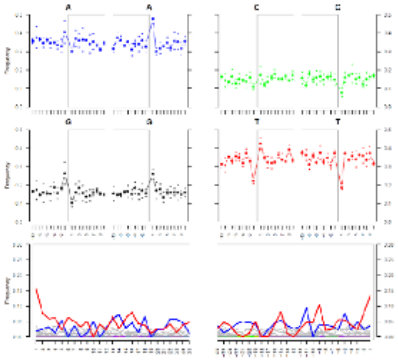


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	1346	684	1396	75916	529
Reads aligned	122	73	518	34150	41

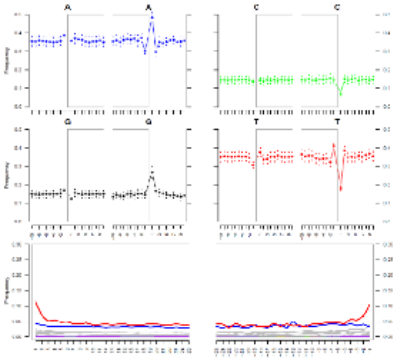
Alignment was correct for B17-D5. Alignment not perfect and bad for B17-D7, B17-D6, B17-D10 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded and/or the selected reference genome was not optimal.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference, mainly for B17-D5. The signal caught in this study for this family is probably of ancient origin.

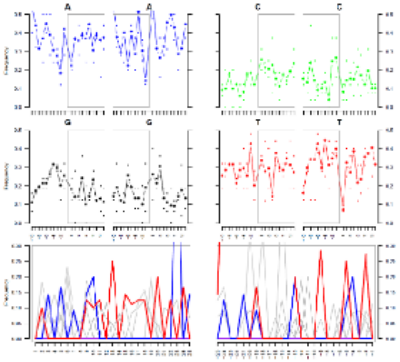
B17-D10



B17-D5



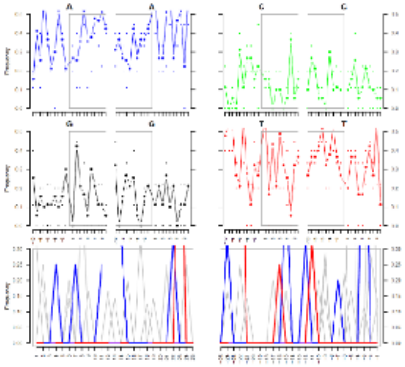
B17-D3



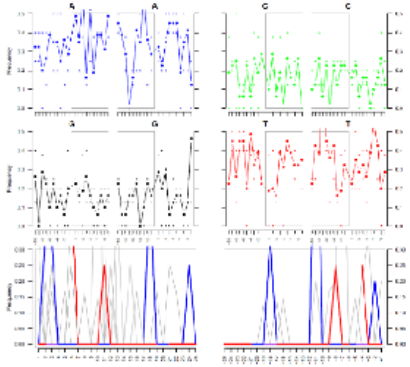
Mapdamage2 plots – Fungi – Glomerellaceae signal

Reference used: *Colletotrichum acutatum* mitochondrion - NC_027280.1 (NCBI)

B17-D7



B17-D6

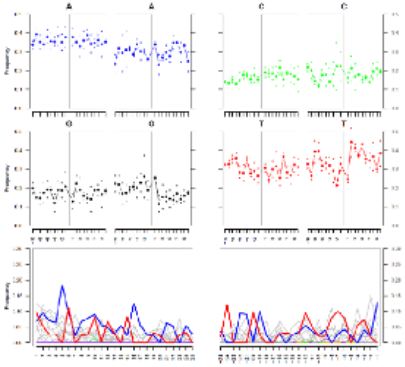


	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Reads extracted for alignment	180	117	1247	7652	2760
Reads aligned	14	13	225	783	398

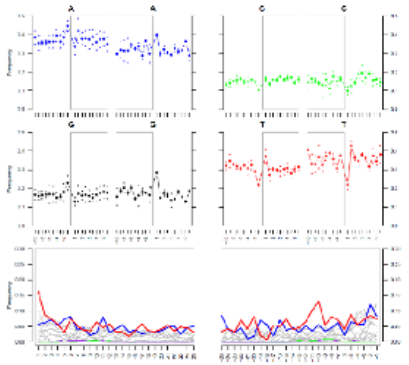
Alignment was correct for B17-D5. Alignment not perfect and bad for B17-D7, B17-D6, B17-D10 and B17-D3 as fewer DNA reads were assigned to this family in those samples and/or DNA is too degraded and/or the selected reference genome was not optimal.

C to T substitutions are visible on both extremities of the DNA fragments sequenced in our samples when compared to a modern reference, mainly for B17-D5. The signal caught in this study for this family is probably of ancient origin.

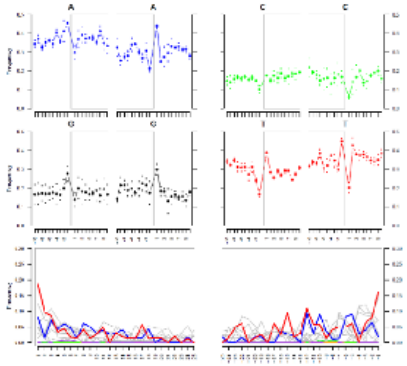
B17-D10



B17-D5

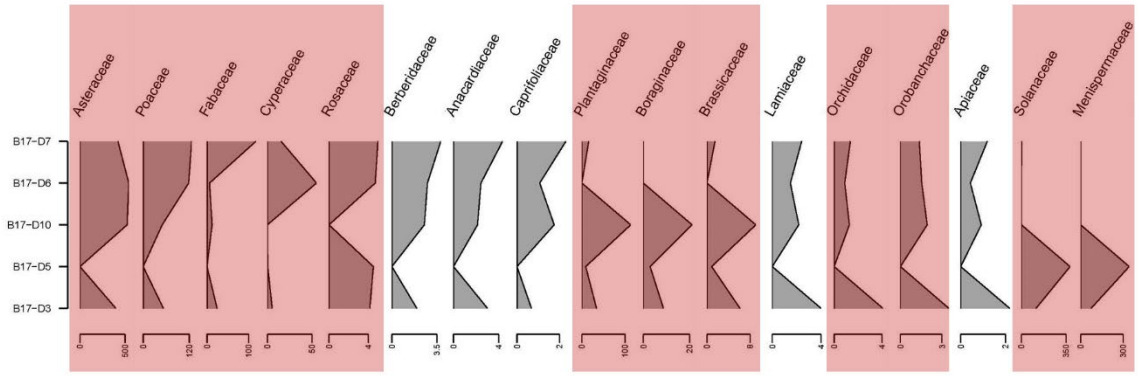


B17-D3

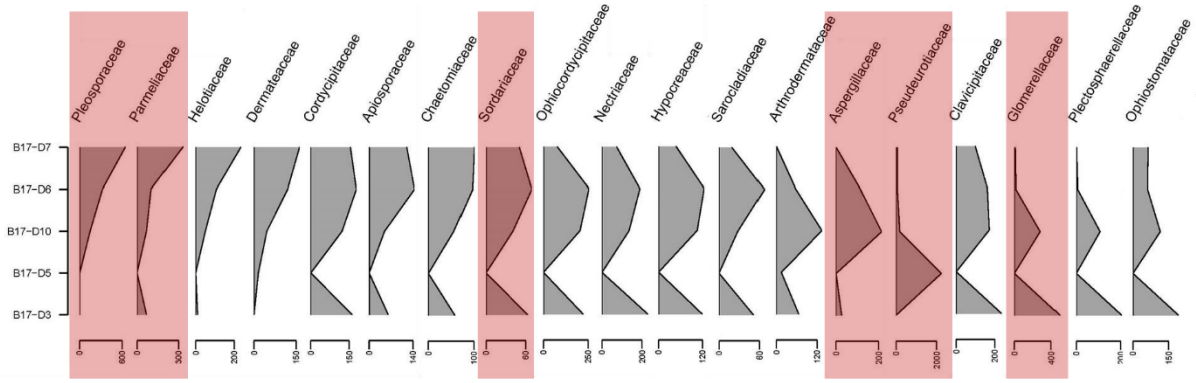


Supplementary Figure 3.3 | Positive mean centred plots. The sample with the minimum count is fixed at 0 count and the number of additional reads are plotted for each sample. Such plot is used to highlight and identify specific patterns between samples and facilitate the selection of families to investigate in m ore detail. Families are agenced from left to right according to observed pattern from younger to older sample. Families highlighted in red are plotted in the bubble chart figure (Figure 4.2) and discussed in the manuscript.

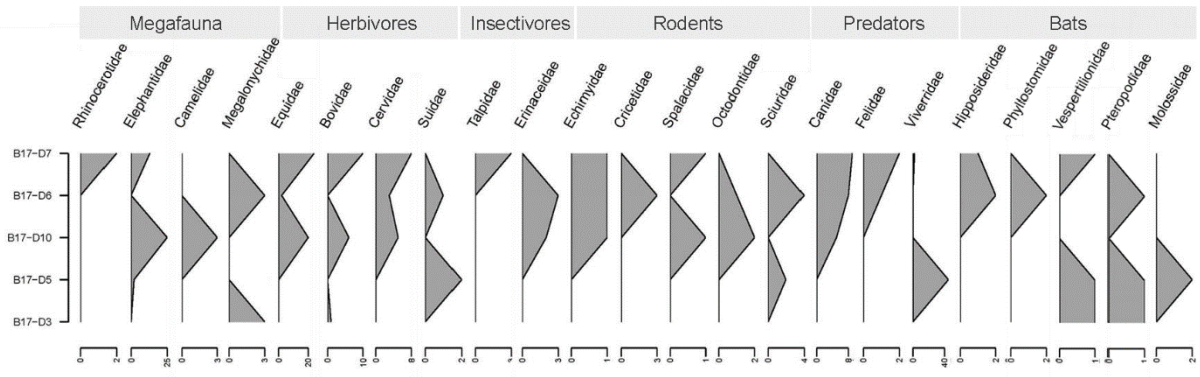
Viridiplantae:



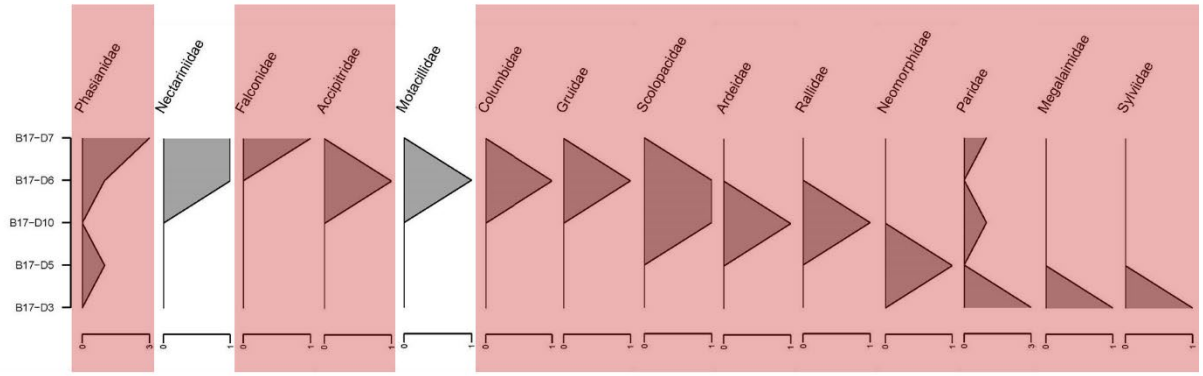
Fungi:



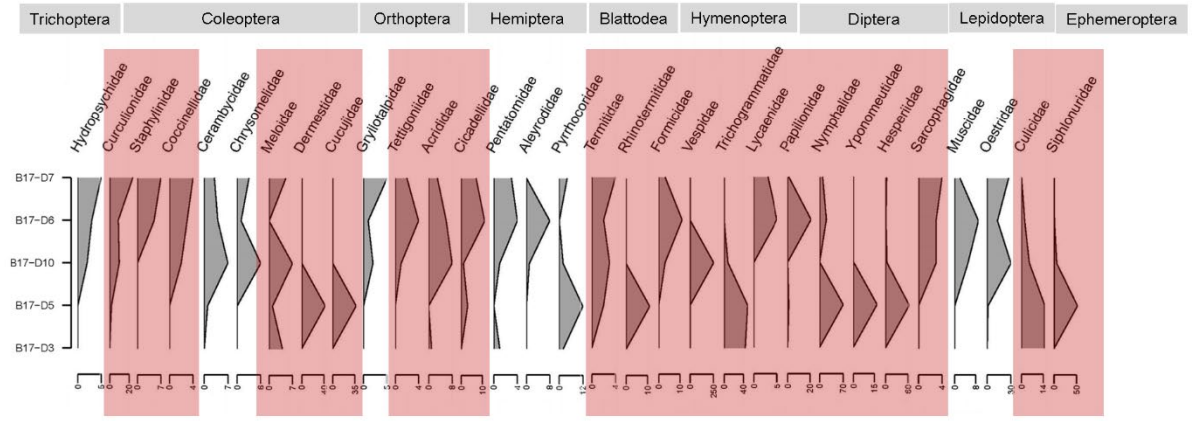
Mammalia:



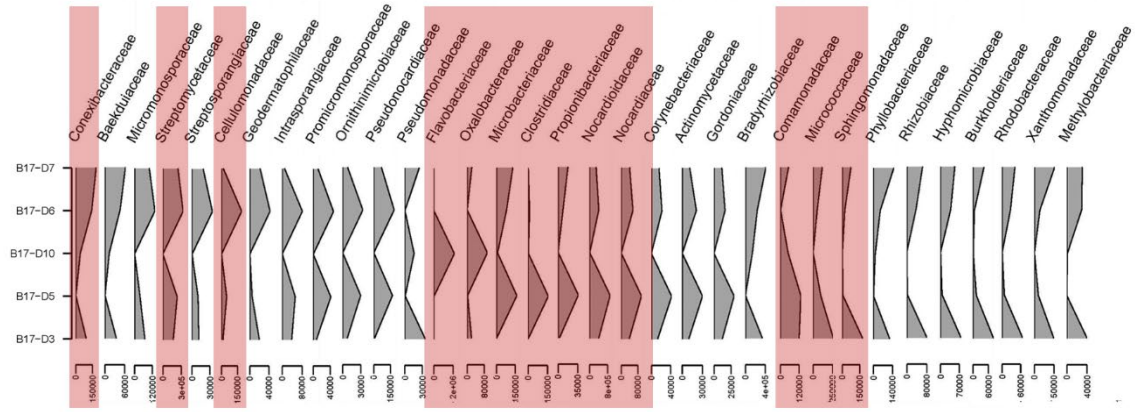
Aves:



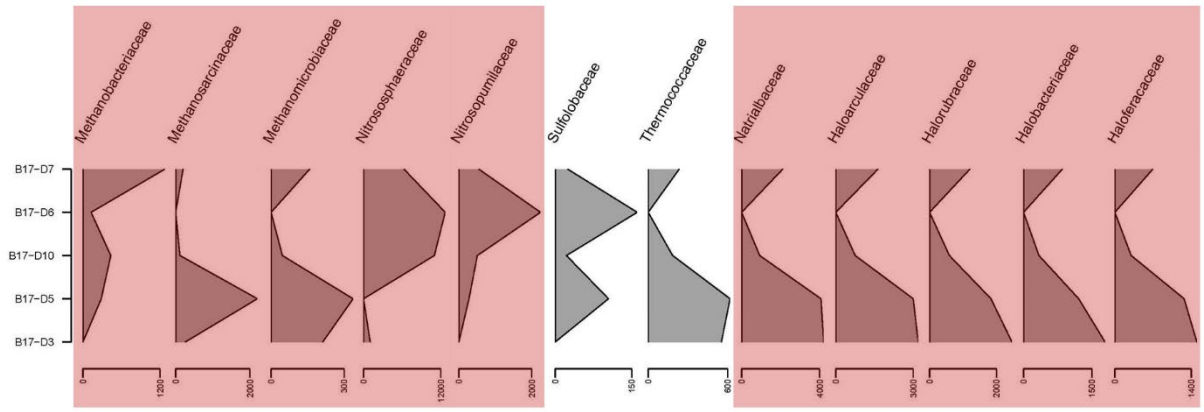
Insecta:



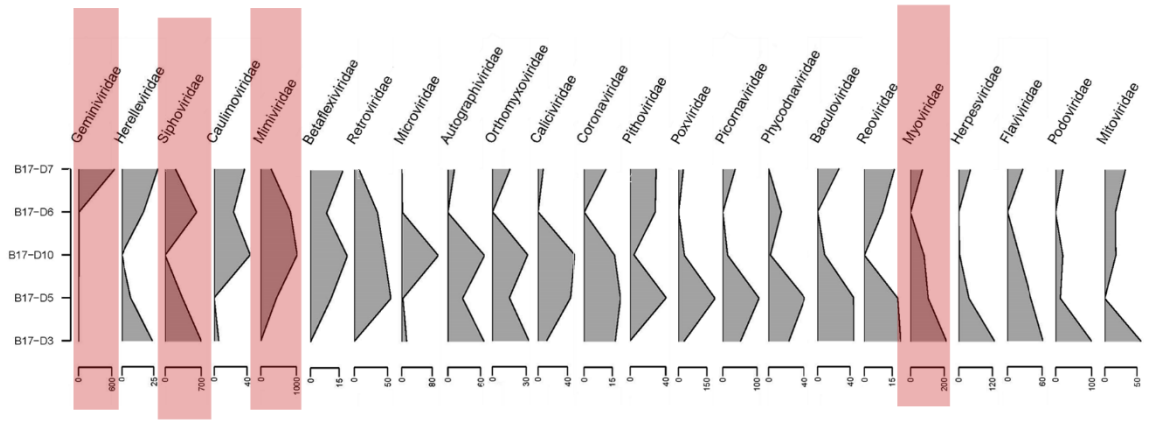
Bacteria:



Archaea:



Virus:



Supplementary Table 3.1 | Sequences assigned to plants after the OBITOOLs pipeline and filtering for the best database to use. Filtering out of blanks, filtering of sequences assigned at 100% and filtering of unlikely taxa. Sequences with less than 10 reads per samples were removed, PCR replicates were merged and sequences with similar assignment were merged. More information can be found with the published article: https://doi.org/10.1002/edn3.336

	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3	Extraction blank	PCR blank	All samples
Total	132276	114236	215265	165083	107979	1340842	0	734839
Total selected for investigation	123196	106816	207924	131344	85451	11	0	654731

Sequences excluded:

family	genus	species	Identity	count	Samples	Blanks	sequence
Poaceae	NA	NA	1.00	1344795	11411	1333384	atccgtgttttgagaaaacaaggggttctcgaactagaatacaaaggaaaag
NA	NA	NA	1.00	4477	4477	0	Atccgtgttttgagaaaaccaggaggttctcgaactagaatacaaaggaaaag
Anacardiaceae	*Schinus*	*terebinthifolia*	1.00	163	0	163	Atcctattttatgagaacaaaaacaaacaaggggtcttaacgggagaaaaag
Menyanthaceae	*Menyanthes*	*trifoliata*	1.00	60	59	1	Atcccgttttccgaaaacaaacaaaggttcagaaagcgaaaataaaaaag
Poaceae	NA	NA	1.00	18	15	3	Atccgtgttttgagaaaacaaaggggttctcgaactagaatacaaaggaaaag

Final table before resampling:

family	genus	species	total	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Asteraceae	NA	NA	421603	78769	84890	150643	51168	56133
Boraginaceae	*Eritrichium*	NA	46272	0	0	13872	28868	3532
Plantaginaceae	*Plantago*	NA	34380	1243	0	28329	3927	881
Fabaceae	*Oxytropis*	NA	24095	17764	1991	1661	2377	302
Poaceae	*Festuca*	NA	23541	6915	9558	5750	682	636
Salicaceae	NA	NA	13845	0	0	0	8204	5641
Papaveraceae	*Papaver*	NA	10756	0	0	0	7512	3244
Asteraceae	*Hulteniella*	*integrifolia*	9961	1614	2058	3987	1690	612

family	genus	species	total	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Rosaceae	*Dryas*	NA	8487	0	0	0	8266	221
Caryophyllaceae	*Eremogone*	NA	7956	5530	1806	257	310	53
Fabaceae	*Astragalus*	NA	7948	6389	958	601	0	0
Poaceae	NA	NA	7349	1589	1029	445	1503	2783
Pinaceae	*Cedrus*	NA	5876	0	0	0	0	5876
Rosaceae	*Potentilla*	NA	5263	230	408	0	3295	1330
Caryophyllaceae	*Eremogone*	*capillaris*	5243	2064	748	90	1843	498
Rosaceae	NA	NA	2856	0	0	0	2856	0
Cyperaceae	Carex	NA	1898	79	1748	0	71	0
Amaranthaceae	*Monolepis*	*asiatica*	1641	0	0	0	1641	0
Apiaceae	NA	NA	1559	0	0	0	1559	0
Orobanchaceae	*Pedicularis*	NA	1509	0	0	0	1509	0
Caryophyllaceae	*Eremogone*	*procera*	1394	407	204	0	609	174
Brassicaceae	*Smelowskia*	NA	1043	0	0	1043	0	0
Asteraceae	*Tephroseris*	NA	900	0	0	0	0	900
Polygonaceae	*Rumex*	NA	867	0	0	0	0	867
Poaceae	*Koeleria*	*asiatica*	780	96	684	0	0	0
Asteraceae	*Artemisia*	NA	727	125	444	158	0	0
Orobanchaceae	*Pedicularis*	*pacifica*	566	0	0	0	566	0
Caryophyllaceae	*Silene*	*linnaeana*	433	0	0	0	433	0
Brassicaceae	NA	NA	400	0	0	400	0	0
Asteraceae	*Artemisia*	*scopulorum*	386	64	99	174	49	0
Papaveraceae	*Papaver*	*paucistaminum*	251	0	0	0	0	251
Brassicaceae	*Alyssum*	*obovatum*	210	45	0	165	0	0
Fabaceae	*Astragalus*	*striatus*	163	0	0	0	0	163
Poaceae	*Puccinellia*	NA	161	0	0	0	161	0
Caryophyllaceae	*Silene*	NA	107	13	0	69	25	0
Orobanchaceae	*Pedicularis*	*schistostegia*	86	0	0	0	86	0
Rosaceae	*Rubus*	NA	55	0	0	0	55	0
Asteraceae	*Chersodoma*	*jodopappa*	52	0	0	0	0	52
Asteraceae	*Inula*	NA	49	0	0	0	0	49

family	genus	species	total	B17-D7	B17-D6	B17-D10	B17-D5	B17-D3
Asteraceae	*Echinops*	*onopordum*	48	0	48	0	0	0
Asteraceae	*Leontodon*	*hispidus*	42	0	0	0	0	42
Asteraceae	*Euryops*	NA	41	0	0	0	0	41
Asteraceae	*Uechtritzia*	*kokanica*	35	35	0	0	0	0
Brassicaceae	*Transberingia*	*bursifolia*	32	0	0	32	0	0
Asteraceae	*Vernonia*	*colorata*	24	0	0	0	0	24
Caryophyllaceae	*Cerastium*	*maximum*	22	0	0	0	0	22
Equisetaceae	*Equisetum*	NA	20	0	0	0	20	0
Asteraceae	*Trichocline*	*caulescens*	11	0	0	0	0	11
Brassicaceae	*Draba*	NA	11	0	0	11	0	0
Rosaceae	*Potentilla*	*tetrandra*	11	0	0	0	11	0
Asteraceae	*Artemisia*	*armeniaca*	10	0	0	10	0	0
Menyanthaceae	*Menyanthes*	*trifoliata*	10	0	0	0	0	10

Supplementary Table 3.2: Because of the large amount of data generated with the shotgun analysis, here are listed only the general statistics of the sequencing runs. More information that is detailed can be found online with the published article: https://doi.org/10.1002/edn3.336

Sequencing on Illumina HiSeq Platform (2 x 125 bp):

fastp option: -m (merge paired end) and --low-complexity-filter

library	depth (m)	before total reads	before GC content	reads passed filters	% reads passed filters (merged)	duplication rate	insert size peak	reads after filtering (merged)	% reads after filtering (max 50% - merged)	after GC content
B17-D3	49.5	20505514	53.54	20104808	98.0	2.75	38	8549328	41.69	61.23
B17-D5	47.6	59802866	51.62	58686566	98.1	3.75	50	24380101	40.77	56.04
B17-D6	2.15	59883418	47.34	58602324	97.9	4.31	58	24275292	40.54	48.91
B17-D7	2.3	50130748	50.13	49002482	97.7	3.81	48	19369453	38.64	55.20
B17-D10	25.8	46461704	46.75	45347992	97.6	4.17	57	17742113	38.19	49.21
extraction blank	/	12434560	40.99	11383322	91.5	27.64	32	42715	0.34	50.75
library blank	/	6217068	39.66	5408170	87.0	30.92	32	13715	0.22	50.70

Sequencing on Illumina Novaseq Sp Platform (2 x 100 bp):

fastp option: -m (merge paired end) and --low-complexity-filter

library	depth (m)	before total reads	before GC content	reads passed filters	% reads passed filters (merged)	duplication rate	insert size peak	reads after filtering (merged	% reads after filtering (max 50% - merged)	after GC content
B17-D3	49.5	407595512	58.68	405241460	99.4	9.88	38	158311526	38.84	62.51
B17-D5	47.6	363847752	56.15	361418234	99.3	9.17	47	131656480	36.18	57.57
B17-D6	2.15	364941924	52.33	362581890	99.4	11.31	59	131381665	36.00	51.95
B17-D7	2.3	391164750	56.15	388488524	99.3	12.59	48	135158042	34.55	57.74
B17-D10	25.8	360045840	52.09	357324688	99.2	11.94	48	120315326	33.42	52.24
extraction blank	/	8188244	58.51	8115350	99.1	5.22	32	361447	4.41	53.78
library blank	/	10796798	57.38	10612738	98.3	1.85	32	847276	7.85	55.44

Supplementary Information: Detailed Material & Methods

Isolation of sedimentary ancient DNA

DNA isolation, Polymerase chain reaction (PCR) setup and library preparations were performed in the paleogenetic laboratory of the Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research in Potsdam, Germany. This lab is dedicated to ancient DNA isolation and PCR setup and is located in a building devoid of any molecular genetic laboratories. The lab surfaces are cleaned frequently with DNA-ExitusPlusTM (AppliChem, Germany) and subjected to nightly UV-irradiation. All laboratory work was performed in a UVC/T-M-AR DNA/RNA cleaner-box (BIOSAN, Latvia). DNA isolations, PCR setups and library preparations were performed on different days using dedicated UVC/T-M-AR DNA/RNA cleaner-box (BIOSAN, Latvia), sets of pipettes and equipment. Further precautions to reduce contamination included UV-irradiation of PCR chemicals in thin-walled PCR reaction tubes, including PCR buffer, BSA, MgSO4 and DEPC-treated water, for 10 min in a UV crosslinker (recommendations of Champlot et al., 2010).

Each sample was prepared for DNA isolation and they were extracted within one extraction experiment including one extraction blank. Total DNA was isolated from between 4.5g to 10.6g of sample material using the DNeasy PowerMax Soil kit (Qiagen, Hilden, Germany). We added to 1.2 ml of C1 buffer, 0.4 ml of 2 mg/mL Proteinase K (VWR International, Germany) and 0.5 ml of 1 M dithiothreitol (DTT, VWR International, Germany) to the samples, mixed the solution for 10 minutes on a vibratory mixer (VortexGenie2, Mo Bio Laboratories, USA) and incubated overnight at 56°C in a rotating mixer. All subsequent steps were performed according to Qiagen's manufacturer's instructions. For final elution 2mL elution buffer (C6 buffer) was applied to the filter membrane, incubated for 10 minutes at room temperature and then centrifuged for 3 minutes at 2500x g. The DNA concentrations were estimated with the dsDNA BR Assay and the Qubit® 4.0 fluorometer (Invitrogen, USA) using 2 µL of the extracts. The extracts with a concentration lower than 3ng µL-1 were concentrated with Genejet PCR purification kit (Thermo Fisher, Germany). All samples were diluted to a final concentration of 3ng µL-1.

Metabarcoding approach

The PCR reactions were performed with the trnL g and h primers (Taberlet et al., 2007). Both primers were modified on the 5’ end by unique 8 bp tags which varied from each other

in at least five base pairs to distinguish sample affiliation after sequencing (Binladen et al., 2007). Additionally, 8bp tags were elongated by NNN to improve cluster detection on the sequencing platform (De Barba et al., 2014). The PCR reactions of a final volume of 25 µL contained 1.25U Platinum® Taq High Fidelity DNA Polymerase (Invitrogen, USA), 1×HiFi buffer, 2 m MgSO4, 0.25 mM mixed dNTPs, 0.8 mg Bovine Serum Albumin (VWR, Germany), 0.2 mM of each primer and 3 µL DNA (diluted to 3ng µL-1). PCRs were carried out in a Professional Basic Thermocycler (Biometra, Germany) with initial denaturation at 94°C for 5 min, followed by 50 cycles of 94°C for 30 s, 50°C for 30 s, 68°C for 30 s and a final extension at 72°C for 10 min. One no template control (NTC) was included in each PCR to trace possible contamination. For each extraction three PCR replicates with different tag combinations were performed. PCR success was validated on 2% agarose (Carl Roth GmbH & Co. KG, Germany) gels. The PCR products were purified using the MinElute PCR Purification Kit (Qiagen, Germany), following the manufacturer's recommendations. Elution was carried out twice with the elution buffer incubated 10 minutes at room temperature to a final volume of 30 µL. The DNA concentrations were estimated with the dsDNA BR Assay and the Qubit® 2.0 fluorometer (Invitrogen, USA) using 2 µL of the purified PCR products. To avoid bias based on differences in DNA concentration between samples, all replicates were pooled in equimolar concentrations. The amplification of extraction blanks and PCR NTCs were included in the sequencing run, using a standardized volume of 5 µL, even though they were negative in the PCRs. The sequencing results of extraction blanks and PCR controls are reported in the Supplement. Fasteris SA sequencing service (Switzerland) performed the paired-end sequencing on one-tenth of a full lane of the Illumina HiSeq platform (2×125 bp) with an expected output of 10 million reads.

Shotgun approach

The DNA libraries were prepared following the single-stranded DNA library preparation protocol of Gansauge et al. (2017) except that the ligation of the second adapter (CL53/CL73) took place in a rotating incubator (Schulte et al., 2020). For the first steps of library construction including dephosphorylation and adaptor ligation we used 30ng of extracted and concentrated DNA. After adapter fill and washing steps libraries were quantified with qPCR as described by Gansauge & Meyer (2013). qPCR results were used to estimate the needed PCR cycles for indexing PCR preventing amplification of libraries

into the saturating phase. Twenty-four µL of the prepared DNA library were amplified with 13 PCR cycles as described in Gansauge & Meyer (2013) using unique indexing primer combinations (P5_1 –P5_6 and P7_91 –P7_96. Fragment length distribution of libraries was controlled on with D1000 Screentape on an Agilent 4200 Tapestation system (Agilent Technologies, USA). The five libraries were pooled in equimolar ratios to a final pool of 10 nM with the blanks accounting for a molarity of 20% compared to the samples. The sequencing of the pool was performed by Fasteris SA sequencing service (Switzerland) using a modified forward sequencing primer CL72 as described in Gansauge & Meyer (2013). The sequencing results of extraction blanks are reported in the Supplement. After the first sequencing on one lane of an Illumina HiSeq 2500 platform (2 x 125 bp), we wanted to increase the sequencing depth especially for the deepest samples and sequenced a second time the pool on an Illumina Novaseq SP platform (2 x 100 bp).

Bioinformatic processing - Metabarcoding

Filtering, sorting and taxonomic assignments of the metabarcoding sequences were performed with the OBITools package (Boyer et al., 2016). The implemented function *illuminapairedend*, merges forward and reverse reads, then ngsfilter was used for demultiplexing and sample sorting. Sequences shorter than 10bp and with fewer read count than 10 were excluded with obigrep command. Duplicated sequences were merged with obiuniq while keeping the read count per within the different samples. Sequences occurring with fewer than 10 read counts were excluded from the dataset as they are probably artefacts. Further, obiclean was used to exclude sequence variants probably attributed to PCR or sequencing errors (Boyer et al., 2016). Two reference databases were used for taxonomic assignments (Epp et al., 2015). The first one, ArctBorBryo, is based on the quality-checked and curated Arctic and Boreal vascular plant and bryophyte reference libraries (Soininen et al., 2015; Sønstebø et al., 2010, Willerslev et al., 2014). The second is based on the EMBL Nucleotide Database standard sequence release 138 (Kanz et al., 2005; http://www.ebi.acuk/embl/). The sequences were assigned to taxon names based on sequence similarity to each of the reference databases using the function ecotag.

Final taxonomic assignments were determined by selecting the best identity given by embl or ArctBorBryo database, if both assignments showed the same sequence identity, the taxonomic name was selected from ArctBorBryo database, because of its specificity to arctic and boreal vegetation. Only sequences assigned with a best identity of 100% and that

are present in at least two replicates or samples were used for this study. The sequence types assigned to a lower taxonomic level than family level were removed. Detected sequences in the extraction blanks and PCR NTCs were removed from the sample dataset, if they were present at abundance higher than 0.2% of total reads per sample (**Supplementary Table 3.2**). Sequence data of the PCR replicates were merged and sequences assigned to the same taxa were merged.

References

Binladen, J., Wiuf, C., Gilbert, M. T. P., Bunce, M., Barnett, R., Larson, G., ... & Willerslev, E. (2005). Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics*, 172(2), 733-741. https://doi.org/10.1534/genetics.105.049718

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

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*(9), e13042. https://doi.org/10.1371/journal.pone.0013042

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Molecular ecology resources, 14(2), 306-323. https://doi.org/10.1111/1755-0998.12188

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. https://doi.org/10.1016/j.quascirev.2015.03.027

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., ... & Apweiler, R. (2005). The EMBL nucleotide sequence database. Nucleic acids research, 33(suppl_1), D29-D33. https://doi.org/10.1093/nar/gki098

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

Gansauge, M. T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., ... & Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic acids research, 45(10), e79-e79. https://doi.org/10.1093/nar/gkx033

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

Sønstebø, J. H., Gielly, L., Brysting, A., Elven, R., Edwards, M., Haile, J., ... & Brochmann, C. (2010). A minimalist DNA barcoding approach for reconstructing past Arctic vegetation and climate. Molecular Ecology Resources, 10, 1009-1018.

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

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

APPENDIX 4: MANUSCRIPT IV

Plant Sedimentary Ancient DNA From Far East Russia Covering the Last 28,000 Years Reveals Different Assembly Rules in Cold and Warm Climates

Authors

Sichao Huang[1,2], Kathleen R. Stoof-Leichsenring[1], Sisi Liu[1,3], **Jérémy Courtin**[1,2], Andrei A. Andreev[1], Luidmila. A. Pestryakova[4] and Ulrike Herzschuh[1,2,3*]

Affiliation

[1]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Potsdam, Germany,

[2]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany,

[3]Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany,

[4]Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, Yakutsk, Russia

***Correspondence**

Ulrike Herzschuh, ulrike.herzschuh@awi.de.

Status

Published in *Frontiers in Ecology and Evolution*. doi: 10.3389/fevo.2021.763747

ABSTRACT

Woody plants are expanding into the Arctic in response to the warming climate. The impact on arctic plant communities is not well understood due to the limited knowledge about plant assembly rules. Records of past plant diversity over long time series are rare. Here, we applied sedimentary ancient DNA metabarcoding targeting the P6 loop of the chloroplast trnL gene to a sediment record from Lake Ilirney (central Chukotka, Far Eastern Russia) covering the last 28 thousand years. Our results show that forb-rich steppe-tundra and dwarf-shrub tundra dominated during the cold climate before 14 ka, while deciduous erect-shrub tundra was abundant during the warm period since 14 ka. Larix invasion during the late Holocene substantially lagged behind the likely warmest period between 10 and 6 ka, where the vegetation biomass could be highest. We reveal highest richness during 28–23 ka and a second richness peak during 13–9 ka, with both periods being accompanied by low relative abundance of shrubs. During the cold period before 14 ka, rich plant assemblages were phylogenetically clustered, suggesting low genetic divergence in the assemblages despite the great number of species. This probably originates from environmental filtering along with niche differentiation due to limited resources under harsh environmental conditions. In contrast, during the warmer period after 14 ka, rich plant assemblages were phylogenetically overdispersed. This results from a high number of species which were found to harbor high genetic divergence, likely originating from an erratic recruitment process in the course of warming. Some of our evidence may be of relevance for inferring future arctic plant assembly rules and diversity changes. By analogy to the past, we expect a lagged response of tree invasion. Plant richness might overshoot in the short term; in the long- term, however, the ongoing expansion of deciduous shrubs will eventually result in a phylogenetically more diverse community.

Keywords: sedimentary ancient DNA (sedaDNA), metabarcoding, phylogenetic and taxonomic plant diversity, Arctic Russia, Siberia, holocene, glacial, treeline.

INTRODUCTION

Recent warming has triggered vegetation changes in the Arctic (Wilson and Nilsson, 2009). Woody taxa have expanded their range further north (Kharuk et al., 2006; Pearson et al., 2013) albeit at a rather low rate compared to the corresponding increase in temperature (Kruse et al., 2018). Patterns of taxonomic richness are one of the indicators for biodiversity monitoring of arctic plants (Khitun et al., 2016). Richness reflects the extinction and colonization of all taxa in a community, and thus shows a more general signal compared to the response of individual taxa. Some modern observations suggest an increased richness in the arctic tundra related to tree expansion under a warming climate (Rupp et al., 2001; Danby et al., 2011), whereas other studies suggest a decline (Walker et al., 2006).

The increase in richness could result from ecological facilitation, which brings mutual beneficial interactions between trees or supporting the establishment of other plants (Carcaillet et al., 2012), whereas the decline of richness would result from competition or inhibition from trees to herbs and forbs (Mod et al., 2016). Patterns of phylogenetic diversity are of interest as they can indicate changes in assembly rules under warmer and colder climate conditions. Low phylogenetic diversity reflects the co-occurrence of genetically close species. In contrast, a community with high phylogenetic diversity hosts an assembly of genetically distant spe cies (Webb, 2000; Webb and Pitman, 2002). On the one hand, phylogenetically more varied plant species coexist when more soil nutrients are available as a result of warming (Chu and Grogan, 2010; Zhu et al., 2020). On the other hand, climate warming may decrease phylogenetic diversity (Thuiller et al., 2011) through environmental filtering due to drought (Dorji et al., 2012). Accordingly, the major rules of the assembly process under the ongoing warming are not well understood, which makes it difficult to predict the future vegetation composition as well as taxonomic and phylogenetic diversity.

Conventional studies of vegetation history are mostly based on pollen analysis (Moser and MacDonald, 1990; Anderson and Brubaker, 1994; Seppä et al., 2002; Andreev et al., 2011). Many Eurasian pollen records show lower palynological richness during the cold and dry Late Glacial (transition from the Last Glacial to the Holocene, approximately 14–10 ka) compared with the warm and wet Holocene (Feurdean et al., 2011; Chytrý et al., 2017). However, some studies indicate a relatively high pollen-taxa diversity during the Late Glacial (Blarquez et al., 2014; Reitalu et al., 2015) and a decreasing diversity during the

Holocene (Blarquez et al., 2014). Sedimentary pollen records from north-eastern Europe show that climate warming led to an increase in overall plant richness, but a decrease in phylogenetic diversity during the post-glacial period (Reitalu et al., 2015). Our knowledge about past changes in phylogenetic diversity is still very scarce, especially for Siberia, and thus our understanding of plant assembly rules in terms of warming is insufficient.

Knowledge gaps about millennial-scale composition and diversity changes may, to some extent, arise from the limited taxonomic resolution and taphonomic peculiarities of pollen records. For example, *Larix*, the treeline-forming genus in most of Siberia is poorly reflected in pollen records (Sjögren et al., 2008; Binney et al., 2011; Niemeyer et al., 2017). Another restriction of pollen records is that they reflect regional vegetation signals instead of local signals. Pollen records from a European subalpine lake reported increased richness in vegetation which disperses over long distances, while macroremain assemblages analyzed from the same core displayed decreased richness, probably resulting from the reduced tree-cover (Blarquez et al., 2013). Plant DNA preserved in lake sediments can be a reliable proxy for vegetation composition and diversity (Bálint et al., 2018). Modern studies have confirmed that sedimentary DNA from high-latitude lakes reliably reflect the vegetation surrounding the lake (Niemeyer et al., 2017; Alsos et al., 2018) and can be used to trace temporal changes in plant diversity (Zimmermann et al., 2017; Parducci et al., 2019; Liu et al., 2020; Rijal et al., 2021). The universal, plant-specific and short barcode marker of the P6 loop of the chloroplast is used for sedimentary ancient DNA (sedaDNA) studies (Taberlet et al., 2007).

Plant DNA metabarcoding has revealed that Siberian plant diversity was highest during pre-Last Glacial Maximum (pre- LGM, 50–25 ka) and lowest during the Last Glacial Maximum (LGM, 25–15 ka) (Willerslev et al., 2014; Epp et al., 2018; Clarke et al., 2019). This was speculatively related to changes in forb abundance and the impact of megafauna (Willerslev et al., 2014). However, Clarke et al. (2019) hypothesize that the growth of shrubs during the Late Glacial led to competition between arctic taxa, resulting in reduced taxon diversity. Our study investigates the plant assembly rules by assessing changes in vegetation composition as well as taxonomic and phylogenetic diversity as inferred from sedaDNA from Lake Ilirney (Chukotka, Far East Russia) since 28 ka. We aim (1) to reconstruct vegetation composition changes; (2) to investigate changes in taxonomic alpha diversity and its relationship with vegetation compositional changes, particularly during

woody taxa invasion; and (3) to explore changes in phylogenetic diversity and how this reflects vegetation composition and taxonomic diversity.

MATERIALS AND METHODS

Study Area

Lake Ilirney (67°21′N, 168°19′E, 421 m a.s.l.) is a glacially formed lake located in the autonomous region of Chukotka (Far East Russia, **Figure 1**). The lake is bounded by the Anadyr Mountains (up to 1,790 m a.s.l.) to the north. According to the meteorological station at Ilirney, mean annual temperature is -13.5°C, and mean January and July temperatures are -33.4 and 12.1°C, respectively. Annual precipitation is ~200 mm (Menne et al., 2012). As an area with unique vegetation composition that harbors many endemic species (Kuzmina et al., 2011), Chukotka is a hotspot for biodiversity in the Russian Arctic. It is a key region for understanding both Quaternary glaciation history and transcontinental species migrations as it lies at the far eastern edge of Asia bordering the Bering Sea (Kuzmina et al., 2011), which is known for being a glacial refugium during the LGM (Brubaker et al., 2005). Chukotka is characterized by a comparatively pristine environment including wild herbivore mammals which have survived human impact.

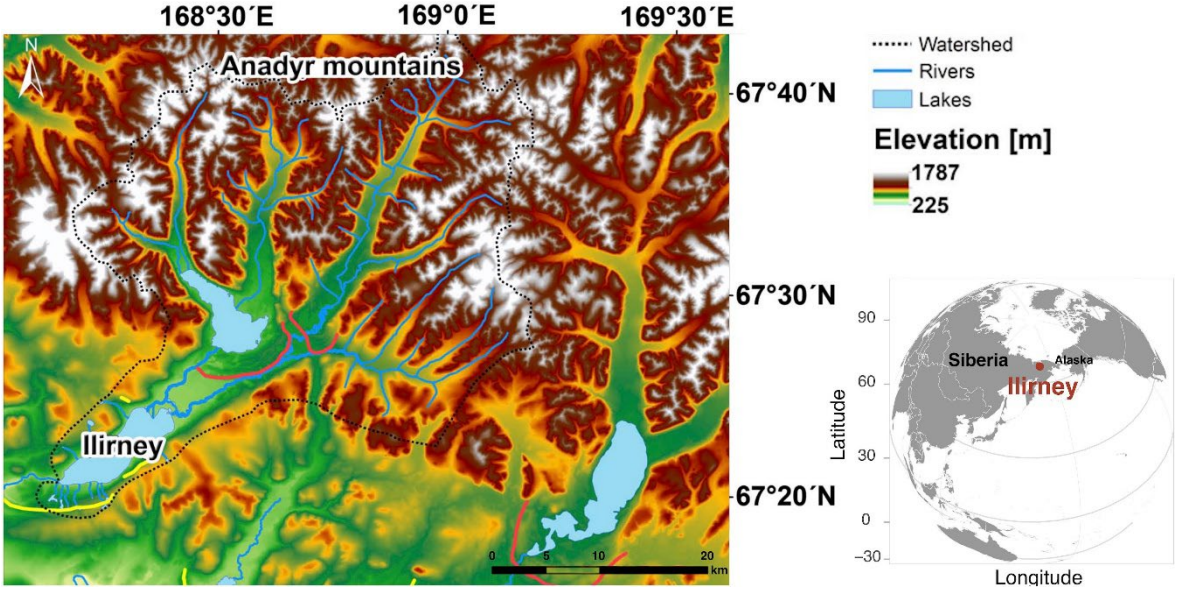


Figure 1 | Overview of Ilirney area in the Russian Arctic and the study location. The figure is adapted from Vyse et al. (2020).

Lake Ilirney is situated at the tundra-taiga ecotone, which is characterized by a mixture of forest tundra, shrub tundra, and open tundra. Stands of *Larix cajanderi* are common in the lake vicinity as well as low and dwarf shrubs such as *Salix* spp., *Betula nana*, *Ledum palustre*, *Vaccinium* spp., and *Empetrum nigrum*. Elevational transitions between forest tundra and open tundra are often covered by patchy dwarf pine (*Pinus pumila*). Open tundra on the gentle slopes (540–700 m a.s.l) is graminoid-rich including hummock (*Eriophorum vaginatum*) and non-hummock (*Carex* spp.) tundra with dense moss cover in the ground layer. Intermediate elevations (700–900 m a.s.l.) are dominated by *Dryas octopetala* and lichens, accompanied by herbs from the Fabaceae, Orobanchaceae, Poaceae, Rosaceae, and Saxifragaceae. The vegetation at higher elevations becomes barren.

Fieldwork and Sampling

Sediment core "16-KP-01-L02" was obtained from Lake Ilirney (67.34148, 168.30443) in summer 2016. The maximum water depth is 44 m in the southwestern basin and 25 m in the northeastern basin (Vyse et al., 2020). The coring was accomplished using a modified UWITEC gravity coring device deployed from a catamaran. We used a cable winch fixed on the catamaran to support a precise position of the coring (Andreev et al., 2021). The 235 cm-long sediment core was cut into 1-m- long pieces and transferred to the Alfred Wegener Institute for Polar and Marine Research (AWI) in Potsdam, Germany where it was stored at 4◦C prior to sub-sampling.

The core segments were cut into two halves. One was stored as an archive at 4◦C and the other was opened for subsampling. The core subsampling was performed in the climate chamber of the Helmholtz Centre Potsdam - German Research Centre for Geosciences (GFZ). The subsampling environment was under strict hygienic rules to prevent contamination with modern DNA. The surface in the climate chamber was cleaned with DNA Exitus PlusTM (VWR, Germany) and purified water. All sampling tools were cleaned regularly when taking each sample (Champlot et al., 2010). Sediment at either end of each core section was removed as they were in contact with the plastic tube and considered unsterile (Parducci et al., 2017). DNA samples were taken every 2 cm with 5 mL disposable syringes, in which the anterior caps were previously cut off with sterile scalpels. In total, 58 DNA samples were collected into 8 mL sterile tubes (Sarstedt) across the entire core. The collected samples were then stored at −20◦C.

Dating and Chronology

The age-depth model of core 16-KP-01-L02 has been published in Andreev et al. (2021). It is based on Accelerator Mass Spectrometry (AMS) radiocarbon dating of seven bulk total organic carbon samples (**Supplementary Table 1**). The samples were 14C dated in Mini Carbon Dating System (MICADAS) laboratory at AWI Bremerhaven (Germany). Alternative dating techniques were applied to constrain the old carbon reservoir effect (Andreev et al., 2021). The ages were modeled using the package "Rbacon" (Blaauw and Christen, 2011) in R 2.4.1[1] using the IntCal13 calibration curve (Reimer et al., 2013). To have a robust chronology of Lake Ilirney, the age model of our core was compared with the age model of another sediment core from the same lake. This core is named EN18208 and its age model has been published by Vyse et al. (2020) with a higher resolution. The two cores have a distance of approximately 250 m and have highly homogenous sediment patterns based on acoustic profiles (Andreev et al., 2021). Thus the modeled calibrated ages from core EN18208 were transferred to core 16- KP-01-L02 (Vyse et al., 2020). The age-depth plot, calibrated and modeled ages, and correlation between two cores based on the data derived from X-ray fluorescence analysis are presented in **Supplementary Figure 1**.

Genetic Laboratory Works

DNA isolation and polymerase chain reaction (PCR) setup were performed in the paleogenetic laboratory of AWI in Potsdam, Germany. This lab is dedicated to ancient DNA environment and is located in a building devoid of any modern molecular genetic work. For hygiene maintenance, the lab is equipped with an antechamber and separate rooms and individual UV-working station for DNA extraction and PCR set-up. Generally, the lab is cleaned by nightly UV-irradiation with ceiling UV-lamps.

DNA Extraction

The 58 samples selected for DNA extraction range from 1 to 235 cm depth. DNA isolation was performed using the DNeasy PowerMax Soil Kit (Qiagen, Germany). The DNA extraction was divided into 10 batches. To check for chemical contamination, one extraction blank was included for each extraction batch. Approximately 3 g of sediment sample, 15 mL bead solution, 1.2 mL C1 buffer, 400 µL of 2 mg/L proteinase K (VWR

[1] https://www.r-project.org

International) and 100 µL of 5 M dithiothreitol were prepared in a bead tube for each sample. The samples were incubated at 56◦C in a rocking shaker overnight. The subsequent steps were carried out according to the instructions of the manufacturer by Qiagen, and the final volume of the isolated DNA was 2 mL.

DNA concentration was measured with Qubit 4.0 fluorometer (Invitrogen, Carlsbad, CA, United States) using the Qubit dsDNA BR Assay Kit (Invitrogen, Carlsbad, CA, United States). Depending on the initial DNA concentrations we purified a volume of 600 or 1,000 µL using the GeneJET PCR Purification KIT (Thermo Scientific, Carlsbad, CA, United States) and eluted twice with 15 or 25 µL elution buffer, respectively. Then we repeated DNA concentration measurement with Qubit. Finally, DNA was diluted to 3 ng/µL. The DNA extracts and aliquots were stored at −20◦C.

Polymerase Chain Reaction

Polymerase chain reaction protocols were set up with plant- specific *g* and *h* primers of the *trn*L gene (Taberlet et al., 2007). To distinguish the samples after sequencing, both forward and reverse primers were modified on the 5t end by adding eight unique randomized nucleotides (Binladen et al., 2007). The primers were elongated by three additional unidentified bases (NNN) for improving cluster detection on Illumina sequencing platforms (Barba et al., 2014). The PCR setup contained 1.25U Platinum^{R} Taq High Fidelity DNA Polymerase (Invitrogen, United States), 1 × HiFi buffer (Invitrogen, United States), 2 mM $MgSO_4$, 0.25 mM mixed dNTPs, 0.8 mg Bovine Serum Albumin (VWR, Germany) 0.2 mM of each primer and 3 µL of DNA template (with a concentration of 3 ng/µL).

All Post-PCR steps were conducted in a dedicated Post- PCR area located in a different building and strictly separated from the paleogenetic laboratories. PCRs were run in the Post-PCR area and were performed in a TProfessional Basic Thermocycler (Biometra, Germany). The PCRs were carried out in three replicates using different primer tag combinations. Twenty-one PCR reactions were performed, and 9–11 sediment samples were included in each reaction. PCR no template control (NTC) and the DNA extraction blank (blank) were run alongside each reaction. In total there were 216 PCR products including replicates, NTCs, and extraction blanks. The expected amplification products were validated by 2% agarose (Carl Roth GmbH and Co. KG, Germany) gel electrophoresis.

Purification, Pooling, and Sequencing

The PCR products were purified using the MinElute PCR Purification Kit (Qiagen, Germany), following the kit manufacturer's instructions. The purified PCR products were eluted to a final volume of 20 µL. The DNA concentrations were quantified with the Qubit R dsDNA BR Assay Kit (Invitrogen, United States) using 1 µL of the purified DNA. All samples were then pooled equimolarly to a final concentration of approximately 60 ng in 1,711.80 µL. NTCs and extractions blanks were standardized to a volume of 10 µL and were also included in the pool. Pools were sent to by Fasteris SA sequencing service (Switzerland), who prepared the DNA library following the Metafast PCR-free protocol and performed parallel high-throughput paired-end (2 x 150 bp) amplicon sequencing on the Illumina NextSeq 500 platform (Illumina Inc.). The sequencing at Fasteris was done in parallel with other sequencing projects on the same flow cell, but none of which targeted plant DNA amplifications.

Processing the Sequence Data

Sequence data were processed using the OBITools package (Boyer et al., 2016). The raw forward and reverse reads were assembled to a single file using *illuminapairedend*. Based on an exact match to the tag combinations, the sequences were assigned to samples using *ngsfilter*. Identical sequences were then deduplicated using *obiuniq*. Sequences with less than 10 read counts were removed using *obigrep* to reduce possible artifacts. As a final denoising step, *obiclean* was used to exclude further sequence variants that likely originate from PCR and sequencing errors (Boyer et al., 2016). The reference used for taxonomic assignment is based on the curated Arctic and Boreal vascular plant and bryophytes database. It contains 1,664 vascular plants and 486 bryophyte species (Willerslev et al., 2014; Soininen et al., 2015). The processed sequences were assigned using *ecotag* by searching for possible matches within the reference library.

Further Sequence Data Filtering

To have a reliable taxonomic assignment of the sequences but not to overestimate the diversity, the data were further filtered (**Supplementary Table 2**). Blanks and NTCs accounted for 1,115,522 reads (2.5% of the data before filtering). More than 80% of those reads were assigned to *Dryas* and Saliceae. They occurred in the blanks of three PCR

batches (PCR batch 01, 02, and 21 in **Supplementary Table 3**), with the reads two to three levels of magnitude lower than the reads in the samples. The affected samples are between 28 and 23 ka. However, we kept the data from the two contaminated batches, as we identified the counts in the rest of the samples from these batches that have similar patterns as their corresponding replicates. After excluding extraction blanks and NTCs, the PCR replicates were summed up. Finally, we excluded sequences that were assigned to submerged aquatics and bryophytes. To obtain normalized count data, the true counts were rarefied 100 times based on the minimum number of counts (84,343) across the samples[2]. The true and rarefied data can be found in **Supplementary Data 1**.

Statistical Analyses

Statistical analyses were computed in R 3.6.0[3]. Relative count proportions of taxa within a sample were calculated to visualize taxa abundance for all ages. We applied a double-square root transformation on the relative proportions of the sedaDNA dataset to mitigate the effect of overrepresented and rare sequences (Zimmermann et al., 2017), followed by a Chord transformation using method = "norm" in decostand (Legendre and Borcard, 2018). Analyses were performed in the "vegan" and "rioja" packages (Juggins, 2009). The transformed proportions were converted to Euclidean dissimilarity distances by vegdist. Hierarchical clustering was CONISS constrained using the function chclust. Zonation was evaluated by the broken-stick model using *bstick*. Detrended correspondence analysis (DCA) was applied to the dataset using *decorana*. Results showed that a length of the first DCA axis < 3, suggesting linear relationships among variables (ter Braak and Šmilauer, 2002). The principal component analysis (PCA) was then calculated based on covariance matrix (scale = FALSE). We used *rda* and restricted to those taxa that occur at a minimum proportion of 0.2% and in at least three samples.

Richness was calculated for each of the rarefied datasets and then averaged. However, richness can depend on how rare taxa are. It is highly sensitive to sampling effort and relative abundance (Roswell et al., 2021). Therefore, Hill's N1 and N2 (Hill, 1973) have been considered, as they both explicitly include the relative abundance in the calculation (Roswell et al., 2021). They were calculated for each rarefied dataset using the function

[2] https://github.com/StefanKruse/R_Rarefaction
[3] https://www.r-project.org

diversity in "vegan" and then averaged. The classic Shannon and Simpson indexes do not measure the same quantities (Roswell et al., 2021). They also have different scales in terms of species gain and loss (Tuomisto, 2010). One solution to overcome this is to use the diversity metric developed by Hill (1973), which calculates the exponential of Shannon's index as Hill's N1 and the inverse of Simpson's index as Hill's N2. They depend on the exponent, which can vary from counting all species equally to emphasizing the species that are most common (Roswell et al., 2021).

$$Hill's\ N1 = e^{-\sum_{i=1}^{S} p_i \ln(p_i)}$$

$$Hill's\ N2 = \frac{1}{\sum_{i=1}^{S} (p_i)^2}$$

In the above formulars for Hill's N1 and N2 (Hill, 1973), S = number of species in the sample S equals to the number of species in the sample and

$$p_i = \frac{\text{Number of individuals i}}{\text{Total number of individuals in the sample}}$$

Evenness was calculated as Hill's N1 divided by log(richness). Error bars are the minimum and maximum values of the rarefied data.

The taxonomic uniqueness and the taxa that contribute the most to the compositional differences were evaluated by local contribution to beta diversity (LCBD) and species contribution to beta diversity (SCBD). They were calculated using *beta.div* and the Euclidean method in the "adespatial" package (Dray et al., 2017) in both abundance- and occurrence-based data.

$$\text{LCBD}_i = \sum_{j=1}^{p} s_{ij} \Big/ \sum_{i=1}^{n}\sum_{j=1}^{p} s_{ij}$$

$$\text{SCBD}_j = \sum_{i=1}^{n} s_{ij} \Big/ \sum_{i=1}^{n}\sum_{j=1}^{p} s_{ij}$$

In the above formulars for LCBD and SCBD (Legendre and Cáceres, 2013), s represents an $n \times p$ rectangular matrix. We use index i for sampling units and index j for species. y_{ij} means an individual value in a data table containing the presence-absence or the abundance values of p species (column vectors $y_1, y_2, \ldots, y_p$) observed in n sampling units (row vectors $x_1, x_2, \ldots x_n$), and s_{ij} is the square of the difference between the y_{ij} value and the mean value of the corresponding jth species:

$$s_{ij} = (y_{ij} - \bar{y}_j)^2$$

The phylogenetic distances were retrieved from the implemented mega-trees of vascular plants in the "V.Phylomaker" package (Jin and Qian, 2019). Extracted genetic distances were used to calculate the phylogenetic alpha diversity and beta diversity. For visualization of phylogenetic relatedness, we calculated a phylogenetic tree including plant taxa involved in our study using the function *phylo.maker* and the "scenario 1" approach (Jin and Qian, 2019). The following phylogenetic analyses were performed in the "picante" package (Kembel et al., 2010). We supplied the phylogenetic tree with the plant proportion data using *match.phylo.comm*. The cophenetic distance was calculated using *cophenetic*. The net relatedness index (NRI) and nearest taxon index (NTI) were used to quantify the degree of phylogenetic relatedness (Webb, 2000). We created 999 null communities by randomly assembling communities using an "independent swap" algorithm. The null communities were used to compare the standardized effect sizes of mean phylogenetic distance (MPD) and mean nearest phylogenetic taxon distance (MNTD) to their observed communities implemented with *ses.mpd* and *ses.mntd*, respectively. We obtained the NRI by multiplying the estimates of MPD (mpd.obs.z) by 1 and NTI by multiplying the estimate of MNTD (mntd.obs.z) by 1. A significant and positive NTI or NRI indicates the phylogenetic clustering of the taxa occurs. In contrast, a significant and negative NTI or NRI indicates the phylogenetic overdispersion in a community (Webb, 2000).

We calculated the inter-community mean pairwise distance using the function *comdist*. The phylogenetic local contribution to beta diversity (pLCBD) was estimated using the function *LCBD.comp* in the "adespatial" package (Dray et al., 2017). A high pLCBD value corresponds to the phylogenetic uniqueness of the taxa in an assemblage compared to other assemblages. The Pearson's correlation between species composition, taxonomic, and

phylogenetic diversity was calculated using *corr.test* in the "psych" package (Revelle, 2018).

RESULTS

Overview of the Sequencing Data and Taxonomic Composition

The final dataset of 58 samples contains 21,697,725 reads and 158 amplicon sequence variants (ASVs) with 100% identify to reference taxa. Twelve ASVs are identified to the family including subfamily level, 79 ASVs are identified to genus or tribe and subtribe level, and 67 to species level. Aside from a few outliers, the number of reads per sample is of the same order of magnitude. The first, second and third quartiles of the reads are 261,465, 319,778, and 467,465, respectively. 79.5% of the total counts are from shrubs and dwarf shrubs and 20.4% are from herbaceous plants. *Picea*, *Larix*, *Populus,* and *Ulmus* are the only tree taxa in our dataset and occur with low counts.

According to the clustering and zonation results, the plant assemblage referred by aDNA can be divided into three zones (**Figure 2**). There are 19 unique taxa in Zone 1 (28–19 ka) and 21 in Zone 3 (14–0 ka) (**Supplementary Table 4**). No unique taxa are detected in Zone 2 (18–14 ka). Zone 1 features prevalent and diverse dwarf shrubs and herbaceous plants, including *Cassiope tetragona*, *Dryas* sp., *Papaver* sp., Anthemideae 1, and *Saxifraga* sp. 2. Saxifragaceae and Asteraceae are the most diverse families in Zone 1. Saliceae and Betulaceae show strong between-zone variation by increasing greatly from Zone 1 to Zone 2. There is an overall reduction of plant taxa in Zone 2. Despite harboring 76% of the taxa seen in the LGM, Zone 3 is dominated by woody taxa from the Salicaceae, Betulaceae, and Ericaceae families. Shrubs such as *Alnus alnobetula*, *Populus,* and *Ribes* start to become dominant at 11 ka, with *Alnus alnobetula* showing a high relative abundance between 9 and 6 ka. There is a noticeable rise in the diversity of dwarf shrubs in Zone 3. In total, 87% (137 taxa) are detected during 28–14 ka, of which 73% (100 taxa) persist through the expansion of the coverage of dwarf and erect shrubs during 14–0 ka, and only 56% (88 taxa) remain during the maximum expansion of shrubs and trees (9–6 ka).

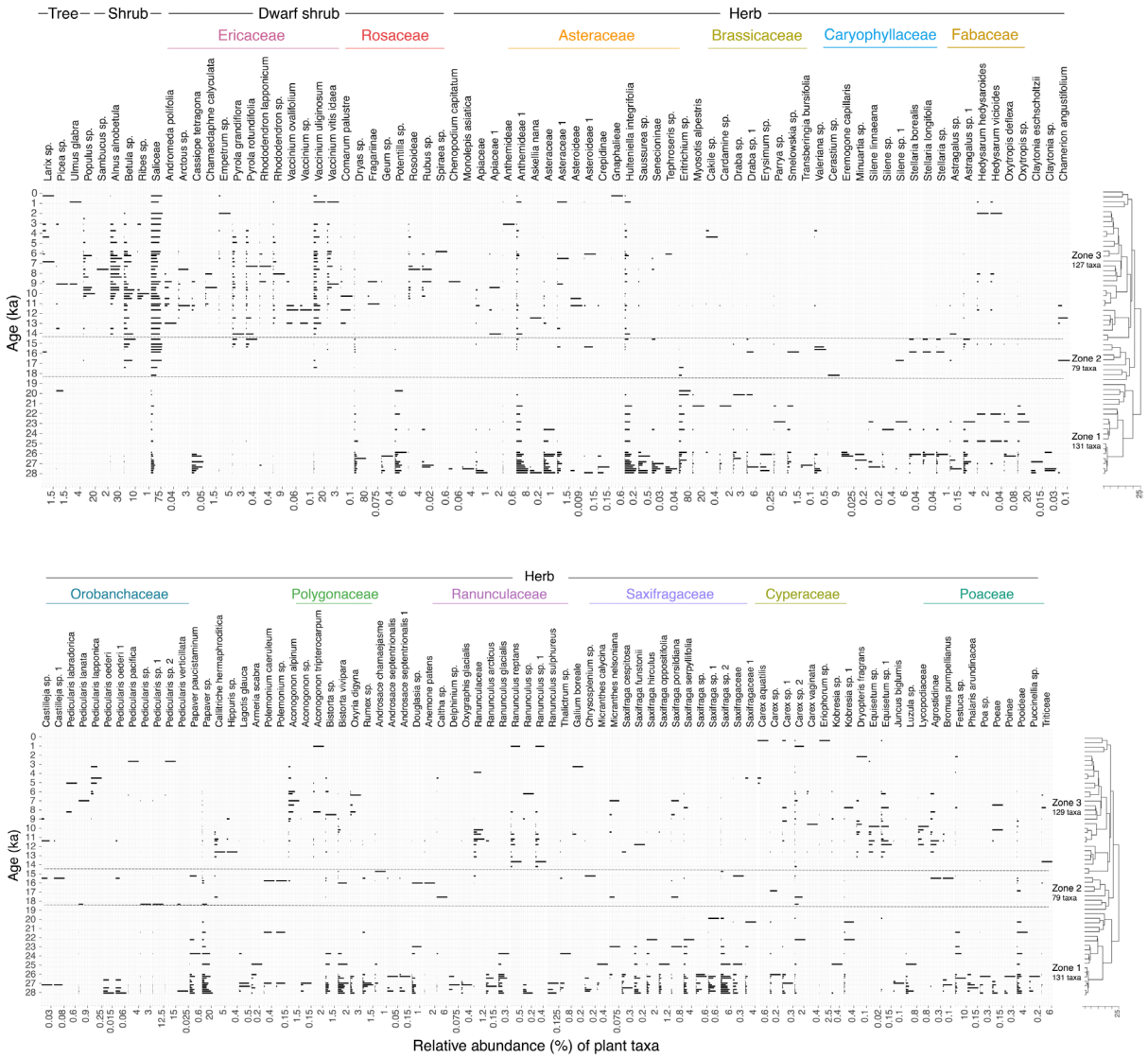


Figure 2 | Stratigram of taxa recovered from sedimentary ancient DNA samples showing the relative proportions as a percentage of the plant taxa in each sample (horizontal bars). The plot is split into the upper and lower parts. Zones are derived based on a depth-constrained CONISS cluster analysis. Numbers of plant taxa occurring in each zone are indicated.

The PCA biplot (**Figure 3**) of the first and second PC axes explain, respectively, 39.7% and 8.0% of the variance in the dataset. The ordination shows a clear separation of samples older and younger than 14 ka. Samples from 28 to 14 ka are mostly dispersed toward the positive end of the PC1 axis, with the highest loading on PC1 given by *Eritrichium* sp. and *Papaver* sp. Along the positive PC1 axis, samples since 14 ka form a cluster, characterized by high values of *Alnus alnobetula* and *Vaccinium uliginosum*.

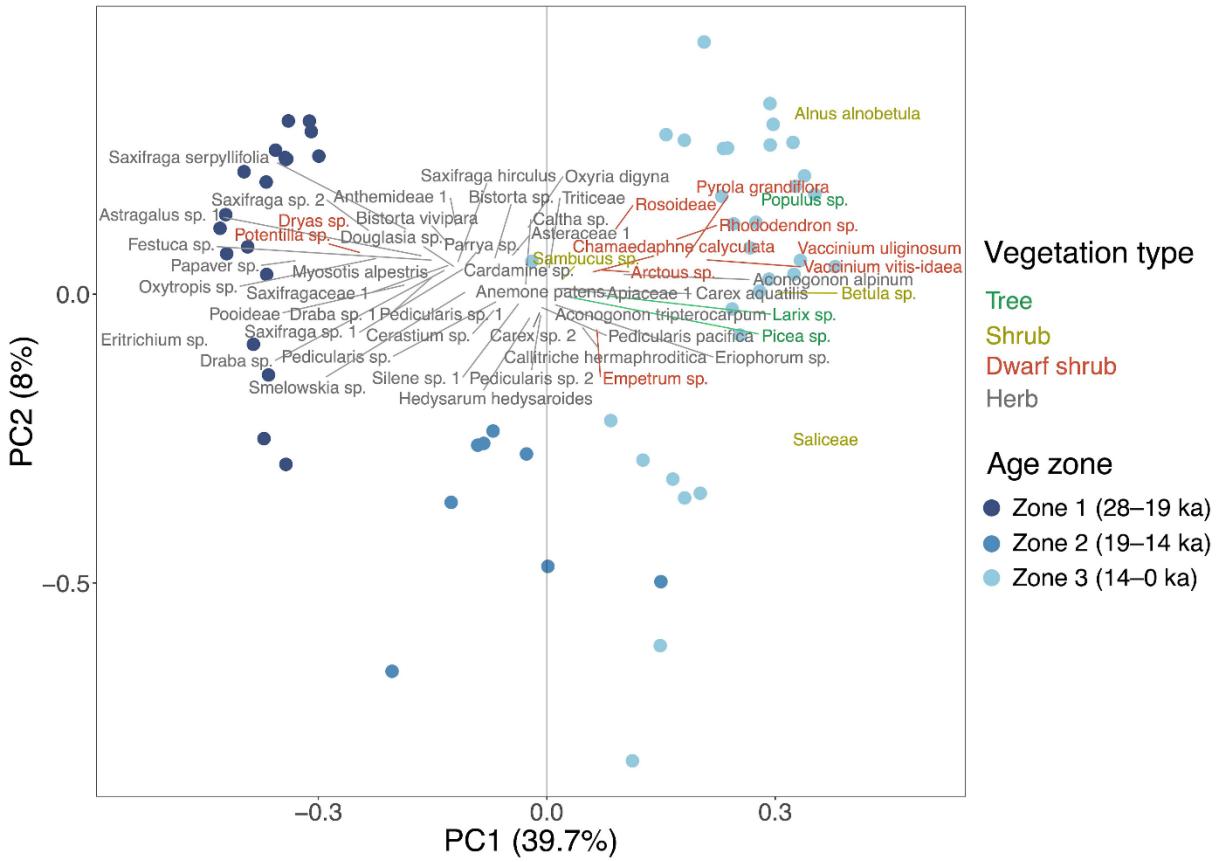


Figure 3 | Principal component analysis (PCA) biplot of plant taxa from sedimentary ancient DNA, showing a subset of the dominant taxa (maximum relative abundance > 20% and occurrence > 3) which explain most of the variance in the dataset. Sample ages are color-graded based on zones, which are estimated by a depth-constrained CONISS cluster analysis.

Taxonomic Alpha and Beta Diversity

We infer the highest species richness in the plant assemblages from 28 to 23 ka and a second richness peak during 13–9 ka (**Figure 4**). The richness of every single repetition of the experiment is similar and has a significant positive correlation to the one using summed repeats (**Supplementary Figure 3** and **Supplementary Table 5**), indicating that the difference in the PCR replicates has minor influence on the diversity in our data. Overall, we find almost no relationship between the rarefied richness and the number of reads based on the un-rarefied data (r = 0.26, p = 0.05), although the lowest read numbers during the post-LGM co-occur with the lowest richness. Similar to richness, the diversity indices and evenness have high values before 21 ka (**Figure 4**), but the second peaks of diversity and evenness lag the second richness peak by 3,000 years (9–6 ka).

The abundance-weighted LCBD (**Figure 4**) indicate that plant populations are taxonomically unique during 24–18 ka as well as from 3 to 2 ka. *Eritrichium* sp., Saliceae,

and *Alnus alnobetula* contribute most to beta diversity (**Supplementary Figure 4A**). In contrast, the occurrence-based LCBD (**Supplementary Figure 2**) indicate significant uniqueness during 28–23 and 13–9 ka, that is, indicating times of high richness.

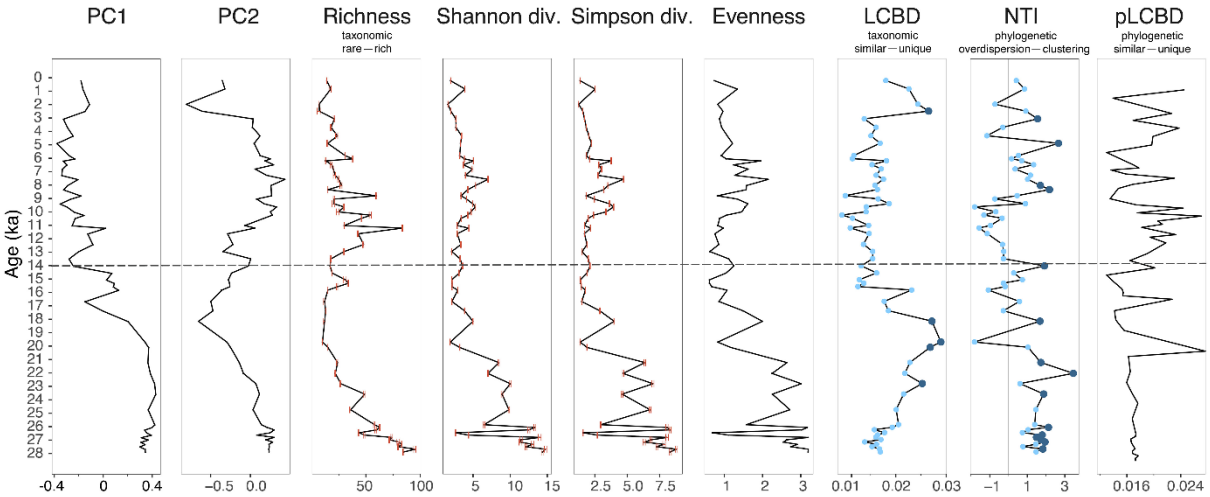


Figure 4 | Plant assemblage composition and taxonomic and phylogenetic diversity derived from sedimentary ancient DNA. The measurements from left to right are first and second axes of a principal component analysis (PC1 and PC2), taxonomic richness, Hill's N1 (exponential of Shannon diversity), Hill's N2 (inverse Simpson diversity) and evenness, abundance-weighted local contributions to beta diversity (LCBD), abundance-weighted nearest taxon index (NTI), and phylogenetic abundance-weighted local contributions to beta diversity (pLCBD). Maximum and minimum values of the 100 resampling runs are indicated as error bars in red. Significance of NTI and LCBD are colored in dark blue when the values are significant ($p < 0.05$) and light blue when the values are not significant ($p > 0.05$).

Phylogenetic Alpha and Beta Diversity

The phylogenetic tree (**Figure 5**) enables an estimation of phylogenetic diversity indices. The abundance- (**Figure 4**) and occurrence-based NTI, as well as NRI (**Supplementary Figure 2**), show overall similar patterns. Most of the NTI values are statistically significant and positive ($p < 0.05$) between 28 and 21 ka and between 9 and 5 ka in both abundance- and occurrence- based data, indicating a phylogenetic clustering in the plant assemblages during these periods. Although not significant, the NTI values are lowest and negative during 14–10 ka, suggesting a tendency of phylogenetic dispersion during that period.

Both abundance- (**Figure 4**) and occurrence-based (**Supplementary Figure 2**) pLCBD have in general higher values during 14–0 than 28–0 ka (0.016 +- 0.0027 during 28–14ka and 0.018 +- 0.0037 during 14–0 ka in abundance-based pLCBD; 0.015 +- 0.0036 during 28–14 ka and 0.019 +- 0.0058 during 14–0 ka in occurrence-based pLCBD). This suggests that the 14–0 ka period tend to contribute more strongly to the phylogenetic beta diversity and the plant assemblages have higher phylogenetic uniqueness at that time.

Relationship Between Taxonomic Composition and Phylogenetic Diversity

The correlation of indices for phylogenetic alpha diversity, taxonomic composition, and phylogenetic beta diversity of the abundance- and occurrence-based data are summarized in **Table 1**. In addition to the correlation over the last 28 k years, we report the results for the 28–14 and 14–0 ka periods.

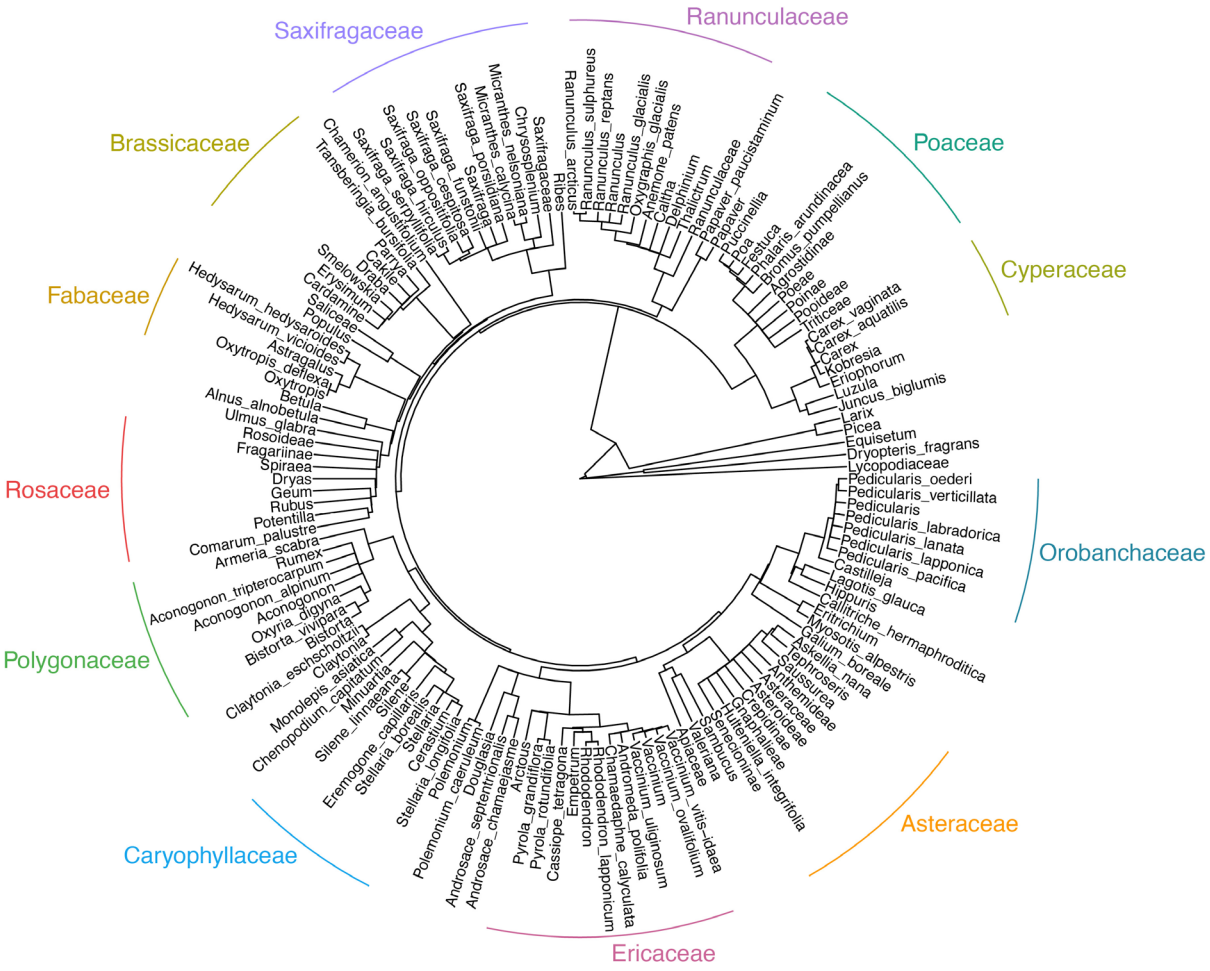


Figure 5 | Phylogenetic tree of plant taxa from sedimentary ancient DNA, demonstrating the evolutionary relationship of the presented plant taxa from subarctic Far East Russia covering the last 28 thousand years.

We merged the 28–19 and 19–14 ka zones as the cold and dry climate during 19–14 ka is more closely aligned to the climate during 28–19 ka (Andreev et al., 2021). Such a division makes the sample numbers between the two periods are more even (27 samples for 28–14 ka; 32 samples for 14–0 ka). There are 137 taxa in the records of 28–14 ka, where 40.9% (56 taxa) are identified to species level and 92% (126 taxa) are at genus including below genus levels. Similarly, there are 129 taxa in the records of 14–0 ka, where 39.5% (51 taxa)

are identified to species level and 90.7% (117 taxa) are at genus including below genus levels.

Table 1 | Pearson's correlation between plant assemblage composition and taxonomic and phylogenetic diversity in the records inferred from sedimentary ancient DNA (sedaDNA) from subarctic Far East Russia.

	28–0 ka	28–14 ka	14–0 ka
richness ~ evenness	$r = 0.54^{***}$	$r = 0.67^{**}$	$r = -0.31$
richness ~ PC1	$r = 0.52^{***}$	$r = 0.47^{*}$	$r = 0.58^{**}$
richness ~ PC2	$r = 0.52^{***}$	$r = 0.87^{***}$	$r = 0.27$
richness ~ abundance-based NTI	$r = 0.16$	$r = 0.46^{*}$	$r = -0.48^{*}$
richness ~ occurrence-based NTI	$r = 0.27^{*}$	$r = 0.49^{*}$	$r = -0.37^{*}$
richness ~ abundance-weighted LCBD	$r = -0.37^{**}$	$r = -0.83^{***}$	$r = -0.72^{***}$
richness ~ occurrence-based LCBD	$r = 0.98^{***}$	$r = 0.99^{***}$	$r = 0.96^{***}$
abundance-weighted NTI ~ abundance-weighted pLCBD	$r = -0.63^{***}$	$r = -0.45^{***}$	$r = -0.69^{***}$
occurrence-based NTI ~ occurrence-based pLCBD	$r = -0.72^{***}$	$r = -0.58^{**}$	$r = -0.70^{***}$

*Statistical significance: *p < 0.05; **p < 0.005; ***p < 0.001.*
ka = thousand years before present.

NTI values are positively correlated with richness during 28– 14 ka ($r = 0.46$, $p < 0.05$) but are negatively correlated with richness during 14–0 ka ($r = -0.48$, $p < 0.05$). This indicates that rich plant assemblages are associated with closely related species during the cool period of 28–14 ka, but with more distantly related species afterward. Species richness is negatively related to the abundance-weighted LCBD ($r = -0.37$, $p < 0.005$) for the whole record, whereas a strong positive relationship is found between species richness and occurrence-based LCBD ($r = 0.98$, $p < 0.001$). This suggests that the rich plant assemblages are taxonomically unique only in the incidence data. When the abundance-weight of the species is considered, rich assemblages are taxonomically even. Nearest taxon index is negatively associated with pLCBD in both abundance- and occurrence-based data since 28 ka ($p < 0.001$), suggesting that the phylogenetically unique assemblages are more common within genetically divergent species.

DISCUSSION

Vegetation History Revealed by Sedimentary Ancient DNA

Our results indicate that forb-dominated steppe tundra occurred from 28 to 19 ka and was followed by Saliceae-shrub tundra during 19–14 ka. Since 14 ka, Ericaceae dwarf-shrubs expanded in the area, and the lowlands of the Lake Ilirney catchment were covered by deciduous erect shrubs including *Betula* and *Alnus*. By mapping the above-ground biomass of recent and historical times in central Chukotka, regions with more erect shrubs are found with the highest above-ground biomass (Shevtsova et al., 2020, 2021). Based on this

correlation, the erect shrubs in our records likely reached a maximum density during 10–6 ka. Since the mid-Holocene, open *Larix* forest expanded, several thousand years later than the maximum vegetation density and warming.

Zone 1 (28–19 ka) is characterized by forb-dominated steppe. Forbs such as Asteraceae, Saxifragaceae, *Dryas,* and *Eritrichium* (Boraginaceae) are common in the lake's vicinity. This agrees with other ancient DNA-based studies in south- eastern Siberia (Courtin et al., 2021) and west Beringia of similar age (Willerslev et al., 2014). Our results suggest a dominant role for *Eritrichium* during the glacial period. As an alpine cushion nurse plant, *Eritrichium* may have facilitated diversity under harsh environments through interactions with other species (Cavieres and Badano, 2009). Shrubs such as *Betula*, *Alnus alnobetula*, Saliceae, and Ericaceae are absent or occur only at very low relative abundances. This contrasts with pollen records that show a relatively high percentage of shrubs from the same core (Andreev et al., 2021), which likely reflect a long- distant rather than a catchment-scale signal. As with previous sedaDNA evidence (Courtin et al., 2021), we do not find any indication of a grass-dominated landscape as suggested by pollen records (Guthrie, 2013), also because grasses such as Poaceae and Cyperaceae can be underrepresented in the *trn*L metabarcoding data (Alsos et al., 2018). Despite the evidence of mammoth steppe in other plant metabarcoding records from northeastern Siberia during Last Glacial (Courtin et al., 2021), the taxonomic composition in our data does not match the description of the mammoth steppe, which are diverse and vastly covered by grasses, forbs, and sedges (Johnson, 2009). The formation of the forb-dominated assemblages in our records was thus likely not shaped by mammals.

Zone 2 (18–14 ka) is significantly different from Zone 1 in its assemblage composition. Compared to Zone 1, forb taxa such as Asteraceae, Brassicaceae, and Cyperaceae greatly reduce their relative abundances, whereas dwarf birch and dwarf willow increase. This is probably related to climate amelioration during the Late Glacial period (Lozhkin et al., 2007). An increased representation of shrubs may also be related to a local signal of meltwater along the glacial channels (Whittaker, 1993). The presence of *Betula* in the wider region is also confirmed from 18.7 ka by pollen records from the same core, though at low abundance (Andreev et al., 2021).

Zone 3 (14–0 ka) witnesses a shift from dwarf-shrub tundra to erect-shrub tundra and open forest with an Ericaceae understory. High relative abundance of deciduous erect shrubs

(*Alnus*, *Betula,* and *Ribes*) and deciduous trees (*Populus*) occurred between 10 and 6 ka. This is also reported by other Siberian and Beringian pollen records (Szeicz and MacDonald, 2001; Velichko et al., 2002; Anderson and Lozhkin, 2015) and likely relates to increased summer warmth and moisture (Mann et al., 2002). The period from 13 to 9 ka shows slightly reduced erect shrub values which might be related to cold events (Lozhkin, 1993; Kokorowski et al., 2008). Interestingly, the late Holocene is characterized by an invasion of *Larix*, although sporadically and at low relative abundance. This implies *Larix* invaded the area after the Holocene Climatic Optimum; however, this needs further investigation.

It appears that there is an overrepresentation of certain plant taxa in our sedaDNA record, for example, Saliceae, which is not dominant in the pollen records from Ilirney (Andreev et al., 2021). Willows preferentially grow along rivers and channels. Hence they have a higher chance of their vegetative remains being transported to the center of a lake compared with upland taxa (Lozhkin et al., 2001). This invests massively in their roots and creeping stems, which increases their DNA preservation potential (Pedersen et al., 2013).

By analogy to the past, future warming will likely result in increased summer-green shrub growth for upland areas replacing forb communities, which has already been observed via remote- sensing data of the last 20 years in the catchment (Shevtsova et al., 2020). Furthermore, future changes in catchment hydrology may have a strong impact on the vegetation grow along the waters, such as willows. Our finding of delayed forest response to warming implies that tree invasion strongly lags ongoing and future warming.

Patterns of Taxonomic Alpha Diversity and Their Relationship to Assemblage Composition

The change in species richness shows that the greatest diversity of plant taxa occurred about 28 thousand years ago (**Figure 4**) and was even higher than the maximum post-LGM richness, which occurred during 13–9 ka, and not, as expected, during the time of the Holocene vegetation “optimum” (as defined by the highest PC1 values) during 10–6 ka.

The specific geographical setting of our study area in central Chukotka probably favored higher biodiversity during the pre- LGM. The north-eastern Siberia was widely believed to be largely ice-free during most of the Pleistocene (Svendsen et al., 2004; Brubaker et al.,

2005; Jakobsson et al., 2016). Meanwhile, the species-area relationship suggests a positive association between area and species richness (Connor and McCoy, 1979). Hence the enormous land area could have allowed more taxa to thrive during the pre-LGM and provided a glacial refugium for a variety of vegetation types during the LGM (Brubaker et al., 2005). We find no DNA evidence for the proposed shrub and herb vegetation during the pre-LGM, which is a presumed characteristic of the mammoth steppe (Johnson, 2009; Jørgensen et al., 2012). Hence, pre-LGM plant richness might not be related to the mammalian influence on vegetation (Sandom et al., 2014).

The LGM flora is largely a subset of pre-LGM flora. We find a decrease in richness compared to the pre- LGM, which is mainly led by the dramatic reduction of forb diversity, in particular for Asteraceae, Brassicaceae, and Cyperaceae. This probably results from the local extinctions of thermophilic species in response to the cooling into LGM (Stuart and Lister, 2007), or due to the potential breakdown of herbivore-vegetation interactions by declining mammalian populations during the LGM (Willerslev et al., 2014).

Interestingly, the pre-LGM forb diversity did not re-establish during the early Late Glacial, which suggests a delayed taxa invasion. The shift from herbaceous to woody plants during 18– 14 ka may hinder or slow the establishment of a species-rich forb assemblage via an increase in shrub height and density. But could also be due to the overrepresentation of woody plants like Salicaceae as discussed in section "Vegetation History Revealed by sedaDNA," which might out-compete with forbs, leading to the lower detection of forbs. The second richness peak between 13 and 9 ka also coincided with a slight reduction in the relative abundance of erect deciduous shrubs, while the relative abundance of dwarf shrubs (e.g., Ericaceae and Rosaceae) and herbs (e.g., Asteraceae and Ranunculaceae) increased.

Evenness, indicating a uniform distribution of taxa, and richness are positively correlated for the 28–14 ka interval in our record. In contrast, the high evenness during 10–6 ka is accompanied by relatively low richness resulting in a weak negative correlation between evenness and richness during the last 14 kyr (**Table 1**). We assume that the densification of deciduous woodlands formed by *Betula*, *Alnus alnobetula,* and *Populus* during the mid-Holocene may have resulted in a disconnect between richness and evenness by inhibiting the establishment of a diverse herb and dwarf-shrub flora, causing the dominance of a rather species-poor undergrowth.

Unlike the pollen data, the results derived from DNA metabarcoding are only semi-quantitative. The biases are usually caused by coverage and resolution of the PCR primer, by the completeness of the reference databases, and by DNA taphonomy. Several sources of bias may influence the diversity of the inferred vegetation. We used a short DNA marker in this study, which may limit the detection of taxa with relatively longer sequences, as shorter barcode sequences are more likely to be preserved and thus amplified in PCR. For example, *Carex* and *Eriophorum* have sequences longer than 80 bp (Alsos et al., 2018), causing detection to only genus or family level (Epp et al., 2015; Sjögren et al., 2016). Homopolymer versions of the species may be present within genera such as *Vaccinium*, *Pyrola*, *Ranunculus*, and *Pedicularis*. The restricted resolution of the marker may also have weakened the reliability of assessment in some families such as Asteraceae and Poaceae (Sønstebø et al., 2010). Family-specific markers such as ITS primers (Willerslev et al., 2014) are needed for reducing the biases in chloroplast primers. Further alternatives to resolve the PCR biases are to use metagenomic approaches such as shotgun sequencing or capture approaches of enriching barcode regions (Gauthier et al., 2019). Nevertheless, the current richness is very similar to the richness at genus and family levels (**Supplementary Figure 5A**). Despite potential problems such as homopolymers and marker resolution, our pattern of richness is not affected as demonstrated by the highly significant correlations between richness at the different taxonomic ranks (**Supplementary Table 6**).

Furthermore, sedaDNA diversity signals from Lake Ilirney may be impacted by changes in taphonomy (Giguet-Covex et al., 2019). For example, large fluvial input and high sedimentation rate from glacier meltwater during the Late Glacial as a result of increased erosion of organic-poor sediments may have led to a dilution of DNA concentration and thus an apparent richness decline (Pawłowska et al., 2016). While the reference database we used for taxonomic assignment contains plants from the circum- arctic (Willerslev et al., 2014), the establishment of a reference database for Beringia is therefore a future task.

Our record indicates that low relative abundance of shrubs may have enabled a rich forb-dominated steppe-tundra vegetation assemblage during the pre-LGM while the dominance of deciduous erect shrubs during 10–6 ka coincides with low richness, but by an evenly distributed assemblage. Our data imply that, at the catchment scale, the widespread expansion of shrubs in the currently warming Arctic (Chapin et al., 2005; Walker et al., 2006) may represent a regime shift for local plant diversity. By analogy to the past, the

ongoing “Arctic greening” might result in a reduction in plant richness in the long term. The dominant taxa as identified from the sedaDNA are also common in the landscape of the study area according to modern observations (Shevtsova et al., 2020).

Relationship Between Richness and Phylogenetic Alpha and Beta Diversity

Our exploration of the changes in phylogenetic diversity using plant sedaDNA data reveals that species richness and phylogenetic clustering are positively related during the dry and cold 28–14 ka period but negatively during the wet and warm 14–0 ka period, indicating that rules of species assembly differed substantially during these periods.

The statistically significant and positive NTI values for 28–14 ka suggest phylogenetic clustering (**Figure 4** and **Table 1**). Closely related taxa have more similar ecological preferences and often share similar environmental tolerances (Cavender-Bares et al., 2009; Mayfield and Levine, 2010), although some ubiquitous taxa such as *Carex* can inhabit a wide range of ecosystems. Plant assembly during the pre-LGM consists of closely related sequences assigned to forb families such as Asteraceae and Saxifragaceae (**Figure 2**). Phylogenetic clustering likely happened through a selection of climatically pre-adapted species or even adaptation to cope with severe glacial conditions (Marx et al., 2017; Kubota et al., 2018).

A positive correlation between NTI and richness indicates that rich plant assemblages were composed of closely related species during 28–14 ka. A common assumption is that genetically closely related species tend to compete more intensely than their distantly related peers, resulting in reduced coexistence (Webb, 2000; Cahill et al., 2008). However, competition also increases the opportunities for ecological facilitation through niche differentiation (Gioria and Osborne, 2014). Some closely related arctic forbs have been found to interact positively (Kolář et al., 2013). For example, species of Asteraceae can coexist with their sister clades through stabilizing niche differences (Godoy et al., 2014).

There is a tendency toward an overall increase in the phylogenetic distance between 14 and 10 ka compared with the 28–14 ka interval, indicating a shift from phylogenetic clustering before 14 ka to phylogenetic overdispersion during 14– 10 ka (**Figure 4**). Climate amelioration after 14 ka supported the invasion of many taxa in the catchment of Lake Ilirney. Species recruitment was therefore likely to be rather erratic and neither

environmental nor biological filtering was shaping the assemblage, resulting in a phylogenetically overdispersed assemblage between 14 and 10 ka. Though species arrival depends not only on climate but also on the distance to the glacial refugia (Alsos et al., 2015) and on the presence of migration barriers (Crump et al., 2019), which could result in a delayed arrival. Non- angiosperms such as conifers and ferns were infrequent before 14 ka but became more common after 14 ka. Their influence on the NTI is, however, minor (**Supplementary Figure 6** and **Supplementary Table 7**). Despite the potential problems such as homopolymers and marker resolution, the patterns of NTI are not affected as we find highly similar NTI results at both the genus and family levels (**Supplementary Figures 5B,C** and **Supplementary Table 6**).

A negative correlation between NTI and richness, as found after 14 ka, indicates a pattern that rich plant assemblages are distantly related. Following the stress gradient hypothesis (Bertness and Callaway, 1994), less stressful environments during 14–10 ka were characterized by erect-shrub tundra and open forest composed of distantly related species. These species interact positively while biotic filtering is of minor importance. Such beneficial ecological facilitation may have supported the maintenance of high productivity between 14 and 10 ka (Zhang et al., 2016). Increased NTI values during 10– 6 ka are accompanied by reduced richness. This is probably due to an unstable co-existence within closely related shrubs. After the massive expansion of shrubs and trees during the early Holocene, co-occurring shrub species are more distantly related due to competitive exclusion. Alternatively, the closely related shrubs can have negative effects on the diversity of the understory by competition for light, water, and nutrients (Barbier et al., 2008). Further experiments are needed to validate whether competition or facilitation dominates when close relatives of arctic shrubs coexist.

Our results indicate that ecologically unique plant assemblages are rich when considering only presence and absence in the data (**Figure 4** and **Table 1**). Species of intermediate abundance contribute most to the occurrence-based beta diversity (**Supplementary Figure 4B**) as they show strong abundance variations (Heino and Grönroos, 2016; Silva et al., 2018). A pattern of negative correlation between NTI and pLCBD during 28–0 ka is not surprising. Phylogenetically close species tend to share similar ecological requirements and are therefore often found in the same community (Mayfield and Levine, 2010; Kamilar et

al., 2014). These assemblages then exhibit genetic clustering (Shooner et al., 2018) and become phylogenetically even less unique.

The changes in phylogenetic diversity before and after 14 ka differ from those shown by taxonomic diversity, suggesting that inferred assembly rules can be hampered if the analyses are based solely on taxonomic or solely on phylogenetic diversity (Chai et al., 2016). Using a combination of taxonomic data and phylogenetic data gives a more complete representation of assemblage structure. With respect to ongoing warming, our results imply, that, in the short term, many species might invade the area leading to rich phylogenetic overdispersed assemblages. However, expansion of shrubs might cause a reduction in plant taxonomic richness in the long term.

DATA AVAILABILITY STATEMENT

Depths and the corresponding ages of the samples for this study can be found in Pangaea (https://doi.pangaea.de/10.1594/ PANGAEA.925767). The raw sequencing data, related scripts for analysing and final sequence dataset with relative abundances and taxonomic assignments, can be found in Dryad (https://doi.org/ 10.5061/dryad.sbcc2fr4k).

AUTHOR CONTRIBUTIONS

SH: conceptualization (lead), formal analysis (lead), and writing – original draft and review and editing (lead). KS- L: conceptualization (supporting), formal analysis (supporting), resources (supporting), and writing – review and editing (supporting). SL, JC, and AA: writing – review and editing (supporting). LP: resources (supporting) and writing – review and editing (supporting). UH: conceptualization (lead), resources (lead), writing – review and editing (lead), and supervision (lead). All authors contributed to the article and approved the submitted version.

FUNDING

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant No. 772852, GlacialLegacy to UH), the Chinese Scholarship Council (CSC) (Grant No. 201708080102 to SH), and Deutsche Forschungsgemeinschaft (DFG) (Grant No. EP 98/3-1 to UH). The expedition is part of a joint Russian-German Expedition.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021. 763747/full#supplementary-material

REFERENCES

Alsos, I. G., Ehrich, D., Eidesen, P. B., Solstad, H., Westergaard, K. B., Schönswetter, P., et al. (2015). Long-distance plant dispersal to North Atlantic islands:colonization routes and founder effect. *AoB Plants* 7:lv036. doi: 10.1093/aobpla/ plv036

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., et al. (2018). Plant DNA metabarcoding of lake sediments: how does it represent the contemporary vegetation. *PLoS One* 13:e0195403. doi: 10.1371/journal.pone. 0195403

Anderson, P. M., and Brubaker, L. B. (1994). Vegetation history of northcentral Alaska: a mapped summary of late-quaternary pollen data. *Quat. Sci. Rev.* 13, 71–92. doi: 10.1016/0277-3791(94)90125-2

Anderson, P. M., and Lozhkin, A. V. (2015). Late quaternary vegetation of chukotka (Northeast Russia), implications for Glacial and Holocene environments of Beringia. *Quat. Sci. Rev.* 107, 112–128. doi: 10.1016/j. quascirev.2014.10.016

Andreev, A. A., Raschke, E., Biskaborn, B. K., Vyse, S. A., Courtin, J., Böhmer, T., et al. (2021). Late Pleistocene to Holocene vegetation and climate changes in northwestern Chukotka (Far East Russia) deduced from lakes Ilirney and Rauchuagytgyn pollen records. *Boreas* 50, 652–670. doi: 10.1111/bor.12521

Andreev, A. A., Schirrmeister, L., Tarasov, P. E., Ganopolski, A., Brovkin, V., Siegert, C., et al. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during late Quaternary inferred from pollen records. *Quat. Sci. Rev.* 30, 2182–2199. doi: 10.1016/j.quascirev.2010.12.026

Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., et al. (2018). Environmental DNA time series in ecology. *Trends Ecol. Evol.* 33, 945–957. doi: 10.1016/j.tree.2018.09.003

Barba, M. D., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., et al. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol. Ecol. Resour.* 14, 306–323. doi: 10.1111/1755-0998.12188

Barbier, S., Gosselin, F., and Balandier, P. (2008). Influence of tree species on understory vegetation diversity and mechanisms involved—A critical review for temperate and boreal forests. *Forest Ecol. Manag.* 254, 1–15. doi: 10.1016/ j.foreco.2007.09.038

Bertness, M. D., and Callaway, R. (1994). Positive interactions in communities. *Trends Ecol. Evol.* 9, 191–193. doi: 10.1016/0169-5347(94)90088-4

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., et al. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2:e197. doi: 10.1371/journal.pone.0000197

Binney, H. A., Gething, P. W., Nield, J. M., Sugita, S., and Edwards, M. E. (2011). Tree line identification from pollen data: beyond the limit?: Tree line identification from pollen data. *J. Biogeogr.* 38, 1792–1806. doi: 10.1111/j.1365- 2699.2011.02507.x

Blaauw, M., and Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal.* 6, 457–474. doi: 10. 1214/11-ba618

Blarquez, O., Carcaillet, C., Frejaville, T., and Bergeron, Y. (2014). Disentangling the trajectories of alpha, beta and gamma plant diversity of North American boreal ecoregions since 15,500 years. *Front. Ecol. Evol.* 2:6. doi: 10.3389/fevo. 2014.00006

Blarquez, O., Finsinger, W., and Carcaillet, C. (2013). Assessing paleo-biodiversity using low proxy influx. *PLoS One* 8:e65852. doi: 10.1371/journal.pone.0065852 Boyer, F., Mercier, C., Bonin, A., Bras, Y. L., Taberlet, P., and Coissac, E. (2016). obitools: a unix−inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* 16, 176–182. doi: 10.1111/1755-0998.12428

Brubaker, L. B., Anderson, P. M., Edwards, M. E., and Lozhkin, A. V. (2005). Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *J. Biogeogr.* 32, 833–848. doi: 10.1111/j.1365-2699. 2004.01203.x

Cahill, J. F., Kembel, S. W., Lamb, E. G., and Keddy, P. A. (2008). Does phylogenetic relatedness influence the strength of competition among vascular plants? *Perspect. Plant Ecol. Evol. Syst.* 10, 41–50. doi: 10.1016/j.ppees.2007.10.001

Carcaillet, C., Hörnberg, G., and Zackrisson, O. (2012). Woody vegetation, fuel and fire track the melting of the Scandinavian ice-sheet before 9500calyrBP. *Quat. Res.* 78, 540–548. doi: 10.1016/j.yqres.2012.08.001

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., and Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715. doi: 10.1111/j.1461-0248.2009.01314.x

Cavieres, L. A., and Badano, E. I. (2009). Do facilitative interactions increase species richness at the entire community level? *J. Ecol.* 97, 1181–1191. doi: 10.1111/j. 1365-2745.2009.01579.x

Chai, Y., Yue, M., Liu, X., Guo, Y., Wang, M., Xu, J., et al. (2016). Patterns of taxonomic, phylogenetic diversity during a long-term succession of forest on the Loess Plateau, China: insights into assembly process. *Sci. Rep.* 6:27087. doi: 10.1038/srep27087

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., and Geigl, E.-M. (2010). An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5:e13042. doi: 10.1371/journal.pone.0013042

Chapin, F. S., Sturm, M., Serreze, M. C., McFadden, J. P., Key, J. R., Lloyd, A. H., et al. (2005). Role of Land-Surface changes in arctic summer warming. *Science* 310, 657–660. doi: 10.1126/science.1117368

Chu, H., and Grogan, P. (2010). Soil microbial biomass, nutrient availability and nitrogen mineralization potential among vegetation-types in a low arctic tundra landscape. *Plant Soil* 329, 411–420. doi: 10.1007/s11104-009-0167-y

Chytrý, M., Horsák, M., Syrovátka, V., Danihelka, J., Ermakov, N., German, D. A., et al. (2017). Refugial ecosystems in central Asia as indicators of biodiversity change during the Pleistocene–Holocene transition. *Ecol. Indic.* 77, 357–367. doi: 10.1016/j.ecolind.2016.12.033

Clarke, C. L., Edwards, M. E., Gielly, L., Ehrich, D., Hughes, P. D. M., Morozova, L. M., et al. (2019). Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Sci. Rep.* 9:19613. doi: 10.1038/ s41598-019-55989-9

Connor, E. F., and McCoy, E. D. (1979). The statistics and biology of the species- area relationship. *Am. Nat.* 113, 791–833. doi: 10.1086/283438

Courtin, J., Andreev, A. A., Raschke, E., Bala, S., Biskaborn, B. K., Liu, S., et al. (2021). Vegetation changes in southeastern siberia during the late pleistocene and the Holocene. *Front. Ecol. Evol.* 09:625096. doi: 10.3389/fevo.2021.625096

Crump, S. E., Miller, G. H., Power, M., Sepúlveda, J., Dildar, N., Coghlan, M., et al. (2019). Arctic shrub colonization lagged peak postglacial warmth: molecular evidence in lake sediment from Arctic Canada. *Glob. Change Biol.* 25, 4244– 4256. doi: 10.1111/gcb.14836

Danby, R. K., Koh, S., Hik, D. S., and Price, L. W. (2011). Four decades of plant community change in the Alpine Tundra of Southwest Yukon, Canada. *Ambio* 40, 660–671. doi: 10.1007/s13280-011-0172-2

Dorji, T., Totland, O., Moe, S. R., Hopping, K. A., Pan, J., and Klein, J. A. (2012). Plant functional traits mediate reproductive phenology and success in response to experimental warming and snow addition in Tibet. *Global Change Biol.* 19, 459–472. doi: 10.1111/gcb.12059

Dray, S., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., et al. (2017). *Adespatial: Multivariate Multiscale Spatial Analysis. R Package Version 0.0-9*.

Epp, L. S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., et al. (2015). Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quat. Sci. Rev.* 117, 152–163. doi: 10.1016/j.quascirev. 2015.03.027

Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., et al. (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

Feurdean, A., Ta˘ma¸s, T., Tan¸ta˘u, I., and Fa˘rca¸s, S. (2011). Elevational variation in regional vegetation responses to late-glacial climate changes in the Carpathians. *J. Biogeogr.* 39, 258–271. doi: 10.1111/j.1365-2699.2011.02605.x

Gauthier, M., Konecny−Dupré, L., Nguyen, A., Elbrecht, V., Datry, T., Douady, C., et al. (2019). Enhancing DNA metabarcoding performance and applicability with bait capture enrichment and DNA from conservative ethanol. *Mol. Ecol. Resour.* 20, 79–96. doi: 10.1111/1755-0998.13088

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., et al. (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

Gioria, M., and Osborne, B. A. (2014). Resource competition in plant invasions: emerging patterns and research needs. *Front. Plant Sci.* 5:501. doi: 10.3389/fpls. 2014.00501

Godoy, O., Kraft, N. J. B., and Levine, J. M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* 17, 836–844. doi: 10.1111/ele.12289

Guthrie, R. D. (2013). *Frozen Fauna of the Mammoth Steppe: The Story of Blue Babe*. Chicago, IL: University of Chicago Press.

Heino, J., and Grönroos, M. (2016). Exploring species and site contributions to beta diversity in stream insect assemblages. *Oecologia* 183, 151–160. doi: 10. 1007/s00442-016-3754-7

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432. doi: 10.2307/1934352

Jakobsson, M., Nilsson, J., Anderson, L., Backman, J., Björk, G., Cronin, T. M., et al. (2016). Evidence for an ice shelf covering the central Arctic Ocean during the penultimate glaciation. *Nat. Commun.* 7:10365. doi: 10.1038/ncomms10365

Jin, Y., and Qian, H. (2019). V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* 42, 1353–1359. doi: 10.1111/ ecog.04434

Johnson, C. N. (2009). Ecological consequences of Late Quaternary extinctions of megafauna. *Proc. R. Soc. B Biol. Sci.* 276, 2509–2519. doi: 10.1098/rspb.2008. 1921

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., et al. (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.

Juggins, S. (2009). *). Rioja: An R Package for the Analysis of Quaternary Science Data*. Newcastle: Newcastle University Library.

Kamilar, J. M., Beaudrot, L., and Reed, K. E. (2014). The influences of species richness and climate on the phylogenetic structure of African Haplorhine and strepsirrhine primate communities. *Int. J. Primatol.* 35, 1105–1121. doi: 10. 1007/s10764-014-9784-2

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166

Kharuk, V. I., Ranson, K. J., Im, S. T., and Naurzbaev, M. M. (2006). Forest- tundra larch forests and climatic trends. *Russ. J. Ecol.* 37, 291–298. doi: 10.1134/ s1067413606050018

Khitun, O. V., Koroleva, T. M., Chinenko, S. V., Petrovsky, V. V., Pospelova, E. B., Pospelov, I. N., et al. (2016). Applications of local floras for floristic subdivision and monitoring vascular plant diversity in the Russian Arctic. *Arct. Sci.* 2, 103–126. doi: 10.1139/as-2015-0010

Kokorowski, H. D., Anderson, P. M., Mock, C. J., and Lozhkin, A. V. (2008). A re- evaluation and spatial analysis of evidence for a Younger Dryas climatic reversal in Beringia. *Quat. Sci. Rev.* 27, 1710–1722. doi: 10.1016/j.quascirev.2008.06.010

Kolář, F., Lučanová, M., Vít, P., Urfus, T., Chrtek, J., Fér, T., et al. (2013). Diversity and endemism in deglaciated areas: ploidy, relative genome size and niche differentiation in the *Galium pusillum* complex (Rubiaceae) in Northern and Central Europe. *Ann. Bot.* 111, 1095–1108. doi: 10.1093/aob/mct074

Kruse, S., Gerdes, A., Kath, N. J., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., et al. (2018). Dispersal distances and migration rates at the arctic treeline in Siberia – a genetic and simulation based study. *Biogeosciences* 16, 1211–1224. doi: 10.5194/bg-2018-267

Kubota, Y., Kusumoto, B., Shiono, T., and Ulrich, W. (2018). Environmental filters shaping angiosperm tree assembly along climatic and geographic gradients. *J. Veg. Sci.* 29, 607–618. doi: 10.1111/jvs.12648

Kuzmina, S. A., Sher, A. V., Edwards, M. E., Haile, J., Yan, E. V., Kotov, A. V., et al. (2011). The late Pleistocene environment of the Eastern West Beringia based on the principal section at the Main River, Chukotka. *Quat. Sci. Rev.* 30, 2091–2106. doi: 10.1016/j.quascirev.2010.03.019

Legendre, P., and Borcard, D. (2018). Box–Cox−chord transformations for community composition data prior to beta diversity analysis. *Ecography* 41, 1820–1824. doi: 10.1111/ecog.03498

Legendre, P., and Cáceres, M. D. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecol. Lett.* 16, 951–963. doi: 10.1111/ele.12141

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., and Herzschuh, U. (2020). Holocene vegetation and plant diversity changes in the north-eastern Siberian treeline region from pollen and sedimentary ancient DNA. *Front. Ecol. Evol.* 8:560243. doi: 10.3389/fevo.2020.560243

Lozhkin, A. V. (1993). Geochronology of late quaternary events in Northeastern Russia. *Radiocarbon* 35, 429–433. doi: 10.1017/s0033822200060446

Lozhkin, A. V., Anderson, P. M., Matrosova, T. V., and Minyuk, P. S. (2007). The pollen record from El'gygytgyn Lake: implications for vegetation and climate histories of northern Chukotka since the late middle Pleistocene. *J. Paleolimnol.* 37, 135–153. doi: 10.1007/s10933-006-9018-5

Lozhkin, A. V., Anderson, P. M., Vartanyan, S. L., Brown, T. A., Belaya, B. V., and Kotov, A. N. (2001). Late quaternary paleoenvironments and modern pollen data from Wrangel Island (Northern Chukotka). *Quat. Sci. Rev.* 20, 217–233. doi: 10.1016/s0277-3791(00)00121-9

Mann, D. H., Peteet, D. M., Reanier, R. E., and Kunz, M. L. (2002). Responses of an arctic landscape to Lateglacial and early Holocene climatic changes: the importance of moisture. *Quat. Sci. Rev.* 21, 997–1021. doi: 10.1016/s0277- 3791(01)00116-0

Marx, H. E., Dentant, C., Renaud, J., Delunel, R., Tank, D. C., and Lavergne, S. (2017). Riders in the sky (islands): using a mega-phylogenetic approach to understand plant species distribution and coexistence at the altitudinal limits of angiosperm plant life. *J. Biogeogr.* 44, 2618–2630. doi: 10.1111/jbi.13073

Mayfield, M. M., and Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities: phylogeny and coexistence. *Ecol. Lett.* 13, 1085–1093. doi: 10.1111/j.1461-0248.2010.01509.x

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* 29, 897–910.

Mod, H. K., Heikkinen, R. K., Roux, P. C., Väre, H., and Luoto, M. (2016). Contrasting effects of biotic interactions on richness and distribution of vascular plants, bryophytes and lichens in an arctic–alpine landscape. *Polar Biol.* 39, 649–657. doi: 10.1007/s00300-015-1820-y

Moser, K. A., and MacDonald, G. M. (1990). Holocene vegetation change at Treeline North of Yellowknife, Northwest Territories, Canada. *Quat. Res.* 34, 227–239. doi: 10.1016/0033-5894(90)90033-h

Niemeyer, B., Epp, L. S., Stoof−Leichsenring, K. R., Pestryakova, L. A., and 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

Parducci, L., Alsos, I. G., Unneberg, P., Pedersen, M. W., Han, L., Lammers, Y., et al. (2019). Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from Southern Sweden. *Front. Ecol. Evol.* 7:189. doi: 10.3389/fevo.2019.00189

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

Pawłowska, J., Zaja¸czkowski, M., Ła¸cka, M., Lejzerowicz, F., Esling, P., and Pawlowski, J. (2016). Palaeoceanographic changes in Hornsund Fjord (Spitsbergen, Svalbard) over the last millennium: new insights from ancient DNA. *Clim. Past* 12, 1459–1472. doi: 10.5194/cp-12-1459-2016

Pearson, R. G., Phillips, S. J., Loranty, M. M., Beck, P. S. A., Damoulas, T., Knight, S. J., et al. (2013). Shifts in Arctic vegetation and associated feedbacks under climate change. *Nat. Clim. Change* 3, 673–677. doi: 10.1038/nclimate1858

Pedersen, M. W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., et al. (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

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., et al. (2013). IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55, 1869–1887. doi: 10.2458/azu_js_rc.55.16947

Reitalu, T., Gerhold, P., Poska, A., Pärtel, M., Väli, V., and Veski, S. (2015). Novel insights into post-glacial vegetation change: functional and phylogenetic diversity in pollen records. *J. Veg. Sci.* 26, 911–922. doi: 10.1111/jvs.12300

Revelle, W. R. (2018). *Psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University.

Rijal, D. P., Heintzman, P. D., Lammers, Y., Yoccoz, N. G., Lorberau, K. E., Pitelkova, I., et al. (2021). Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. *Sci. Adv.* 7:eabf9557. doi: 10.1126/sciadv.abf9557

Roswell, M., Dushoff, J., and Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos* 130, 321–338. doi: 10.1111/oik.07202

Rupp, T. S., Chapin, F. S. III, and Starfield, A. M. (2001). Modeling the influence of topographic barriers on treeline advance at the forest-Tundra ecotone in Northwestern Alaska. *Clim. Change* 48, 399–416. doi: 10.1023/a: 1010738502596

Sandom, C. J., Ejrnaes, R., Hansen, M. D. D., and Svenning, J.-C. (2014). High herbivore density associated with vegetation diversity in interglacial ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4162–4167. doi: 10.1073/pnas.1311014111

Seppä, H., Nyman, M., Korhola, A., and Weckström, J. (2002). Changes of treelines and alpine vegetation in relation to post-glacial climate dynamics in northern Fennoscandia based on pollen and chironomid records. *J. Quat. Sci.* 17, 287– 301. doi: 10.1002/jqs.678

Shevtsova, I., Heim, B., Kruse, S., Schrder, J., Troeva, E. I., Pestryakova, L. A., et al. (2020). Strong shrub expansion in tundra-taiga, tree infilling in taiga and stable tundra in central Chukotka (north-eastern Siberia) between 2000 and 2017. *Environ. Res. Lett.* 15:085006. doi: 10.1088/1748-9326/ab9059

Shevtsova, I., Herzschuh, U., Heim, B., Schulte, L., Stünzi, S., Pestryakova, L. A., et al. (2021). Recent above-ground biomass changes in central Chukotka (Russian Far East) using field sampling and Landsat satellite data. *Biogeosciences* 18, 3343–3366. doi: 10.5194/bg-18-3343-2021

Shooner, S., Davies, T. J., Saikia, P., Deka, J., Bharali, S., Tripathi, O. P., et al. (2018). Phylogenetic diversity patterns in Himalayan forests reveal evidence for environmental filtering of distinct lineages. *Ecosphere* 9:e02157. doi: 10.1002/ ecs2.2157

Silva, P. G., Hernández, M. I. M., and Heino, J. (2018). Disentangling the correlates of species and site contributions to beta diversity in dung beetle assemblages. *Divers. Distrib.* 24, 1674–1686. doi: 10.1111/ddi.12785

Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Merkel, M. K. F., et al. (2016). Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. *New Phytol.* 213, 929–941. doi: 10.1111/nph.14199

Sjögren, P., van der Knaap, W. O., Huusko, A., and van Leeuwen, J. F. N. (2008). Pollen productivity, dispersal, and correction factors for major tree taxa in the Swiss Alps based on pollen-trap results. *Rev. Palaeobot. Palynol.* 152, 200–210. doi: 10.1016/j.revpalbo.2008.05.003

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., et al. (2015). Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. *PLoS One* 10:e0115335. doi: 10.1371/journal. pone.0115335

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., et al. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol. Ecol. Resour.* 10, 1009–1018. doi: 10.1111/j. 1755-0998.2010.02855.x

Stuart, A. J., and Lister, A. M. (2007). Patterns of late quaternary megafaunal extinctions in Europe and northern Asia. *Cour. Forschung. Senckenb.* 259:287.

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., et al. (2004). Late Quaternary ice sheet history of northern Eurasia. *Quat. Sci. Rev.* 23, 1229–1271. doi: 10.1016/j.quascirev.2003.12.008

Szeicz, J. M., and MacDonald, G. M. (2001). Montane climate and vegetation dynamics in easternmost Beringia during the Late Quaternary. *Quat. Sci. Rev.* 20, 247–257. doi: 10.1016/s0277-3791(00)00119-0

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., et al. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35:e14. doi: 10.1093/nar/gkl938

ter Braak, C. J. F., and Šmilauer, P. (2002). *CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (version 4.5)*. Ithaca, NY: Microcomputer power.

Thuiller, W., Lavergne, S., Roquet, C., Boulangeat, I., Lafourcade, B., and Araujo, M. B. (2011). Consequences of climate change on the tree of life in Europe. *Nature* 470, 531–534. doi: 10.1038/nature09705

Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* 33, 23–45. doi: 10.1111/j.1600-0587.2009.06148.x

Velichko, A. A., Catto, N., Drenova, A. N., Klimanov, V. A., Kremenetski, K. V., and Nechaev, V. P. (2002). Climate changes in East Europe and Siberia at the Late glacial–holocene transition. *Quat. Int.* 91, 75–99. doi: 10.1016/s1040- 6182(01)00104-5

Vyse, S. A., Herzschuh, U., Andreev, A. A., Pestryakova, L. A., Diekmann, B., Armitage, S. J., et al. (2020). Geochemical and sedimentological responses of arctic glacial Lake Ilirney, Chukotka (Far East Russia) to palaeoenvironmental change since ∼51.8 ka BP. *Quat. Sci. Rev.* 247:106607. doi: 10.1016/j.quascirev. 2020.106607

Walker, M. D., Wahren, C. H., Hollister, R. D., Henry, G. H. R., Ahlquist, L. E., Alatalo, J. M., et al. (2006). Plant community responses to experimental warming across the tundra biome. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1342–1346. doi: 10.1073/pnas.0503198103

Webb, C. O. (2000). Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am. Nat.* 146, 145–155. doi: 10.1086/303378

Webb, C. O., and Pitman, N. C. A. (2002). Phylogenetic balance and ecological evenness. *Syst. Biol.* 51, 898–907. doi: 10.1080/10635150290102609

Whittaker, R. J. (1993). Plant population patterns in a glacier foreland succession: pioneer herbs and later-colonizing shrubs. *Ecography* 16, 117–136. doi: 10.1111/ j.1600-0587.1993.tb00064.x

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., et al. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506, 47–51. doi: 10.1038/nature12921

Wilson, S. D., and Nilsson, C. (2009). Arctic alpine vegetation change over 20 years. *Global Change Biol.* 15, 1676–1684. doi: 10.1111/j.1365-2486.2009.01896.x

Zhang, L., Mi, X., and Shao, H. (2016). Phylogenetic relatedness influences plant interspecific interactions across stress levels in coastal ecosystems: a meta- analysis. *Estuar. Coasts* 39, 1669–1678. doi: 10.1007/s12237-016-0104-2

Zhu, J., Zhang, Y., Yang, X., Chen, N., Li, S., Wang, P., et al. (2020). Warming alters plant phylogenetic and functional community structure. *J. Ecol.* 108, 2406–2415. doi: 10.1111/1365-2745.13448

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., et al. (2017). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences* 14, 575–596. doi: 10.5194/bg-14-575-2017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit " Biodiversity Changes In Siberia Between Quaternary Glacial And Interglacial Stages - Exploring The Potential Of *seda*DNA", selbständig verfasst und keine anderen als die von mir angegebenen Hilfsmittel genutzt habe. Wörtlich übernommene Sätze oder Satzteile sind als Zitate belegt. Diese Dissertation wird erstmalig an der Universität Potsdam eingereicht, dem Verfahren zu Grunde liegende Promotionsordnung ist mir bekannt.

Potsdam, 30.06.2022

Jérémy Courtin