UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics
Department of Arctic and Marine Biology

Sedimentary ancient DNA: Exploring methods of ancient DNA analysis for different taxonomic groups

Youri Lammers

A dissertation for the degree of Philosophiae Doctor

February 2020

Sedimentary ancient DNA: Exploring methods of ancient DNA analysis for different taxonomic groups.

Youri Lammers

A dissertation for the degree of Philosophiae Doctor
Tromsø, Norway, February 2020

UiT – The Arctic University of Norway
Faculty of Bioscience, Fisheries and Economics
Department of Arctic and Marine Biology

Supervisors

Professor Inger Greve Alsos
Tromsø Museum
UiT – The Arctic University of Norway
N-9037 Tromsø, Norway

Professor Nigel Gilles Yoccoz
Faculty of Bioscience, Fisheries and Economics
Department of Arctic and Marine Biology
UiT – The Arctic University of Norway
N-9037 Tromsø, Norway

Maître de conférences (Associate professor) Eric Coissac
CNRS, Laboratoire d'Ecologie Alpine (LECA)
Université Grenoble Alpes
Grenoble 38000, France

Cover image: Lake Endletvatn, Andøya, Norway. (Photo: Youri Lammers)

Table of contents

Acknowledgements

I'm grateful to those that made this PhD project possible. First my supervisors, Inger, Eric and Nigel, without whom this project could not have happened. Inger for both her supervision and guidance into the exciting world of palaeoecology and ancient DNA. Eric for all the help he provided with both the metabarcoding analysis and particularly the assembly of the organellar genomes as part of the PhyloNorway project. Nigel for his useful insights regarding the filtering procedures and data interpretation.

I'd like to thank all my co-authors and collaborators that I worked with over the past few years. In particular: Charlotte, for all her metabarcoding contributions and discussions regarding both the methods used and the life of a PhD student in general. Tony, for helping out with the more complex geological side of *sed*aDNA. Christer and Emilia, for their invaluable insights regarding the worm ecology. Laura, for fruitful discussions regarding the future of *sed*aDNA. Pete, for his contributions and ideas to extract every bit of information from the shotgun sequence data acquired as well as ways to process the metabarcode data. Finally, Dilli, Pete, Xabi, Sandra and Kristin for the adventurous fieldwork that was undertaken as part of the ECOGEN project.

Furthermore, I would like to thank the various teachers for the courses followed. In particular Galina and the rest of the ForBio research school, for the relevant courses on DNA target enrichment, phylogenetics and botany. Eric, Pierre and Fred, for valuable metabarcoding discussions either during the courses or other occasions. Further thanks go out to the countless people on stackoverflow, biostars and other forums, for their useful Python, R and Unix solutions, without whom the DNA analysis would have been far more of a headache.

Further thanks go out to all the people at the botany department of the Tromsø museum. Marie for introducing me into the ancient DNA laboratory, as well as her and Iva's help with the PhyloNorway project. Torbjørn, for making sure that there was always coffee ready in the morning. Arve, Martin, Kristian and Heini, for tolerating my attempts at the Norwegian language. Sandra, Dan, Aya and Per for dealing with my occasional cries of frustration as some analysis or piece of code inevitably failed. (Sorry!)

Finally, my thanks go out to all my friends and family who supported me throughout my journey up to the far north into a world of wonderful science. With particular thanks to my parents and sister for their continued support and valuable contributions in the form of cheese, waffles and liquorice.

List of papers

The following papers are included in my PhD thesis:

Paper I: Alsos IG, **Lammers Y**, Yoccoz NG, Jørgensen T, Sjögren P, Gielly L, Edwards ME. 2018. Plant DNA Metabarcoding of Lake Sediments: How Does It Represent the Contemporary Vegetation. *PloS One.* **13**: e0195403.

Paper II: Clarke CL, Edwards ME, Brown AG, Gielly L, **Lammers Y**, Heintzman PD, Ancin-Murguzur FJ, Bråthen K, Goslar T, Alsos IG. 2019. Holocene Floristic Diversity and Richness in Northeast Norway Revealed by Sedimentary ancient DNA (sedaDNA) and Pollen. *Boreas*. **48**: 299– 316.

Paper III: **Lammers Y**, Clarke CL, Erséus C, Brown AG, Edwards ME, Gielly L, Haflidason H, Mangerud J, Rota E, Svendsen JI, Alsos IG. 2019. Clitellate Worms (Annelida) in Lateglacial and Holocene sedimentary DNA records from the Polar Urals and Northern Norway. *Boreas*. **48**: 317– 329.

Paper IV: Parducci L, Alsos IG, Unneberg P, Pedersen MW, Han L, **Lammers Y**, Salonen JS, Väliranta MM, Slotte T, Wohlfarth B. 2019. Shotgun Environmental DNA, Pollen, and Macrofossil Analysis of Lateglacial Lake Sediments From Southern Sweden. *Frontiers in Ecology and Evolution*. **7**: 189.

Paper V: Alsos IG, Sjögren P, Brown AG, Gielly L, Merkel MKF, Paus A, **Lammers Y**, Edwards ME, Alm T, Leng M, Goslar T, Langdon CT, Bakke J, van der Bilt WGM. Last Glacial Maximum environmental condition of Andøya, a northern ice-edge ecological "hotspot". *Manuscript*

Paper VI: **Lammers Y**, Heintzman PD, Alsos IG. Environmental palaeogenomic reconstruction of an Ice Age algal population. *Manuscript*

Contributions

Paper I: Study conceptualization: IGA, NGY, MEE. Data curation: IGA, YL. Formal analysis: IGA, YL, NGY, LG. Funding acquisition: IGA LG, MEE. Investigation: IGA, NGY, TJ, PS, LG. Project administration: IGA Supervision: IGA. Visualization: IGA, YL, NGY. Writing – original draft: IGA. Writing – review & editing: IGA, YL, NGY, TJ, PS, LG, MEE.

Paper II: CLC, AGB, IGA and FJAM cored Lake Uhca Rohči. CLC extracted the DNA and LG performed the amplifications and ran OBITools. CLC analysed the DNA data with contributions from YL, PDH and

IGA. AGB described the sediment properties, provided the geomorphological background and compiled data on Pinus pollen influx rates. Radiocarbon dating was performed by TG. CLC constructed the age-depth models and performed XRF, LOI, magnetic susceptibility and pollen analysis. MEE contributed to pollen analysis. IGA, KAB and MEE provided ecological background. CLC drafted the first version of the manuscript of which all co-authors commented on.

Paper III: YL analysed the data and wrote the first version of the manuscript; CLC, AGB and IGA carried out fieldwork on Varanger and JIS, HH and JM in the Polar Urals; CLC extracted the DNA; LG amplified the DNA and ran OBITools; CE and ER provided taxonomic and ecological data; the project was devised by IGA / MEE and JIS / JM / HH. All authors contributed in various ways to the final version of the manuscript.

Paper IV: LP, TS, and BW designed the research. LH, BW, and LP organized and performed the coring. LH, PU, MWP, MMV, and JSS performed the lab research. LP, TS, BW, PU, MWP, YL, IGA, MMV, and JSS analysed the data. LP, BW, IGA, and MWP wrote the paper with final contributions from MMV, JSS, PU, YL, LH, and TS.

Paper V: IGA, MEE and TA planned, designed and organised the research, IGA, AP, LG and PS carried out the coring, LG, IGA, MKFM and PS performed the DNA analysis, YL performed the bioinformatics analyses, AP and CTL performed the pollen analysis, PS performed the macrofossil analysis, PS, ML, AGB and performed the geochemical analyses, WB and JB carried out Georadar (GPR), constructed GIS maps and models, and contributed to respective paper sections, IGA, AGB and PS organized the data and wrote the manuscript with input from all co-authors.

Paper VI: YL analysed the data and wrote the first version of the manuscript; IGA carried out the fieldwork and extracted the DNA; PDH helped out with the analysis of the sequence data; All authors contributed in various ways to the final version of the manuscript

Authors

Torbjørn Alm (TA), Inger Greve Alsos (IGA), Francisco Javier Ancin-Murguzur (FJAM), Jostein Bakke (JB), Willem van der Bilt (WB), Antony G. Brown (AGB), Kari-Anne Bråthen (KAB), Charlotte L. Clarke (CLC), Mary E. Edwards (MEE), Christer Erséus (CE), Ludovic Gielly (LG), Tomasz Goslar (TG), Haflidi Haflidason (HH), Lu Han (LH), Peter D. Heintzman (PDH), Tina Jørgensen (TJ), Cathrine T. Langdon (CTL), Melanie Leng (ML), Jan Mangerud (JM), Marie Kristine Føreid Merkel (MKFM), Laura Parducci

(LP), Aage Paus (AP), Mikkel W. Pedersen (MWP), Emilia Rota (ER), J. Sakari Salonen (JSS), Per Sjögren (PS), Tanja Slotte (TS), John Inge Svendsen (JIS), Per Unneberg (PU), Minna M. Väliranta (MMV), Barbara Wohlfarth (BW), Nigel Gilles Yoccoz (NGY)

Summary

Sedimentary ancient DNA (*sed*aDNA) from lake sediments is a promising tool for studying past communities, but there are still uncertainties regarding the methods. How does the lake *sed*aDNA obtained reflect the actual community present surrounding the lake? Are there differences in detection between various groups of organisms? How does the widely applied metabarcoding compare to shotgun sequencing of *sed*aDNA? How do either *sed*aDNA methods compare to the traditional palaeoecological methods? How should we treat and filter the obtained data and what more can we do with it? This thesis aims to answer the above questions and evaluate the capability of *sed*aDNA methods. For this, six different papers are presented, that utilize either metabarcoding or shotgun sequencing of various organisms from both contemporary and ancient sediments.

The aim of Paper I was to use contemporary environmental DNA in combination with vegetation surveys to better understand how and which groups of plants contribute material to the sedimentary DNA records. From 11 lakes in northern Norway, that differed in size, depth as well as surrounding vegetation, top sediments were collected. In addition, for each of these lakes the local vegetation was recorded. For each taxon in the vegetation, the abundance was scored, as well as the location relative to the lake. Metabarcoding was carried out on the top sediments with plant specific primers and the results were compared to the known species in the vegetation. Access to the known vegetation also allowed for further exploration of the effects that bioinformatic filtering has on the data. The results indicated that the taxa growing in close vicinity of the lake (<2 m) or within the lake, were more often detected than taxa only occurring further away in the catchment. In addition, dominant taxa in the vegetation were more often detected in the sedimentary DNA results compared to rarer taxa. The exception were aquatic taxa, for which even the rarer taxa were often detected in the results, making metabarcoding a promising tool for their identification. Finally, the results indicated that there is no good way to filter the data, where there always exists a trade-off between the removal of false positives while retaining as many true positives.

For Paper II, the vegetation through time was studied for Lake Ucha Rohči, Varanger Peninsula, Norway, based on a sediment core dating from 10,700 to 3300 cal a BP. This study combined metabarcoding of plant specific primers with pollen analysis for the identification of the flora. Both methods identified a shrub-tundra vegetation as the main community type present throughout the Holocene, which is comparable to the vegetation found today. However, some thermophilous taxa, such as *Callitriche hermaphroditica*, were detected during the Early Holocene, indicating warmer than present temperatures. Both *sed*aDNA and pollen differ in the total taxa detected, where,

*sed*aDNA could identify 118 taxa compared to 39 for pollen. Furthermore, *sed*aDNA could identify taxa to lower taxonomic ranks and was capable of detecting more insect pollinated forbs and aquatic taxa compared to pollen, which was mainly dominated by wind dispersed taxa.

Clitellate worms were the focus of Paper III, though originally not the taxonomic group we set out to detect. Two sediment cores, the Varanger core from Paper II, as well as a sediment core from Lake Bolshoye Shchuchye, Polar Urals, Russia, were subjected to metabarcoding with universal mammalian primers. The majority of the taxa detected, besides reindeer in six samples, were worm species. The lack of mammals is primarily caused by their relatively low biomass in the environment, which resulted in less template material contributed to the sediments for metabarcoding detection. The worm taxa on the other hand could be more abundant in the lake catchment. The metabarcoding of worms itself was possible due to the conserved nature of the selected metabarcode region, where, even though there are a few primer mismatches, worms could be amplified in the absence of sufficient mammalian template. These results indicate that worm amplification was possible from *sed*aDNA, even when suboptimal primers were used, which opens the door to further palaeoecological studies where worms can be used as a potential proxy for temperature, soil moisture and acidity.

Paper IV explored the potential of shotgun sequencing of *sed*aDNA for the reconstruction of past vegetation. A Lateglacial sediment core from Hässeldala Port, southern Sweden, was analysed with both *sed*aDNA, pollen and macrofossils. The combined proxies identified the presence of a boreal to subarctic climate during the Allerød, that persisted throughout the Younger Dryas into the Preboreal. While some taxa, such as *Betula*, *Salix*, *Myriophyllum* and *Nymphaea* could be identified with all methods, the majority of taxa were only detected with one or two proxies. Out of the three different methods, the highest number of taxa were identified with shotgun sequencing, which was primarily due to the detection of new taxa that were introduced in the Younger Dryas. A similar trend was not detected by the other proxies. The observed DNA results could be due to the reference database used, which had a large component of reference sequences from northern Norwegian taxa and thus might have been better suited to identify the taxa present during the colder Younger Dryas. Overall, shotgun sequencing was capable of detecting taxa that were otherwise overlooked by traditional methods, though a large part of the sequence data remained unused, either because it originated from other organisms, or because the appropriate plant reference genomes were not available.

Lake Øvre Æråsvatnet on northern Andøya, Norway was the study site for Paper V. Andøya has been of particular interest for palaeoecological studies, as it was partly unglaciated during the Last Glacial Maximum (LGM) and could thus provide a refugia for species survival. Three sediment cores from the lake were analysed with metabarcoding using plant specific primers as well as pollen, macrofossils and geochemistry. The presence of thermophilous taxa in both *sed*aDNA, pollen and macrofossils during the LGM indicated at least some warmer phases. Furthermore, a high organic content detected throughout the LGM suggest a high inflow of nutrients into the lake, likely originating from a local bird colony, as indicated by stable isotope analysis and the finds of bird bones. The *sed*aDNA results from this period, however, were problematic, as they were dominated by algal DNA, possibly obscuring the terrestrial plant signal.

The final *sed*aDNA study (Paper VI) deviated from the other studies in that it applied deep shotgun sequencing to two LGM samples, rather than a full sediment record. The samples used originated from the same sediment cores as Paper V, which were known to contain an abundance of the *Nannochloropsis* microalgae. Shotgun sequencing allowed for the identification of the most abundant *Nannochloropsis* species present in the sediment record. Furthermore, the amount of *Nannochloropsis* sequence material obtained allowed for reconstruction of the palaeo organellar genomes. Phylogenetic placement of both the reconstructed organellar genomes as well as *rbcL*, a traditional barcode marker, identified the Andøya variants as *N.* cf. *limnetica*. Furthermore, the coverage obtained for the organellar genomes allowed for the estimation of the present haplogroups. At least two main variants of *N.* cf. *limnetica* could be detected for both samples, based on either single variants or linked sites. This study demonstrates the potential of shotgun sequencing of *sed*aDNA for both phylogenetic and population genetic analysis.

The above papers demonstrate that the *sed*aDNA results obtained from lake sediments reflect the local community. *sed*aDNA gives more comprehensive results for plants than for animals, most likely due to difference in biomass between the two groups. Furthermore, both metabarcoding and shotgun sequencing are capable of reconstructing past communities. However, metabarcoding, for now, remains more efficient in the identification of multiple taxa, while shotgun sequencing performs better when it can be utilized to explore phylogenetic or population genetic questions.

Introduction

Palaeo records provide a vital source of information for studying the effects of a changing climate on species distributions (Seppä and Birks 2001, Willis and Birks 2006, Alsos *et al*. 2016). In addition, they can provide insight into human influence (Giguet-Covex *et al*. 2014, Pansu *et al*. 2015), as well as extinctions (Haile *et al*. 2009, Graham *et al*. 2016, Brüniche-Olsen *et al*. 2018) and survival (Westergaard *et al*. 2011, Parducci *et al*. 2012). Commonly used methods for the study of palaeo records include pollen and macrofossil studies (Odgaard 1999, Birks and Birks 2000). These methods, however, suffer from some limitations, such as pollen data representing a regional signal due to an overabundance of wind dispersed taxa (Eide *et al*. 2006, Birks and Birks 2000) and macrofossils being relatively scarce (Birks 2003). During the last decade sedimentary ancient DNA (*sed*aDNA) became an interesting alternative to the conventional methods (Willerslev *et al*. 2003, Haile *et al*. 2007, Parducci *et al*. 2017), first via DNA metabarcoding and later shotgun sequencing methods.

DNA barcoding has proven to be an invaluable tool for the identification of species (Hebert *et al*. 2003). The method relies on the amplification of smaller conserved barcode regions, that still contain enough taxonomic information for species identification. Typical barcode regions are several hundred base pairs (bp) in length and include: the COI gene for animals (Hebert *et al*. 2003), *rbcL* and *matK* for plants (CBOL Plant Working Group *et al*. 2009) and the Internal Transcribed Spacer (ITS) for Fungi (Schoch *et al*. 2012). DNA metabarcoding takes the barcoding technique a step further by amplifying barcode regions from bulk or environmental samples, rather than a single individual (Taberlet *et al*. 2012, Taberlet *et al*. 2018). The amplified pools of barcodes can be sequenced in their entirety due to the advances made in sequencing techniques, which now provides the sequence depth required for bulk identifications.

Metabarcoding of environmental samples requires different markers than traditional barcoding, as the environmental DNA (eDNA) is commonly degraded and the longer template molecules required for traditional barcoding could either be scarce or absent (Blum *et al*. 1997, Barnes *et al*. 2014). Thus a whole range of shorter markers exist for the metabarcoding of various taxa, including: plants (Taberlet *et al*. 2007, Willerslev *et al*. 2014), mammals (Boessenkool *et al*. 2012, Giguet-Covex *et al*. 2014), fungi (Buée *et al*. 2009, Tedersoo *et al*. 2018), fishes (Valentini *et al*. 2016, Hänfling *et al*. 2016) and insects (Epp *et al*. 2012, Elbrecht *et al*. 2016). The above method can be applied to ancient sediments in order to reveal past ecological information (Jørgensen *et al*. 2012, Pedersen *et al*. 2015). However, ancient environmental DNA brings additional challenges compared to eDNA, as the available DNA is further reduced (Allentoft *et al*. 2012).

One of the main advantages of metabarcoding comes from having access to large reference collections, such as the general NCBI nucleotide database, the Barcode of Life Database (BoLD) or the SILVA database for ribosomal sequences. These reference collections allow for either exact or close identifications of the taxa present. Furthermore, reliable community identifications can be obtained with sequence counts between 10,000 and 50,000 sequences, which allows for pooling of multiple samples into a single sequence library in order to reduce sequencing costs (Binladen *et al*. 2007, Valentini *et al*. 2009). Though some problems exist in the form of PCR artefacts, chimeric sequences and tag jumps (Schloss *et al*. 2011, Schnell *et al*. 2015). In particular, homopolymer repeats can cause polymerase slippage and introduce additional artificial barcode variants (Taberlet *et al*. 2012). The above problems require careful bioinformatic filtering and data interpretation.

An alternative approach to *sed*aDNA is via shotgun sequencing of the sediments. This method does not rely on amplification of a target barcode region, but instead sequences all material present (Pedersen *et al*. 2016, Graham *et al*. 2016). The lack of amplification gives it some advantages over metabarcoding. First, the lack of amplification allows for a less biased abundance estimation (Ziesemer *et al*. 2015). Second, since the method captures all sequences present, sequencing can be carried out without *a priori* knowledge of the sample. Finally, the data can be used to pursue various genetic questions. Though the lack of targeting can also be problematic as the vast majority of the sequence data originates from single celled organisms, such as bacteria, with multicellular life making up a relatively small part (Pedersen *et al*. 2016, Slon *et al*. 2017). A high sequence depth is thus required to ensure that sufficient material is retrieved for the taxonomic group of interest.

Shotgun sequencing is quite reliant on the availability of reference material. As the sequences can originate from any part of the genome, a broad reference set of nuclear and organellar reference genomes is desirable (Pedersen *et al*. 2015). Full nuclear genomes however are costly to generate and are mainly available for taxa that are of economic or particular scientific interest (Parducci *et al*. 2017). More reference material is available for organellar genomes, as they are easier to obtain. Furthermore, with sequencing projects such the PhyloAlps and PhyloNorway projects, the number of available organellar sequences is rapidly increasing. However, the majority of the *sed*aDNA shotgun sequences are derived from the nuclear genome (Rauwolf *et al*. 2010). Full nuclear genomes will therefore allow for a more complete identification of the sequences present. However, with time, the reference issue will become less problematic as more genomes become available for a wider variety of taxa.

Contamination is a prevalent issue when working with ancient material and both *sed*aDNA methods have some ways to deal with it. A common solution is the incorporation of negative control samples during the various steps of sampling and extraction (De Barba *et al*. 2014, Pedersen *et al*. 2015). These controls can be used during the filtering of the data and allow for the identification and removal of the contaminant taxa. In addition, for metabarcoding, blocking primers can be designed for common contaminant species, such as human, to inhibit their amplification (Boessenkool *et al*. 2012). While for shotgun sequencing ancient DNA damage patterns can be inspected for authentication (Briggs *et al*. 2007, Jónsson *et al*. 2013). Ancient DNA damage patterns are caused by deamination of cytosine bases into uracil bases, which will be interpreted as thymine bases post sequencing. As a result, characteristic C/T, or the reverse complement G/A, substitutions are observed in the sequence data for true ancient material.

Regardless of the *sed*aDNA method used, first ancient sediments need to be collected and sampled for DNA. Not every type of sediment is suitable for aDNA studies, as the available DNA in any environmental system degrades over time (Pedersen *et al*. 2015). Stable, cold and dark conditions are desirable for long-term ancient DNA preservation, such as cave, lake or permafrost sediments (Hansen *et al*. 2006, Sønstebø *et al*. 2010, Pansu *et al*. 2015, Slon *et al*. 2017). DNA preservation in these conditions is possible due to the extracellular DNA binding to either negatively charged silicates, clays, or organic compounds (Taberlet *et al*. 2012), which can stabilize the molecules over longer periods of time. Once the sediments are obtained, these ancient DNA molecules can be released for *sed*aDNA analysis.

Distinct differences are noted when *sed*aDNA results are compared to either pollen or macrofossils. Pollen, especially those which are wind dispersed, represent a regional signal (Rousseau *et al*. 2006, Hicks 2006, Parducci *et al*. 2015). Furthermore, identifications based on pollen are often not resolved to species level and thus cannot be used for finer reconstructions (Birks and Birks 2000). Macrofossils on the other hand provide a more local signal, as they are unlikely to be transported over larger distances (Eide *et al*. 2006, Birks and Birks 2000). Furthermore, depending on the type of macrofossil, species level identifications are possible (Birks and Birks 2000). The main problem with macrofossils is that they are relatively scarce and enough material for full ecosystem reconstruction might not be available from a sediment record (Birks 2003). The results obtained with *sed*aDNA from lake sediments tend to overlap with results obtained by macrofossil studies, indicating a local origin of the material (Jørgensen *et al*. 2012, Alsos *et al*. 2016), though some contradicting observations

have been made for different sediment types, such as peat bogs, which noted a greater difference between *sed*aDNA and macrofossils (Parducci *et al*. 2015). Finally, depending on the *sed*aDNA method used, species level identifications can be obtained for various taxa (Sønstebø *et al*. 2010, Epp *et al*. 2015, Zimmermann *et al*. 2017).

The above methods play an important role for the identification of taxa present in palaeo records, however, *sed*aDNA has the potential to answer more elaborate questions. Via the sequencing of more variable sites, the resolution of the identification can be improved to population level (Parducci *et al*. 2017). Tracking populations over time will allow for more sophisticated environmental models. In addition, improved abundance estimations will allow for the detection of finer within population changes (Slon *et al*. 2017, Søe *et al*. 2018). Finally, genomic information derived from *sed*aDNA can be used for phylogenetic and evolutionary studies, as more and more sequence information is extracted and potential larger genomic regions are reconstructed (Seersholm *et al*. 2016, Slon *et al*. 2017).

Thesis aims

This thesis consists out of six different papers, but all have a unifying theme in eDNA or *sed*aDNA from lake sediments. It sets out to answer the following questions:

1) How does the eDNA derived from contemporary lake sediments reflect the local vegetation? (Paper I)
2) How effective are metabarcoding and shotgun sequencing methods, compared to traditional palaeoecological methods, in the identification of plants from sediment records? (Papers II, IV and V)
3) How effective is metabarcoding of *sed*aDNA for the detection of animals? (Paper III)
4) How reliably can we derive phylo- and population genomic information from *sed*aDNA shotgun datasets? (Paper VI)

Methods

Study organisms

A range of different organisms were studied in the papers for this thesis and they include: vascular plants (Papers I, II, IV, V and VI), mammals (Papers III and VI), clitellate worms (Paper III) and algae (Papers V and VI). Some of these organisms were specifically targeted due to the usage of metabarcode primers, such as the vascular plants in Papers I, II and V, as well as the mammals in Paper III. While others were detected as metabarcoding "bycatch", such as the worms in Paper III and the algae in paper V. The shotgun sequence datasets, by their nature, contained sequences from all taxa, though in Paper IV, only the vascular plant component was analysed due to the usage of a constrained reference dataset containing chloroplast genomes. The shotgun data in Paper VI, was subjected to a wider panel of reference genomes, which allowed for more taxonomically diverse identifications.

Site and material collection

Sites

All data in this thesis originated from lake sediments from either northern Norway, Sweden or the Russian Polar Urals (Figure 1). Paper I utilized top sediments from 11 lakes located in the counties of Nordland, Troms and Finnmark, Norway. The lakes were selected such that they covered a wide range of lake sizes, depths, altitudes and surrounding vegetation types. In addition, the lakes themselves had limited in- and outflow and were relatively undisturbed by human activity.

Papers II and III used sedimentary data from Lake Uhca Rohči on the Varanger peninsula, Finnmark, Norway. In addition, Paper III also utilized data from lake Lake Bolshoye Shchuchye, located in the Polar Urals, Russia. There is a stark contrast between both sites. Where Uhca Rohči is relatively small (<1 ha), Bolshoye Shchuchye is the largest (1200 ha) and deepest (134 meters deep) lake in the Polar Urals (Svendsen *et al*. 2019). Furthermore, the surrounding areas are different as well. Uhca Rohči is located in the relative flat Komagdalen river valley, while Bolshoye Shchuchye is surrounded by mountainous terrain characterised by sharp peaks.

Paper IV used Lateglacial lake sediments from Hässeldala Port, Blekinge province in southeast Sweden. The lake filled in during the early Holocene and turned into a peat bog. The older Lateglacial lake sediments themselves are represented in a 1 meter long record (Wohlfarth *et al*. 2017), that is currently fairly close to the surface due to peat cutting activities.

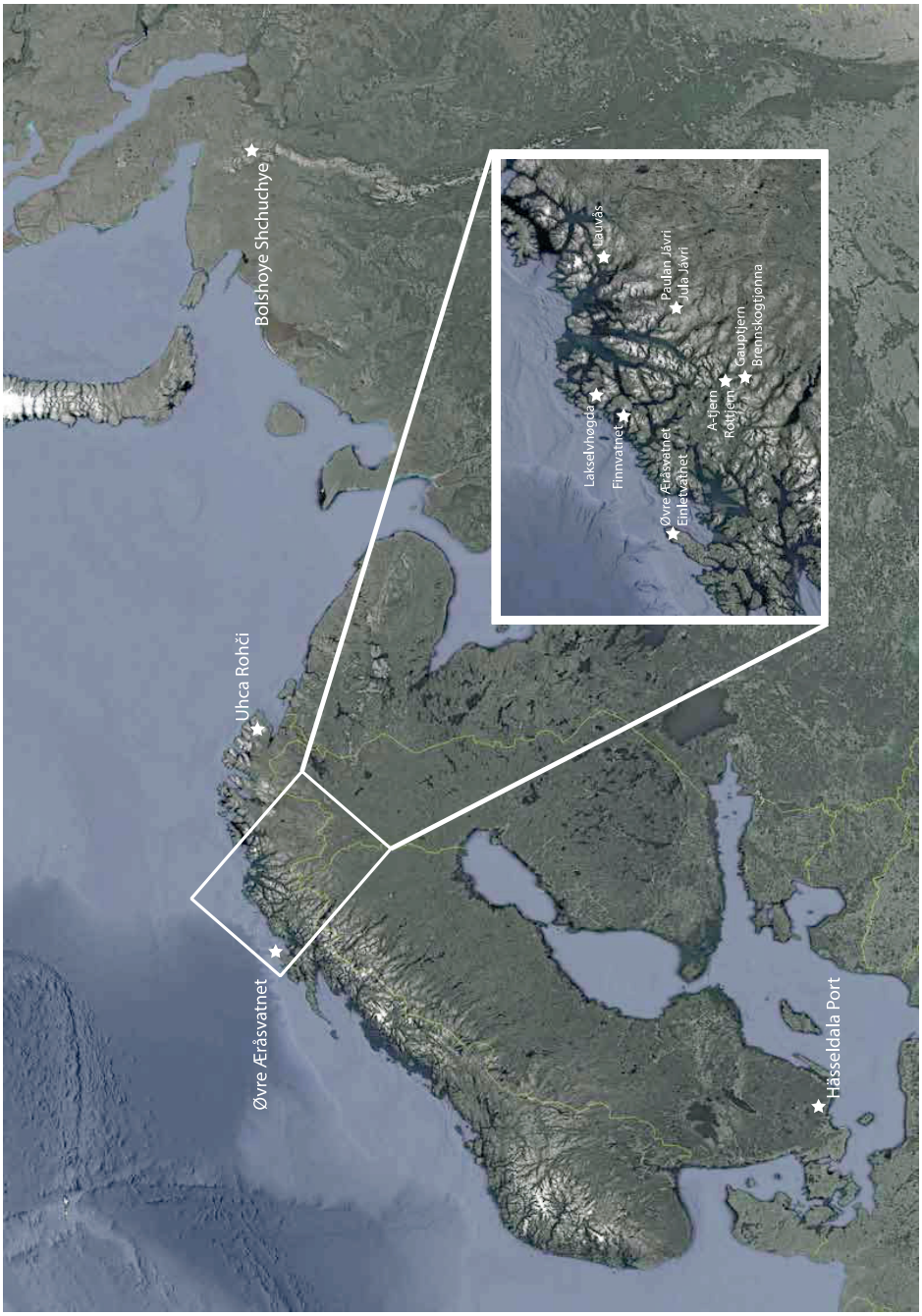


Figure 1: Map of the sites studied in this thesis.

The final site is Lake Øvre Æråsvatnet on the island Andøya, north-western Norway. Sediment cores, that covered the LGM and the early Holocene, from this site were analysed for Papers V and VI. Furthermore, top samples from this site were also part of the contemporary eDNA comparison in Paper I. Andøya is of special interest, as the island was partially unglaciated during the LGM and is thus provides opportunities to study glacial survival.

Coring

The sediment samples used in the presented studies were collected with a variety of different coring methods. The modern lake surface samples were collected with a Kajak corer, that utilized several equidistant spaced tubes that were lowered into the soft top sediments via gravity, before they were retrieved.

The deeper sediments for Øvre Æråsvatnet, Uhca Rohči, Bolshoye Shchuchye and Hässeldala Port were collected with different methods, as gravity alone was not sufficient for reaching the deeper layers. The Uhca Rohči core and one of the Øvre Æråsvatnet cores were collected with a modified Nesje corer (Figure 1; Nesje 1992) while the Bolshoye Shchuchye core was collected with a UWITEC Piston Corer. Both methods differ in execution but rely on the same principle of forcing the core into the sediment by hammering the top with a weight. The Russian corer used for Hässeldala Port and the Geonor corer used for the other two Øvre Æråsvatnet cores were based on pressure, either by forcing the core into the sediment by hand in case of the former, or a jack and solid extension rods for the latter.

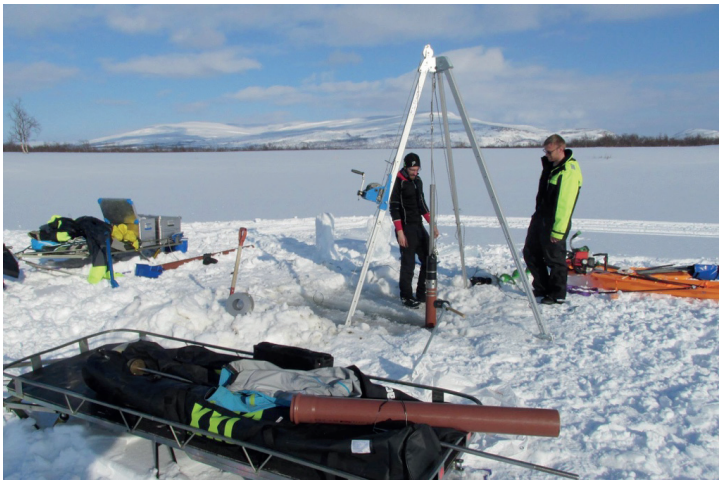


Figure 1: Coring from ice in northern Norway using a Nesje system. (Photo: Peter D. Heintzman)

Sampling of the sediments

Ancient DNA is due to its very nature both fragmented and present in low concentrations (Briggs *et al*. 2007, Pedersen *et al*. 2015), therefore contamination with modern material was a considerable risk during sampling. To limit the chance of contamination, sediment cores were opened in sterile ancient DNA laboratories. These laboratories, as well as the tools used for sampling, were kept clean via bleaching of the surfaces and exposure to UV-light. In addition, the usage of bench controls during sampling gave an indication of aerosol contamination.

Lithology and dating

Loss-on-Ignition (LOI) values were obtained for all palaeo records (Papers II-VI). The LOI values were obtained by drying smaller subsamples overnight at temperatures between 100-105 °C, weighting the dry samples and burning them afterwards at 550 °C. The burned remains were weighted again after which the LOI values could be computed as the percent dry-weight lost after burning (Heiri *et al*. 2001).

Qualitative geochemical analysis was performed on the Øvre Æråsvatnet, Uhca Rohči and Bolshoye Shchuchye cores and was carried out via X-ray fluorescence (XRF) scanning. XRF is based on the ionization of the elements present in a sample and the measurement of the photons released when the electrons collapse to lower electron shell. A spectra of the detected photons was translated to the elements present and their abundances, either as a proportion of the elements present, or normalized against a single element (Röhl and Abrams 2000, Rothwell *et al*. 2015).

Radiocarbon dating was performed on the sediment cores from Øvre Æråsvatnet, Uhca Rohči and Bolshoye Shchuchye. The dating itself was determined via accelerated mass spectrometry of macrofossils, such as bryophytes, seeds, or leaf and bark fragments. The acquired radiocarbon dates were calibrated via terrestrial calibration curves (Reimer *et al*. 2013) and age-depth models were constructed via the Bayesian based modeling software Bacon (Blaauw and Andrés Christen 2011). Dating of the Hässeldala Port sediments was based on alignment of the LOI curves to an older, already radiocarbon dated, core from the same site (Wohlfarth *et al*. 2017).

Sampling, DNA extraction and sequencing

DNA extraction

Sediment samples contain complex mixtures of organic molecules, both DNA or otherwise and inorganic compounds (Miller *et al*. 1999). Various DNA extraction protocols exist in order to free up

the DNA molecules present in the sediment bound to silica, clays, humic acids and remove potential inhibiting compounds that could interfere with downstream molecular methods (England *et al*. 2004, McKee *et al*. 2015).

The two DNA extraction protocols used were either based on the PowerMax Soil DNA extraction kit (MO BIO Laboratories, Carlsbad, CA, USA) (Papers I-V), or the phosphate extraction protocol (Taberlet *et al*. 2012) (Paper VI). The protocols differ in the type of DNA targeted. The PowerMax variants were inclusive in terms of intra- and extracellular DNA obtained due to the inclusion of cell lysis steps. The phosphate protocol on the other hand forwent cell lysis to retrieve extracellular DNA. Additional differences were found in the amount of sediment material required for extraction, number of steps, the resulting DNA concentrations and potential extent of inhibitors present.

Similar to the subsampling of the sediment cores, all DNA extractions, as well as the PCR preparations and shotgun library constructions were carried out in dedicated ancient DNA laboratories, to avoid the potential contamination of modern material. The inclusion of DNA extraction negatives allowed for the estimation of the contamination present in the ancient laboratories, as well as in the extraction reagents used themselves (Champlot *et al*. 2010, Parducci *et al*. 2017).

Metabarcoding

Two types of metabarcoding PCRs were carried out for the presented papers, either targeting and amplifying plants (Papers I,II and V) or mammals (Paper III). The plant metabarcoding was carried out with the universal plant *g* and *h* primers (Taberlet *et al*. 2007), which targeted a ~15-120 bp long fragment of the plastid *trnL* UAA P6 loop region. The mammals were amplified with the universal mammalian MamP007F and MamP007R primers (Giguet-Covex *et al*. 2014), which targeted a ~60-80 bp long fragment of the mitochondrial 16S rRNA. Since the mammalian primers were capable of amplifying human, the MamP007_B_Hum1 human blocking primer was included to inhibit amplification (Giguet-Covex *et al*. 2014). The relative short fragments that both primers targeted made them particularly suitable for the amplification of the fragmented ancient DNA. Furthermore, the high number of chloroplasts and mitochondria per cell, increased the potential amount of target molecules in the sediment compared to nuclear markers (Thomsen and Willerslev 2015).

Both the plant and mammalian PCR reactions were carried out with 45 PCR cycles, which, while high, is necessary when working with ancient DNA to ensure the detection of low abundance template

molecules (Epp *et al*. 2012). For each sample, six (Paper I) to eight (Papers II, III and V) PCR replicates were carried out to ensure both validation of the detections and increase the chance of observing rarer amplicons (Ficetola *et al*. 2015). In addition to the amplification of the samples, both negative and positive control samples were included, to detect the presence of contaminants and validate the success of the PCR reaction respectively. All primers contained additional unique tag sequences that allowed for pooling of multiple PCR products into a single sequence library, which aided in reducing the sequencing costs (Binladen *et al*. 2007, Valentini *et al*. 2009). Post amplification and pooling, the material was incorporated into sequence libraries and sequenced on either the Illumina HiSeq 2500 platform at 2x125 bp (Papers I,II and III) or the Illumina NextSeq at 2x150 bp (Paper V).

Shotgun sequencing

Compared to the metabarcoding procedure, shotgun sequencing is more straightforward. The DNA extracts themselves were directly incorporated into sequencing libraries, without any initial amplification (Ziesemer *et al*. 2015, Graham *et al*. 2016). This process does put more importance on the DNA extracts, as the library preparation could be susceptible to inhibition. Similar to metabarcoding, the inclusion of negative controls could again aid in the detection of contamination. Both the libraries for Papers IV and VI were sequenced on an Illumina HiSeq 2500, but the former was sequenced at 1x100 bp, while the latter at 2x80 bp.

DNA analysis

Metabarcode analysis and identification

Analysis of the metabarcoding sequence data for Papers I, II, III and V, consisted of a number of different steps: merging of the paired-end sequence data, demultiplexing of tagged PCR products, collapsing, clustering and error correction of the amplicon data and finally the identification of the metabarcode sequences. Several software packages and pipelines were available for these analysis, such as: QIIME (Caporaso *et al*. 2010), mothur (Schloss *et al*. 2009) and OBITools (Boyer *et al*. 2016), where the latter was used for the analysis of the metabarcode data presented in this thesis.

The first step after paired-end sequencing of the metabarcode data was the merging of the forward and reverse sequences. The resulting dataset still contained a mixture of the different samples and PCR replicates that were pooled together, which required demultiplexing. The demultiplexing step used the known primer and unique tag information for each of the samples to identify from which sample a given sequence originated from. The sample information was subsequently added to the sequence metadata and the primer and tag sequences were removed from the amplicon sequence.

Post merging and demultiplexing, the sequence files still contained several million sequences. To reduce downstream file sizes and computation time, identical sequences were collapsed into a single sequence. The sample metadata was tallied up upon collapsing, so that the number sequences and the origin of the samples were preserved in the resulting dataset. Following the collapsing, a series of filtering steps were included to remove PCR and sequencing artefacts. First both singleton sequences (those that only occurred once) and sequences shorter than 10 bp (shorter than either the expected plant or mammalian amplicon size) were removed. The resulting dataset was then analysed for the presence of PCR artefacts, which could be detected by looking for the co-occurrence of highly similar sequences in a sample (Boyer *et al*. 2016). The original sequence was present in the DNA extract from the beginning of the PCR reaction and has an assumed high sequence abundance. The PCR errors on the other hand originated partway through the amplification process and thus underwent less amplification cycles which resulted in a lower abundance compared to the original sequence. Problems arose when dealing with certain taxa that could co-occur in a sample and that had comparable barcode sequences, which could resemble a PCR error during the error correction. In addition, different types of PCR errors could occur at different rates depending on the sequence composition, error type and the polymerase used (Nichols *et al*. 2018). Given these issues, the error correction carried out was quite conservative, so that no true positives were lost during this step, though at the cost of additional false positives.

After error correction the remaining sequences were identified via comparison with the reference databases. The plant metabarcodes were identified with two different databases, a curated set containing regional arctic and boreal sequences (Sønstebø *et al*. 2010, Willerslev *et al*. 2014, Soininen *et al*. 2015) and all sequences available in the NCBI nucleotide database. The usage of the NCBI reference allowed for the identification of either rarer metabarcodes or contaminant sequences, both of which could be absent from the curated dataset. The mammalian metabarcodes were exclusively identified with the NCBI set, as no curated reference set was available for the used marker.

Post identification, the two different identified datasets for the plant metabarcoding were merged together. The metabarcode results were processed further in R, were additional filtering was applied. First, only identifications with a 100% match to the reference databases were retained, to reduce erroneous taxonomic assignments. Second, a minimum read cut-off was applied to each observation, as low read (<=2) observations could be the result of tag or library index swaps,

background contamination or uncorrected PCR artefacts (De Barba *et al*. 2014). Additional filtering was carried out based on the total number of observations (samples or repeats), the minimum observations per sample and by comparing the sample data to that of the controls. The effects of these different filtering steps were thoroughly investigated in Paper I.

Shotgun analysis

The analysis of shotgun *sed*aDNA allowed for more varied approaches compared to metabarcoding, thus the analysis carried out for Papers IV and VI differed from each other. As Paper VI used paired-end sequence data, the sequences were merged prior to analysis. The next step was the removal of short or otherwise problematic sequences. Short sequences were expected from shotgun sequencing of *sed*aDNA, but they could be problematic for identification purposes, as they could match against multiple organisms (Schubert *et al*. 2012). Similar issues existed for sequences that were of low complexity and/or contained longer homopolymer stretches. Thus these sequences were removed with tools such as cutadapt (Martin 2011) or the SGA preprocess function (Simpson and Durbin 2012).

Identification of shotgun sequence data differed from metabarcoding, as not a single locus was amplified, instead the data could have originated from any part of the genome. Therefore, it was important to have a broad and inclusive database. For Paper IV the data was identified via all published chloroplast genomes in GenBank, as well as the available chloroplast genomes from the PhyloNorway project. The data from Paper VI was identified both via the NCBI nucleotide database, the available organellar genomes and 37 different nuclear genomes representing taxa that were either expected in the region or were used as exotic controls. Post identification, the results were either summarized with tools such as MEGAN (Huson *et al*. 2016), or subjected to further filtering in case of the nuclear genomes.

Beyond taxonomic identifications, shotgun data could aid in the validation of the ancient data, via inspection of characteristic ancient DNA deamination patterns. mapDamage (Jónsson *et al*. 2013) was used to analyse shotgun sequence data that was aligned to a reference genome and inspect the C/T and G/A mismatches that occurred towards the ends of the sequences. In addition, when enough coverage was present for a taxon in Paper VI, the shotgun data was utilized to reconstruct the genomic sequences present in the sample (Slon *et al*. 2017, Søe *et al*. 2018). The above methods however required both the presence of relevant reference sequences as well as enough sequence coverage in the shotgun data to work and thus were only applied in Paper VI.

Pollen and macrofossil identification and vegetation surveys

Pollen and macrofossil identification

Pollen analysis was carried out for Papers II, IV and V, while macrofossil analysis was performed for Papers IV and V. Pollen identification was based on the shape of the pollen themselves, as well as outer features present on the pollen wall. Prior to identification, the pollen first were isolated from the sediments in which they occurred, either via washing and filtering or chemical digestion of the sediments (Berglund and Ralska-Jasiewiczowa 1986). The resulting pollen concentrations were mounted in silicon oil or glycerol and were inspected via microscopy, where grouping and identifications occurred via published keys and reference collections (Faegri and Iversen 1989, Moore *et al*. 1991). Macrofossil identification was based on subsampling the core and sieving of the resulting sediment subsamples. As the macrofossils could originate from different plants or different parts, identification could be problematic and thus required the usage of extensive reference collections for correct identifications.

Vegetation survey

Vegetations surveys of both the surrounding areas around a lake, as well as the within lake vegetation were carried out for Paper I (Figure 2). The vegetation data allowed for estimation of the true positives obtained in the environmental DNA results. For each of the modern sites, the surrounding and within lake vegetation was identified and abundance values were assigned for each taxon, ranging from: rare (observed in only a few places), scattered (low abundance, but found in multiple parts), common (common throughout the vegetation, but not the main component) and dominant (majority of the biomass, usually shrubs or trees).

Figure 2: The aquatic vegetation survey. (Photo: Inger Greve Alsos)

Results and discussion

How does eDNA reflect the vegetation?

Based on the contemporary eDNA study carried out in Paper I, the majority of the taxa detected indicated a local signal. Seventy three percent of the taxa detected based on eDNA were either aquatic or growing within two meters of the lake's edge. Taxa that were exclusively growing further away in the catchment were only represented in 11% of the eDNA records, with the remaining 16% being taxa that were undetected in the vegetation survey. In addition, it was found that the dominant or common taxa in the vegetation were more likely to be detected based on eDNA. Rare taxa on the contrary were more difficult to detect based on eDNA and either didn't appear in the sequence data or were indistinguishable from background contamination. The results for the plant families were varied. Groups such as Salicaceae, Betulaceae and Lycopodiaceae were commonly detected based on eDNA for the lakes in which they were recorded in the vegetation. However, Cyperaceae and Asteraceae specifically were hardly detected based on eDNA. Overall, 35% of the barcode data could be identified to species level, while 39% to genus level and 26% to higher taxonomic levels. Finally, quite some differences could be observed between the different lakes studied, where the percentage of taxa detected in the eDNA that matched to the vegetation varied between 13% and 49%.

The vegetation survey allowed for a direct comparison between the present flora and the obtained eDNA results, which in addition allowed for the exploration of the effects of the bioinformatic filtering. A total of 244 true positive records were obtained in the raw data, where a true positive is defined as a taxon that was either recorded in the surrounding vegetation or could be expected in the region. Furthermore, 181 false positives occurred in the raw data, defined as taxa that were neither recorded in the vegetation, nor expected in the region. Various ways to filter the data were explored, in order to remove false positive records, while retaining as many true positives as possible. The final selected filter criteria retained 75% of the true positives, while removing 90% of the false positives, which presented one of the better true to false positive ratios. Still, some true positives were removed, such as *Pinus*, who's sequences could be retained with different filter criteria, but at the cost of introducing more false positives. These results indicated that there is no single optimal way to filter the data and that different criteria must be applied to different experiments, depending on the sequence depth, number of PCR repeats, the controls used, and the barcode diversity obtained.

Some taxa that were expected in the region, were only reported based on eDNA records, but were not recorded in the vegetation surveys. These taxa included aquatic species that were likely growing in deeper parts of the lakes and smaller plants which could have been easily overlooked by the vegetation surveys. The high detection rate for the dominant taxa could be explained by their greater biomass compared to the rarer taxa. The high biomass would lead to more template DNA being contributed to the sediments which would result in a higher detection rate (Yoccoz *et al*. 2012, Barnes and Turner 2016). The exception to the above were the aquatic taxa, who were all well detected regardless of their scored abundance in the vegetation. This is most likely because there was no transport issue for their DNA to the lake sediments. The discrepancy observed for some taxa in the vegetation and their detection rate could be the result of primer biases and amplification success. Asteraceae are problematic due to primer mismatches and a general low biomass (Sønstebø *et al*. 2010). While groups such as Poaceae and Cyperaceae were problematic in this study, possibly due to a lack of material, however, both groups have been proven to work based on the used *trnL* P6 loop primers (Alsos *et al*. 2016, Zimmermann *et al*. 2017). Other taxa, such as the various *Salix* species, could only be identified to family level, as the barcode region does not have enough discriminatory power to identify the different species. Finally, some lakes performed worse than others, with only a handful of taxa observed based on eDNA and relatively few sequences surviving the filtering criteria used. There was no obvious difference between the lakes that performed well or had issues based on the lake characteristics, such as depth, catchment size or vegetation types present. A potential explanation could have been a considerable presence of algae in these lakes. The algal DNA could have competed with the DNA of terrestrial plants and thus lead to a lower vascular plant detection rate.

How does *sed*aDNA from plants compare to traditional palaeoecological methods?

The metabarcoding of plants from the Lake Uhca Rohči sediment core in Paper II identified a total of 118 different taxa, while pollen analysis on the other hand identified 39 different taxa. (Table 1.) The difference between the two methods was due to *sed*aDNA being able to identify a number of taxa to either genus or species level in families that are problematic based on pollen identifications, such as Poaceae, Cyperaceae, Caryophyllaceae and Callitrichaceae. Pollen analysis on the other hand could identify some additional groups such as the algae *Pediastrum*, which could not be amplified with the used *trnL* P6 loop primers. Both the metabarcode and pollen results indicated a relative stable flora during the Holocene. Where most plants came in early or were already present prior to the oldest sample and remained detected throughout the record. One exception was the presence of several

thermophilous taxa during the early and middle Holocene, such as *Limosella aquatica, Rhododendron tomentosum* and *Callitriche hermaphroditica*, which currently have more southern distributions and indicated a warmer climate than is present today. The remainder of the taxa indicated an early establishment of shrub-tundra vegetation, dominated by *Salix, Betula*, *Empetrum* and *Vaccinium* species, which are still present in the area today.

The shotgun *sed*aDNA data generated from the Lateglacial Hässeldala Port lake sediments in Paper IV was capable of identifying 51 different taxa. This number is a bit higher compared to the 45 and 22 taxa detected via pollen and macrofossil identification respectively. (Table 1.) Only four different taxa could be detected with all methods: *Betula*, *Salix*, *Myriophyllum cf. alterniflorum* and Nymphaea, with most taxa only being observed by one method. Though more overlap between the different methods could be observed at higher taxonomic levels, such as Rosaceae and Cyperaceae. Furthermore, a varying number of taxa were detected across the different time periods and identification methods. The *sed*aDNA results indicated that 31% of the taxa were new arrivals at the site during the Younger Dryas. Compared to the 2% and 12% indicated by the pollen and macrofossils respectively. However, all methods indicated that a large constituent of the taxa were already established during the Allerød and persisted during the Younger Dryas into the Preboreal.

Metabarcoding of the Lake Øvre Æråsvatnet sediment core in Paper V identified a total of 45 different taxa. Pollen and macrofossil analysis identified 60 and 19 taxa respectively. (Table 1.) Two major zones were identified based on the vegetation detected. First the lower half of the sediment record, that covered the LGM and extended up to 14,200 cal a BP, was rather species poor. The main taxon identified based on *sed*aDNA was the microalgal genus *Nannochloropsis,* with only a few scattered occurrences of *Papaver*, Apiaceae*,* Poaceae and *Potamogeton*. For the same period, pollen and macrofossil consisted primarily out of Poaceae and *Papaver*. The presence of some of these thermophilous taxa in the lower half of the sediment record, detected by all proxies, indicated at least some warmer phases during the LGM. The top half of the sediment record, ranging from 14,200 to 8200 cal a BP, contained a more species rich set of taxa. All proxies could detect taxa such as *Betula, Salicaceae* and *Filipendula*, while in particular, good results were obtained for aquatic taxa such as *Callitriche*, *Isoëtes*, *Menyanthes* and *Myriophyllum* based on *sed*aDNA. The proxies obtained in the top half suggested low arctic or shrub tundra conditions that eventually turned into a shrub birch forest around 11,700 cal a BP.

Paper	Analysis	Total taxa identified	Species	Genus	Higher than Genus
II	Metabarcoding	118	41%	47%	12%
	Pollen	39	5%	72%	23%
IV	Shotgun sequencing	51	22%	31%	47%
	Pollen	45	32%	44%	24%
	Macrofossils	22	45%	46%	9%
V	Metabarcoding	45	33%	53%	14%
	Pollen	60	22%	55%	23%
	Macrofossils	19	47%	32%	21%

Table 1: Overview of the number of taxa identified per paper and method, as well as a breakdown of the number of identifications per taxonomic rank.

Differences that were observed between the pollen and *sed*aDNA results in Paper II, and IV and V could be due to the different signals that these methods represent, i.e. a more regional signal for pollen (Rousseau *et al*. 2006, Hicks 2006, Parducci *et al*. 2015). However, there were also differences in the taxonomic resolution between the methods. Certain groups were problematic to identify according to either the metabarcode used or based on the pollen morphology. In addition, certain wind dispersed species could dominate the pollen results and potentially masked the presence of rarer plants. Regardless of these differences, similar vegetation trends were still observed between the pollen and *sed*aDNA results obtained in Paper II, which is in line with earlier studies (Parducci *et al*. 2015, Niemeyer *et al*. 2017).

The macrofossil results from Paper IV and V indicated some larger differences between it and the *sed*aDNA results, which was unexpected given how, based on previous studies, a reasonable overlap was observed between the methods (Jørgensen *et al*. 2012, Alsos *et al*. 2016). Some of these issues could be explained by either the *sed*aDNA method used or the DNA extracted. Only a small component of the shotgun sequence data (0,0003%) could be identified in Paper IV, which primarily yielded identifications to higher taxonomic groups. These issues were most likely due to the limited reference database used for the identifications. The reference only included chloroplast genomes, while it could be expected that the majority of the sequences material originates from the nuclear genome (Rauwolf *et al*. 2010). A more complete reference database that included several nuclear genomes would probably have resulted in both improved identifications to lower taxonomic ranks and more sequences assigned to various taxa.

The metabarcode results for the lower half of the cores in Paper V were rather species poor and were dominated by the microalgal genus *Nannochloropsis.* The high LOI and stable isotope values obtained for Lake Øvre Æråsvatnet, as well as the finds of bird bones indicated the presence of a nearby bird colony. The expected inflow of nutrients from bird guano could have resulted in eutrophication of the in-lake ecosystem and thus lead to algal blooms. Similar to the results obtained in Paper I, the overabundance of algal material could have swamped out the DNA from terrestrial sources and thus resulted in poorer metabarcode identifications. Shotgun sequencing that was carried out for some of these samples in Paper VI yielded no additional identifications for non-algal or microbial taxa, suggesting that both *sed*aDNA methods struggle to overcome this swamping effect. The results for Papers IV and V indicate that the different methods can be more complementary depending on the sediment conditions and *sed*aDNA methods used and that for now, no single method is able to detect all the taxa present.

Some improvements were detected based on *sed*aDNA over traditional identification methods. Both the metabarcode records in Papers II and V yielded abundant identifications for aquatic taxa. Most likely because the aquatic taxa contributed more material to the sediments and were thus easier to detect with *sed*aDNA (Sjögren *et al*. 2017). A similar conclusion was drawn from the contemporary eDNA results obtained in Paper I. The reliable aquatic detections provided by *sed*aDNA are potentially an important tool for temperature reconstructions, as it has been demonstrated that aquatic taxa give informative temperature approximations partly due to their ability to rapidly colonize new areas (Väliranta *et al*. 2015). Furthermore, shotgun sequencing was capable of detecting a high number of Younger Dryas introductions in Paper IV. This was most likely the result of the reference dataset used, which although it only consisted of chloroplast genomes, contained several genomes generated by the PhyloNorway project. The species included in the PhyloNorway project were primarily from northern Norway or were circumpolar in their distribution. The used set of reference genomes thus could have been better suited to identify the taxa that arrived during the colder Younger Dryas.

What animals can we detect based on metabarcoding of *sed*aDNA?

The *sed*aDNA results obtained in Paper III are complex given the mammal specific marker used and the taxa detected. Reindeer could be detected in both sediment records, but only in six different samples. Human on the other hand was detected throughout both cores and negative controls, even though a blocking primer was used to limit amplification. The most dominant taxonomic group, in

terms of the number of species detected, were clitellate worms, with nine different taxa. Eight of the worm species detected were either known to be cold tolerant or were described for more southern regions but could have occurred in the Polar Ural or Varanger areas during warmer climatic phases.

The poor mammal results were most likely due to the relatively low biomass that they have, compared to other taxonomic groups such as plants or invertebrates. This would also explain the presence of human, which could be detected even when a specific human blocking primer was used. Human amplification, though delayed, would eventually pick up as there was little other template material available (Boessenkool *et al*. 2012). The primers used here have been able to detect mammals based on studies that used *sed*aDNA from archaeological sites, where the amount of mammalian template is artificially increased due to human influence (Giguet-Covex *et al*. 2014). While other studies applied mammalian metabarcoding on different sediment types, such as permafrost, that perhaps preserved more mammalian material, but also utilized different metabarcoding primers (Haile *et al*. 2009, Willerslev *et al.* 2014). The detection of mammals from lake *sed*aDNA could potentially benefit from either a different maker, that co-amplifies less off-target taxa or is specific for a taxon of interest and is unable to amplify human. In addition, improved human blocking strategies, with multiple blocking primers compared to single one used in this study, could potentially result in more informative non-human detections.

Amplification of worms was unexpected given the mammal specific primers used. However, upon closer inspection, the 16S primers used were relatively conserved across all animals and could potentially amplify various off-target taxa. For worms, there were only two mismatches on average between the mammalian primers and the primer binding sites of the various clitellate worm families. As a result, the primers were capable of amplifying the off-target worm taxa, especially in the absences of sufficient mammalian template material (Schloss *et al.* 2011, Brown *et al.* 2015). The results obtained in Paper III indicated that metabarcoding “bycatch” could be informative and should not be outright ignored. Furthermore, it proves that the amplification of worms was possible from ancient sediments, which could be an interesting target for future studies as the various worm taxa could serve as a useful proxy for soil development and humidity (Edwards and Lofty 1977, Beylich and Graefe 2009).

Can we derive phylo- and population genomic data from shotgun sequencing?

The shotgun results for Andøya generated in Paper VI were in line with the metabarcode results from the same periods presented in Paper V. Both the mapping analysis as well as the metagenomic

approach identified the microalgae *Nannochloropsis* as the most abundant eukaryotic genus, with *N. limnetica* as the most likely species present in the lake. The ancient nature of the algae could be confirmed through the presence of deaminated sites. The only other taxon present in high sequences numbers was human, which was identified as modern contamination due to a lack of ancient DNA damage patterns.

The amount of sequence material present for *Nannochloropsis* allowed for the reconstruction of the palaeo organellar genomes. Based on these, additional consensus sequences were generated for each sample that incorporated the high and low frequency variants present. These sequences represented the extreme ends of the variation present in the samples but did not necessarily correspond to set haplotypes. The phylogenies between the reconstructed organellar genomes and the *Nannochloropsis* reference sequences indicated that the Andøya *Nannochloropsis* variant is evolutionary closest to *N. limnetica*. Similar results were obtained for a more traditional phylogeny using the plastid *rbcL* barcode marker that could be extracted from the reconstructed chloroplast genomes. However, some distance remained present between the *N. limnetica* references genomes and the reconstructed Andøya variant, which likely represented an unknown variant and was thus referred to as *N.* cf. *limnetica*.

The average proportions identified for the alternative alleles for both samples and reconstructed genomes was between 0.39 and 0.43, which suggest that there were at least two common *Nannochloropsis* variants present in Andøya. Inspection of linked variant sites on single sequences again identified the presence of two main haplotypes, though a number of locations were detected that supported the presence of three haplotypes. Furthermore, limited differences were observed between the variants for both samples, which suggests a relative population stability.

Previous palaeo organellar reconstructions were exclusively for mitochondrial genomes from either cave sediments, archaeological middens or latrines (Seersholm *et al*. 2016, Slon *et al*. 2017, Søe *et al*. 2018). The results presented here indicated that it was possible to reconstruct organellar genomes from lake *sed*aDNA based on shotgun sequence data. In addition, it represented the first reconstructed palaeo chloroplast genome. Even though two main haplogroups could be detected for both samples and organellar genomes, it remained unclear how many variants there were actually present. Part of the difficulty stems from the fact that the phasing for the single variants, as well as the different linked sites, was unknown. If all variable sites present were fully independent, the actual number of variants could be far higher. These issues were partly related to the relatively

limited amount of reference genomes available for the various *Nannochloropsis* species and strains. Additional reference information, especially for the species present, could help with the phasing of the variants and result in a more accurate estimation of the variants present.

sed*aDNA outlook and future methods*

One issue identified with metabarcoding is the variable taxonomic resolution between different groups. For example, the commonly used *trnL* P6 loop struggles with groups such as Poaceae, Cyperaceae and Salicaceae (Yoccoz *et al*. 2012). It seems unlikely that a single marker will solve these issues, but the combination of different markers could potentially shore up each other's taxonomic weaknesses. Such an alternative marker is required to work on the short ancient fragments, which rules out some commonly used environmental DNA markers. An example of a potential second marker would be a short variable part of the ITS region, which has proven to work for plants from ancient sediments (Willerslev *et al*. 2014). A second issue is that of the amplification of undesirable taxa, either through the amplification of unexpected off-target taxa, such as algae via plant metabarcoding or contaminant species like human when amplifying mammals. For the former, alternative markers that are more specific for the group of interest could help with the amplification of the desired taxonomic group, while limiting the amplification off-target taxa. For the latter, more elaborate blocking strategies might be used, such as dual blocking primers, or redesigning the primer sequences themselves.

Metabarcoding and shotgun sequencing have their own unique strengths and weaknesses, in terms of marker specificity, abundance estimation, primer reliance and authentication of the results. However, an intermediate approach exists. Target capture, or target enrichment, seeks to combine the specificity of metabarcoding with the abundance estimation and ancient DNA authentication provided by shotgun sequencing. Depending on the capture probe sequences, either a number of conserved regions can be targeted in order to identify a range of taxa, akin to metabarcoding (Dowle *et al*. 2016, Wilcox *et al*. 2018), or alternatively probes can be used to enrich for larger genomic regions or entire organellar genomes (Carpenter *et al*. 2013, Enk *et al*. 2014, Schmid *et al.* 2017, Slon *et al*. 2017). Target capture, however, requires access to usable reference sequences for both the probe design, as well as the identification of the sequenced material. The dependence on reference material is similar to the needs of shotgun sequencing. As more and more genomes become available, either organellar sequences through projects like the PhyloAlps or PhyloNorway projects, or nuclear genomes, both the identifications for shotgun sequencing, as well as the ability to design new metabarcode markers or capture probes will improve.

Any of the available *sed*aDNA methods could potentially be used for further genomic exploration of *sed*aDNA. For example, population estimations and tracking of populations through time and space. The above has been demonstrated based on metabarcoding from environmental DNA, but with relative long markers (Sigsgaard *et al*. 2016, Parsons *et al*. 2018), or based on target capture of macrofossils (Schmid *et al*. 2017). Shorter markers, shotgun sequencing or target capture should allow for application of the above methods in *sed*aDNA, as was demonstrated based on the amplification of species specific markers for Siberian larches (Epp *et al*. 2018). In addition, unidentified, or possibly extinct populations or taxa can be identified from sedimentary ancient DNA. Reconstruction of either standard markers or larger genomic regions will allow for phylogenetic placement of these taxa, without the need to find macrofossil remains (Seersholm *et al*. 2016, Slon *et al*. 2017). Finally, through the use of either different markers, or deep shotgun sequencing, whole ecosystems can be identified, rather than having a limited focus on a single group, though both more reference material as well as ecological knowledge for each species is required.

Conclusion

The presented papers in this thesis show the potential of *sed*aDNA for both different methods as well as taxonomic groups. The majority was carried out on plants (Papers I, II, IV and V). Based on the comparisons with either vegetation or pollen in Papers I and II, *sed*aDNA was capable of identifying the major vegetation components present. The different proxies seem more complementary in Papers IV and V, where there was some overlap between the methods, but some taxa were only identified by one or two methods. Part of the issue experienced might be related to the *sed*aDNA methods used or the sedimentary material present. The incorporation of different metabarcode regions, or expansion of the genomic reference material for shotgun sequencing should shore up some of the weaknesses detected for *sed*aDNA and allow for more robust identifications. Reliable detection and identification of mammals remains problematic due to low biomass and DNA contribution to the sediments. More elaborate blocking of human material or targeting of either shorter or species-specific fragments will potentially improve these results. Though the detection of worms demonstrates that they, as well as other invertebrates, could potentially be of more use for palaeoenvironmental reconstructions. Finally, there are numerous new research opportunities for *sed*aDNA that seek to utilize population genomic and phylogenetic methods, which will allow for more elaborate studies of species and populations.

References

Allentoft, M.E., Collins, M., Harker, D., *et al*. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings. Biological sciences / The Royal Society 279, 4724–4733.

Alsos, I.G., Sjögren, P., Edwards, M.E., *et al*. (2016) Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene 26, 627–642.

Barnes, M.A. & Turner, C.R. (2016) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17, 1–17.

Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L. & Lodge, D.M. (2014) Environmental conditions influence eDNA persistence in aquatic systems. Environmental science & technology 48, 1819–1827.

Berglund, B.E. & Ralska-Jasiewiczowa, M. (1986) Handbook of Holocene Palaeoecology and Palaeohydrology, Chichester: John Wiley & Sons.

Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PloS one 2, e197.

Birks, H.H. (2003) The importance of plant macrofossils in the reconstruction of Lateglacial vegetation and climate: examples from Scotland, western Norway, and Minnesota, USA. Quaternary Science Reviews 22, 453–473.

Birks, H.H. & Birks, H.J.B. (2000) Future uses of pollen analysis must include plant macrofossils. Journal of Biogeography 27, 31–35.

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

Beylich, A. & Graefe, U. (2009) Investigations of annelids at soil monitoring sites in Northern Germany: reference ranges and time-series data. Soil organisms 81, 175–196.

Blum, S.A.E., Lorenz, M.G. & Wackernagel, W. (1997) Mechanism of Retarded DNA Degradation and Prokaryotic Origin of DNases in Nonsterile Soils. Systematic and Applied Microbiology 20, 513–521.

Boessenkool, S., Epp, L.S., Haile, J., Bellemain, E., Edwards, M., Coissac, E., Willerslev, E. & Brochmann, C. (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. Molecular ecology 21, 1806–1815.

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

Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America 104, 14616–14621.

Brown, S.P., Veach, A.M., Rigdon-Huss, A.R., Grond, K., Lickteig, S.K., Lothamer, K., Oliver, A.K. & Jumpponen, A. (2015) Scraping the bottom of the barrel: are rare high throughput sequences artifacts? Fungal Ecology 13, 221–225.

Brüniche-Olsen, A., Jones, M.E., Burridge, C.P., Murchison, E.P., Holland, B.R. & Austin, J.J. (2018) Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil. Journal of Biogeography 45, 963–976.

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

Caporaso, J.G., Kuczynski, J., Stombaugh, J., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. Nature methods 7, 335–336.

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., *et al*. (2013) Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. American journal of human genetics 93, 852–864.

CBOL Plant Working Group, Hollingsworth, P.M., Forrest, L.L., *et al*. (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106, 12794–12797.

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, e13042.

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, 306–323.

Dowle, E.J., Pochon, X., C Banks, J., Shearer, K. & Wood, S.A. (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. Molecular ecology resources 16, 1240–1254.

Edwards, C.A. & Lofty, J.R. (1977) Biology of Earthworms, London: Chapman and Hall.

Eide, W., Birks, H.H., Bigelow, N.H., Peglar, S.M. & Birks, H.J.B. (2006) Holocene forest development along the Setesdal valley, southern Norway, reconstructed from macrofossil and pollen evidence. Vegetation History and Archaeobotany 15, 65–85.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.-N., Coissac, E.,

Boyer, F. & Leese, F. (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4, e1966.

England, L.S., Vincent, M.L., Trevors, J.T. & Holmes, S.B. (2004) Extraction, detection and persistence of extracellular DNA in forest litter microcosms. Molecular and cellular probes 18, 313–319.

Enk, J.M., Devault, A.M., Kuch, M., Murgha, Y.E., Rouillard, J.-M. & Poinar, H.N. (2014) Ancient Whole Genome Enrichment Using Baits Built from Modern DNA. Molecular Biology and Evolution 31, 1292–1294.

Epp, L.S., Boessenkool, S., Bellemain, E.P., *et al*. (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. Molecular ecology 21, 1821–1833.

Epp, L.S., Gussarova, G., Boessenkool, S., *et al*. (2015) Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. Quaternary Science Reviews 117, 152–163.

Epp, L.S., Kruse, S., Kath, N.J., Stoof-Leichsenring, K.R., Tiedemann, R., Pestryakova, L.A. & Herzschuh, U. (2018) Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. Scientific reports 8, 17436.

Faegri, K. & Iversen, J. (1989) Textbook of Pollen Analysis, Chichester: John Wiley and Sons.

Ficetola, G.F., Pansu, J., Bonin, A., *et al*. (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular ecology resources 15, 543–556.

Giguet-Covex, C., Pansu, J., Arnaud, F., *et al*. (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nature Communications 5.

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

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

Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M.T.P., Nielsen, R., Munch, K., Ho, S.Y.W., Shapiro, B. & Willerslev, E. (2007) Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? Molecular Biology and Evolution 24, 982–989.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A. &

Winfield, I.J. (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular ecology 25, 3101–3119.

Hansen, A.J., Mitchell, D.L., Wiuf, C., Paniker, L., Brand, T.B., Binladen, J., Gilichinsky, D.A., Rønn, R. & Willerslev, E. (2006) Crosslinks Rather Than Strand Breaks Determine Access to Ancient DNA Sequences From Frozen Sediments. Genetics 173, 1175–1179.

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

Heiri, O., Lotter, A.F. & Lemcke, G. (2001) Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. Journal of paleolimnology 25, 101–110.

Hicks, S. (2006) When no pollen does not mean no trees. Vegetation History and Archaeobotany 15, 253–261.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J. & Tappu, R. (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLOS Computational Biology 12, e1004957.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F. & Orlando, L. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684.

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

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

McKee, A.M., Spear, S.F. & Pierson, T.W. (2015) The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biological Conservation 183, 70–76.

Miller, D.N., Bryant, J.E., Madsen, E.L. & Ghiorse, W.C. (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Applied and environmental microbiology 65, 4715–4724.

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

Nesje, A. (1992) A Piston Corer for Lacustrine and Marine Sediments. Arctic and Alpine Research 24, 257.

Nichols, R.V., Vollmers, C., Newsom, L.A., Wang, Y., Heintzman, P.D., Leighton, M., Green, R.E. & Shapiro, B. (2018) Minimizing polymerase biases in metabarcoding. Molecular ecology resources.

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, e46–e62.

Odgaard, B.V. (1999) Fossil pollen as a record of past biodiversity. Journal of Biogeography 26, 7–17.

Pansu, J., Giguet-Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P. & Choler, P. (2015) Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Molecular ecology 24, 1485–1498.

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, 924–942.

Parducci, L., Jørgensen, T., Tollefsrud, M.M., *et al*. (2012) Glacial survival of boreal trees in northern Scandinavia. Science 335, 1083–1086.

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

Parsons, K.M., Everett, M., Dahlheim, M. & Park, L. (2018) Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. Royal Society open science 5, 180537.

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., *et al*. (2015) Ancient and modern environmental DNA. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 370, 20130383.

Pedersen, M.W., Ruter, A., Schweger, C., *et al*. (2016) Postglacial viability and colonization in North America's ice-free corridor. Nature 537, 45–49.

Rauwolf, U., Golczyk, H., Greiner, S. & Herrmann, R.G. (2010) Variable amounts of DNA related to the size of chloroplasts III. Biochemical determinations of DNA amounts per organelle. Molecular genetics and genomics: MGG 283, 35–47.

Reimer, P.J., Bard, E., Bayliss, A., *et al*. (2013) IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon 55, 1869–1887.

Röhl, U. & Abrams, L.J. (2000) High-resolution, downhole, and nondestructive core measurements from Sites 999 and 1001 in the Caribbean Sea: application to the Late Paleocene Thermal Maximum. Proceedings of the Ocean Drilling Program, 165 Scientific Results.

Rothwell, R.G., Guy Rothwell, R. & Croudace, I. w. (2015) Micro-XRF Studies of Sediment Cores: A Perspective on Capability and Application in the Environmental Sciences. In: Dordrecht: Springer, pp. 1–21.

Rousseau, D.-D., -D. Rousseau, D., Schevin, P., Duzer, D., Cambon, G., Ferrier, J., Jolly, D. & Poulsen, U. (2006) New evidence of long distance pollen transport to southern Greenland in late spring. Review of Palaeobotany and Palynology 141, 277–286.

Schloss, P.D., Gevers, D. & Westcott, S.L. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one 6, e27310.

Schloss, P.D., Westcott, S.L., Ryabin, T., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75, 7537–7541.

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W. & Alvarez, N. (2017) HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods in Ecology and Evolution 8, 1374–1388.

Schnell, I.B., Bohmann, K. & Gilbert, M.T.P. (2015) Tag jumps illuminated--reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular ecology resources 15, 1289–1303.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., André Levesque, C., Chen, W. & Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences of the United States of America 109, 6241–6246.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A.S., Willerslev, E., Krogh, A. & Orlando, L. (2012) Improving ancient DNA read mapping against modern reference genomes. BMC genomics 13, 178.

Seersholm, F.V., Pedersen, M.W., Søe, M.J., *et al*. (2016) DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. Nature communications 7, 13389.

Seppä, H. & Birks, H.J.B. (2001) July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: pollen-based climate reconstructions. The Holocene 11, 527–539.

Sigsgaard, E.E., Nielsen, I.B., Bach, S.S., *et al*. (2016) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nature ecology & evolution 1, 4.

Simpson, J.T. & Durbin, R. (2012) Efficient de novo assembly of large genomes using compressed data structures. Genome Research 22, 549–556.

Sjögren, P., Edwards, M.E., Gielly, L., Langdon, C.T., Croudace, I.W., Merkel, M.K.F., Fonville, T. &

Alsos, I.G. (2017) Lake sedimentary DNA accurately records 20 Century introductions of exotic conifers in Scotland. The New phytologist 213, 929–941.

Slon, V., Hopfe, C., Weiß, C.L., *et al*. (2017) Neandertal and Denisovan DNA from Pleistocene sediments. Science 356, 605–608.

Søe, M.J., Nejsum, P., Seersholm, F.V., *et al*. (2018) Ancient DNA from latrines in Northern Europe and the Middle East (500 BC-1700 AD) reveals past parasites and diet. PloS one 13, e0195481.

Soininen, E.M., Gauthier, G., Bilodeau, F., *et al*. (2015) Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. PloS one 10, e0115335.

Sønstebø, J.H., Gielly, L., Brysting, A.K., *et al*. (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Molecular ecology resources 10, 1009–1018.

Svendsen, J.I., Færseth, L.M.B., Gyllencreutz, R., *et al*. (2019) Glacial and environmental changes over the last 60 000 years in the Polar Ural Mountains, Arctic Russia, inferred from a high-resolution lake record and other observations from adjacent areas. Boreas 48, 407–431.

Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. (2018) Environmental DNA, Oxford: Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular ecology 21, 2045–2050.

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

Taberlet, P., Prud'Homme, S.M., Campione, E., *et al*. (2012) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. Molecular ecology 21, 1816–1820.

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

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

Valentini, A., Miquel, C., Nawaz, M.A., *et al*. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Molecular ecology resources 9, 51–60.

Valentini, A., Taberlet, P., Miaud, C., *et al*. (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular ecology 25, 929–942.

Väliranta, M., Salonen, J.S., Heikkilä, M., *et al*. (2015) Plant macrofossil evidence for an early onset of the Holocene summer thermal maximum in northernmost Europe. Nature Communications 6.

Westergaard, K.B., Alsos, I.G., Popp, M., Engelskjøn, T., Flatberg, K.I. & Brochmann, C. (2011) Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. Molecular ecology 20, 376–393.

Wilcox, T.M., Zarn, K.E., Piggott, M.P., Young, M.K., McKelvey, K.S. & Schwartz, M.K. (2018) Capture enrichment of aquatic environmental DNA: A first proof of concept. Molecular ecology resources 18, 1392–1401.

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

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

Willis, K.J. & Birks, H.J.B. (2006) What is natural? The need for a long-term perspective in biodiversity conservation. Science 314, 1261–1265.

Wohlfarth, B., Muschitiello, F., Greenwood, S.L., Andersson, A., Kylander, M., Smittenberg, R.H., Steinthorsdottir, M., Watson, J. & Whitehouse, N.J. (2017) Hässeldala - a key site for Last Termination climate events in northern Europe. Boreas 46, 143–161.

Yoccoz, N.G., Bråthen, K.A., Gielly, L., *et al*. (2012) DNA from soil mirrors plant taxonomic and growth form diversity. Molecular ecology 21, 3647–3655.

Ziesemer, K.A., Mann, A.E., Sankaranarayanan, K., *et al*. (2015) Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. Scientific reports 5, 16498.

Zimmermann, H.H., Raschke, E., Epp, L.S., Stoof-Leichsenring, K.R., Schwamborn, G., Schirrmeister, L., Overduin, P.P. & 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, 575–596.

Paper I

PLOS ONE

RESEARCH ARTICLE

Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation

Inger Greve Alsos[1]*, Youri Lammers[1], Nigel Giles Yoccoz[2], Tina Jørgensen[1], Per Sjögren[1], Ludovic Gielly[3,4], Mary E. Edwards[1,5]

1 Tromsø Museum, University of Tromsø –The Arctic University of Norway, Tromsø, Norway, **2** Department of Arctic and Marine Biology, University of Tromsø –The Arctic University of Norway, Tromsø, Norway, **3** University Grenoble Alpes, LECA, Grenoble, France, **4** CNRS, LECA, Grenoble, France, **5** Geography and Environment, University of Southampton, Highfield, Southampton, United Kingdom

* inger.g.alsos@uit.no

OPEN ACCESS

Citation: Alsos IG, Lammers Y, Yoccoz NG, 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(4): e0195403. https://doi.org/10.1371/journal.pone.0195403

Editor: Christopher Carcaillet, Ecole Pratique des Hautes Etudes, FRANCE

Received: December 21, 2017

Accepted: March 21, 2018

Published: April 17, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except the raw data files, which are available from Dryad using the following DOI: 10.5061/dryad.g72v731.

Funding: The work was supported by the Research Council of Norway (grant nos. 213692/F20, 250963/F20 and 230617/E10 to Alsos). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Metabarcoding of lake sediments have been shown to reveal current and past biodiversity, but little is known about the degree to which taxa growing in the vegetation are represented in environmental DNA (eDNA) records. We analysed composition of lake and catchment vegetation and vascular plant eDNA at 11 lakes in northern Norway. Out of 489 records of taxa growing within 2 m from the lake shore, 17–49% (mean 31%) of the identifiable taxa recorded were detected with eDNA. Of the 217 eDNA records of 47 plant taxa in the 11 lakes, 73% and 12% matched taxa recorded in vegetation surveys within 2 m and up to about 50 m away from the lakeshore, respectively, whereas 16% were not recorded in the vegetation surveys of the same lake. The latter include taxa likely overlooked in the vegetation surveys or growing outside the survey area. The percentages detected were 61, 47, 25, and 15 for dominant, common, scattered, and rare taxa, respectively. Similar numbers for aquatic plants were 88, 88, 33 and 62%, respectively. Detection rate and taxonomic resolution varied among plant families and functional groups with good detection of e.g. Ericaceae, Roseaceae, deciduous trees, ferns, club mosses and aquatics. The representation of terrestrial taxa in eDNA depends on both their distance from the sampling site and their abundance and is sufficient for recording vegetation types. For aquatic vegetation, eDNA may be comparable with, or even superior to, in-lake vegetation surveys and may therefore be used as an tool for biomonitoring. For reconstruction of terrestrial vegetation, technical improvements and more intensive sampling is needed to detect a higher proportion of rare taxa although DNA of some taxa may never reach the lake sediments due to taphonomical constrains. Nevertheless, eDNA performs similar to conventional methods of pollen and macrofossil analyses and may therefore be an important tool for reconstruction of past vegetation.

Competing interests: The authors would like to mention that LG is one of the co-inventors of patents related to g-h primers and the subsequent use of the P6 loop of the chloroplast trnL (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Environmental DNA (eDNA), DNA obtained from environmental samples rather than tissue, is a potentially powerful tool in fields such as modern biodiversity assessment, environmental sciences, diet, medicine, archaeology, and paleoecology [1–4]. Its scope has been greatly enlarged by the emergence of metabarcoding: massive parallel next generation DNA sequencing for the simultaneous molecular identification of multiple taxa in a complex sample [5]. The advantages of metabarcoding in estimating species diversity are many. It is cost-effective, it has minimal effect on the environment during sampling, and data production (though not interpretation) is independent of the taxonomic expertise of the investigator [4, 6]. It may even out-perform traditional methods in the detection of individual species [7, 8]. Nevertheless, the discipline is still in its infancy, and we know little about the actual extent to which species diversity is represented in the eDNA records [9, 10]. This study assesses representation of modern vegetation by eDNA from lake sediments.

DNA occurs predominantly within cells but is released to the environment upon cell membrane degradation [4]. It may then bind to sediment components such as refractory organic molecules or grains of quartz, feldspar and clay [11]. It can be detected after river transport over distances of nearly 10 km [9, 12]. When released into the environment, degradation increases exponentially [9, 13], so eDNA from more distant sources is likely to be of low concentration in a given sample. Once in the environment, preservation ranges from weeks in temperate water, to hundreds of thousands years in dry, frozen sediment [4]. Preservation depends on factors such as temperature, pH, UV-B levels, and thus lake depth [14–16]. Even when present, many factors affect the probability of correct detection of species in environmental samples, for example: the quantity of DNA [8, 17], the DNA extraction and amplification method used [7, 18], PCR and sequencing errors, as well as the reference library and bioinformatics methods applied [4, 18–20]. If preservation conditions are good and the methods applied adequate, most or all species present may be identified and the number of DNA reads may even reflect the biomass of species [6, 7, 21], making this a promising method for biodiversity monitoring.

When applied to late-Quaternary sediments, eDNA analysis may help disclose hitherto inaccessible information, thus providing promising new avenues of palaeoenvironmental reconstruction [22, 23]. Lake sediments are a major source of palaeoenvironmental information [24] and, given good preservation, DNA in lake sediments can provide information on biodiversity change over time [4, 22, 25]. However, sedimentary ancient DNA is still beset by authentication issues [2, 10]. For example, the authenticity and source of DNA reported in several recent studies have been questioned [26–30]. As with pollen and macrofossils [31, 32], we need to understand the source of the DNA retrieved from lake sediments and know which portion of the flora is represented in DNA records.

The P6 loop of the plastid DNA *trn*L (UAA) intron [33] is the most widely applied marker for identification of vascular plants in environmental samples such as Pleistocene permafrost samples [34–36], late-Quaternary lake sediments [15, 22, 27, 37–41], sub-modern or modern lake sediments [42], animal faeces [43, 44], and sub-modern or modern soil samples [6, 45]. While some studies include comparator proxies to assess the ability of DNA to represent species diversity (e.g., [35, 41, 46, 47], only one study has explicitly tested how well the floristic composition of eDNA assemblages reflect the composition of extant plant communities [6], and similar tests are urgently needed for lake sediments. Yoccoz *et al*. found most common species and some rare species in the vegetation were represented in the soil eDNA at a subarctic site in northern Norway. The present study attempts a similar vegetation-DNA calibration in relation to lake sediments.

We retrieved sedimentary eDNA and recorded the vegetation at 11 lakes that represent a gradient from boreal to alpine vegetation types in northern Norway. We chose this area because DNA is best preserved in cold environments and because an almost complete reference library is available for the relevant DNA sequences for arctic and boreal taxa [34, 36]. Our aims were to 1) increase our understanding of eDNA taphonomy by determining how abundance in vegetation and distance from lake shore affect the detection of taxa, and 2) examine variation in detection of DNA among lakes and taxa. Based on this, we discuss the potential of eDNA from lake sediments as a proxy for modern and past floristic richness.

Materials and methods

Study sites

Eleven lakes were selected using the following criteria: 1) lakes size within the range of lakes studied for pollen in the region and with limited inflow and outflow streams; 2) a range of vegetation types from boreal forest to alpine heath was represented; and 3) lakes sediments are assumed to be undisturbed by human construction activity (Figs 1 and 2). Six of the lakes were selected also for the availability of pollen, macro and/or ancient DNA analyses [27, 48–52]. Data on catchment size, altitude, yearly mean temperature, mean summer temperature and yearly precipitation were gathered using NEVINA (http://nevina.nve.no/) from the Norwegian Water Resources and Energy Directorate (NVE, https://www.nve.no). Lake size was calculated using http://www.norgeibilder.no/. Number and size of inlets and outlets were noted during fieldwork.

Vegetation surveys

We attempted to record all species growing within 2 m from the lakeshore. This was a practically achievable survey, and data are comparable among sites. Aquatics were surveyed from the

Fig 1. Location of the studied lakes in Norway.

https://doi.org/10.1371/journal.pone.0195403.g001

Fig 2. Study lakes in northern Norway. a) A-tjern, b) Brennskogtjørna, c) Einletvatnet, d) Finnvatnet, e) Gauptjern, f) Jula Jávrí, g) Lakselvhøgda, h) Lauvås, i) Øvre Æråsvatnet, j) Paulan Jávri, k) Rottjern, l) Tina Jørgensen sampling surface sediments with Kajak corer. Photo: I.G. Alsos.

https://doi.org/10.1371/journal.pone.0195403.g002

boat using a "water binocular" and a long-handled rake, while rowing all around smaller lakes and at least half way around the three largest lakes. We also surveyed a larger part of the catchment vegetation. For this, we used aerial photos (http://www.norgeibilder.no) to identify polygons of relatively homogeneous vegetation (including the area within 2 m). In the field we surveyed each polygon and classified observed species giving them the following abundance scorers: rare (only a few ramets), scattered (ramets occur throughout but at low abundance), common (common throughout but not the most abundant ones), or dominant (making up the majority of the biomass of the field, shrub or tree layer). The area covered and intensity of these broad-scale vegetation surveys varied among lakes due to heterogeneity of the vegetation, catchment size and time constraints. They mainly represent the vegetation within 50 m of the lakeshore. Sites were revisited several times during the growing season to increase the detection rate. For each lake our dataset consisted of a taxon list for 1) the <2-m survey, 2) the extended

Table 1. Characteristics of lakes where vegetation surveys and lake sediment DNA analyses were performed.

Lakes	District	Habitat type	Catchment area (km^2)	Alt. (m a.s.l.)	Lake size (ha)	Water depth (m)	Yearly mean (˚C)	Summer mean (˚C)	Yearly prec. (mm)	Inlets	N lat.	E lat.
A-tjern[a]	Dividalen	Mixed forest/mire, tall herbs	0.17	125	1.70	5.5	-0.8	6.9	636	3	68.996	19.486
Brennskogtjønna	Dividalen	Pine forest, heath	1.20	311	10.64	20.0	-0.9	6.4	457	2	68.859	19.594
Einletvatnet	Andøya	Mires, patches of birch forest	1.26	35	27.00	4 (6.7)	3.7	8.8	1025	5 minor	69.258	16.071
Finnvatnet	Kvaløya	Birch forest/mire	0.20	158	0.86	2.0	2.7	7.9	1005	3–4 minor	69.778	18.612
Gauptjern	Dividalen	Sub-alpine mixed forest, tall and low herbs	0.07	400	0.79	4.0	-0.9	6.5	451	2	68.856	19.618
Jula Jávri[c]	Kåfjorddalen	Alpine heath and mire	1.05	791	0.04	1.7	-3.6	3.9	670	2–5 minor	69.365	21.099
Lakselvhøgda	Ringvassøya	Alpine heath and mire, scattered birch forest	0.06	143	0.77	2.0	2.5	7.2	977	0	69.927	18.846
Lauvås	Ringvassøya	Heath, mire and mesic herb birch forest	0.41	4	0.71	3.3	2.7	7.5	971	2	69.946	18.860
Øvre Æråsvatnet	Andøya	Mires and birch forest, conifers planted	3.60	43	24.00	9.5	3.4	8.3	1027	3	69.256	16.034
Paulan Jávri	Kåfjorddalen	Alpine heath	0.56	746	0.22	2.0	-3.7	3.7	662	1+1 minor	69.399	21.015
Rottjern[b]	Dividalen	Mixed forest, tall herbs	0.96	126	1.91	3.0	-0.3	7.6	619	2	68.983	19.477

All lakes are in northern Norway. Water depth given for sampling site in the centre of the lake; deepest point in brackets if different. "Summer" is May-September, "Alt." is altitude, "prec." is precipitation, and "N lat." and "E. lat" are northern and eastern latitude, respectively. Mixed forest is forest dominated by birch but with some Pine.

[a]Named A-tjern in Jensen& Vorren 2008. Named "Vesltjønna" on NEVINA but this name is not official.

[b]Named B-tjern in Jensen& Vorren 2008, but later official named Rottjern.

[c]Catchment area could not be calculated using NEVINA so this was done in http://norgeskart.no. Temperature and precipitation data were taken from the nearby Goulassaiva.

https://doi.org/10.1371/journal.pone.0195403.t001

survey consisting of observations from <2 m and the polygons, 3) an abundance score based on the highest abundance score from any polygon at that lake. Taxonomy follows [53, 54].

Sampling lake sediments

Surface sediments were collected from the centres of the lakes between September 21st and October 1st, 2012, using a Kajak corer (mini gravity corer) modified to hold three core tubes spaced 15 cm apart, each with a diameter of 3 cm and a length of 63 cm (Fig 2, Table 1). The core tubes were washed in Deconex22 LIQ-x and bleached prior to each sampling. The top 8 cm sediments were extruded in field. Samples of ca. 25 mL were taken in 2-cm increments and placed in 50-ml falcon tubes using a sterilized spoon. All samples were frozen until extraction.

DNA extraction and amplification

For each lake, we analysed the top 0–2 cm of sediment separately from two of the three core tubes (n = 22). Twenty extra samples from lower in the cores were also analysed. The main down-core results will be presented in a separate paper in which we compare eDNA records

with the pollen analyses by [49]. Taxa that were only identified from lower levels in the cores are noted in S1 Table. Samples were thawed in a refrigerator over 24–48 hours, and 4–10 g were subsampled for DNA. The 42 samples and 6 extraction negative controls underwent extraction at the Department for Medical Biology, University of Tromsø, in a room where no previous plant DNA work had been done. A PowerMax Soil DNA Isolation kit (MO BIO Laboratories, Carlsbad, CA, USA) was used following the manufacturer´s instructions, with water bath at 60˚C and vortexing for 40 min.

All PCRs were performed at LECA (Laboratoire d'ECologie Alpine, University Grenoble Alpes), using the *g* and *h* universal plant primers for the short and variable P6 loop region of the chloroplast *trn*L (UAA) intron [33]. Primers include a unique flanking sequence of 8 bp at the 5' end (tag, each primer pair having the same tag) to allow parallel sequencing of multiple samples [55, 56].

PCR and sequencing on an Illumina 2500 HiSeq sequencing platform follows [41]. DNA amplifications were carried out in 50 μl final volumes containing 5 μl of DNA sample, 2 U of AmpliTaq Gold DNA Polymerase (Life Technologies, Carlsbad, CA, USA), 15 mM Tris-HCl, 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.2 μM each primer and 8 μg Bovine Serum Albumin. All PCR samples (DNA and controls) were randomly placed on PCR plates. Following the enzyme activation step (10 min at 95˚C), PCR mixtures underwent 45 cycles of 30 s at 95˚C, 30 s at 50˚C and 1 min at 72˚C, plus a final elongation step (7 min at 72˚C). using six PCR negative controls and two positive controls, and six different PCR replicates for each of the 56 samples, giving a total of 336 PCR samples, of which 216 represent the upper 0–2 cm. Equal volumes of PCR products were mixed (15 μl of each), and ten aliquots of 100 μl of the resulting mix were then purified using MinElute Purification kit (Qiagen GmbH, Hilden, Germany). Purified products were then pooled together before sequencing; 2×100+7 paired-end sequencing was performed on an Illumina HiSeq 2500 platform using TruSeq SBS Kit v3 (FASTERIS SA, Switzerland).

DNA sequences analysis and filtering

Initial filtering steps were done using OBITools [57] following the same criteria as in [41, 42] (S2 Table). We then used *ecotag* program [57] to assign the sequences to taxa by comparing them against a local taxonomic reference library containing 2445 sequences of 815 arctic [34] and 835 boreal [36] vascular plant taxa; the library also contained 455 bryophytes [44]. We also made comparisons with a second reference library generated after running *ecopcr* on the global EMBL database (release r117 from October 2013). Only sequences with 100% match to a reference sequence were kept. We excluded sequences matching bryophytes as we did not include them in the vegetation surveys. We used BLAST (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/blast/) to check for potential wrong assignments of sequences.

When filtering next-generation sequencing data, there is a trade-off between losing true positives (TP, sequences present in the samples and correctly identified) and retaining false positives (FP, sequences that originate from contamination, PCR or sequencing artefacts, or wrong match to database) [17, 20, 58]. We therefore assessed the number of TP and FP when applying different last step filtering criteria. We initially used two spatial levels of comparison with the DNA results: i) data from our vegetation surveys and ii) the regional flora (i.e., species in the county of Nordland and Troms as listed by the Norwegian Bioinformation Centre (http://www.biodiversity.no/). For any lake, both datasets are likely incomplete, as inconspicuous species may be lacking in the regional records [59] and our vegetation surveys did not include the entire catchment area. Nevertheless, the exercise is useful for evaluating how many FPs and TPs are lost by applying different filtering criteria. We defined true positives as

sequences that matched a species recorded in the vegetation surveys at the same lake, being aware that this is an under-representation, as the vegetation surveys likely missed species. We defined false positives as species recorded neither in the vegetation surveys nor the regional flora. We tested the effect of different rules of sequence removal: 1) found as $\leq$1,$\leq$5 or $\leq$10 reads in a PCR repeat, 2) found as $\leq$1,$\leq$2 or $\leq$3 PCR repeats for a lake sample, 3) occurring in more than one of 72 negative control PCR replicates, 4) on average, higher number of PCR repeats in negative controls than in sample, and 5) on average a higher number of reads in negative controls than in samples (S2 Table). The filtering criteria resulting in overall highest number of true positives kept compared to false positives lost were applied to all lakes. These were removing sequences with less than 10 reads, less than 2 PCR repeats in lake samples, and on average a lower number of reads in lake samples than in negative controls.

Data analyses and statistics

After data filtering, we compared taxon assemblages from DNA amplifications with the taxa recorded in the vegetation surveys. To make this comparison, taxa in the vegetation surveys were lumped according to the taxonomic resolution of the P6 loop (S1 Table), and the comparison was done at the lowest resultant taxonomic level. The majority of results explore only presence/absence (taxa richness); quantitative data are given in tables (including Supporting Information).

Multivariate ordinations (Correspondence Analysis and Non-symmetric Correspondence Analysis, the latter giving more weight to species present in more lakes; [60, 61]), were run independently on the vegetation data (present/absent using only taxa recorded within 2m) and eDNA data (present/absent). The similarity between ordinations of vegetation and eDNA data was assessed using Procrustes analysis [62], as implemented in the functions procrustes() and protest() in R library vegan [63].

To estimate the percentages of false negatives and positives in the DNA data and in the vegetation survey, we used the approach described in [64]. If we define the probability of a DNA false positive as p_{DNA_0}, the detectability by DNA as p_{DNA_1}, the detectability in the vegetation survey as p_{VEG_1}, and the probability that a species is present as p_{OCC}, we can state that the four probabilities of observing Presence(1)/Absence(0) in the DNA and Vegetation are as follows:

$$\begin{aligned} &Prob(\text{DNA} = 0, \text{Vegetation} = 0) \\ &\quad = (1 - \text{Pocc})(1 - \text{PDNA\_0}) + \text{Pocc}(1 - \text{PDNA\_1})(1 - \text{PVEG\_1}) \end{aligned} \quad (1)$$

In this case, if the species is absent in both the DNA and vegetation, it is either absent with probability (1- p_{OCC}) and no false positive has occurred with probability (1- p_{DNA_0}), or it is present with probability p_{OCC}, but was not detected both in the DNA with probability (1- p_{DNA_1}) and in the vegetation with probability (1- p_{VEG_1}).

$$Prob(\text{DNA} = 0, \text{Vegetation} = 1) = \text{Pocc}(1 - \text{PDNA\_1})\text{PVEG\_1} \quad (2)$$

In this case, the species is present, not detected in DNA but detected in the vegetation survey.

$$Prob(\text{DNA} = 1, \text{Vegetation} = 0) = (1 - \text{Pocc})\text{PDNA\_0} + \text{Pocc}\text{PDNA\_1}(1 - \text{PVEG\_1}) \quad (3)$$

In this case, the species is either absent and is a false DNA positive, or is present, detected by DNA but not in the vegetation survey.

$$Prob(\text{DNA} = 1, \text{Vegetation} = 1) = \text{Pocc}\text{PDNA\_1}\text{PVEG\_1} \quad (4)$$

In this case, the species is present and is detected both in the DNA and the vegetation survey.

We assumed the four probabilities varied only among lakes, not among species. We also restricted the analyses to species that were detected at least once using DNA, because for species that were never detected using eDNA, different processes might be important. For p_{DNA_1}, we also considered a model assuming a logistic relationship between p_{DNA_1} and lake characteristics, such as lake depth or catchment area, that is: $\text{logit}(p_{DNA\_1}) = b_0 + b_1$ Lake Covariate. We fitted these models using Bayesian methods, using uninformative priors (uniform distributions on the [0,1] interval) for the false positive/negative rates for DNA, and an informative prior for the detectability in the vegetation survey (uniform prior on the [0.8,1] interval, as detectability was high in the vegetation survey, but we had no repeated surveys or time to detection available to estimate it). We used the R package rjags to run the MCMC simulations [64]. Model convergence was assessed using the Gelman-Rubin statistics [65], values of which were all ~1.0.

Results

Vegetation records

The vegetation surveys provided 2316 observations of 268 taxa, including hybrids, subspecies, and uncertain identifications. Of these, 97 taxa share sequences with one or more other taxa (e.g., 20 taxa of *Carex* and 15 of *Salix*). Another nine taxa were not in the reference library (e.g. *Cicerbita alpina*), and eight taxa could not be matched due to incomplete identification in the vegetation survey. Eight taxa of *Equisetum* were filtered out due to short sequence length. This left 171 taxa that could potentially be recognized by the technique we used (S1 Table). For the 11 sites, between 31 and 58 taxa were potentially identifiable (Table 2), and this value was positively correlated with vegetation species richness ($y = 0.67x+10.3$, $r^2 = 0.93$, $p<0.0001$, $n = 11$). Taxonomic resolution at species level was 77–93% (mean 88%) and 65–79% (mean 74%) for the <2 m and extended (i.e., combined) vegetation surveys, respectively.

Table 2. Number of records in vegetation and eDNA per lake.

Lake	Raw reads per sample	Reads after filtering per sample	Veg. <2 m	Identifiable Veg.<2 m	Tot. DNA	eDNA match Veg.	% Veg. <2 m detected in eDNA	% eDNA detected in Veg.	Additional identifiable extended surveys	Additional eDNA Veg match extended survey
A-tjern	706 954	280 277	56	51	30	25	49	83	14	1
Brennskogtjønna	919 672	584 537	75	58	23	17	29	74	15	2
Einletvatnet	700 805	411 923	59	50	27	22	44	82	18	1
Finnvatnet	516 878	31 288	47	40	16	10	25	63	13	3
Gauptjern	673 977	279 752	47	45	22	17	38	77	18	3
Jula Jávri	669 351	161 871	36	31	11	4	13	36	31	2
Lakselvhøgda	613 386	4 880	41	37	10	9	24	90	14	1
Lauvås	250 979	3 453	44	41	12	7	17	58	27	5
Øvre Æråsvatnet	744 618	340 976	64	54	24	20	37	83	40	2
Paulan Jávri	747 665	178 532	43	40	17	10	25	59	34	2
Rottjern	580 970	222 649	47	42	25	17	41	68	24	3
Sum	7 125 255	2 500 138	559	489	217	158			248	25
Mean	647750	227285	50.8	44.5	19.7	14.4	31.1	70.3	22.5	2.3
Highest/lowest	3.7	169.3	2.1	1.9	3	6.3	3.8	2.5	3.1	5

Taxa in the vegetation surveys (Veg.), number of taxa that could potentially be identified with the applied molecular marker used and available reference database, and taxa actually identified in the eDNA. The results are given for vegetation surveys <2 m from lakeshore (including aquatics) and for additional taxa recorded in extended surveys. Raw reads refer to all reads assigned to samples (S1 Table). The ratio between the highest and lowest value on each category is given as a indicator of variation among lakes.

https://doi.org/10.1371/journal.pone.0195403.t002

Of 489 records <2 m from the lakeshore, the majority were rare (148) or scattered (146) in the vegetation; fewer were common (131) or dominant (64). An additional 245 observations of 46 taxa came from >2 m from the lakeshore (156 rare, 68 scattered, 19 common and 2 dominant).

Molecular data

The numbers of sequences matching entries in the regional arctic-boreal and EMBL-r117 databases were 227 and 573 at 98% identity, respectively. For sequences matching both databases, we retained the arctic-boreal identification; this resulted in 11,236,288 reads of 301 sequences having 100% sequence similarity with the reference libraries and at least 10 reads in total (S2 Table). There were 244 and 181 records of sequences (each sequence occurring in 1–11 of the lakes) that with certainty could be defined as true or false positive, respectively (see methods). We found no combination of filtering criteria that only filtered out the false positives without any loss of true positives (S3 Table). The best ratio was obtained when retaining sequences that were on average more common in samples than in negative controls, plus with at least two PCR replicates in one sample and at least 10 reads per PCR replicate. Applying these criteria filtered out 163 false positives leaving only three false positive taxa (Annonaceae, Meliaceae and Solanaceae) recorded in total 18 times in the 11 lakes. These were then removed as obvious contamination. However, it also removed 61 (25%) true positives, e.g., *Pinus*, which had high read numbers at lakes in pine forest and low ones at lakes where it is probably brought in as firewood, but which also occurred with high read numbers in two of the negative controls (S4 Table). After this final filtering, 2,500,138 reads of 56 unique sequences remained. Sequences matching to the same taxa in the reference library were merged, resulting in 47 final taxa (Table 3). Taking into account that some species within some genera shared sequences, for example *Carex* and *Salix*, these may potentially represent 81 taxa (S1 Table).

The read numbers are sum of two DNA extractions with 6 PCR replicates for each. All read numbers are after the filtering steps in S2 Table. Note that the records of *Chamaeodaphne calyculata* are likely to represent PCR or sequencing errors of *Andromeda polyfolia* (S1 Appendix). For taxa only recorded in vegetation and/or filtered out of the eDNA records, see S1 Table. The lakes names are A-tjern (A-tj), Brennskogtjørna (Bren), Einletvatnet (Einl), Finnvatnet (Finn), Gauptjern (Gaup), Jula Jávri (Jula), Lakselvhøgda (Laks), Lauvås (Lauv), Øvre Æråsvatnet (Ovre), Paulan Jávri (Paul), and Rottjern (Rott).

In our positive control, 7 out of 8 species were detected in all replicates (S5 Table). Only *Aira praecox*, which was added with the lowest DNA concentration, could not be detected. This indicates that the PCR and sequencing was successful for taxa with an extracted DNA concentration of ≥0.03 ng/μL (S5 Table).

The gain in number of taxa when analysing two cores instead of one was 2.5±1.2 per lake. All data presented here are based on the upper 0–2 cm of sediment of two cores combined (but not from deeper levels as these were not sampled at all sites). This gave an average of 19.7 ±6.9 taxa (range 10–30) per lake (Table 2). Samples from below 2-cm depth provide an additional 14 records of 42 taxa, some not recorded in 0–2 cm samples (S1 Table).

Detection of taxa in eDNA

Of the 217 eDNA records, the majority matched taxa recorded within 2 m of the lake shore (Fig 3A). Higher proportions of dominant or common taxa were detected in DNA compared with scattered or rare ones (Fig 3B). Most dominant taxa, such as *Betula*, *Empetrum nigrum*, *Vaccinium uliginosum*, and *Salix*, were correctly detected at most or all lakes (Table 3), whereas some were filtered out (*Equisetum* spp., *Pinus sylvestris*, many *Poa*, S1 Table). Of dominants,

Table 3. Read numbers per taxa and per lake, and the sum per taxa for all lakes.

Family	Taxa	A-tj	Bren	Einl	Finn	Gaup	Jula	Laks	Lauv	Ovre	Paul	Rott	Sum
Asteraceae	*Crepis paludosa*									455			455
Betulaceae	*Alnus incana*	48 183	117 855			40 802					131	15 710	222 681
Betulaceae	*Betula* spp.	126 727	120 369	40 991	5 630	101 688		144	32	31 639	3 263	16 283	446 766
Caryophyllaceae	*Sagina* sp.		46	10	37		18			10	24		145
Cornaceae	*Chamaepericlymenum suecicum*			338									338
Cupressaceae	*Juniperus communis*	261	752						45		27		1 085
Cyperaceae	*Carex lasiocarpa*	47				76						84	207
Cyperaceae	*Carex* spp.	34		48	33	72							187
Dryopteridaceae	*Dryopteris* spp.	10 088	16 947	6 406	6 781	5 882	87	1 886	1 141	6 252	216	5 239	60 925
Ericaceae	*Andromeda polifolia*	191		235		244			23			310	1 003
Ericaceae	*Calluna vulgaris*			1 384	357								1 741
Ericaceae	*Cassiope tetragona*	181	86								163		430
Ericaceae	*Chamaedaphne calyculata*	31		29		46						41	147
Ericaceae	*Empetrum nigrum*	3 466	12 736	2 266	4 714	2 807	6 813		14	3 149	13 507	1 758	51 230
Ericaceae	*Oxycoccus microcarpus*					538							538
Ericaceae	*Phyllodoce caeruela*	1 386	305				165						1 856
Ericaceae	*Vaccinium vitis-idaea/myrtillus*	2 005	2 042	916	308	1 286				189	815	394	7 955
Ericaceae	*Vaccinium uliginosum*	1 073	2 325	1 045		2 726	431	30		1 233	1 014	873	10 750
Geraniaceae	*Geranium sylvaticum*									68	145		213
Haloragaceae	*Myriophyllum alterniflorum*	11 389		273 929						226 753			512 071
Isoetaceae	*Isoetes* spp.			27 136						14 411			41 547
Lentibulariaceae	*Utricularia minor*											893	893
Lycopodiaceae	*Huperzia selago*	783	710						10		27	195	1 725
Lycopodiaceae	Lycopodiaceae	9 226	32 590	1 016	2 360	4 285	299	270	217	1 196	5 082	3 381	59 922
Menyanthaceae	*Menyanthes trifoliata*	26 842	467	17 384	1 173	18 978		98	871	378		42 408	108 599
Nymphaeaceae	*Nuphar pumila*											63 844	63 844
Plantaginaceae	*Callitriche hermaphroditica*			951			5 598						6 549
Plantaginaceae	*Hippuris vulgare*			238						107			345
Poaceae	*Festuca* spp.	30			2 724								2 754
Polygonaceae	*Oxyria digyna*										429		429
Polypodiaceae	*Athyrium* sp.	6 266	33 588	10 557	2 098	1 258		743	539	10 851	1 239	466	67 605
Potamogetonaceae	*Potamogeton praelongus*	1 754								254		9 268	11 276
Potamogetonaceae	*Potamogeton* sp.	28		19 281						12 817		1 335	33 461
Potamogetonaceae	*Stuckenia filiformis*	4 964	183			7 023						246	12 416
Ranunculaceae	*Caltha palustris*			1 131						5 080			6 211
Rosaceae	*Comarum palustre*	258				1 058				222			1 538
Rosaceae	*Dryas octopetala*		750				37					394	1 181
Rosaceae	*Filipendula ulmaria*	850		957		2 293				2 520		6 019	12 639
Rosaceae	*Rubus chamaemorus*	1 453		75		197						317	2 042
Rosaceae	*Sorbus aucuparia*	1 198	894	1 915		1 953				1 468			7 428
Salicaceae	*Populus tremula*	2 009	1 671		1 225		27			1 152		48 201	54 285
Salicaceae	*Salicaceae*	4 488	182 354	1 212	246	68 186	148 060	141		15 658	149 450	2 542	572 337
Saxifragaceae	*Saxifraga aizoides*		585	30									615

(*Continued*)

Table 3. (Continued)

Family	Taxa	A-tj	Bren	Einl	Finn	Gaup	Jula	Laks	Lauv	Ovre	Paul	Rott	Sum
Saxifragaceae	*Saxifraga oppositifolia*		922										922
Sparganiaceae	*Sparganium* spp.				958			258	74				1 290
Thelypteridaceae	*Phegopteris connectilis*	4 776	13 594	1 104	1 357	100		546	132	2 085	1 014	366	25 074
Woodsiaceae	*Gymnocarpium dryopteris*	10 290	42 766	1 339	1 287	18 254	336	764	355	3 029	1 986	2 082	82 488
Sum DNA reads		280 277	584 537	411 923	31 288	279 752	161 871	4 880	3 453	340 976	178 532	222 649	2 500 138
DNA and vegetation < 2m	Vegetation <2m and potentially >2m												
DNA and vegetation > 2m	Vegetation only > 2m												
DNA only	No DNA, no vegetation												

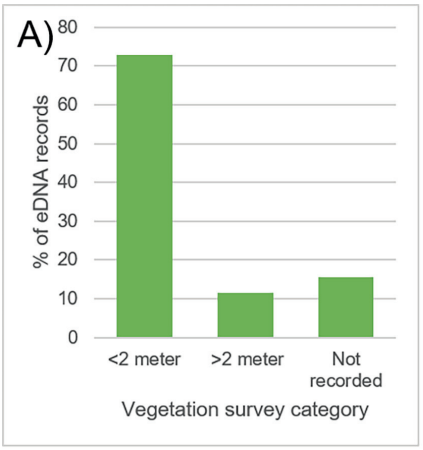
https://doi.org/10.1371/journal.pone.0195403.t003

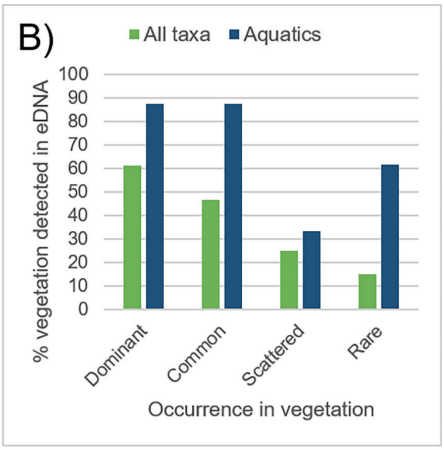
only two *Juncus* and two *Eriophorum* species were not recorded. Many taxa that were rare or scattered were filtered out (S1 and S4 Tables).

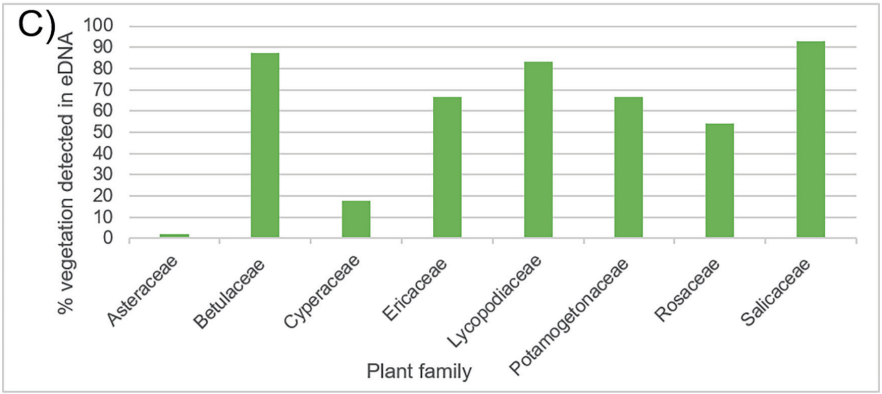
Detection success and taxonomic resolution in the eDNA varied among families (Table 3, Fig 3C). High success and resolution characterise Ericaceae and Rosaceae as they were identified to species level and successfully detected at most sites. Ferns (Dryopteridaceae, Thelypteridaceae, Woodsiaceae) and club mosses (Lycopodiaceae) were almost always detected, even when only growing >2 m from the lake shore. Aquatics (Haloragaceae, Lentibulariaceae, Menyanthaceae, Numphaeaceae, Plantaginaceae, Potamogetonaceae, Sparganiaceae) were also well detected, often also when not recorded in the vegetation surveys. Deciduous trees and shrubs (Betulaceae, Salicaceae) were also correctly identified at most lakes although often at genus level. In contrast, Poaceae and Cyperaceae, which were common to dominant around most lakes, were underrepresented in the DNA records. Juncaceae and Asteraceae, which were present at all lakes, although mainly scattered or rare, were mainly filtered out due to presence in only one PCR repeat or only in samples from 2–8 cm depth (S1–S4 Tables).

The numbers of taxa recorded in vegetation, in eDNA, and as match between them varied two- to six-fold among lakes (Table 2, Fig 3D). Jula Jávri had the lowest match between eDNA and vegetation with only four taxa in common. Lakselvhøgda and Lauvås had extremely low read numbers after filtering. For Lauvås, Finnvatnet and Lakselvhøgda, 84%, 30% and 20%, respectively, of raw reads were allocated to algae. If we assume that a big unidentified sequence cluster also represents algae, this increases to 69% for Lakselvhøgda, where a 15–20 cm algal layer was observed across most of the lake bottom. A lake-bottom algal layer was also observed at Jula Jávri, and in this we suspect that an unidentified cluster of 170,772 reads was algae. In most other lakes, algal reads were 3–15% (0.2% in Brennskogtjern, the lake with highest numbers of reads after filtering; algal data not shown).

Thirty-three records of 17 DNA taxa did not match vegetation taxa at a given lake (Table 3). These include taxa that are easily overlooked in vegetation surveys due to minute size (e.g., *Sagina* sp.), or only growing in deeper parts of the lake (e.g., *Potamogeton praelongus*). Other taxa are probably confined to ridge-tops of larger catchments, which lay outside the survey areas (e.g., *Cassiope tetragona* and *Dryas octopetala*). Two tree species that occur as shrubs or dwarf shrubs at their altitudinal limits, *Alnus incana* and *Populus tremula*, were found in the DNA at high-elevation sites. Also, ferns were detected at several sites where they were not observed in the vegetation surveys. On balance, most mismatches probably relate to plants being overlooked in the vegetation surveys or growing outside the survey area, whereas *Chamaedaphne calyculata* likely represents a false positive (Table 3, S1 Appendix).


A)
% of eDNA records
80
70
60
50
40
30
20
10
0
<2 meter
>2 meter
Not recorded
Vegetation survey category


B)
All taxa
Aquatics
% vegetation detected in eDNA
100
90
80
70
60
50
40
30
20
10
0
Dominant
Common
Scattered
Rare
Occurrence in vegetation


C)
% vegetation detected in eDNA
100
90
80
70
60
50
40
30
20
10
0
Asteraceae
Betulaceae
Cyperaceae
Ericaceae
Lycopodiaceae
Potamogetonaceae
Rosaceae
Salicaceae
Plant family

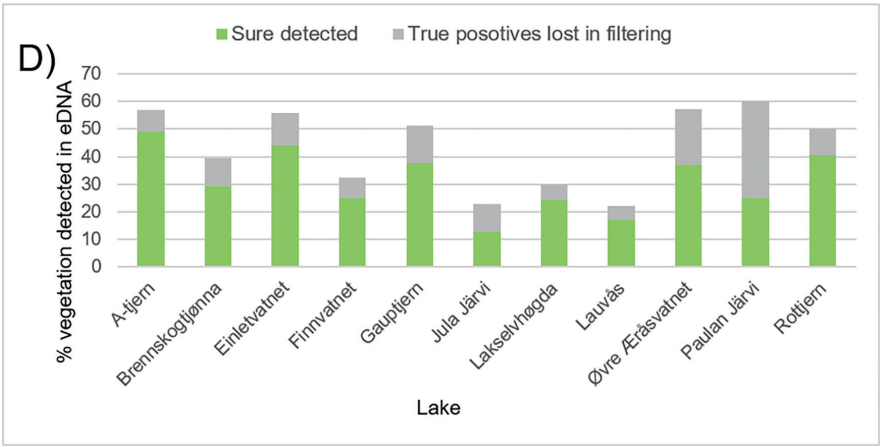

D)
Sure detected
True posotives lost in filtering
% vegetation detected in eDNA
70
60
50
40
30
20
10
0
A-tjern
Brennskogtjønna
Einletvatnet
Finnvatnet
Gauptjern
Jula Järvi
Lakselvhøgda
Lauvås
Øvre Æråsvatnet
Paulan Järvi
Rottjern
Lake

Fig 3. Match between records of taxa in the sedimentary eDNA in relation to vegetation surveys. a) Number of records in the sedimentary eDNA in relation to vegetation survey distance. b) Percentage records in eDNA in relation to abundance in vegetation surveys. c) Variation in percentage data among families with >11 eDNA records. d) Variation in percentage of taxa detected among lakes. Percentages in b), c) and d) refers to percentage of taxa recorded in the vegetation that potentially could be identified with the DNA barcode used. Note that DNA of more taxa were likely recorded but filtered out (S1–S4 Tables)–these numbers are only shown in Fig b).

https://doi.org/10.1371/journal.pone.0195403.g003

The multivariate ordinations gave similar results for the vegetation and eDNA records with the only lake from Pine forest, Brennskogtjønna, and one of the two alpine lakes, Paulan Jávri, clearly distinguished on the first axis, whereas the lakes with varying cover of birch forest were in one cluster (Fig 4A and 4B). The other alpine lake, Jula Jávri, was only distinguished on the vegetation, probably due to the low number of taxa identified in the eDNA of this lake (Table 2). Percentages of variation explained by the first two axes were similar for the two analyses (CA Vegetation: Axis 1, λ = 0.50, 20.4%, Axis 2, λ = 0.37, 15.1%; eDNA: Axis 1, λ = 0.24, 18.9%, Axis 2, λ = 0.24, 18.5%). The Procrustes analyses indicated a good similarity between vegetation and eDNA (CA Correlation = 0.53, P = 0.099; NSCA Correlation = 0.59, P = 0.045).

Probability of detecting taxa in vegetation and DNA records

The posterior probability that all local taxa were recorded during the vegetation survey varied from 0.85–0.95 (S6 Table). Thus, on average, about three species may have been overlooked at each lake. The posterior probability that taxa recorded in the vegetation surveys and detected

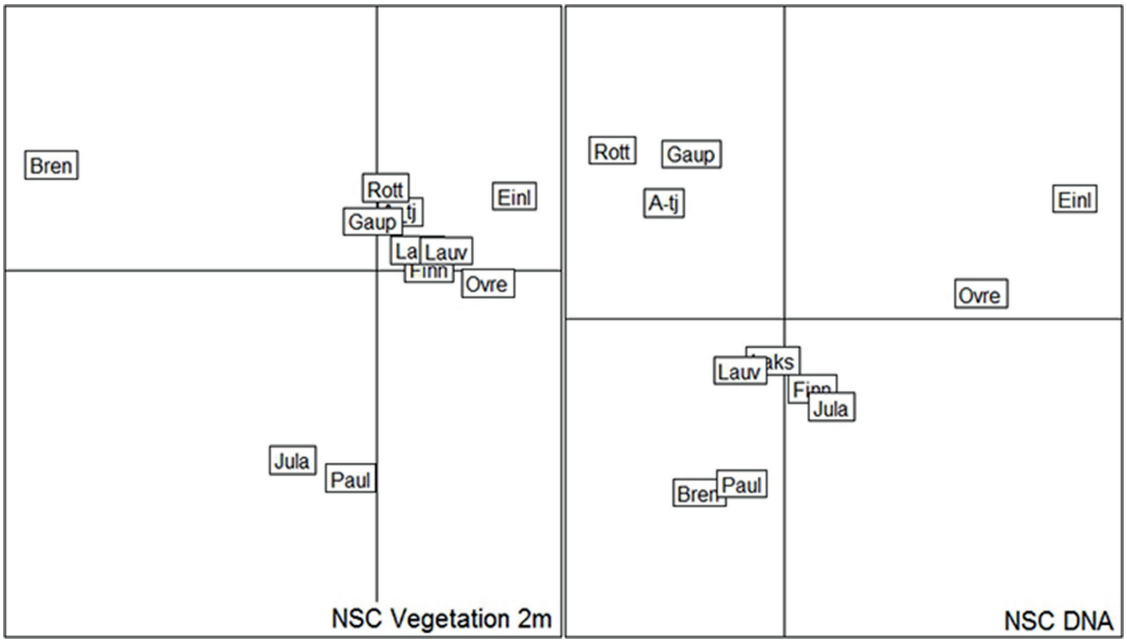


Fig 4. Multivariate ordination (non symmetric correspondence analysis; NSC) of the 11 lakes. The ordination is based on taxa recorded in the vegetation (a) and eDNA (b). Note that lakes in tall forbs birch/pine mixed forest (A-tjern, Rottjern, Gauptjern are clustered together in both plots; so are also Einletvatnet and Øvre Æråsvatnet (both mire/birch forest at the island Andøya), whereas some lake with poorer DNA records show some differences in clustering.

https://doi.org/10.1371/journal.pone.0195403.g004

at least once by eDNA were also recorded in the DNA in a given lake (true positives) was 0.33–0.90, whereas the posterior probability of any DNA records representing a false positive varied from 0.06–0.33 per lake (S6 Table). There was evidence that the probability of detecting a species using eDNA (p_{DNA_1}) was higher for deeper lakes (slope b_1 = 0.58, 95% CI = [0.20; 0.98], Fig 5). Not surprisingly a similar effect was found for lake size (slope b_1 = 0.25 [0.10, 0.41]) as lake size and depth were highly correlated (r = 0.81). Catchment area (b_1 = 0.06 [-0.15, 0.27]) and mean annual temperature (b_1 = -0.03 [-0.14, 0.08]) did not appear to influence probability of detection by eDNA.

Discussion

Taking into account the limitation of taxonomic resolution due to sequence sharing or taxa missing in the reference library, we were able to detect about one third of the taxa growing in the immediate vicinity of the lake using only two small sediment samples from the lake centre. The large number of true positives lost (S1 Appendix) suggests that this proportion may be further improved. Nevertheless, the current approach was sufficient to distinguish the main vegetation types.

Taphonomy of environmental plant DNA

The high proportion of taxa in the <2 m survey detected with eDNA than in the extended surveys indicates that eDNA is mainly locally deposited. The observation of taxa not recorded in

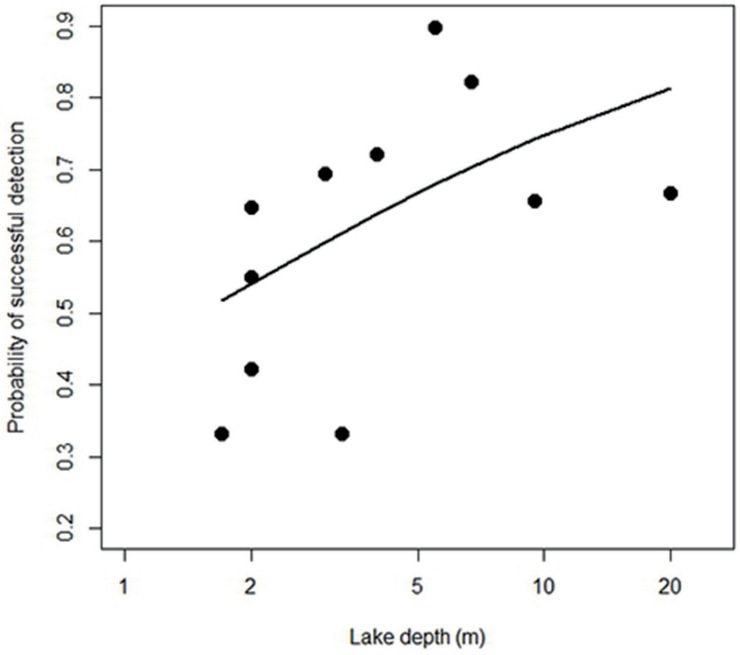


Fig 5. Lake depth versus detection probability. Relationship between lake depth and probability that a species present in the vegetation and detected at least once by eDNA is detected by eDNA in a given lake. The relationship is modelled as a logit function and back-transformed to the probability scale.

https://doi.org/10.1371/journal.pone.0195403.g005

the vegetation surveys but common in the region (Fig 4, S1 Table) indicates that some DNA does originate from some hundreds of meters or even a few km distant. Indeed, a higher correlation between catchment relief and total eDNA ($R^2 = 0.42$) than eDNA matching records in the vegetation ($R^2 = 0.34$), may suggest that runoff water from snow melt or material blown in also contributes. Thus, the taphonomy of eDNA may be similar to that of macrofossils [66, 67], except that eDNA may also be transported freely or via non-biological particles (e.g. fine mineral grains) [9]. From other studies, pollen does not appear to contribute much to local eDNA records [15, 35, 37, 42, 47]. This is probably due to its generally low biomass compared with stems, roots and leaves, and to the resilience of the sporopollenin coat, which requires a separate lysis step in extraction of DNA [68].

The higher proportion of eDNA taxa that matched common or dominant taxa in the vegetation, compared with taxa that were rare or scattered, was as expected, as higher biomass should be related to a greater chance for deposition and preservation in the lake sediments [9]. Yoccoz *et al.* [6] found the same in their comparison of soil eDNA with standing vegetation. While some dominant taxa were filtered out in our study, their DNA was mainly present (S1 Appendix, S1–S4 Tables), and most dominant taxa were recorded in all PCR replicates (not shown). Thus, for studies where the focus is on detecting dominant taxa, running costs may be reduced by performing fewer PCR replicates.

Variation among lakes

The variation among lakes seen in DNA-based detection of taxa shows that even when identical laboratory procedures are followed, the ability to detect taxa can vary. Our sample size of 11 lakes does not allow a full evaluation of the reasons for this variation. Factors such as low pH or higher temperature may increase DNA degradation [16], but the two lakes with lowest numbers of reads after filtering in our study, Lakselvhøgda and Lauvås, had pH values close to optimal for DNA preservation (7.2 and 6.8, respectively, I.G. Alsos and A.G. Brown, pers. obs. 2016), and variation in temperature was low among our sites. The lack of an inflowing stream at Lakselvhøgda may reduce the supply of eDNA, but Lauvås has two inflows. For these two lakes, and to a lesser extent Finnvatnet, we suspect high algal abundance might have caused PCR competition [69]. PCR competition may also occurred in samples from Jula Jávri, but in this case we were not able to identify the most dominant cluster of sequences. These lakes are also small and shallow. Variation among eDNA qualities has also been observed in a study of 31 lakes on Taymyr Peninsula in Siberia [70]. We suspect that high algae production may be a limiting factor as we also have seen poor aDNA results in samples with high Loss on Ignition values, but this should be studied further. A potential solution to avoid PCR competition may be to design a primer to block amplification of algae as has been done for human DNA in studies of mammals eDNA [71].

Variation among taxa

The variation we observed among plant families, both in taxonomic resolution and likelihood of detection, is a general problem when using generic primers [45, 72, 73]. For example, the poor detection of the Cyperaceae may be due to the long sequence length of *Carex* and *Eriophorum* (>80 bp), and most studies only detect it at genus or family level [38, 42, 74]. The low representation of Asteraceae may be due to its rare or scattered representation in the vegetation and/or its poor amplification. While some studies successfully amplify Asteraceae [15, 37, 38, 42, 75], others do not, even when other proxies indicate its presence in the environment [46]. This may be due to the high percentage of Asteraceae taxa that have a one base-pair mismatch in the reverse primer [34]. Poaceae, which has no primer mismatch, is regularly

detected in ancient DNA studies [15, 36–38, 41], and was present in nine lakes, although most records were filtered out due to occurrence in negative controls. To avoid any bias due to primer match and potentially increase the overall detection of taxa, one solution would be to use family-specific primers, such as ITS primers developed for Cyperaceae, Poaceae, and Asteraceae [36]. Alternatively, shotgun sequencing could be tested as this minimizes PCR biases [76, 77].

The common woody deciduous taxa *Betula* and *Salix*, as well as most common dwarf shrubs such as *Andromeda polifolia*, *Empetrum nigrum*, and *Vaccinium uliginosum*, were correctly detected in most cases. They are also regularly recorded in late-Quaternary lake-sediment samples [15, 25, 37, 41, 70, 74]. These are ecologically important taxa in many northern ecosystems, and their reliable detection in eDNA could be expected to extend to other types of samples, e.g., samples relating to herbivore diet [44].

The general over-representation of spore plants in eDNA among taxa only found >2 m from the lake and those not recorded in the catchment vegetation raises the question as to whether eDNA can originate from spores. Spore-plant DNA is well represented in some studies [42, 78], is lacking in other studies [15, 37] and has been found as an exotic in one study [41]. As with pollen, the protective coat and low biomass of spores suggest that they are an unlikely source of the eDNA. This inference is supported by clear stratigraphic patterns shown by fern DNA in two lake records from Scotland. Records are ecologically consistent with other changes in vegetation, whereas spores at the same sites show no clear stratigraphy [42]. Preferential amplification could be an alternative explanation, but this is not likely as the amplification of fern DNA from herbarium specimens is poor [34]. It is possible that in some cases, including this study, we are detecting the minute but numerous gametophytes present in soil, which would not be visible in vegetation surveys.

Aquatic taxa were detected in all lakes, and they have been regularly identified in eDNA analyses of recent [42] and late-Quaternary lake sediments [15, 37, 38]. eDNA may be superior to vegetation surveys in some cases, e.g., *Potamogeton praelongus*, which is characteristic of deeper water https://www.brc.ac.uk/plantatlas/) and was likely overlooked in surveys due to poor visibility. *Callitriche hermaphroditica* was observed in two lakes (Einleten and Jula javri), whereas *C. palustris* was observed at Einleten. We cross-checked the herbarium voucher and the DNA sequence and both seems correct, so potentially both were present but detected only in either eDNA or vegetation surveys. Overall, eDNA appears to detect aquatic plants more efficiently than terrestrial plants, which is not unexpected as the path from plant to sediment is short.

The use of eDNA for reconstruction of present and past plant richness

In contrast to water samples, from which eDNA has been shown to represent up to 100% of fish and amphibian taxa living in a lake [7, 79], one or two small, surficial sediment samples do not yield enough DNA to capture the full richness of vascular plants growing around a lake; the same limitation may apply in attempts to capture Holocene mammalian richness [22]. This is likely due to taphonomic limitations affecting preservation and transport on land, as aquatics were generally well detected. Also, surface samples are typically flocculent and represent a short time span, e.g. a few centimetre may represent 10–25 years ([49]; pers. obs.). Increasing the amount of material analysed, the amount of time sampled (by combining the top several cm of sediment), and/or the number of surface samples may improve detection rates for species that are rare, have low biomass and/or grow at some distance from the lake. In this study we identified more taxa when we used two surface samples and/or material from deeper in the sediment cores. Nevertheless, taphonomic constraints may mean that DNA of

some species rarely reaches the lake sediment. On the technical side, both improvements in laboratory techniques and in bioinformatics could increase detection of rare species. In this study, DNA of many of the rarer taxa was recorded but was filtered out. As the rarest species are also difficult to detect in vegetation surveys [59], combining conventional and DNA-based surveys may produce optimal estimates of biodiversity.

The potential taxonomic resolution (i.e., for eDNA taxa to be identified to species level) was similar or higher than that for macrofossils [80] or pollen [81, 82]. The potential taxonomic resolution of any of these methods depends on how well the local flora is represented in the available reference collection/library, site-specific characteristics, such as the complexity and type of the vegetation [34, 82], and the morphological or genetic variation displayed by different taxonomic groups. In our case, only 3% of the taxa found in the vegetation surveys were missing in the reference database which likely improved the resolution. To reach 100% resolution, a different genetic marker is needed to avoid the problem of identical sequences. Using longer barcodes may improve resolution [45, 83] and may work for modern samples, but for taxa with cpDNA sharing as e.g. *Salix*, nuclear regions should be explored. For ancient samples with highly degraded DNA, taxonomic resolution may potentially be increased by using a combining several markers, hybridization capture RAD probe techniques, or full-genome approach [77, 84–86].

The actual proportion of taxa in the vegetation detected in the eDNA records (average 28% and 18% for <2 m and extended surveys, respectively, not adjusting for taxonomic resolution) is similar to the results of various macrofossil [80, 87–89] and pollen studies [81, 82]. This contrasts with five previous studies of late-Quaternary sediments that compared aDNA with macrofossils and seven that did so with pollen; these showed rather poor richness in aDNA compared to other approaches (reviewed in [10]). We think a major explanation may be the quality and size of available reference collections/libraries, as the richness found in studies done prior to the publication of the boreal reference library (e.g. [15, 27, 35, 37]) was lower than in more recent studies, including this one [42, 70, 90, 91]. The variation in laboratory procedures, the number and size of samples processed and the number of replicates also affect the results [4, 82, 86]. Nevertheless, the correlation between eDNA and vegetation found in the Procrustes analyses show that the current standard of the method is sufficient to detect major vegetation types.

Conclusion

Our study supports previous conclusions that eDNA mainly detects vegetation from within a lake catchment area. Local biomass is important, as dominant and common taxa showed the highest probability of detection. For aquatic vegetation, eDNA may be comparable with, or even superior to, in-lake vegetation surveys. Lake-based eDNA detection is currently not good enough to monitor modern terrestrial plant biodiversity because too many rare species are overlooked. The method can, however, detect a similar percentage of the local flora as is possible with macrofossil or pollen analyses. As many true positives are lost in the filtering process, and as even higher taxonomic resolution could be obtained by adding genetic markers or doing full genome analysis, there is the potential to increase detection rates. Similarly, results will improve as we learn more about how physical conditions influence detection success among lakes, and how sampling strategies can be optimized.

Supporting information

S1 Appendix. Comments on true and false positives.
(DOCX)

S1 Table. All taxa recorded in the vegetation surveys. (<2 m and/or larger surveys) at 11 lakes in northern Norway. Number refers to the highest abundance recorded among 2–17 vegetation polygons in the larger vegetation surveys (1 = rare, 2 = scattered, 3 = frequent, and 4 = dominant). Thus, 2316 records were combined to give one vegetation record per species and lake, in total 1000 records. Taxa match represent taxa that could potentially be identified by the molecular method used: ND = no data in reference library, ID incomp = could not be identified in DNA because the vegetation is incomplete identified, <12 bp = filtered out in initial filtering steps due to short sequence length. Max = the maximum abundance score observed at any of the lakes. The lakes names are A-tjern (A-tj), Brennskogtjørna (Bren), Einletvatnet (Einl), Finnvatnet (Finn), Gauptjern (Gaup), Jula Jávri (Jula), Lakselvhøgda (Laks), Lauvås (Lauv), Øvre Æråsvatnet (Ovre), Paulan Jávri (Paul), and Rottjern (Rott). See colour codes below. Hatched colour refer to DNA-vegetation match at higher taxonomic level (e.g. *Salix*).
(DOCX)

S2 Table. Number of sequence reads remaining after each filtering step for 42 samples from modern lake sediment collected in northern Norway, 6 extraction negative controls, 6 PCR negative controls and 2 PCR positive controls. Six individually tagged PCR repeats were run for each sample, giving a total of 336 PCR samples. Numbers of sequences and unique sequences are given for applying the criteria to all sequences.
(DOCX)

S3 Table. Effect of different filtering criteria on the number of true positives. True positive (TP, defined as species also detected in vegetation surveys thus lower than the numbers given in Table 2) and False Positives (FP, defined as species not found in the regional flora; including 15 potential food plants) per lake and in total. The criteria used in this study, which gives the highest ratio between TP kept and FP lost, is shown in bold. 1) Minimum number of reads in lake samples, 2) minimum number of PCR repeats in lake samples, 3) "1" if occurring in more than 1 PCR repeat of negative control samples, 4) "1" if number of PCR repeats in lake sample > PCR repeats in negative control samples, 5) "1" if mean number of reads in lake samples > mean number of reads in negative control samples. The lakes names are A-tjern (A-tj), Brennskogtjørna (Bren), Einletvatnet (Einl), Finnvatnet (Finn), Gauptjern (Gaup), Jula Jávri (Jula), Lakselvhøgda (Laks), Lauvås (Lauv), Øvre Æråsvatnet (Ovre), Paulan Jávri (Paul), and Rottjern (Rott).
(XLSX)

S4 Table. Taxa removed during filtering. All DNA reads that have 100% match to the reference libraries and have been removed during the second last step of filtering (see S2 Table).
(XLSX)

S5 Table. Retrieval of positive controls from raw orbitool output file. The file consisted of 12706536 reads of 581 sequences, 98% match). Note that not all taxa used in the positive controls were present in the reference library but they match to closely related taxa.
(DOCX)

S6 Table. The probability of detection in eDNA and vegetation. The probability that all taxa in the vegetation were recorded (Vegetation), and that the DNA records represents true and false positives. Mean probability, standard deviation (SD) are given for each lake.
(DOCX)

Acknowledgments

We thank Marie Kristine Føreid Merkel, Veronica Rystad, Premasany Kanapathippillai, Chris Ware, Sarah Lovibond, Martin Årseth-Hansen, Torbjørn Alm and Antony G. Brown for field assistance, Department for Medical Biology at UiT for use of laboratory for extraction, Frederic Boyer for raw data handling, Torbjørn Alm for help with identifying plant specimens, and H. John B. Birks for valuable comments on an earlier draft of this manuscript.

Author Contributions

Conceptualization: Inger Greve Alsos, Nigel Giles Yoccoz, Mary E. Edwards.

Data curation: Inger Greve Alsos, Youri Lammers.

Formal analysis: Inger Greve Alsos, Youri Lammers, Nigel Giles Yoccoz, Ludovic Gielly.

Funding acquisition: Inger Greve Alsos, Ludovic Gielly, Mary E. Edwards.

Investigation: Inger Greve Alsos, Nigel Giles Yoccoz, Tina Jørgensen, Per Sjögren, Ludovic Gielly.

Project administration: Inger Greve Alsos.

Supervision: Inger Greve Alsos.

Visualization: Inger Greve Alsos, Youri Lammers, Nigel Giles Yoccoz.

Writing – original draft: Inger Greve Alsos.

Writing – review & editing: Inger Greve Alsos, Youri Lammers, Nigel Giles Yoccoz, Tina Jørgensen, Per Sjögren, Ludovic Gielly, Mary E. Edwards.

References

1. Orlando L, Cooper A. Using ancient DNA to understand evolutionary and ecological processes. Annu Rev Ecol Evol Syst. 2014; 45(1):573–98. https://doi.org/10.1146/annurev-ecolsys-120213-091712
2. Brown TA, Barnes IM. The current and future applications of ancient DNA in Quaternary science. J Quat Sci. 2015; 30(2):144–53. https://doi.org/10.1002/jqs.2770
3. Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, et al. Ancient and modern environmental DNA. Philos Trans R Soc London Ser B. 2015; 370(1660):20130383. Epub 2014/12/10. https://doi.org/10.1098/rstb.2013.0383 PMID: 25487334
4. Thomsen PF, Willerslev E. Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv. 2015; 183:4–18. https://doi.org/10.1016/j.biocon.2014.11.019
5. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012; 21(8):2045–50. https://doi.org/10.1111/j.1365-294X.2012.05470.x PMID: 22486824
6. Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, et al. DNA from soil mirrors plant taxonomic and growth form diversity. Mol Ecol. 2012; 21(15):3647–55. https://doi.org/10.1111/j.1365-294X.2012.05545.x PMID: 22507540
7. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol. 2016; 25(4):929–42. https://doi.org/10.1111/mec.13428 PMID: 26479867
8. Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, et al. Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char Salvelinus fontinalis. Biol Conserv. 2016; 194:209–16. http://dx.doi.org/10.1016/j.biocon.2015.12.023.
9. Barnes MA, Turner CR. The ecology of environmental DNA and implications for conservation genetics. Conserv Genet. 2016; 17(1):1–17. https://doi.org/10.1007/s10592-015-0775-4
10. Birks HJB, Birks HH. How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? New Phytol. 2016; 209:499–506. https://doi.org/10.1111/nph.13657 PMID: 26402315

11. Torti A, Lever MA, Jørgensen BB. Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Marine Genomics. 2015; 24, Part 3:185–96. http://dx.doi.org/10.1016/j.margen.2015.08.007.
12. Deiner K, Altermatt F. Transport distance of invertebrate environmental DNA in a natural river. PLoS ONE. 2014; 9(2):e88786. https://doi.org/10.1371/journal.pone.0088786 PMID: 24523940
13. Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. Environmental conditions influence eDNA persistence in aquatic systems. Environmental Science & Technology. 2014; 48 (3):1819–27. https://doi.org/10.1021/es404734p PMID: 24422450
14. Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjær KH, et al. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Mol Ecol. 2012; 21:1966–79. https://doi.org/10.1111/j.1365-294X.2011.05261.x PMID: 21917035
15. Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, et al. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. Mol Ecol. 2013; 22:3511–24. https://doi.org/10.1111/mec.12298 PMID: 23587049
16. Strickler KM, Fremier AK, Goldberg CS. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biol Conserv. 2015; 183(0):85–92. http://dx.doi.org/10.1016/j.biocon.2014.11.038.
17. Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, et al. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol Ecol Res. 2015; 15(3):543–56. https://doi.org/10.1111/1755-0998.12338 PMID: 25327646
18. Deiner K, Walser J-C, Mächler E, Altermatt F. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biol Conserv. 2015; 183:53–63. http://dx.doi.org/10.1016/j.biocon.2014.11.018.
19. Lahoz-Monfort JJ, Guillera-Arroita G, Tingley R. Statistical approaches to account for false-positive errors in environmental DNA samples. Mol Ecol Res. 2016; 16:673–85. https://doi.org/10.1111/1755-0998.12486 PMID: 26558345
20. Nguyen NH, Smith D, Peay K, Kennedy P. Parsing ecological signal from noise in next generation amplicon sequencing. New Phytol. 2015; 205(4):1389–93. https://doi.org/10.1111/nph.12923 PMID: 24985885
21. Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, et al. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol Ecol Res. 2016; 16(1):29–41. https://doi.org/10.1111/1755-0998.12433 PMID: 26032773
22. Giguet-Covex C, Pansu J, Arnaud F, Rey P-J, Griggo C, Gielly L, et al. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. Nat Commun. 2014; 5. https://doi.org/10.1038/ncomms4211 PMID: 24487920
23. Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, Huang Y-T, et al. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. J Quat Sci. 2014; 29(7):610–26. https://doi.org/10.1002/jqs.2740
24. Smol JP, Birks HJ, Last WM, editors. Terrestrial, algal, and siliceous indicators. Heidelberg: Springer; 2001.
25. Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, et al. Postglacial viability and colonization in North America's ice-free corridor. Nature. 2016; 537:45–9. https://doi.org/10.1038/nature19085 PMID: 27509852
26. Birks HH, Giesecke T, Hewitt GM, Tzedakis PC, Bakke J, Birks HJB. Comment on "Glacial survival of boreal trees in northern Scandinavia". Science. 2012; 338(6108):742. https://doi.org/10.1126/science.1225345 PMID: 23139314
27. Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, et al. Glacial survival of boreal trees in northern Scandinavia. Science. 2012; 335(6072):1083–6. https://doi.org/10.1126/science.1216043 PMID: 22383845
28. Parducci L, Edwards ME, Bennett KD, Alm T, Elverland E, Tollefsrud MM, et al. Response to Comment on "Glacial Survival of Boreal Trees in Northern Scandinavia". Science. 2012; 338(6108):742. https://doi.org/10.1126/science.1225345
29. Smith O, Momber G, Bates R, Garwood P, Fitch S, Pallen M, et al. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. Science. 2015; 347(6225):998–1001. https://doi.org/10.1126/science.1261278 PMID: 25722413
30. Weiß CL, Dannemann M, Prüfer K, Burbano HA. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. eLife. 2015; 4. https://doi.org/10.7554/eLife.10005 PMID: 26525598

31. Birks HH, Bjune AE. Can we detect a west Norwegian tree line from modern samples of plant remains and pollen? Results from the DOORMAT project. Veg Hist Archaeobot. 2010; 19(4):325–40. https://doi.org/10.1007/s00334-010-0256-0

32. Jackson ST. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. Quat Sci Rev. 2012; 49:1–15. http://dx.doi.org/10.1016/j.quascirev.2012.05.020.

33. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 2007; 35(3):e14. https://doi.org/10.1093/nar/gkl938 PMID: 17169982

34. Sønstebø JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J, et al. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Mol Ecol Res. 2010; 10(6):1009–18. https://doi.org/10.1111/j.1755-0998.2010.02855.x PMID: 21565110

35. Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, Rasmussen M, et al. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Mol Ecol. 2012; 21(8):1989–2003. https://doi.org/10.1111/j.1365-294X.2011.05287.x PMID: 22590727

36. Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, et al. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature. 2014; 506(7486):47–51. https://doi.org/10.1038/nature12921 PMID: 24499916

37. Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, et al. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quat Sci Rev. 2013; 75(0):161–8. http://dx.doi.org/10.1016/j.quascirev.2013.06.006.

38. Boessenkool S, McGlynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, et al. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. Conserv Biol. 2014; 28(2):446–55. https://doi.org/10.1111/cobi.12195 PMID: 24372820

39. Pansu J, Giguet-Covex C, Ficetola GF, Gielly L, Boyer F, Zinger L, et al. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. Mol Ecol. 2015; 24:1485–98. https://doi.org/10.1111/mec.13136 PMID: 25735209

40. Paus A, Boessenkool S, Brochmann C, Epp LS, Fabel D, Haflidason H, et al. Lake Store Finnsjøen–a key for understanding Lateglacial/early Holocene vegetation and ice sheet dynamics in the central Scandes Mountains. Quat Sci Rev. 2015; 121:36–51. http://dx.doi.org/10.1016/j.quascirev.2015.05.004.

41. Alsos IG, Sjögren P, Edwards ME, Landvik JY, Gielly L, Forwick M, et al. Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene. 2016; 26(4):627–42. https://doi.org/10.1177/0959683615612563

42. Sjögren P, Edwards ME, Gielly L, Langdon CT, Croudace IW, Merkel MKF, et al. Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytol. 2017; 213(2):929–41. https://doi.org/10.1111/nph.14199 PMID: 27678125

43. De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, et al. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol Ecol Res. 2014; 14(2):306–23. https://doi.org/10.1111/1755-0998.12188 PMID: 24128180

44. Soininen EM, Gauthier G, Bilodeau F, Berteaux D, Gielly L, Taberlet P, et al. Highly overlapping diet in two sympatric lemming species during winter revealed by DNA metabarcoding. Plos One. 2015; 10: e0115335. https://doi.org/10.1371/journal.pone.0115335 PMID: 25635852

45. Fahner NA, Shokralla S, Baird DJ, Hajibabaei M. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. PLoS ONE. 2016; 11(6):e0157505. https://doi.org/10.1371/journal.pone.0157505 PMID: 27310720

46. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, et al. Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in Zoology. 2009; 6:16. https://doi.org/10.1186/1742-9994-6-16 PMID: 19695081

47. Parducci L, Väliranta M, Salonen JS, Ronkainen T, Matetovici I, Fontana SL, et al. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. Philos Trans R Soc London Ser B. 2015; 370(1660):20130382. https://doi.org/10.1098/rstb.2013.0382 PMID: 25487333

48. Alm T. Øvre Æråsvatn–palynostratigraphy of a 22,000 to 10,000 B.P. lacustrine record on Andøya, Northern Norway. Boreas. 1993; 22:171–88.

49. Jensen C, Kunzendorf H, Vorren K-D. Pollen deposition rates in peat and lake sediments from the *Pinus sylvestris* L. forest-line ecotone of northern Norway. Review of Palaeobotany and Palynology. 2002; 121(2):113–32. https://doi.org/10.1016/s0034-6667(02)00077-5

50. Jensen C, Kuiper JGJ, Vorren K-D. First post-glacial establishment of forest trees: early Holocene vegetation, mollusc settlement and climate dynamics in central Troms, North Norway. Boreas. 2002; 31 (3):285–301. https://doi.org/10.1111/j.1502-3885.2002.tb01074.x

51. Jensen C, Vorren K-D. Holocene vegetation and climate dynamics of the boreal alpine ecotone of northwestern Fennoscandia. J Quat Sci. 2008; 23(8):719–43. https://doi.org/10.1002/jqs.1155

52. Vorren TO, Vorren K-D, Aasheim O, Dahlgren KIT, Forwick M, Hassel K. Palaeoenvironment in northern Norway between 22.2 and 14.5 cal. ka BP. Boreas. 2013: 876–95 https://doi.org/10.1111/bor.12013

53. Elven R. J. Lid & D.T. Lid. Norsk flora. 7th edition. Oslo: Det Norske Samlaget; 2005.

54. Elven R, Murray DF, Razzhivin VY, Yurtsev BA. Annotated checklist of the Panarctic Flora (PAF). Vascular plants. Natural History Museum, University of Oslo: CAFF/University of Oslo; 2011 [cited 2013]. Available from: http://nhm2.uio.no/paf/.

55. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, et al. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE. 2007; 2(2):e197. https://doi.org/10.1371/journal.pone.0000197 PMID: 17299583

56. Valentini A, Miquel C, Nawaz M, Bellemain E, Coissac E, Pompanon F, et al. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Res. 2009; 24(2):110–7.

57. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. OBITOOLS: a unix-inspired software package for DNA metabarcoding. Mol Ecol Res. 2016; 16(1):176–82. https://doi.org/10.1111/1755-0998.12428 PMID: 25959493

58. Yoccoz NG. The future of environmental DNA in ecology. Mol Ecol. 2012; 21(8):2031–8. https://doi.org/10.1111/j.1365-294X.2012.05505.x PMID: 22486823

59. Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, et al. Using niche-based models to improve the sampling of rare species. Conserv Biol. 2006; 20(2):501–11. https://doi.org/10.1111/j.1523-1739.2006.00354.x PMID: 16903111

60. Pélissier R, Couteron P, Dray S, Sabatier D. Consistency between ordination techniques and diversity measurements: Two strategies for species occurrence data. Ecology. 2003; 84(1):242–51. https://doi.org/10.1890/0012-9658(2003)084[0242:CBOTAD]2.0.CO;2

61. Greenacre M. Correspondence analysis of raw data. Ecology. 2010; 91(4):958–63. https://doi.org/10.1890/09-0239.1 PMID: 20462111

62. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia. 2001; 129(2):169–78. https://doi.org/10.1007/s004420100720 PMID: 28547594

63. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.4–3. 2017.

64. Chambert T, Miller DAW, Nichols JD. Modeling false positive detections in species occurrence data under different study designs. Ecology. 2015; 96(2):332–9. https://doi.org/10.1890/14-1507.1 PMID: 26240854

65. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; 7:434–55.

66. Jackson ST, Booth RK. Validation of pollen studies. In: Elias SA, editor. Encyclopedia of Quaternary science. London: Elsevier; 2007. p. 2413–22.

67. Dieffenbacher-Krall AC. Plant macrofossil methods and studies: Surface samples, taphonomy, representation. In: Elias SA, editor. Encyclopedia of Quaternary Science. Oxford: Elsevier; 2013. p. 684–9.

68. Kraaijeveld K, de Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, et al. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Res. 2015; 15(1):8–16. https://doi.org/10.1111/1755-0998.12288 PMID: 24893805

69. Piñol J, Mir G, Gomez-Polo P, Agustí N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Mol Ecol Res. 2015; 15(4):819–30. https://doi.org/10.1111/1755-0998.12355 PMID: 25454249

70. Niemeyer B, Epp LS, Stoof-Leichsenring KR, Pestryakova LA, Herzschuh U. A comparison of sedimentary DNA and pollen from lake sediments in recording vegetation composition at the Siberian treeline. Mol Ecol Res. 2017. https://doi.org/10.1111/1755-0998.12689 PMID: 28488798

71. Boessenkool S, Epp LS, Haile J, Bellemain EVA, Edwards M, Coissac E, et al. Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. Mol Ecol. 2011:1806–1815. https://doi.org/10.1111/j.1365-294X.2011.05306.x PMID: 21988749

72. CBOL PWG, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. A DNA barcode for land plants. Proceedings of the National Academy of Sciences. 2009; 106(31):12794–7. https://doi.org/10.1073/pnas.0905845106 PMID: 19666622

73. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE. 2011; 6(5):e19254. https://doi.org/10.1371/journal.pone.0019254 PMID: 21637336

74. Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, Schrøder-Nielsen A, et al. Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. Quat Sci Rev. 2015; 117(0):152–63. http://dx.doi.org/10.1016/j.quascirev.2015.03.027.

75. Hiiesalu I, ÖPik M, Metsis M, Lilje L, Davison J, Vasar M, et al. Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. Mol Ecol. 2011; 21:2004–16. https://doi.org/10.1111/j.1365-294X.2011.05390.x PMID: 22168247

76. Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, et al. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol Ecol Res. 2014; 14(5):966–75. https://doi.org/10.1111/1755-0998.12246 PMID: 24606032

77. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol. 2016:1423–8. https://doi.org/10.1111/mec.13549 PMID: 26821259

78. Wilmshurst JM, Moar NT, Wood JR, Bellingham PJ, Findlater AM, Robinson JJ, et al. Use of pollen and ancient DNA as conservation baselines for offshore islands in New Zealand. Conserv Biol. 2014; 28 (1):202–12. https://doi.org/10.1111/cobi.12150 PMID: 24024911

79. Lopes CM, Sasso T, Valentini A, Dejean T, Martins M, Zamudio KR, et al. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Mol Ecol Res. 2016:904–914. https://doi.org/10.1111/1755-0998.12643 PMID: 27987263

80. Allen JRM, Huntley B. Estimating past floristic diversity in montane regions from macrofossil assemblages. J Biogeogr. 1999; 26(1):55–73. https://doi.org/10.1046/j.1365-2699.1999.00284.x

81. Meltsov V, Poska A, Odgaard BV, Sammul M, Kull T. Palynological richness and pollen sample evenness in relation to local floristic diversity in southern Estonia. Review of Palaeobotany and Palynology. 2011; 166(3–4):344–51. http://dx.doi.org/10.1016/j.revpalbo.2011.06.008.

82. Birks HJB, Felde VA, Bjune AE, Grytnes J-A, Seppä H, Giesecke T. Does pollen-assemblage richness reflect floristic richness? A review of recent developments and future challenges. Review of Palaeobotany and Palynology. 2016; 228:1–25. http://dx.doi.org/10.1016/j.revpalbo.2015.12.011.

83. Lamb EG, Winsley T, Piper CL, Freidrich SA, Siciliano SD. A high-throughput belowground plant diversity assay using next-generation sequencing of the trnL intron. Plant and Soil. 2016; 404(1):361–72. https://doi.org/10.1007/s11104-016-2852-y

84. Schmid S, Genevest R, Gobet E, Suchan T, Sperisen C, Tinner W, et al. HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. Methods in Ecology and Evolution. 2017; 8(10):1374–88. https://doi.org/10.1111/2041-210X.12785

85. Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, et al. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS ONE. 2016; 11(3):e0151651. https://doi.org/10.1371/journal.pone.0151651 PMID: 26999359

86. Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, et al. Transley Reviews: Ancient plant DNA from lake sediments. New Phytol. 2017; 214(3):924–42. https://doi.org/10.1111/nph.14470 PMID: 28370025

87. Dunwiddie PW. Macrofossil and pollen representation of coniferous trees in modern sediments from Washington. Ecology. 1987; 68(1):1–11. https://doi.org/10.2307/1938800

88. McQueen DR. Macroscopic plant remains in recent lake sediments. Tuatara. 1969; 17(1):13–9.

89. Drake H, Burrows CJ. The influx of potential macrofossils into Lady Lake, north Westland, New Zealand. New Zealand Journal of Botany. 1980; 18(2):257–74. https://doi.org/10.1080/0028825X.1980.10426924

90. Zimmermann H, Raschke E, Epp L, Stoof-Leichsenring K, Schirrmeister L, Schwamborn G, et al. 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. 2017; 8(10):273. https://doi.org/10.3390/genes8100273 PMID: 29027988

91. Zimmermann HH, Raschke E, Epp LS, Stoof-Leichsenring KR, Schwamborn G, Schirrmeister L, et al. 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. 2017; 14 (3):575–96. https://doi.org/10.5194/bg-14-575-2017

Paper II

BOREAS An international journal of Quaternary research

Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (*sed*aDNA) and pollen

CHARLOTTE L. CLARKE, MARY E. EDWARDS, ANTONY G. BROWN, LUDOVIC GIELLY, YOURI LAMMERS, PETER D. HEINTZMAN, FRANCISCO JAVIER ANCIN-MURGUZUR, KARI-ANNE BRÅTHEN, TOMASZ GOSLAR AND INGER G. ALSOS

BOREAS

Clarke, C. L., Edwards, M. E., Brown, A. G., Gielly, L., Lammers, Y., Heintzman, P. D., Ancin-Murguzur, F. J., Bråthen, K.A., Goslar, T. & Alsos, I. G. 2019 (April): Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (*sed*aDNA) and pollen. *Boreas*, Vol. 48, pp. 299–316. https://doi.org/10.1111/bor.12357. ISSN 0300-9483.

We present a Holocene record of floristic diversity and environmental change for the central Varanger Peninsula, Finnmark, based on ancient DNA extracted from the sediments of a small lake (*sed*aDNA). The record covers the period *c*. 10 700 to 3300 cal. a BP and is complemented by pollen data. Measures of species richness, sample evenness and beta diversity were calculated based on *sed*aDNA sampling intervals and 1000-year time windows. We identified 101 vascular plant and 17 bryophyte taxa, a high proportion (86%) of which are still growing within the region today. The high species richness (>60 taxa) observed in the Early Holocene, including representatives from all important plant functional groups, shows that modern shrub-tundra communities, and much of their species complement, were in place as early as *c*. 10 700 cal. a BP. We infer that postglacial colonization of the area occurred prior to the full Holocene, during the Pleistocene-Holocene transition, Younger Dryas stadial or earlier. Abundant DNA of the extra-limital aquatic plant *Callitriche hermaphroditica* suggests it expanded its range northward between *c*. 10 200 and 9600 cal. a BP, when summers were warmer than present. High values of *Pinus* DNA occur throughout the record, but we cannot say with certainty if they represent prior local presence; however, pollen influx values >500 grains cm^{-2} a^{-1} between *c*. 8000 and 7300 cal. a BP strongly suggest the presence of pine woodland during this period. As the site lies beyond the modern tree limit of pine, it is likely that this expansion also reflects a response to warmer Early Holocene summers.

Charlotte L. Clarke (C.Clarke@soton.ac.uk), Mary E. Edwards and Antony G. Brown, Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK; Ludovic Gielly, Laboratoire d'ÉCologie Alpine (LECA), Université Grenoble Alpes, C2 40700 38058, Grenoble Cedex 9, France; Youri Lammers, Peter D. Heintzman, Francisco Javier Ancin-Murguzur and Inger Greve Alsos, Tromsø University Museum, UiT - The Arctic University of Norway, Tromsø NO-9037, Norway; Kari Anne Bråthen, Department of Arctic and Marine Biology, UiT - The Arctic University of Norway, Tromsø NO-9037, Norway; Tomasz Goslar, Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland and Poznań Radiocarbon Laboratory, Foundation of the A. Mickiewicz University, 61-612, Poznań, Poland; *received 31st March 2018, accepted 18th September 2018.*

In the mid- to high-latitudes, the repeated waxing and waning of continental ice sheets throughout the Quaternary has regulated habitat availability for plants (Hultén 1937; Abbott & Brochmann 2003). By around 13 000–11 000 a BP, the Fennoscandian Ice Sheet had retreated from the northeasternmost peninsulas of Finnmark, Norway (Sollid *et al.* 1973; Stokes *et al.* 2014; Hughes *et al.* 2016; Stroeven *et al.* 2016), and newly deglaciated land became accessible for plant colonization and vegetation development (Prentice 1981, 1982). Today, these peninsulas harbour the ecotone from boreal forest to tundra. At and near their northern limit, tree species are particularly sensitive to climate changes; climate variation may cause shifts in tree species ranges that may in turn generate changes in vegetation structure and habitat (Hyvärinen 1976; Barnekow 1999; Bjune *et al.* 2004; Jensen & Vorren 2008). The deglacial history and proximity to the tree-line suggest that records from the region can potentially address key ecological questions: how floristic richness and/or composition is affected by tree-line dynamics and/or Holocene climate variation; whether taxa show discernible migration lags (and thus locally variable postglacial successional sequences); and how today's dominant plant communities assembled over the Holocene.

The understanding of tree-line fluctuations has inspired palynological studies of vegetation history for many years (Hyvärinen 1975; Seppä 1996; Allen *et al.* 2007). Pollen percentage and influx values have been used to track tree-line fluctuations across northeast Finnmark in response to Holocene climate changes (Hyvärinen 1975; Hicks 1994; Seppä 1996; Hicks & Hyvärinen 1999; Høeg 2000; Huntley *et al.* 2013). During the regional Holocene Thermal Maximum (HTM; 8000–6000 cal. a BP; Seppä *et al.* 2009; Huntley *et al.* 2013), the ecotonal boundaries between *Pinus* and *Betula* forests and *Betula* forest with low shrub-tundra had more northerly positions (Høeg 2000). In contrast to tree-line dynamics, less is known about the development and composition of the shrub and herb communities of northern Fennoscandia and their response to Holocene climate changes. Tundra pollen records are often interpreted in terms of broad-scale community dynamics rather than local compositional changes, due to features such as low accumulation rates and the over-representation of woody anemophilous taxa that mute the signal of entomophilous forbs in many records (Lamb & Edwards 1998; Gajewski 2015).

The Lateglacial and Holocene vegetation history of Finnmark has been documented in several pollen and plant

DOI 10.1111/bor.12357

macrofossil records since the 1970s (e.g. Hyvärinen 1975; Prentice 1981, 1982; Høeg 2000; Allen *et al.* 2007; Birks *et al.* 2012; Sjögren & Damm 2019), reflecting interest in understanding events linked to the early deglaciation of the region (Sollid *et al.* 1973; Stokes *et al.* 2014; Hughes *et al.* 2016; Stroeven *et al.* 2016). The initial postglacial landscape supported sparse, open, herbaceous vegetation with some shrubs (Hyvärinen 1975; Prentice 1981, 1982; Seppä 1996). The oldest pollen record in northeast Finnmark is that of Østervatnet, southern Varanger Peninsula, which records the Older Dryas (*c.* 13 900–13 600 cal. a BP) and Younger Dryas (*c.* 13 500–11 500 cal. a BP) climate oscillations, with *Artemisia* replacing *Salix* and Poaceae in the cold stages (Prentice 1981). The end of the Younger Dryas chronozone is distinguished by a rise in *Oxyria*/*Rumex*, *Salix*, Poaceae and Cyperaceae with species-rich meadows colonized by a succession of shrub and tree species in the Holocene (e.g. *Betula*, *Pinus*). More information on the rate of community assemblage and local variation in community development through time will help inform our understanding of the resilience and longevity of the current dominant communities in this area.

Pollen data have been used to reconstruct past changes in floristic diversity and richness, including studies in Scandinavia (Odgaard 1999; Berglund *et al.* 2008a, b; Fredh *et al.* 2012; Reitlau *et al.* 2015). Records from Scandinavia indicate a rapid increase in species richness from the Lateglacial to the Early Holocene (*c.* 12 000–8000 cal. a BP), while spatially and temporally inconsistent trends characterize the Middle to Late Holocene (*c.* 8200 cal. a BP – present; Seppä 1998; Berglund *et al.* 2008a, b; Birks & Birks 2008; Felde *et al.* 2017). These later Holocene trends have been attributed to climate fluctuations, the first appearance of trees at a locality, and/or human impact. A significant ($p < 0.001$) negative relationship was identified between *Pinus* pollen influx and species richness at the boreal site Lake Rautuselkä, northern Finland, reflecting the importance of vegetation density on floristic richness (Seppä 1998). A significant ($p < 0.05$) negative correlation between *Pinus* and *Betula* pollen influx and species richness was also identified at the tundra site Lake Hopseidet on the Nordkinn Peninsula, northeast Norway, which was not reached by northward expansion of the *Pinus* tree-line during the Holocene. High influx of wind-pollinated taxa in the tundra site probably reduced the statistical probability of other, less frequent insect-pollinated herbaceous types being counted (Birks & Line 1992; Seppä 1998). Records from central Scandinavia show that species richness has remained rather stable, with no long-term trends observed over the Holocene (Giesecke *et al.* 2012). Nevertheless, biases resulting from non-linear relationships between pollen and vegetation representation may confound richness estimates derived from pollen (Prentice 1985; Sugita 1994; Odgaard 2001). Furthermore, the relationship between species richness derived from pollen data (palynological richness) and the observed floristic richness in the landscape remains poorly understood (Meltsov *et al.* 2011; Goring *et al.* 2013).

Analysis of sedimentary ancient DNA (*sed*aDNA) has recently emerged as a promising proxy for reconstructing past floral diversity, augmenting information gained from pollen and macrofossil analyses (Jørgensen *et al.* 2012; Epp *et al.* 2015; Alsos *et al.* 2016; Pedersen *et al.* 2016; Parducci *et al.* 2017; Zimmerman *et al.* 2017). When rigorously applied, the analysis of *sed*aDNA from lake sediments can detect more species per sample than other palaeoecological methods (Alsos *et al.* 2016). It also permits the detection of some key plant taxa that are poorly resolved taxonomically by pollen or plant macrofossil analysis alone (Parducci *et al.* 2013; Sjögren *et al.* 2017; Edwards *et al.* 2018). Recent investigation of the representation of contemporary vegetation in the DNA signal of superficial sediments in small lakes with limited inflowing streams from northern Norway revealed that 73 and 12% of the taxa detected in the DNA were recorded in vegetation surveys within 2 and 50 m of the lake shore, respectively (Alsos *et al.* 2018). Thus, analysis of plant *sed*aDNA from small lakes with limited inflowing streams may give a more local signal of vegetation change and floristic richness than records derived from pollen, as wind-dispersed grains tend to be dispersed over long distances, particularly at the northern limit of trees (Rousseau *et al.* 2006).

This study represents the first palaeoecological exploration of Arctic vegetation dynamics in Finnmark using a *sed*aDNA record, in this case from sediments of a small lake on the Varanger Peninsula, northeast Finnmark. We use metabarcoding techniques (Taberlet *et al.* 2012) to develop the *sed*aDNA record, together with X-ray fluorescence (XRF) to determine geochemical element concentrations over time, sedimentological data and pollen analysis. We then compare the results with published pollen records. Finally, we use the *sed*aDNA data to reconstruct Holocene trends in species richness using rarefaction and measures of beta diversity and evenness for all samples and for 1000-year windows.

Study site

The lake (latitude 70°19′6.85348″ N, longitude 30°1′43.83653″ E; Fig. 1) is unnamed on the 1:50 000 Norwegian Topographic Map (Norgeskart; https://www.norgeskart.no). We refer to it here informally as 'Uhca Rohči', or UR, the Sami name for an adjacent river feature. UR is a small lake (<1 ha) in a depression situated at 138 m above sea level (a.s.l.) within the river valley of Komagdalen on the Varanger Peninsula, northeast Finnmark, Norway. The peninsula is a plateau lying between 200 and 600 m a.s.l. with low relief: ridges are formed of Cambrian quartzites and sandstones, while valleys are eroded into shales and mudstones (Siedlecka & Roberts 1992). After Pliocene uplift, the area was affected by sea-level change and subject

to glacial erosion (Fjellanger & Sørbel 2007). UR Lake lies in the middle section of the glaciated southeast-draining Komagdalen valley. As shown in Fig. 1D, the lake vicinity is associated with probable subglacial scour features on the former valley floor, which is now ~3 m above the present river. The valley lies between the Gaissattrinnet and Hovedtrinnet (Younger Dryas) moraines to the southwest and the older Ytre Porsangertrinnet and Korsnestrinnet moraines to the north and east (ice advance to NE, retreat to SW; Sollid *et al.* 1973). A reconstruction of the lower section of the Komagdalen valley by Olsen *et al.* (1996) suggests that the middle valley lies just north (outside) of an ice margin dated to 17 000 cal. a BP as well as a more southerly ice margin associated with the Vardø moraine stage (13 500 a BP) as mapped by Tolgensbakk & Sollid (1981). Cosmogenic dating at the head of Varangerfjorden and Tanafjorden suggests a local retreat age of *c.* 15 400–14 200 cal. a BP, and it is certain that the peninsula was free of glacial ice by 13 000–12 000 cal. a BP (Stokes *et al.* 2014; Stroeven *et al.* 2016). UR Lake lies approximately 60 m above the main (Younger Dryas) postglacial shoreline (75–85 m a.s.l.), which is reflected by a markedly steeper valley floor reach at approximately 13 km downstream (Fig. 1E; Fletcher *et al.* 1993). It is likely that the site received fluvial input from all of the upstream Komagdalen catchment, at least prior to downcutting associated with postglacial isostatic uplift in the later Holocene, and this is important in the interpretation of the *sed*aDNA data.

The present-day climate of the Varanger Peninsula is characterized as sub-Arctic (<10 °C arctic isotherm in July), with annual precipitation between 500 and 800 mm. It is situated within the meeting zone of the westerlies and the sub-polar low pressure system (with polar easterlies) and thus has highly variable weather (Hanssen-Bauer & Tveito 2014). Large local heterogeneity exists due largely to the topography, and summer temperature may vary from 6–12 °C, with corresponding differences in local vegetation (Karlsen *et al.* 2005). Present-day tundra vegetation of the Varanger Peninsula is classified as erect shrub tundra (Virtanen *et al.* 1999; Walker *et al.* 2005), and is dominated by dwarf shrubs, such as *Empetrum nigrum* subsp. *hermaphroditum* and *Betula nana.* Species-rich meadows occur along the wide riparian plains of the Komagelva River where tall shrubs such as *Salix lanata*, *S. hastata* and *S. glauca* form a spatially and temporally diverse vegetation mosaic with mesic forbs such as *Bistorta vivipara*, *Thalictrum alpinum* and *Viola biflora*, and graminoids such as *Avenella flexuosa*, *Deschampsia cespitosa* and *Eriophorum angustifolium* (Ravolainen *et al.* 2013; Bråthen *et al.* 2017). The headwaters of the Komagelva River originate from the north to northwest of the main Komagelva channel, which flows eastwards to Varangerfjord (Fig. 1B). The entire watershed lies outside of the present-day *Pinus* limit (Fig. 1B). UR Lake is one of several small lakes surrounded by the species-rich riparian meadows of the Komagdalen valley, which is one of the principal sites for the Climate Ecological Observatory for Arctic Tundra (Henden *et al.* 2011; Ravolainen *et al.* 2011, 2013; COAT 2018).

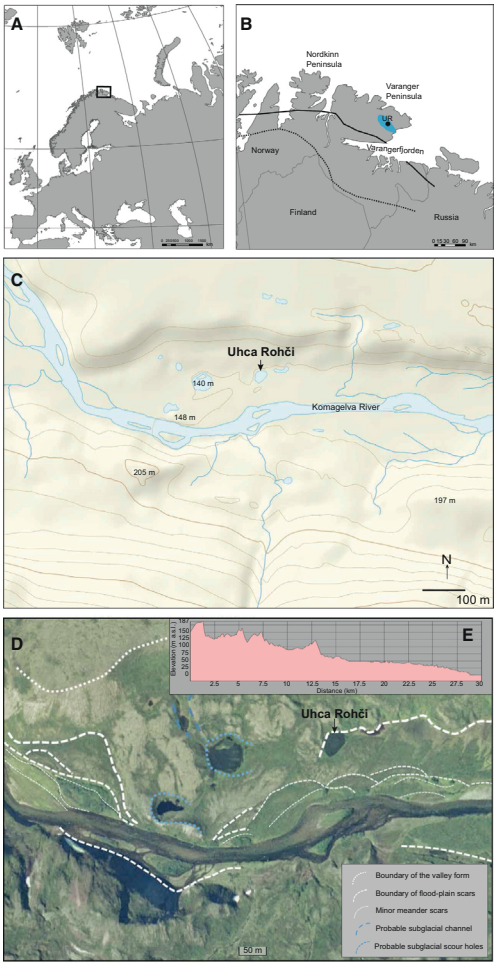


Fig. 1. Location of Uhca Rohči Lake (unofficial name) on the Varanger Peninsula, northeast Finnmark, Norway. A. Map highlighting the location of Finnmark in northeastern-most Norway. B. Varanger Peninsula with watershed of Komagelva River indicated in green. Dashed line indicates the *Pinus* forest limit according to Heikkinen (2005) and the solid black line indicates the tree limit of *Pinus* according to Hustich (1983). Location of Uhca Rohči Lake is indicated by a black circle. C. Map detailing the position of Uhca Rohči Lake within the Komagelva River (Komagdalen) valley (norgeskart.no). D. Geomorphological map created from aerial photography and inset E. valley floor height profile downstream of Uhca Rohči Lake created using Google Earth 2017.

Early Holocene climate in northern Fennoscandia was affected by summer insolation that was higher than present (Berger 1978; Berger & Loutre 1991). Quantitative summer temperature estimates based on aquatic plant macrofossils

suggest an early onset of the HTM, with temperatures ~2 °C warmer than pollen-based estimates between 11 700–7500 cal. a BP (Väliranta *et al.* 2015). However, pollen-based summer temperature reconstructions indicate July temperatures of +1.5±0.5 °C above modern (1961–1990) values in northern Fennoscandia during the regional HTM identified at *c.* 8000–6000 cal. a BP (Møller & Holmeslet 2002; Jensen & Vorren 2008; Seppä *et al.* 2009; Sejrup *et al.* 2016). Northeast Finnmark was characterized by warmer summer temperatures and higher precipitation (Allen *et al.* 2007). Warm conditions were interrupted by a short cold spell at *c.* 8200 cal. a BP (Seppä *et al.* 2009).

Material and methods

Core retrieval and subsampling

A 10-cm-diameter and 2.5-m-long lake sediment core (UR-1) was retrieved in February 2016 from the winter ice surface using a modified Nesje piston-corer (Nesje 1992) with a 4-m-long continuous section of acrylonitrile-butadiene-styrene (ABS) pipe. Total lake depth including winter ice thickness was 2 m, as measured by a single-beam echo-sounder (Echotest II Plastimo) and tape measure. Coring and retrieval of sediments started at 2 m, with the assumption, based on the echo-sounder data, that the top of the core sequence would include surface or near-surface sediments. After retrieval of the sediments, the pipe was cut into 1-m sections and sealed immediately to minimize the risk of contamination by airborne or other modern environmental DNA. The core sections were stored in a shed (2–6 °C) to prevent freezing before transport and stored at 4 °C in the cold room at the Tromsø University Museum (TMU), Norway. Core sections were opened by longitudinal splitting. One half was used for subsampling, and the other half kept for archival purposes. Core UR-1 was subsampled at 1-cm resolution within a dedicated ancient DNA clean-room facility at TMU using sterile tools, a full bodysuit, facemask, and gloves. Following the protocol described by Parducci *et al.* (2017), the outer 10 mm of sediment was avoided or discarded and an ~20-g subsample was retrieved from inside the freshly exposed centre only.

An additional short 50-cm-long and 7-cm-diameter core (UR-2), which included a clear sediment-water interface, was retrieved using a UWITEC gravity corer (UWITEC Corp., Austria) lowered from the surface of the lake ice. Subsampling of core UR-2 was not performed in the clean-room at TMU, as this core was only used for pollen analysis and age-depth model determination (described below).

Lithological and elemental analyses

Subsamples (2 cm^3) were taken for bulk density and loss-on-ignition (LOI) analyses at 3-cm intervals using a volumetric sampler. Samples were weighed in crucibles and dried overnight at 100 °C before dry weight and bulk density were determined (Chambers *et al.* 2011). Samples were then ignited at 550 °C for 2 h, placed in a desiccator to cool to room temperature and reweighed. Total LOI was calculated as the percentage loss of dry weight after ignition (Heiri *et al.* 2001). Magnetic susceptibility and XRF analyses were performed on the archival core halves at the Department of Geosciences, UiT - Arctic University of Norway. Magnetic susceptibility was measured at 1-cm intervals using a Bartington point sensor on the Geotek Ltd. Multi-Sensor Core Logger using a 10-s exposure time. Quantitative element geochemical measurements were performed with an Avaatech XRF core scanner. XRF scanning was performed at 1-cm resolution with the following settings: 10 kV, 1000 μA, 10 s exposure time and no filter. To minimize closed sum effects, we normalized the raw peak area data against Ti, as this element is considered a reliable indicator of allochthonous catchment inputs (Croudace & Rothwell 2015).

Radiocarbon dating and age-depth model construction

Seven samples of terrestrial plant macrofossils from Nesje core UR-1 and an additional two samples from UWITEC core UR-2 were radiocarbon (^{14}C) dated with accelerator mass spectrometry (AMS) at the Poznań Radiocarbon Laboratory (Goslar *et al.* 2004). All radiocarbon ages were calibrated according to the terrestrial IntCal13 curve (Reimer *et al.* 2013), and an age-depth relationship was established using the Bayesian framework calibration software 'Bacon' (v. 2.2; Blaauw & Christen 2011), which was implemented in R v. 3.2.4 (R Core Team 2017).

Pollen analysis

In total, 16 pollen samples were analysed from core UR-1 and an additional two samples from UR-2. Subsamples of 1.5 cm^3 were prepared using standard methods (acid-base-acid-acetolysis; Fægri & Iversen 1989) and were mounted in glycerol. Two *Lycopodium* spore tablets (Batch no. 3862; n = 9666) were added to each sample to calibrate pollen concentration estimation. At least 300 pollen grains of terrestrial taxa were identified per sample using taxonomic keys (Fægri & Iversen 1989) and type material held in the Palaeoecology Laboratory at the University of Southampton. Results are presented as pollen percentages, with trees, shrubs, herbs and graminoids based on the sum of total terrestrial pollen ($\sum$P), and percentages for spores and aquatics based on $\sum P + \sum$spores and $\sum P + \sum$aquatics, respectively.

DNA extraction, amplification, library preparation, and sequencing

DNA extraction, PCR amplification, PCR product pooling and purification, and sequencing follow the protocols of Alsos *et al.* (2016) unless otherwise stated. Within the TMU clean-room facility, DNA was extracted from 80

sediment subsamples and nine negative extraction controls, which consisted of no sediment and were used to monitor for contamination. Aliquots of DNA extracts were then shipped to the Laboratoire d'ÉCologie Alpine (LECA, University Grenoble Alpes, France) for metabarcoding. Each DNA extract and negative extraction control was independently amplified using uniquely tagged generic primers that amplify the *trn*L P6 loop of the plant chloroplast genome (Taberlet *et al.* 2007), a widely applied marker for the identification of vascular plants in environmental samples. Each sample and negative control underwent eight PCR replicates to increase confidence in the results and improve the chance of detecting taxa with small quantities of template in the DNA extracts. We also ran 13 negative PCR controls, consisting of no DNA template. Pooled and cleaned PCR products were then converted to two Illumina-compatible amplicon libraries using the single-indexed, PCR-free MetaFast method (FASTERIS SA, Switzerland). These libraries were then sequenced on the Illumina HiSeq-2500 platform for 2× 125 cycles at FASTERIS.

Sequence analysis and taxonomic assignments

Next-generation sequence data were filtered using the OBITools software package (Boyer *et al.* 2016; http://metabarcoding.org/obitools/doc/index.html) following the protocol and criteria defined by Alsos *et al.* (2016). Taxonomic assignments were performed using the *ecotag* program (Boyer *et al.* 2016) by matching sequences against a local taxonomic reference library comprised of 815 arctic and 835 boreal vascular plant taxa, and 455 bryophytes (Sønstebo *et al.* 2010; Willerslev *et al.* 2014; Soininen *et al.* 2015). In order to minimize any erroneous taxonomic assignments, only taxa with a 100% match to a reference sequence were retained. We further considered a taxon to be undetected in a PCR replicate if it was represented by fewer than 10 reads. Moreover, sequences that displayed higher average reads in negative extraction or PCR controls than lake sediment samples were also removed. Identified taxa were compared with the local flora from Komagdalen (Ravolainen *et al.* 2013), Species Map Service 1.6 (https://artskart1.artsdatabanken.no/Default.aspx), the Norwegian Flora (Elven 2005) and the circumpolar flora (Hultén & Fries 1986). Sequences assigned to taxa not present in northern Scandinavia today were checked against the NCBI BLAST database for multiple or alternative taxonomic assignments (http://www.ncbi.nlm.nih.gov/blast/).

Indices of richness, diversity and evenness

Species diversity was measured by three parameters – beta diversity (β-diversity), richness, and evenness (Magurran 2004; Soininen *et al.* 2012) – and analysed for both the DNA sampling intervals (3-cm resolution) and 1000-year time windows. On average, each 1000-year time window encompassed nine *sed*aDNA samples, whereas time windows of 500 years or less contained too few samples (>4), on average, for estimating long-term changes in diversity. Following Koleff *et al.* (2003), β-diversity was measured using Whittaker's (b_w) index computed using the PAST v. 3.19 software package (Hammer *et al.* 2001). A comparison of richness between samples with different count sizes can be biased, as the chance of detecting rare taxa increases simultaneously with count size (Birks & Line 1992; Brown 1999). We, therefore, rarefied the *sed*aDNA data to estimate the number of vascular plant taxa that would have been detected if the DNA read count had been standardized amongst samples. Rarefaction analysis was performed using the minimum count size in the Vegan (Oksanen *et al.* 2017) package for R (R Core Team 2017). We chose the Simpson evenness index ($E_{1/D}$) to measure DNA sample evenness across the 1000-year time windows, following Meltsov *et al.* (2011). This index of evenness is independent of the number of taxa detected.

Results

Chronology and lithostratigraphy

In total, seven AMS radiocarbon dates were obtained from the UR-1 core (Nesje). Ages span 3330±30 to 9480±50 ^{14}C a, which corresponds to a calibrated weighted-mean range of 3606–10 705 cal. a BP (Table 1). The resulting age-depth model (Fig. 2) suggests a fairly linear sedimentation rate, with the exception of a period of faster accumulation of sediment between 125 and 150 cm depth (as measured from the top of core UR-1; *c.* 7600–8100 cal. a BP). We consider these radiocarbon ages to be reliable for age-depth modelling as they are in the correct stratigraphical sequence, have small errors and are all derived from terrestrial plant macrofossils. It can, therefore, be assumed that the uppermost sediments that correspond to the last *c.* 3000 years are missing from core UR-1. This was most likely caused by a lake-depth measurement error, potentially from the ice layer affecting the echo-sounder, which resulted in the non-retrieval of upper sediments. To test this assumption, two additional AMS radiocarbon dates were obtained from the surface core UR-2 (UWITEC), which displayed a clear sediment–water interface. Radiocarbon ages from these two additional samples (275±30 and 1485±30 ^{14}C a; a calibrated weighted-mean range of 15 and 1433 cal. a BP) confirm that UR-1 is missing the uppermost sediment.

LOI values vary around 20% in the lower part of the core and reach a maximum of 47% in the upper part, indicating a generally increasing organic component with time (Fig. 3). Core UR-1 is divided into three main lithostratigraphical units, labelled A–C from the core base (Figs 2, 3):

Unit UR-1A (234–204 cm; c. 10 700–9900 cal. a BP). Silty-clay with traces of fine sand characterize this unit. Coarse (~2–3-cm thick bands) greenish-brown banding is

Table 1. Radiocarbon ages of plant macrofossil remains from Uhca Rohči Lake shown with their 1σ error, calibrated weighted mean, calibrated median and calibrated 95% confidence age ranges. All radiocarbon ages were calibrated using the IntCal13 curve (Reimer *et al.* 2013).

Core no.	Lab. ID	Depth below sediment surface (cm)	Age±1σ (^{14}C a BP)	Cal. weighted mean age (cal. a BP)	Cal. median age (cal. a BP)	Cal age, 2σ (cal. a BP)	Material
UR-2	Poz-98146	1–2	275±30	15	6	−2 to 150	*Salix* leaves
UR-2	Poz-98147	42–43	1485±30	1433	1417	1305–1598	*Salix* leaves
UR-1	Poz-93338	9–10	3330±30	3606	3592	3479–3819	*Salix* leaves
UR-1	Poz-87278	22–23	3850±40	4246	4244	4059–4408	*Salix* leaves
UR-1	Poz-87277	74–75	5490±50	6280	6287	6040–6432	Charcoal
UR-1	Poz-87276	120–121	6890±50	7715	7717	7572–7842	*Empetrum* wood
UR-1	Poz-93339	140–141	7150±40	8038	8024	7920–8195	*Empetrum* wood
UR-1	Poz-87275	178–179	8240±35	9208	9205	9027–9402	Charcoal
UR-1	Poz-87274	231–232	9480±50	10 706	10 702	10 417–11 030	*Salix* leaves

evident. LOI values range from 15 to 22% (mean 18%) and mean bulk density is 0.4 g cm^{-3}. The unit is characterized by high Ti, K and Fe, suggesting that sedimentation is driven by input of terrigenous minerogenic sediment.

Unit UR-1B (204–15 cm; c. 9900–3800 cal. a BP). A sharp transition into more organic-rich silty clay with small monocot rootlets and wood fragments occurs at 204 cm, marking the base of Unit B. Thin (1–3 mm) laminae of medium-coarse sand alternating with olive-brown organic silty clay comprise this second unit, with silt content decreasing upwards. Higher mean LOI values of 27% and fluctuations in Ti, K and Si characterize Unit B, with peaks in Ca probably related to the presence of inclusions, such as shells, within this unit. Bulk density values range from 0.2 to 0.4 g cm^{-3} and display a general decline through this unit, with the exception of a short-lived interval of higher values between 165 and 138 cm, which may result from sediment compaction after cutting the core sections.

Unit UR-1C (15–0 cm; c. 3800–3300 cal. a BP). This unit comprises mid- to dark-brown silty-clay gyttja with some detritus, including abundant monocot rootlets. A small decline in LOI values from 37 to 33% in this unit suggests a slight rise in terrigenous minerogenic input. The lithology of this unit is similar to that of core UR-2.

DNA analysis

In total, we obtained around 72 000 000 raw reads for the two UR libraries (Table S1). Following post-identification filtering, 118 taxa remained, of which 41% were identified to species level, 47% to the genus and 12% to the family (Figs S1–S3). Of the taxa detected in the *sed*aDNA, 44% are found growing in the Komagdalen valley today and a large proportion (86%) within Finnmark and the Kola Peninsula today (Table S2). Salicaceae and *Pinus* were present in nearly all samples but display variation in the number of PCR replicates (out of eight) in which they were detected; this was unrelated to sample depth/age. The steep drop in *Pinus sed*aDNA in the uppermost three samples (*c.* 3600–3300 cal. a BP) appears to be an artefact of the rapid increase in *Salix sed*aDNA during this time.

Although not present in the study region today, stands of *Pinus* forest occur ~50 km south of the Komagdalen valley (Elven 2005) and scattered trees are observed at the nearby site Østervatnet on the southern Varanger Peninsula (Prentice 1981). In addition, the *sed*aDNA results indicate that Holocene vegetation was dominated by woody taxa such as *Betula*, *Empetrum*, *Vaccinium* spp. and the *Rhododendron tomentosum* complex, the latter of which is also not found in the catchment today. The most common terrestrial forb was *Bistorta vivipara* followed by *Cakile* and Apiaceae, all common in the area today, although *Cakile* occurs mainly at the coast. The most dominant aquatic taxon, *Limosella aquatica*, is restricted to four inner fjord sites in Finnmark today (Alta, Lakselv, Neiden and Pasvik) whereas the second most common aquatic taxon, *Callitriche hermaphroditica*, has a slightly wider distribution in the inner fjord zone and along the main valleys. However, it does not occur within Varangerhalvøya National Park today (Table S2).

Trees and tall shrubs (e.g. *Pinus*, *Betula*, *Empetrum*, Salicaceae) dominate the *sed*aDNA record, accounting for 50% of total DNA reads on average (Fig. 4), followed by total terrestrial forbs (23%) and graminoids (12%; Fig. 5). The percentage dominance of functional groups remains relatively constant across samples, except for a distinct peak in the *sed*aDNA of aquatics between 10 200–9600 cal. a BP. This short period is described by the appearance of *Callitriche hermaphroditica*, *C. palustris*, and *Potamogeton* (Fig. 5). *Callitriche hermaphroditica* is a northern species (<53°N) typically found in shallow lakes and slow-moving rivers; it is on the IUCN Red List (http://www.iucnredlist.org/details/167828/0), and based on current distribution in Finland, it is inferred to indicate a minimum July temperature of 13–14 °C (Väliranta *et al.* 2015). *Callitriche palustris* and *Potamogeton* (not identified to species level) are more common and found in a wide range of aquatic habitats.

Pollen analysis

Pollen analysis detected 39 taxa across the 16 samples analysed from UR Lake, with *Betula* dominating the pollen

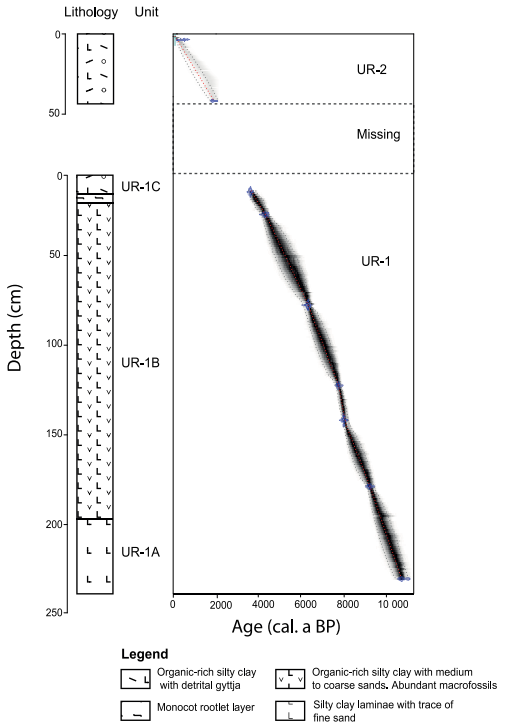


Fig. 2. The age-depth relationship and lithostratigraphical units (labelled UR-1A to C) for Uhca Rohči Lake, Varanger Peninsula. Age-depth relationships for core UR-1 (Nesje) and UR-2 (UWITEC) were analysed independently due to the hiatus remaining unknown. Radiocarbon ages were calibrated following IntCal13 (Reimer *et al.* 2013) and the age-depth model produced using the software Bacon (Blaauw & Christen 2011). Note that only UR-1 was analysed further in this study.

percentages and accounting for 65% of total land pollen (TLP) and influx of 2263 grains cm^{-2} a^{-1}, on average (Fig. 4). Of the 39 taxa, 72% could be identified to genus level, 21% to family level and 7% to species level. Four taxa (Ericales, *Myrica gale*, *Picea* and *Ulmus*) are assumed to be long-distance dispersed based on their current native ranges and their absence in the *sed*aDNA signal (Table S2). Three distinct zones are identified based on the pollen (Fig. 6), numbered from the base as follows:

Ur-Ia (234–190 cm; c. 10 700–9500 cal. a BP) – Betula-Empetrum-Salix zone. This basal pollen assemblage is characterized by high and rising percentages of *Betula* (up to 70%) and *Empetrum* (up to 20%) with high *Salix* (5–15%). Poaceae percentages are up to 17% at the beginning of the zone, decreasing to around 5% near the end of the zone boundary. *Pinus* values remain low at around 5% and influx of 70 grains cm^{-2} a^{-1}. *Filipendula* and Cyperaceae are present at percentages of 5 and 3%, respectively.

Ur-Ib (190–155 cm; c. 9500–8400 cal. a BP) – Betula-Salix-Empetrum zone. This is a short-lived subzone characterized by the rapid decline in values for *Empetrum* (up to 7%) from the previous (Ur-Ia) subzone. *Pinus* values remain consistently low at around 5% and influx of 50 grains cm^{-2} a^{-1}. The Ur-I/Ur/II boundary is defined by a decrease in Empetrum and Salix to low values.

Ur-II (155–80 cm; c. 8400–6400 cal. a BP) – Betula-Pinus-Empetrum zone. Rising *Pinus* values from 5 to 10% (110–860 grains cm^{-2} a^{-1}) accompany an increase in *Betula* values up to a peak of 80% (~1600–8300 grains cm^{-2} a^{-1}) in this zone. *Salix* values are low or zero whilst *Empetrum* values gradually increase.

Ur-III (80–0 cm; c. 6400–3300 cal. a BP) – Betula-Pinus-Empetrum-Poaceae zone. The Ur-II/Ur-III boundary is defined by rising values of Poaceae and *Pinus* and a decline in *Betula*. *Pinus* reaches maximum values in this zone, increasing from around 10% at the start of the zone up to a peak of 18% (200 grains cm^{-2} a^{-1}). *Empetrum* continues to rise through the transition between Ur-II and Ur-III zones coincident with increasing percentages of herbaceous taxa such as Chenopodiaceae and *Rumex*. Poaceae and *Salix* rise slightly. *Isoetes* reaches a maximum value of 10% before declining towards the end of the zone.

Comparison between pollen and aDNA

The combined approach of *sed*aDNA and pollen analysis resulted in 137 taxa of 64 families identified to varying taxonomic levels (Table S2). In total, 20 families were shared between pollen and *sed*aDNA, with poor taxonomic resolution seen for families such as Poaceae, Cyperaceae and Caryophyllaceae based on pollen whilst identification to genus or even species level was possible with *sed*aDNA. Of the 39 taxa detected by pollen analysis, 12 were also identified in the *sed*aDNA to the same taxonomic level. Selected taxa found as pollen are presented in Figs 4, 5 as a percentage of total terrestrial pollen and compared with the results of *sed*aDNA analysis. No algal taxa were detected in the *sed*aDNA record based on the vascular plant *trn*L P6 loop marker, whilst a high abundance of *Pediastrum* was identified throughout the pollen record at UR Lake (Table S2). Both the pollen and *sed*aDNA record are dominated by trees and shrubs, but they differ in terms of the percentage dominance of key taxa. For example, *Betula* accounts for 65% of total terrestrial pollen on average, followed by *Pinus* (8%) and *Empetrum* (7%) whilst *Pinus* and Salicaceae are found to be dominant in the *sed*aDNA, accounting for 21 and 20% of total DNA reads, respectively. The greater dominance of *Empetrum* in the Early Holocene (*c.* 10 700–9500 cal. a BP) revealed by *sed*aDNA is mirrored in the pollen record (Fig. 4). Moreover, the absence of *sed*aDNA belonging to the family Poaceae between *c.* 5500 and 4500 cal. a BP is simultaneous with a rapid and short-lived decline in

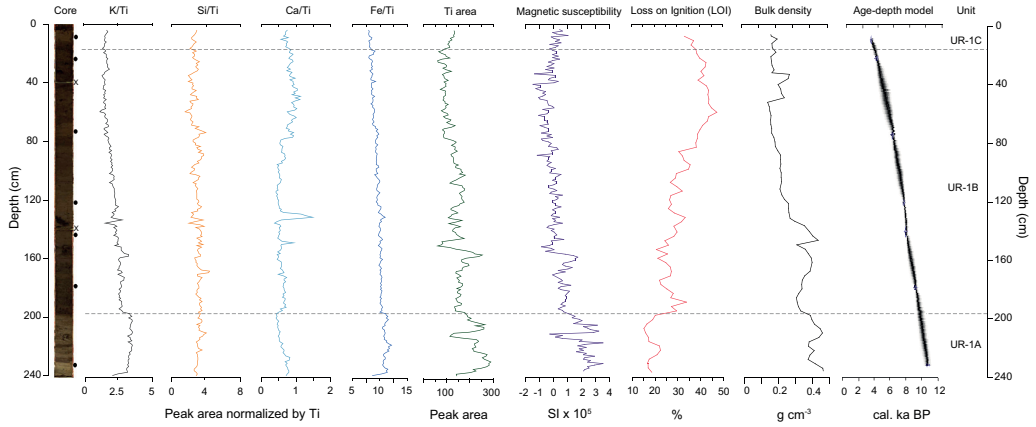


Fig. 3. Sediment properties and element profiles for Uhca Rohči Lake, Varanger Peninsula. Lithostratigraphical units (UR-1A to C) described in the text are indicated. Black circles indicate the depth of ^{14}C samples analysed and black crosses indicate the position of core breaks after splitting the continuous Nesje core (UR-1) sequence. Selected elements measured by XRF are given as a ratio to Ti (see Material and methods). Loss-on-ignition (LOI) is given as a percentage of dry sediment weight.

Poaceae pollen (Fig. 5). Furthermore, two pollen grains of *Callitriche* sp. were found within samples at 198 and 162 cm depth, coincident with the interval of high *sed*aDNA values for *Callitriche hermaphroditica* and *C. palustris* (Fig. 5).

Floristic richness and diversity results

No long-term trends in floristic richness were observed over the time period investigated (Fig. 7) although species richness reconstructed from individual samples shows high-frequency variation. *sed*aDNA detected an anomalously high number of taxa (48 taxa pre-rarefaction) within a single sample at *c.* 6020 cal. a BP (68 cm depth; 39 taxa remaining following rarefaction). Differences in the composition of this sample compared to adjacent samples largely result from the presence of many forb taxa such as *Viola biflora*, *Stellaria borealis*, *Rumex*, *Oxyria digyna*, *Geranium* and *Dryas.* Some bryophyte (e.g. *Andreaea rupestris*, *Dicranum*, *Sphagnum russowii*) and woody taxa (e.g. *Alnus*, *Kalmia procumbens*, *Vaccinium* spp.) are also present but are not found in adjacent samples. There is no clear explanation from the lithological (Fig. 2) or geochemical (Fig. 3) data for the anomalously high floristic richness observed in this sample.

β-diversity calculated based on individual samples displayed variation (SD = 0.09), despite species richness remaining fairly constant amongst samples (Fig. 7A). Whilst the number of taxa detected by *sed*aDNA remains similar amongst samples, the taxonomic composition differed between adjacent samples. Typically, the woody taxa remain a common component of adjacent samples but the herb (forb and graminoid) taxa show sporadic occurrences throughout the record. Merging samples into 1000-year time windows largely removes the effect of these sporadic occurrences, with β-diversity displaying little variation between time windows (Fig. 7B). The number of taxa identified as common between adjacent 1000-year time windows remained consistently high throughout the record, accounting for, on average, 70% of all taxa detected. Six taxa belonging to trees and shrubs (Salicaceae, *Pinus*, *Empetrum*, *Betula*, *Rhododendron tomentosum*, *Vaccinium uligonosum*) seven to forbs (Anthemideae, Asteraceae, Apiaceae, *Bistorta vivipara*, *Comarum palustre*, *Dryas*, *Limosella aquatica*) and three to graminoids (Agrostidinae, *Festuca*, Poaceae) were consistently detected across all time windows. The small variation observed in β-diversity (SD = 0.04) and sample evenness (SD = 0.06) across 1000-year time windows therefore results from the typically sporadic occurrences of the remaining 30% of taxa detected (Fig. 7B).

Floristic diversity and tree-line changes

Rarefied species richness based on pollen (palynological richness) from UR Lake varies between 10 and 24 taxa. Compilation of palynological richness patterns from Lake Rautuselkä and Hopseidet (Seppä 1998) with estimates obtained at UR Lake indicates long-term trends in richness reconstructed from pollen data (Fig. 8). A sharp decline in species richness from 24 to 10 taxa is observed in UR Lake between *c.* 8000–7300 cal. a BP, coincident with maximum values for *Pinus* pollen influx (~110 to 860 grains cm^{-2}a^{-1}), which occur in zone UR-II (Figs 4, 6). No significant relationship (r^2 = 0.08, p > 0.01) between *Pinus* pollen influx and palynological richness was identified at UR Lake, however. Floristic richness reconstructed from *sed*aDNA from UR Lake displays no long-term trends (Fig. 8), indicating richness remains relatively stable with

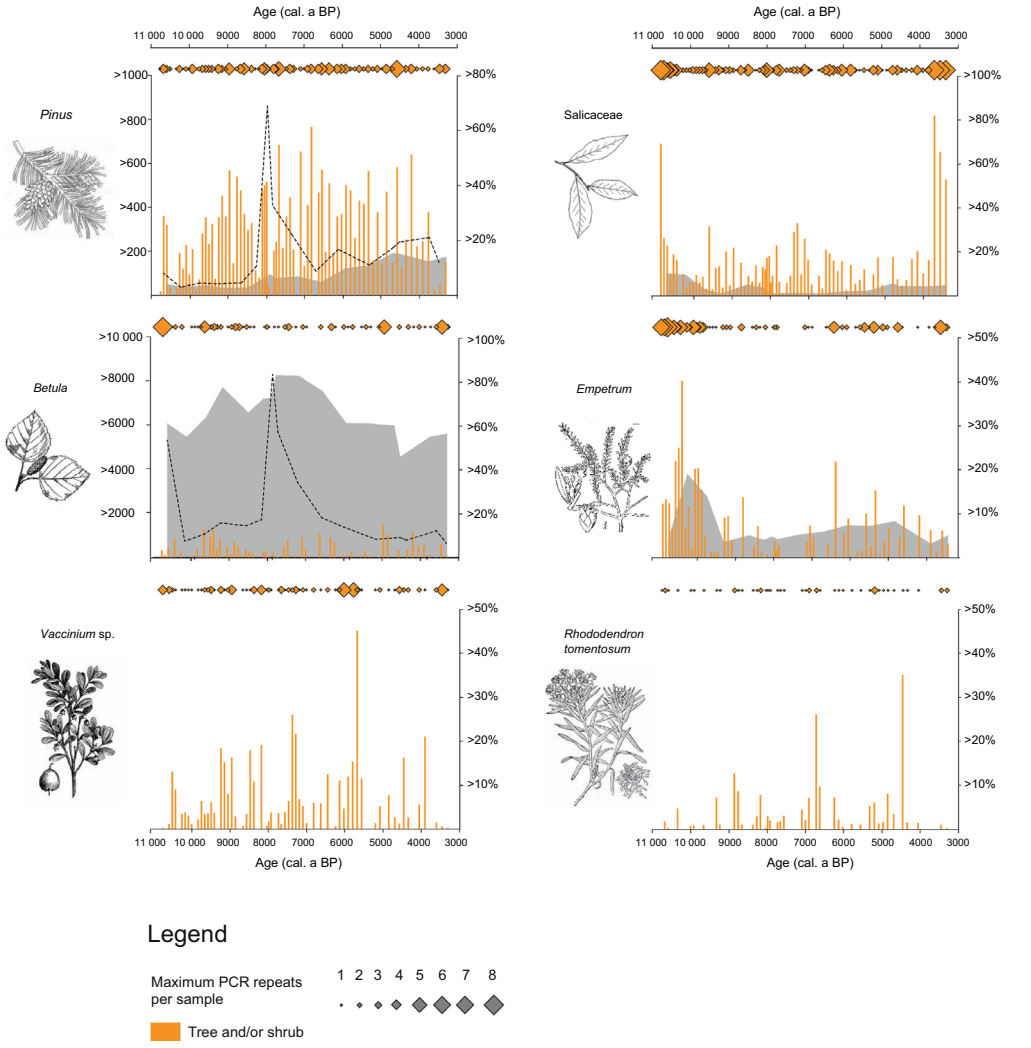


Fig. 4. Selected woody taxa as a percentage of total DNA reads per sample (histogram; right-hand *y*-axis) and maximum number of PCR replicates (diamond symbols) for the Uhca Rohči Lake record. Grey shaded area depicts pollen percentages based on sum of total terrestrial pollen (∑P; right-hand *y*-axis) with pollen influx for *Pinus* and *Betula* indicated by a dashed line (left-hand *y*-axis). Note that the height of the *y*-axis varies amongst panels.

the exception of the anomalously high richness observed at 68 cm depth (*c*. 6000 cal. a BP).

Discussion

Holocene development of Uhca Rohči Lake and the surrounding landscape

The lithostratigraphical record (Figs 2, 3) indicates a continuous input of fine sediment into UR Lake throughout the Holocene. Our results suggest an increase in lake production at *c*. 9900 cal. a BP, with a gradual transition from minerogenic sediments, composed of silty-clay laminae with low LOI and high Ti and magnetic susceptibility, to silty-clay gyttja with gradually increasing values for LOI (Fig. 3). Following this transition, lithostratigraphical properties remain relatively stable for the remainder of the record with only small fluctuations observed in LOI, magnetic susceptibility, and geochemical elements in the uppermost zone. The position of the lake on the Early to

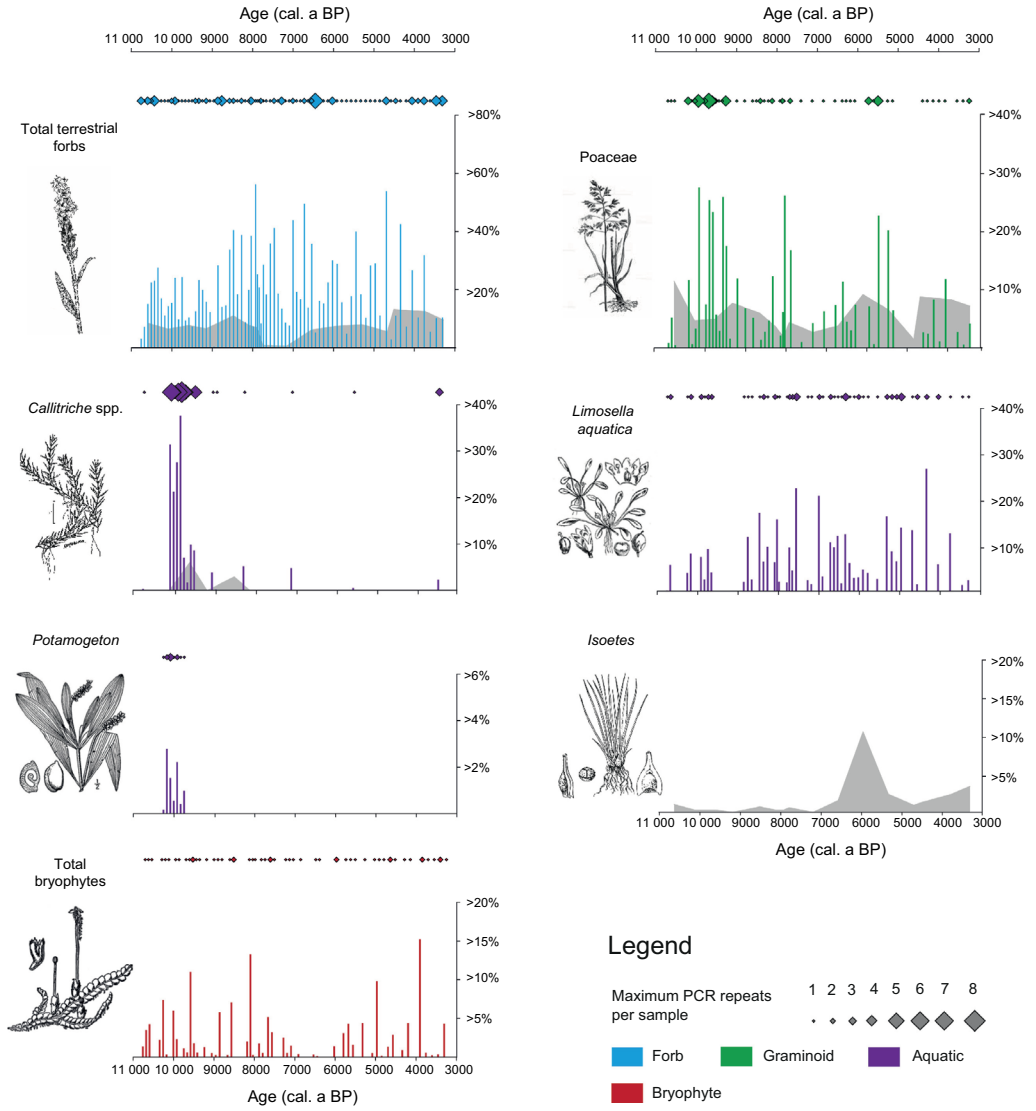


Fig. 5. Selected herbaceous taxa as a percentage of total DNA reads per sample (histogram) and maximum number of PCR replicates (diamond symbols) for the Uhca Rohči Lake record. Grey shaded area depicts pollen percentages based on sum of total terrestrial pollen (∑P). Proportion of aquatics and spores are calculated based on the sum of total terrestrial pollen plus aquatics (∑P + ∑aquatics) or spores (∑P + ∑spores). Pollen percentages for *Callitriche* spp. are presented with a 10× exaggeration. Note that the height of the right-hand *y*-axis varies amongst panels.

Middle Holocene flood-plain (Fig. 1) and laminated nature of the minerogenic sediments in lithostratigraphical unit A and to a lesser extent, unit B (Fig. 3), probably reflect periodic flood events of the Komagelva River prior to river incision forced by continued isostatic uplift (Fletcher *et al.* 1993; Fjellanger & Sørbel 2007). Very local slope wash may also have played a role at this time. Furthermore, the high sand content of these laminae suggests the influence of a relatively high-energy system in these early lithostratigraphical units. Our results suggest UR Lake was isolated from riverine influence at *c.* 3800 cal. a BP, at the transition from silty-clay with thin laminae (unit B) to silty-clay gyttja with detritus (unit C), when the Komagelva River subsequently downcut to an elevation below the lake. Thus, the

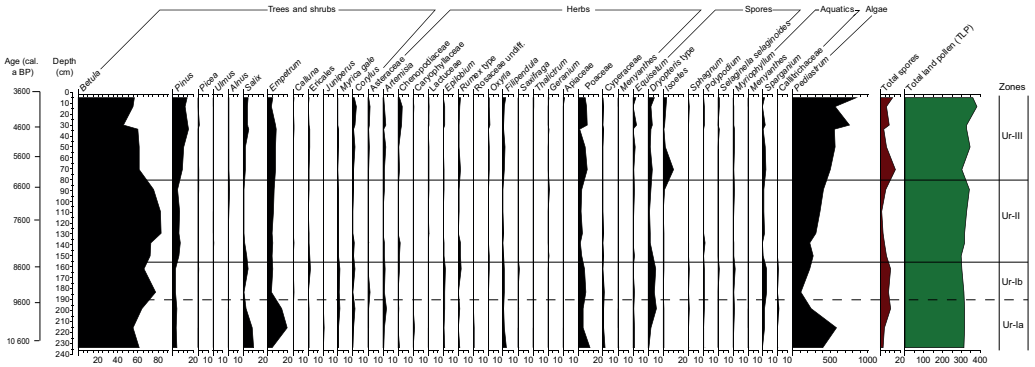


Fig. 6. Percentage pollen diagram for Uhca Rohči Lake with local pollen assemblage zones (LPAZs) Ur-I to Ur-III indicated. Pollen percentages are based on the sum of total terrestrial pollen (∑P). Proportion of aquatics and spores are calculated based on the sum of total terrestrial pollen plus aquatics (∑P + ∑aquatics) or spores (∑P + ∑spores).

source of the *sed*aDNA has probably changed over time, with the Komagelva River delivering some of the DNA to the lake from a larger source area of the upstream catchment during periodic flood events prior to downcutting in the later Holocene. Nevertheless, no distinct change in the taxonomic composition of *sed*aDNA samples is observed between units A, B and C, nor was any clear pattern observed between the presence of banding and/or laminae and the taxonomic composition of samples.

The prominence of aquatic taxa such as *Callitriche hermaphroditica*, *C. palustris* and *Potamogeton* in the *sed*aDNA between *c*. 10 200–9600 cal. a BP (Fig. 5) may reflect particularly good growing conditions in the lake, for example clear water conditions, warmer temperatures and/or more nutrients. In addition, aquatic plants are efficiently dispersed by birds and are therefore likely to show a rapid geographical response to climate change (Birks 2000; Väliranta *et al.* 2015). *Callitriche hermaphroditica*, which is not found in Varangerhalvøya National Park today, occurs in more continental sites in Finnmark and appears to require a minimum July temperature of 13–14 °C (Väliranta *et al.* 2015). The pattern of occurrence of the aquatic taxa suggests a response to warmer-than-present Early Holocene summers.

A handful of other taxa (*Rhododendron tomentosum*, *Limosella aquatica*) that are dominant in the *sed*aDNA signal (Figs 4, 5) do not occur in the region today, with current native ranges more than 50 km south or southeast of the Varanger Peninsula (see above). Thus, the continued presence and dominance of these taxa in the *sed*aDNA signal suggest a warmer climate between, at least, *c*. 10 000–4000 cal. a BP. This is in accordance with the general interpretation from pollen-based temperature reconstructions in northern Fennoscandia, which indicate July temperatures of +1.5±0.5 °C during the HTM (e.g. Møller & Holmeslet 2002; Jensen & Vorren 2008; Seppä *et al.* 2009; Sejrup *et al.* 2016).

Interpreting the major vegetation patterns

Little change is observed in the relative dominance of functional groups between *c*. 10 700 and 3300 cal. a BP, with trees and tall shrubs such as *Pinus*, Salicaceae, *Betula* and *Empetrum* accounting for a high percentage of the terrestrial pollen (mean 87%) and total DNA records (mean 50%; Fig. 4). Wind-pollinated woody taxa (e.g. *Betula*, *Pinus*) are generally over-represented in pollen studies (Prentice 1985; Sugita 1994). Likewise, we note that the abundance of these plant growth forms may be over-represented in our *sed*aDNA, dataset due to polymerase-related biases that generally occur during metabarcoding PCR (Alsos *et al.* 2018; Nichols *et al.* 2018), although calibration against modern vegetation suggests a bias in the opposite direction (Yoccoz *et al.* 2012).

Whilst a large majority (>85%) of the taxa detected in the *sed*aDNA are also found growing in Finnmark and the Kola Peninsula region today, the occurrence of *Pinus* in the *sed*aDNA signal from UR Lake raises questions. *Pinus* is found in nearly every sample, usually in high abundance, yet it is not a major component of the present-day flora of the Varanger Peninsula. The nearest forest stands occur around 50 km south of Varangerfjorden, although some scattered trees are present in the southern Varanger Peninsula (see above; Fig. 1B). The watershed of the Komagelva River, a likely source of *sed*aDNA to UR Lake during the time represented by lithostratigraphical units UR-1A and 1B, is situated outside of the present-day *Pinus* limit (Fig. 1B). The sustained high abundance of *Pinus* throughout the record, including lithostratigraphical unit UR-1C when the lake is presumed to have been isolated from riverine influence, suggests a source of *Pinus sed*aDNA to UR Lake other than the Komagelva River.

It is possible that *Pinus sed*aDNA in UR Lake originates from pollen and thus may indicate long-distance dispersal rather than local growth. Unlike angiosperms, pollen

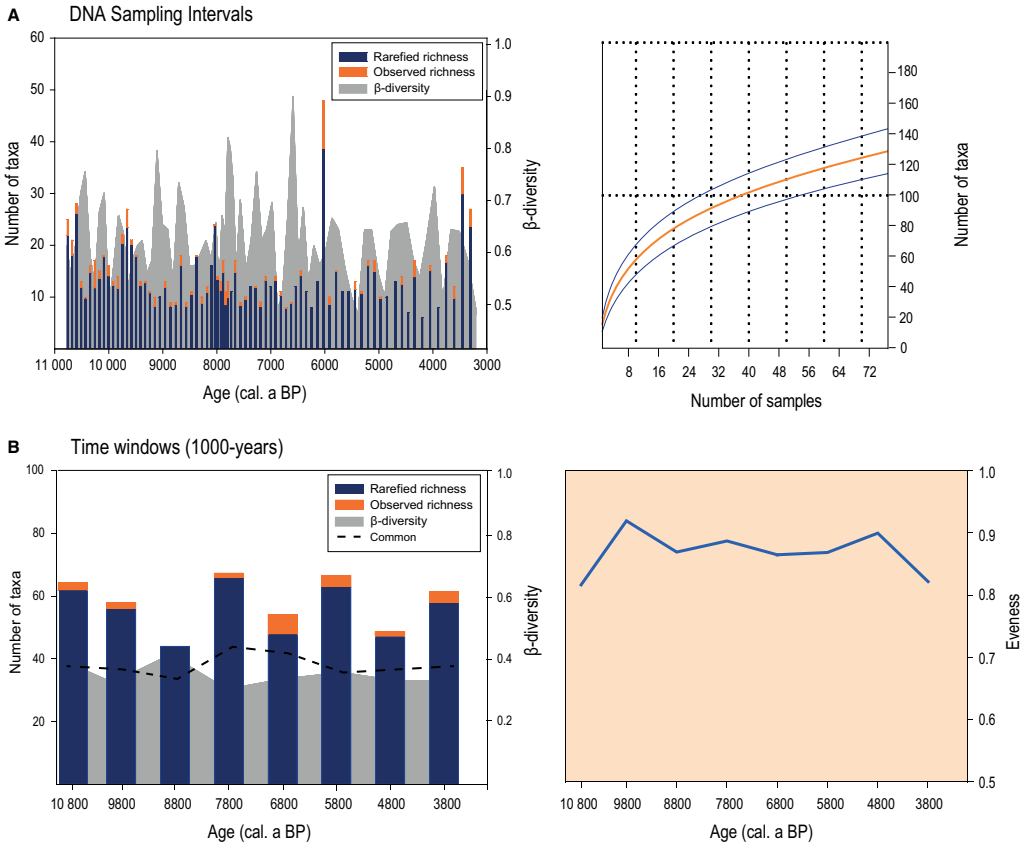


Fig. 7. Measures of species richness, beta diversity and sample evenness for (A) DNA sampling intervals (3-cm resolution) with a sample rarefaction curve presented in the right-hand graph outlining the relationship between sample size and the number of taxa identified (orange line) with 95% confidence intervals (blue lines) and (B) the same data amalgamated into 1000-year time windows for the Uhca Rohči Lake *sed*aDNA record. Dashed line indicates the number of taxa common between 1000-year time windows.

grains derived from gymnosperms contain some chloroplast DNA (cpDNA) within their reproductive cells (Suyama *et al.* 1996; Parducci *et al.* 2005) that, theoretically, could be introduced into the sediment matrix, either naturally or during DNA extraction. However, current thinking from *sed*aDNA studies suggests that DNA extracted from sediments does not derive from pollen grains (Jørgensen *et al.* 2012; Pedersen *et al.* 2016; Sjögren *et al.* 2017; Wang *et al.* 2017), but instead from other components embedded in the sediment matrix (Parducci *et al.* 2017). This is probably due to the generally lower biomass of pollen compared to stems, roots and leaves, and to the resilience of their sporopollenin coats, which requires a separate lysis step in the extraction of DNA (Kraaijeveld *et al.* 2015). The extraction of cpDNA from fossil pollen grains has proven difficult (Parducci *et al.* 2005; Bennett & Parducci 2006), which suggests that consistent detection of pollen-derived cpDNA from the sediment matrix itself is unlikely. Another possibility is contamination. *Pinus* passed the filtering stage, but there was a high number of *Pinus* reads in the negative controls (Table S2). Therefore, our *Pinus* record may support the inference of local presence, but there is enough doubt that other proxy data are required to establish whether pine was locally present during the Holocene.

Pinus was found in all of the pollen samples analysed from UR Lake (Fig. 6; mean 8% TLP; mean concentration 6100 grains g^{-1}). Maximum *Pinus* pollen influx rates (~200–870 grains cm^{-2} a^{-1}) are observed between *c.* 8000–7300 cal. a BP (Table S4). These are comparable to influx values reported from nearby Østervatnet (400–650 grains

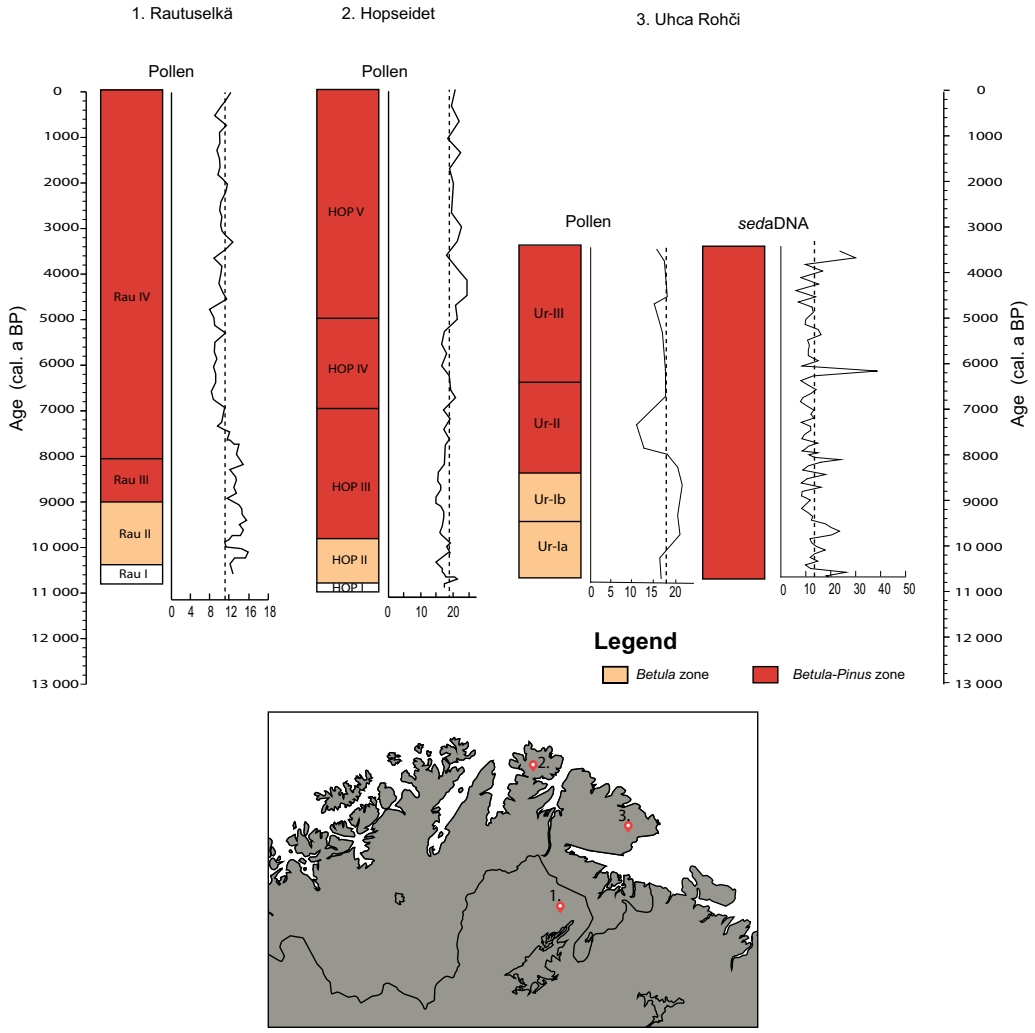


Fig. 8. Compilation of palynological richness for two published records (Rautuselkä and Hopseidet; Seppä 1998) and the Uhca Rohči pollen record (this paper). Floristic richness derived from *sed*aDNA record from Uhca Rohči Lake also indicated (this paper). Dashed line indicates mean species richness. *Betula* and *Betula-Pinus* pollen zones are indicated. Map (inset) details location of sites with 1 = Lake Rautuselkä; 2 = Lake Hopseidet and 3 = Uhca Rohči Lake.

cm^{-2} a^{-1}) *c.* 8000 cal. a BP (Table S3) and Mortensnes (2000 grains cm^{-2} a^{-1}), which is only 46 km to the SW (Høeg 2000). A threshold value of 500 grains cm^{-2} a^{-1} given by Hyvärinen (1985) and Hicks (1994) for indicating pine presence means that from our pollen data, *Pinus* may have been present at the site *c.* 8000–7300 cal. a BP (Table S3). This accords with the northward range expansion across northern Fennoscandia beginning *c.* 8500 cal. a BP (Hyvärinen 1975; Seppä 1996; Huntley *et al.* 2013) that probably reflected Early and Middle Holocene summer warmth, but demonstrates a probable lag in response, compared to aquatic taxa. Maximum *Betula* influx rates (~1300–8300 grains cm^{-2} a^{-1}; Table S4) are also observed during this interval *c.* 8000–7300 cal. a BP (Fig. 4), suggesting a mixed birch-pine forest.

Continuity and stability in Holocene flora

The record obtained from UR Lake reveals stability in the Holocene flora, with over 85% of the total taxa

detected by *sed*aDNA still growing within the catchment and/or in the broader region today (Table S2). Our *sed*aDNA results indicate that the erect shrub-tundra and the riparian grassland communities became established early: dominant taxa of the shrub communities (e.g. Salicaceae, *Betula*, *Empetrum*, *Vaccinium uligonosum*) and riparian meadows (*Bistorta vivipara*, *Caltha*, *Viola biflora*, *Rumex*, *Stellaria longifolia* and *S. borealis*) had appeared by *c*. 10 700 cal. a BP (Figs 4, 5), in accordance with pollen records from nearby Østervatnet and Bergebyvatnet showing a *Salix*-Poaceae assembly from *c*. 12 500 cal. a BP (Prentice 1981, 1982). Presence of these taxa continued to *c*. 3300 cal. a BP and they are common in the catchment today.

Overall, *sed*aDNA permitted identifications at a higher taxonomic resolution than was possible with pollen; the main exception is Salicaceae, which, based on the *trn*L marker, can only be identified to family level due to hybridization and similarity in cpDNA sequences amongst individuals. The *sed*aDNA signal detected taxa from a range of different ecological habitats including shrub tundra, riparian meadows and/or open grassland rich with herbs and bryophytes. It also reveals the dynamics of aquatic taxa. In contrast, the pollen record from UR Lake is largely confined to northern boreal and low arctic taxa (e.g. *Betula*, *Pinus*, *Salix*, *Empetrum*), with a minor signal deriving from herbaceous (e.g. Asteraceae, Chenopodiaceae, *Rumex*, Poaceae) and aquatic taxa (e.g. *Sparganium*, Callitrichaceae). In contrast to the *sed*aDNA record, entomophilous forbs typical of low tundra settings, such as *Bistorta vivipara*, are largely absent or in low percentages in the pollen record from UR Lake (Fig. 6, Table S2). The limited number of pollen samples and dominance of taxa such as *Betula* and *Pinus* probably restrict or mask the presence of low pollen-producing entomophilous taxa in the pollen record from UR Lake.

Direct comparison of the number of taxa contributed by each proxy to overall species richness is not appropriate due to disparity in sampling effort. Although previous comparisons between pollen and *sed*aDNA have shown only partial overlap of taxa (Pedersen *et al.* 2013; Parducci *et al.* 2015), our data demonstrate high similarity in the records for taxa such as *Empetrum*, *Callitriche* and Poaceae, as well as a dominance of *Betula* and *Pinus* (Figs 4, 5). Thus, in contrast to lakes that only receive inflow from catchments/small streams (Alsos *et al.* 2018), the DNA signal in UR Lake, which may have been affected by sediment inputs from the large, upstream catchment, more closely resembles the regional vegetation signal typical of many pollen records (Jacobson & Bradshaw 1981; Rousseau *et al.* 2006).

Postglacial patterns in floristic richness and diversity

Floristic richness reconstructed from *sed*aDNA displays only minor variations over the Holocene interval investigated, particularly when the effect of sporadic occurrences of herb taxa is minimized by merging samples into 1000-year time windows (Fig. 7). High floristic richness (61 taxa) characterizes the earliest time window between 10 700–9800 cal. a BP, and richness remains consistently high throughout the record; small fluctuations may be due to flood-related inputs from the extensive, up-river catchment. No long-term trends in floristic richness are evident, based on either individual samples or 1000-year time windows (Fig. 7). High sample evenness across time windows indicates the persistence of dominant taxa over the Holocene, with a large percentage (>70%) of taxa detected as common amongst 1000-year time windows (Fig. 7B). Our findings derived from *sed*aDNA support the conclusions of Normand *et al.* (2011) and Giesecke *et al.* (2012), which are derived from pollen, that the current distribution of plants in previously glaciated regions established quickly after the onset of the Holocene. For the Holocene at least, there is no evidence for increased floristic richness over time that might be due to the delayed immigration of species, but it should be noted that as the area was deglaciated prior to the end of the Younger Dryas (see above), the earliest part of the record is missing.

Palynological richness from UR Lake and two published records from Lake Rautuselkä and Hopseidet (Seppä 1998) show more variation than floristic richness reconstructed from *sed*aDNA (Fig. 8). Palynological richness displays long-term trends in response to variation in the pollen abundance of anemophilous woody taxa (e.g. *Pinus* and *Betula*). Dominance of taxa producing high amounts of pollen (e.g. *Pinus*, *Picea* and *Betula*) occurs at the expense of entomophilous forbs and leads to a likely underestimation of floristic richness. On the other hand, where low palynological richness coincides with high *Pinus* pollen influx, this may reflect an ecological effect whereby dense forest reduces niche availability for herbaceous taxa (Seppä 1998). Our *sed*aDNA data do not show such patterns, whereas the UR pollen data do, albeit at a coarse sampling resolution. The limitations on the effectiveness of pollen spectra for estimating species richness (Meltsov *et al.* 2011; Goring *et al.* 2013) mean that *sed*aDNA can provide improved estimates; less sensitive to 'swamping' by dominant taxa, as total numbers are not limited by counting time, and all valid reads contribute to the floristic list.

Conclusions

The pollen and *sed*aDNA records from this site on the Varanger Peninsula show a largely consistent pattern. Erect shrub-tundra vegetation was established early in the Holocene. This was soon followed by the establishment of *Betula* forest probably mixed with *Pinus* as indicated by high pollen influx rates of both taxa. This, together with the persistence of *Limosella aquatica* and *Rhododendron tomentosum* through the *sed*aDNA record, indicates a climate warmer than present throughout most of the Early and Middle Holocene.

As species richness reconstructed from pollen data is limited by differential pollen productivity, dispersal and often low taxonomic resolution, alignment with contemporary data often proves difficult if not impossible. Reconstructions of past changes in floristic diversity and richness can be improved by using *sed*aDNA, as it better reflects local floristic composition and diversity. The *sed*aDNA data on floristic richness between *c*. 10 700 and 3300 cal. a BP show that high diversity, richness and sample evenness prevailed across the record, despite known climatic variations in the Holocene. Nevertheless, when considering past and future responses to climate change, important questions remain as to how the crossing of critical climatic thresholds may interact with sub-regional heterogeneity and drivers of vegetation composition and change, including herbivory. Here, as this paper shows, *sed*aDNA studies can make an important contribution.

Acknowledgements. – This study was supported by the Norwegian Research Council (grant nos. 213692/F20 and 250963/F20 to I. G. Alsos) with a Ph.D. studentship for C. L. Clarke provided by the UK Natural Environment Research Council (grant no. NE/L002531/1). We thank the board of the Varanger Peninsula National Park for permission to core (permission 2015/4664 432.3 granted to K. A. Bråthen). We also thank Frédéric Boyer for raw sequence data handling, Catherine Langdon and Rob Scaife for fruitful discussions on the pollen results and Dorothee Ehrich, Kelsey Lorberau and Dilli Prasad Rijal for discussions on the DNA analysis. We thank Vendela K. Lagerholm and Nancy Bigelow for reviewing the manuscript and providing constructive comments. K.A. Bråthen surveyed potential study sites and C. L. Clarke, A. G. Brown, I. G. Alsos and F. J. Ancin-Murguzur cored Uhca Rohči Lake. C. L. Clarke extracted the DNA and L. Gielly performed the amplifications and ran ObiTools. C. L. Clarke analysed the DNA data with contributions from Y. Lammers, P. D. Heintzman and I. G. Alsos. A. G. Brown described the sediment properties, provided the geomorphological background and compiled data on *Pinus* pollen influx rates. Radiocarbon dating was performed by T. Goslar. C. L. Clarke constructed the age-depth models and performed XRF, LOI, magnetic susceptibility and pollen analysis. M. E. Edwards contributed to pollen analysis. I. G. Alsos, K. A. Bråthen and M. E. Edwards provided ecological background. C. L. Clarke drafted the first version of the manuscript of which all co-authors commented on.

References

Abbott, R. J. & Brochmann, C. 2003: History and evolution of the arctic flora: in the footsteps of Eric Hultén. *Molecular Ecology 12*, 299–313.

Allen, J. R. M., Long, A. J., Ottley, C. J., Pearson, D. G. & Huntley, B. 2007: Holocene climate variability in northernmost Europe. *Quaternary Science Reviews 26*, 1432–1453.

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, https://doi.org/10.1371/journal.pone.0195403.

Alsos, I. G., Sjögren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., Coissac, E., Brown, A. G., Jakobsen, L. V., Føreid, M. K. & Pedersen, M. W. 2016: Sedimentary ancient DNA from Lake Skartjørna, Svalbard: assessing the resilience of arctic flora to Holocene climate change. *The Holocene 26*, 627–642.

Barnekow, L. 1999: Holocene tree-line dynamics and inferred climatic changes in the Abisko area, northern Sweden, based on macrofossil and pollen records. *The Holocene 9*, 253–265.

Bennett, K. D. & Parducci, L. 2006: DNA from pollen: principles and potential. *The Holocene 16*, 1031–1034.

Berger, A. L. 1978: Long-term variations of caloric insolation resulting from the earth's orbital elements. *Quaternary Research 9*, 139–167.

Berger, A. & Loutre, M. F. 1991: Insolation values for the climate of the last 10 million of years. *Quaternary Science Reviews 10*, 297–317.

Berglund, B. E., Gaillard, M. J., Björk, L. & Persson, T. 2008a: Long-term changes in floristic diversity in southern Sweden: palynological richness, vegetation dynamics and land-use. *Vegetation History and Archaeobotany 17*, 573–583.

Berglund, B. E., Persson, T. & Björkman, L. 2008b: Late Quaternary landscape and vegetation diversity in a North European perspective. *Quaternary International 184*, 187–194.

Birks, H. H. 2000: Aquatic macrophyte vegetation development in Kråkenes Lake, western Norway, during the late-glacial and early-Holocene. *Journal of Paleolimnology 23*, 7–19.

Birks, H. J. B. & Birks, H. H. 2008: Biological responses to rapid climate change at the Younger Dryas - Holocene transition at Kråkenes, western Norway. *The Holocene 18*, 19–30.

Birks, H. J. B. & Line, J. M. 1992: The use of rarefaction analysis for estimating palynological richness from Quaternary pollen-analytical data. *The Holocene 2*, 1–10, https://doi.org/10.1177/095968369200200101.

Birks, H. H., Jones, V. J., Brooks, S. J., Birks, H. J. B., Telford, R. J., Juggins, S. & Peglar, S. M. 2012: From cool to cool in northernmost Norway: lateglacial and early Holocene multi-proxy environmental and climate reconstructions from Jansvatnet, Hammerfest. *Quaternary Science Reviews 33*, 100–120.

Bjune, A. E., Birks, H. J. B. & Seppä, H. 2004: Holocene vegetation and climate history on a continental-oceanic transect in northern Fennoscandia based on pollen and plant macrofossils. *Boreas 33*, 211–223.

Blaauw, M. & Christen, J. A. 2011: Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis 6*, 457–474.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. 2016: OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources 16*, 176–182.

Bråthen, K. A., Ravolainen, V. T., Stien, A., Tveraa, T. & Ims, R. I. 2017: Rangifer management controls a climate-sensitive tundra state transition. *Ecological Applications 27*, 2416–2427.

Brown, A. G. 1999: Biodiversity from pollen analysis: modern pollen studies and the recent history of a floodplain woodland in S. W. Ireland. *Journal of Biogeography 26*, 19–32.

Chambers, F. M., Beilman, D. W. & Yu, Z. 2011: Methods for determining peat humification and for quantifying peak bulk density, organic matter and carbon content for palaeostudies of climate and peatland carbon dynamics. *Mires and Peat 7*, http://www.mires-and-peat.net/, ISSN 1819-754X.

COAT (Climate-ecological Observatory for Arctic Tundra) 2018: Available at: http://www.coat.no, (accessed 15.01.2018).

Croudace, I. W. & Rothwell, R. G. 2015: *Micro-XRF of Sediment Cores: A Non-destructive Tool for the Environmental Sciences.* 421 pp. Springer, Dordrecht.

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

Elven, R. (red.) 2005: *Norsk Flora (Lid, D. T. & Lid, J. Norwegian Flora)*. 1230 pp. Det Norske Samlaget, Oslo.

Epp, L. S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., Ludikova, A., Hassel, K., Stenøien, H. K., Funder, S., Willerslev, E., Kjær, K. & Brochmann, C. 2015: Lake sediment multitaxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews 117*, 152–163.

Fægri, K. & Iversen, J. 1989: *Textbook of Pollen Analysis.* 328 pp. Wiley, New York.

Felde, V. A., Peglar, S. M., Bjune, A. E., Grytnes, J.-A. & Birks, H. J. B. 2017: Are diversity trends in western Scandinavia influenced by post-glacial dispersal limitation? *Journal of Vegetation Science 29*, 360–370.

Fjellanger, J. & Sørbel, L. 2007: Origin of the palaeic landforms and glacial impact on the Varanger Peninsula, northern Norway. *Norwegian Journal of Geology 87*, 223–238.

Fletcher, C. H., Fairbridge, R. W., Moller, J. & Long, A. 1993: Emergence of the Varanger Peninsula, Arctic Norway, and climate changes since deglaciation. *The Holocene 3*, 116–127.

Fredh, D., Broström, A., Zillén, L., Mazier, F., Rundgren, M. & Lagerås, P. 2012: Floristic diversity in the transition from traditional to modern land-use in southern Sweden A.D. 1800-2008. *Vegetation History and Archaeobotany 21*, 439–452.

Gajewski, K. 2015: Impact of Holocene climate variability on Arctic vegetation. *Global and Planetary Change 133*, 272–287.

Giesecke, T., Wolters, S., Jahns, S. & Brande, A. 2012: Exploring Holocene change in palynological richness in northern Europe- Did postglacial immigration matter? *PLoS ONE 7*, e51624, https://doi.org/10.1371/journal.pone.0051624.

Goring, S., Lacourse, T., Pellatt, M. G. & Mathewes, R. W. 2013: Pollen assemblage richness does not reflect regional plant species richness: a cautionary tale. *Journal of Ecology 101*, 1137–1145.

Goslar, T., Czernik, J. & Goslar, E. 2004: Low-energy ^{14}C AMS in Poznan radiocarbon Laboratory, Poland. *Nuclear Instruments and Methods in Physics Research B 223–224*, 5–11.

Hammer, Ø., Harper, D. A. T. & Ryan, P. D. 2001: PAST: palaeontological statistics software package for education and data analysis. *Palaeontologia Electronica 4*, http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Hanssen-Bauer, I. & Tveito, O. E. 2014: Climate and climate change. *Ottar 4*, 31–36, in Norwegian.

Heikkinen, O. 2005: Boreal forests and northern and upper timberlines. *In* Seppälä, M. (ed.): *The Physical Geography of Fennoscandia*, 185–200. Oxford University Press, New York.

Heiri, O., Lotter, A. F. & Lemcke, G. 2001: Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *Journal of Paleolimnology 25*, 101–110.

Henden, J.-A., Ims, R. A., Yoccoz, N. G., Sørensen, R. & Killengreen, S. T. 2011: Population dynamics of tundra voles in relation to configuration of willow thickets in southern arctic tundra. *Polar Biology 34*, 533–540.

Hicks, S. 1994: Present and past pollen records of Lapland forests. *Review of Palaeobotany and Palynology 82*, 17–35.

Hicks, S. & Hyvärinen, H. 1999: Pollen influx values measured in different sedimentary environments and their palaeoecological implications. *Grana 38*, 228–242.

Høeg, H. I. 2000: Pollen analytical investigations in Finnmark, north Norway. *AmS-Varia 37*, 53–97.

Hughes, A. L. C., Gyllencreutz, R., Lohne, Ø. S., Mangerud, J. & Svendsen, J. I. 2016: The last Eurasian ice sheets- a chronological database and time-slice reconstruction, DATED-1. *Boreas 45*, 1–45.

Hultén, E. 1937: *Outline of the History of Arctic and Boreal Biota during the Quaternary Period*. 168 pp. Lehre J Cramer, New York.

Hultén, E. & Fries, M. 1986: *Atlas of North European Vascular Plants North of the Tropic of Cancer*. 1172 pp. Koeltz Scientific Books, Königstein.

Huntley, B., Long, A. J. & Allen, J. R. M. 2013: Spatio-temporal patterns in Lateglacial and Holocene vegetation and climate of Finnmark, northernmost Europe. *Quaternary Science Reviews 70*, 158–175.

Hustich, I. 1983: Tree-line and tree growth studies during 50 years: some subjective observations. *Nordicana 47*, 181–188.

Hyvärinen, H. 1975: Absolute and relative pollen diagrams from northernmost Fennoscandia. *Fennia - International Journal of Geography* 142, https://fennia.journal.fi/article/view/9189.

Hyvärinen, H. 1976: Flandrian pollen deposition rates and tree-line history in northern Fennoscandia. *Boreas 5*, 163–175.

Hyvärinen, H. 1985: Holocene pollen historyof the Alta area, anisolated pine forest north of the general pine forest region in Fennoscandia. *Ecologia Meditteranea 11*, 69–71.

Jacobson, G. L. & Bradshaw, R. H. W. 1981: The selection of sites for paleovegetational studies. *Quaternary Research 16*, 80–96.

Jensen, C. & Vorren, K.-D. 2008: Holocene vegetation and climate dynamics of the boreal alpine ecotone of northwestern Fennoscandia. *Journal of Quaternary Science 23*, 719–743.

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

Karlsen, S. R., Elvebakk, A. & Johansen, B. 2005: A vegetation-based method to map climatic variation in the arctic-boreal transition area of Finnmark, north-easternmost Norway. *Journal of Biogeography 32*, 1161–1186.

Koleff, P., Gaston, K. J. & Lennon, J. J. 2003: Measuring beta diversity for presence-absence data. *Journal of Animal Ecology 72*, 367–382.

Kraaijeveld, K., de Weger, L. A., Garcia, M. V., Buermans, H., Frank, J., Hiemstra, P. S. & den Dunnen, J. T. 2015: Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources 15*, 8–16.

Lamb, H. F. & Edwards, M. E. 1998: The Arctic. *In* Huntley, B. & Webb, T. III (eds.): *Vegetation History*, 519–555. Kluwer Academic Publishers, Dordrecht.

Magurran, A. E. 2004: *Measuring Biological Diversity*. 256 pp. Blackwell, Oxford.

Meltsov, V., Poska, A., Odgaard, B. V., Sammul, M. & Kull, T. 2011: Palynological richness and pollen sample evenness in relation to local floristic diversity in southern Estonia. *Review of Palaeobotany and Palynology 166*, 344–351.

Møller, J. J. & Holmeslet, B. 2002: *Havets Historie i Fennoskandia og NV Russland*. Available at: http://geo.phys.uit.no/sealev/.

Nesje, A. 1992: A piston corer for lacustrine and marine sediments. *Arctic and Alpine Research 24*, 257–259.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E. & Shapiro, B. 2018: Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources 18*, 927–939.

Normand, S., Ricklefs, R. E., Skov, F., Bladt, J., Tackenberg, O. & Svenning, J. C. 2011: Postglacial migration supplements climate in determining plant species ranges in Europe. *Proceedings of the Royal Society B: Biological Sciences 278*, 3644–3653.

Odgaard, B. V. 1999: Fossil pollen as a record of past biodiversity. *Journal of Biogeography 26*, 7–17.

Odgaard, B. V. 2001: Palaeoecological perspectives on pattern and process in plant diversity and distribution adjustments: a comment on recent developments. *Diversity and Distributions 7*, 197–201.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P. M., Stevens, H. H., Szoecs, E. & Wagner, H. 2017: *Vegan: community ecology package*. Version 2.4-2. R package. Available at: https://CRAN.R-project.org/package5vegan

Olsen, L., Mejdahl, V. & Selvik, S. 1996: Middle and Late Pleistocene stratigraphy, chronology and glacial history in Finnmark, North Norway. *Norges Geologiske Undersøkelse 429*, 1–111.

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*, 924–942.

Parducci, L., Matetovici, I., Fontana, S. L., Bennett, K. D., Suyama, Y., Haile, J., Kjær, K. H., Larsen, N. K., Drouzas, A. D. & Willerslev, E. 2013: Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology 22*, 3511–3524.

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

Parducci, L., Väliranta, M., Salonen, J. S., Ronkainen, T., Matetovici, I., Fontana, S. L., Eskola, T., Sarala, P. & Suyama, Y. 2015: Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions of the Royal Society B 370*, 1660, https://doi.org/10.1098/rstb.2013.0382.

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

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J.,

Kjær, K. H. & Willerslev, E. 2016: Postglacial viability and colonization in North America's ice-free corridor. *Nature 537*, 45–49.
Prentice, H. C. 1981: A Late Weichselian and early Flandrian pollen diagram from Østervatnet, Varanger peninsula, NE Norway. *Boreas 10*, 53–70.
Prentice, H. C. 1982: Late Weichselian and early Flandrian vegetational history of Varanger Peninsula, northeast Norway. *Boreas 11*, 187–208.
Prentice, I. C. 1985: Pollen representation, source area and basin size: toward a unified theory of pollen analysis. *Quaternary Research 23*, 76–86.
R Core Team 2017: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/.
Ravolainen, V. T., Bråthen, K. A., Ims, R. A., Yoccoz, N. G., Henden, J.-A. & Killengreen, S. T. 2011: Rapid, landscape scale responses in riparian tundra vegetation to exclusion of small and large mammalian herbivores. *Basic and Applied Ecology 12*, 643–653.
Ravolainen, V. T., Bråthen, K. A., Ims, R. A., Yoccoz, N. G. & Soininen, E. M. 2013: Shrub patch configuration at the landscape scale is related to diversity of adjacent herbaceous vegetation. *Plant Ecology and Diversity 6*, 257–268.
Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Bronk Ramsey, C., Buck, C. E., Cheng, H., Edwards, R. L., Freidrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatte, C., Heaton, T. J., Hoffman, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M. & van der Pilcht, J. 2013: IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal. BP. *Radiocarbon 55*, 1869–1887.
Reitlau, T., Gerhold, P., Poska, A., Pärtel, M., Väli, V. & Veski, S. 2015: Novel insights into post-glacial vegetation change: functional and phylogenetic diversity in pollen records. *Journal of Vegetation Science 26*, 911–922.
Rousseau, D. D., Schevin, P., Duzer, D., Cambon, G., Ferrier, J., Jolly, D. & Poulsen, U. 2006: New evidence of long-distance pollen transport to Southern Greenland in Late Spring. *Review of Palaeobotany and Palynology 141*, 277–286.
Sejrup, H. P., Seppä, H., McKay, N. P., Kaufman, D. S., Geirsdóttir, Á., de Vernal, A., Renssen, H., Husum, K., Jennings, A. & Andrews, J. T. 2016: North Atlantic-Fennoscandian Holocene climate trends and mechanisms. *Quaternary Science Reviews 147*, 365–378.
Seppä, H. 1996: Post-glacial dynamics of vegetation and tree-lines in the far north of Fennoscandia. *Fennia 174*, 1–96.
Seppä, H. 1998: Postglacial trends in palynological richness in the northern Fennoscandian tree-line area and their ecological interpretation. *The Holocene 8*, 43–53.
Seppä, H., Bjune, A. E., Telford, R. J., Birks, H. J. B. & Veski, S. 2009: Last nine-thousand years of temperature variability in Northern Europe. *Climate of the Past 5*, 523–535.
Siedlecka, A. & Roberts, D. 1992: *The Bedrock Geology of Varanger Peninsula, Finnmark, North Norway: An Excursion Guide*. 45 pp. Norges Geologiske Undersokelse, Trondheim.
Sjögren, P. & Damm, C. 2019: Holocene vegetation change in northernmost Fennoscandia and the impact on prehistoric foragers 12 000–2000 cal. a BP – A review. *Boreas*, *48*, 20–35.
Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Føreid Merkel, M. K., Fonville, T. & Alsos, I. G. 2017: Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. *New Phytologist 213*, 929–941.
Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., Gussarova, G., Bellemain, E., Hassel, K., Stenøien, H. K., Epp, L., Schrøder-Nielsen, A., Brochmann, C. & Yoccoz, N. G. 2015: Highly overlapping winter diet in two sympatric lemming species revealed by DNA metabarcoding. *PLoS ONE 10*, e0115335, https://doi.org/10.1371/journal.pone.0115335.
Soininen, J., Passy, S. & Hillebrand, H. 2012: The relationship between species richness and evenness: a meta-analysis of studies across aquatic ecosystems. *Oecologia 169*, 803–809.
Sollid, J. L., Andersen, S., Hamre, N., Kjeldsen, O., Salvigsen, O., Sturod, S., Tveita, T. & Wilhelmsen, A. 1973: Deglaciation of Finnmark, North Norway. *Norsk Geografisk Tidsskrift, Norwegian Journal of Geography 27*, 233–325.
Sønstebo, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P. & Brochmann, C. 2010: Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources 10*, 1009–1018.
Stokes, C. R., Corner, G. D., Winsborrow, M. C. M., Husum, K. & Andreassen, K. 2014: Asynchronous response of marine-terminating outlet glaciers during deglaciation of the Fennoscandian Ice Sheet. *Geology 42*, 455–458.
Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B. W., Harbor, J. M., Jansen, J. D., Olsen, L., Caffee, M. W., Fink, D., Lundqvist, J., Rosqvist, G. C., Strömberg, B. & Jansson, K. N. 2016: Deglaciation of Fennoscandia. *Quaternary Science Reviews 147*, 91–121.
Sugita, S. 1994: Pollen representation of vegetation in Quaternary sediments: theory and method in patchy vegetation. *Journal of Ecology 82*, 881–897.
Suyama, Y., Kawamuro, K., Kinoshita, I., Yoshimura, K., Tsumura, Y. & Takahara, H. 1996: DNA sequence from a fossil pollen of *Abies* spp from Pleistocene peat. *Genes, Genetics and Systematics 71*, 145–149.
Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. 2012: Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology 21*, 2045–2050.
Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C. & Willerslev, E. 2007: Power and Limitations of the Chloroplast trnL (UAA) Intron for Plant DNA Barcoding. *Nucleic Acids Research 35*, e14, https://doi.org/10.1093/nar/gkl938.
Tolgensbakk, J. & Sollid, J. L. 1981: Vardø. Kvartærgeologisk geomorfologisk kart 2535 IV, M 1:50,000. *Norsk Geografisk Tidsskrift 27*, 233–325.
Väliranta, M., Salonen, J. S., Heikkilä, M., Amon, L., Helmens, K., Klimaschewski, A., Kuhry, P., Kultti, S., Poska, A., Shala, S., Veski, S. & Birks, H. H. 2015: Plant macrofossil evidence for an early onset of the Holocene summer thermal maximum in northernmost Europe. *Nature Communications 6*, 6809, https://doi.org/10.1038/ncomms7809.
Virtanen, R., Oksanen, L. & Razzhivin, V. 1999: Topographic and regional patterns of tundra heath vegetation from northern Fennoscandia to the Taimyr Peninsula. *Acta Botanica Fennica 167*, 29–83.
Walker, D. A., Raynolds, M. K., Daniels, F. J. A., Einarsson, E., Elvebakk, A., Gould, W. A., Katenin, A. E., Kholod, S. S., Markon, C. J., Melnikov, E. S., Moskalenko, N. G., Talbot, S. S. & Yurtsev, B. A. 2005: The Circumpolar Arctic vegetation map. *Journal of Vegetation Science 16*, 267–282.
Wang, Y., Heintzman, P., Newsom, L., Bigelow, N. H., Wooller, M. J., Shapiro, B. & Williams, J. W. 2017: The southern coastal Beringian land bridge: cryptic refugium or pseudorefugium for woody plants during the Last Glacial Maximum? *Journal of Biogeography 44*, 1559–1571.
Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I. G., Bellemain, E., Brysting, A. K., Elven, R., Sønstebø, J. H., Murton, J., Sher, A., Rasmussen, M., Rønn, R., Mourier, T., Cooper, A., Austin, J., Möller, P., Froese, D., Zazula, G., Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R. G., MacPhee, R. D., Gilbert, M. T., Kjær, K. H., Orlando, L., Brochmann, C. & Taberlet, P. 2014: Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature 506*, 47–51.
Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., Von Stedingk, H., Brysting, A. K., Coissac, E., Pompanon, F., Sønstebø, J. H., Miquel, C., Valentini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., Rasmussen, M., Gilbert, M. T. P., Orlando, L., Brochmann, C., Willerslev, E. & Taberlet, P. 2012: DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology 21*, 3647–3655.

Zimmerman, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R. & Schwamborn, G. 2017: Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-east Siberia). *Biogeosciences 14*, 575–596.

Supporting Information

Additional Supporting Information may be found in the online version of this article at http://www.boreas.dk.

Fig. S1. Woody and graminoid taxa detected by *sed*aDNA in Uhca Rohči Lake.

Fig. S2. Forb taxa detected by *sed*aDNA in Uhca Rohči Lake.

Fig. S3. Ferns, bryophytes, clubmoss and aquatic taxa detected by *sed*aDNA in Uhca Rohči Lake.

Table S1. Total DNA read numbers remaining after each filtering step described in the Material and methods section.

Table S2. All taxa detected from *sed*aDNA (orange) and pollen (blue) analyses, sorted according to family within groups: vascular plants, bryophytes and clubmosses, aquatics and algae.

Table S3. Published *Pinus* pollen influx values from Finnmark, northern Norway, (administrative region in brackets) and adjacent sites in Finland.

Table S4. Pinus and *Betula* pollen influx values for Uhca Rohči Lake for cores UR-1 and UR-2.

Paper III

BOREAS An international journal of Quaternary research

Clitellate worms (Annelida) in lateglacial and Holocene sedimentary DNA records from the Polar Urals and northern Norway

YOURI LAMMERS, CHARLOTTE L. CLARKE, CHRISTER ERSÉUS, ANTONY G. BROWN, MARY E. EDWARDS, LUDOVIC GIELLY, HAFLIDI HAFLIDASON, JAN MANGERUD, EMILIA ROTA, JOHN INGE SVENDSEN AND INGER GREVE ALSOS

BOREAS

Lammers, Y., Clarke, C. L., Erséus, C., Brown, A. G., Edwards, M. E., Gielly, L., Haflidason, H., Mangerud, J., Rota, E., Svendsen, J. I. & Alsos, I. G. 2019 (April): Clitellate worms (Annelida) in lateglacial and Holocene sedimentary DNA records from the Polar Urals and northern Norway. *Boreas*, Vol. 48, pp. 317–329. https://doi.org/10.1111/bor.12363. ISSN 0300-9483.

While there are extensive macro- and microfossil records of a range of plants and animals from the Quaternary, earthworms and their close relatives amongst annelids are not preserved as fossils and therefore the knowledge of their past distributions is limited. This lack of fossils means that clitellate worms (Annelida) are currently underused in palaeoecological research, even though they can provide valuable information about terrestrial and aquatic environmental conditions. Their DNA might be preserved in sediments, which offers an alternative method for detection. Here we analyse lacustrine sediments from lakes in the Polar Urals, Arctic Russia, covering the period 24 000–1300 cal. a BP, and NE Norway, covering 10 700–3300 cal. a BP, using a universal mammal 16S rDNA marker. While mammals were recorded using the marker (reindeer was detected twice in the Polar Urals core at 23 000 and 14 000 cal. a BP, and four times in the Norwegian core at 11 000 cal. a BP and between 3600–3300 cal. a BP), worm extracellular DNA 'bycatch' was rather high. In this paper we present the first reported worm detection from ancient DNA. Our results demonstrate that both aquatic and terrestrial clitellates can be identified in late-Quaternary lacustrine sediments, and the ecological information retrievable from this group warrants further research with a more targeted approach.

Youri Lammers (youri.lammers@uit.no) and Inger Greve Alsos, Tromsø Museum, University of Tromsø – The Arctic University of Norway, NO-9037 Tromsø, Norway; Charlotte L. Clarke and Antony G. Brown, Geography and Environment, University of Southampton, Highfield Southampton SO17 1BJ, UK; Christer Erséus, Department of Biological and Environmental Sciences, University of Gothenburg, P.O. Box 463, SE-405 30 Göteborg, Sweden; Antony G. Brown and Mary E. Edwards, Tromsø Museum, University of Tromsø – The Arctic University of Norway, NO-9037 Tromsø, Norway and Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK; Ludovic Gielly, Laboratoire d'Ecologie Alpine (LECA), Université Grenoble Alpes, C2 40700 38058, Grenoble Cedex 9, France; Haflidi Haflidason, Jan Mangerud and John I. Svendsen, Department of Earth Science, University of Bergen, PO Box 7803, NO-5020 Bergen, Norway and Bjerknes Centre for Climate Research, Bergen, Norway; Emilia Rota, Department of Physics, Earth and Environmental Sciences, University of Siena, Via P.A. Mattioli 4, IT-53100 Siena, Italy; *received 30th March 2018, accepted 12th October 2018.*

The fact that earthworms (Clitellata: Megadrili) have an important function in cycling nutrients and structuring soils was famously recognized by Darwin (Darwin 1881). Both earthworms as well as potworms (Clitellata: Enchytraeidae) are used as indicator species in various environmental studies of modern soils and aquatic systems as some are very tolerant of pollution while others are very sensitive (Karaca *et al.* 2010). In theory, they have high potential as indicators due to their known sensitivity to soil conditions including temperature, moisture status, soil texture and pH range (Edwards & Lofty 1977; Beylich & Graefe 2009). However, as soft-bodied organisms, worms rarely get preserved in sediments except as trace fossils and earthworm calcite granules (which can be radiocarbon dated; Canti 2003). Their limited preservation means that worms are currently underused in palaeoecology, even though they can provide valuable ecological information.

DNA barcoding has proven to be an important tool for the identification of species through the amplification and sequencing of small, yet informative, parts of the genome (Hebert *et al.* 2003). The barcoding process was revolutionized with the advent of next-generation sequencing, allowing complex samples such as environmental DNA to be barcoded (metabarcoding; Taberlet *et al.* 2012). Since then metabarcoding has been applied to a wide range of organisms, such as nematodes (Porazinska *et al.* 2009), plants (Taberlet *et al.* 2007; Parducci *et al.* 2017; Zimmermann *et al.* 2017), clitellate worms (Bienert *et al.* 2012; Epp *et al.* 2012; Pansu *et al.* 2015), amphibians and bony fishes (Valentini *et al.* 2015), fungi (Buée *et al.* 2009; Epp *et al.* 2012) and a range of other organisms (Thomsen & Willerslev 2015; Domaizon *et al.* 2017).

After being released into the environment by organisms, extracellular DNA degrades over time, but stabilized smaller fragments can persist over longer periods bound to fine-grained sediment particles or due to low temperatures (Pääbo *et al.* 2004; Willerslev *et al.* 2004; Barnes & Turner 2016). Thus, lake sediments in arctic or mountainous regions are prime locations for the recovery of ancient DNA (Parducci *et al.* 2012; Giguet-Covex *et al.* 2014; Pedersen *et al.* 2016). Metabarcoding of sedimentary ancient DNA (*sedaDNA*; Haile *et al.* 2009) can provide valuable information about past environments and augment traditional methods such as pollen or macrofossils (Pedersen *et al.* 2013; Parducci *et al.* 2015;

DOI 10.1111/bor.12363

Zimmermann *et al.* 2017) and is of particular interest for taxa that leave limited traces in the fossil record such as worms (Domaizon *et al.* 2017). Metabarcoding efforts targeting enchytraeid worms in ancient permafrost have been attempted before, but unlike those in modern soils, they yielded no results (Epp *et al.* 2012), suggesting that detection of worms in *sed*aDNA is not as straightforward as for other taxa that have been explored.

In this study, we set out to analyse mammalian DNA from lateglacial and Holocene lake sediments for faunal reconstruction, but because of the low retrieval of mammalian DNA and the unexpected clitellate DNA barcoding 'bycatch', we explore the potential for DNA-based clitellate palaeoecology.

Study sites

Polar Urals

Lake Bolshoye Shchuchye is located in the northernmost Polar Ural Mountains of Arctic Russia (latitude 67°53′24″N, longitude 66°18′53″E, altitude 221 m a.s.l.; Fig. 1). Bolshoye Shchuchye is an elongated lake (12 km long, 1 km wide) located in a NW–SE orientated valley with a maximum water depth of 136 m in its central part (Svendsen *et al.* 2019) and up to 160 m of lacustrine sediments in the central and northern parts (Haflidason *et al.* 2019). The lake is flanked by steep rock faces but the terrain is more open towards its north side, resulting in a total catchment area of 215 km^2 (Svendsen *et al.* 2019). The bedrock consists of Proterozoic-Cambrian basaltic and andesitic rocks in the eastern and northwestern parts of the catchment and Ordovician quartzite and phyllitic rocks in the southwestern catchment (Dushin *et al.* 2009). The Polar Urals remained mostly ice-free during the Last Glacial Maximum (LGM), except for cirque glaciers or minor valley glaciers (Svendsen *et al.* 2004). Current climate conditions are cold and continental with mean summer temperatures of 7 °C (Solomina *et al.* 2010).

Varanger Peninsula

The Uhca Rohči lake (70°19′07″N, 30°01′44″E; Fig. 1) is unnamed on the 1:50 000 Norwegian Topographic Map (Norgeskart; https://www.norgeskart.no), but we use this name as the lake is located close to a river site with the local Sami name 'Uhca Rohči'. Uhca Rohči is a small lake (<1 ha) in a depression situated at 138 m a.s.l. within the river valley of Komagdalen on the Varanger Peninsula, northeast Finnmark, Norway. The Varanger Peninsula is a low-relief plateau (200–600 m a.s.l.) moulded from the Proterozoic paleic surface (pre-Quaternary erosion surface) by marine and glacial processes (Siedlecka & Roberts 1992). Relief is strongly controlled by rock type and structure, with the ridges being formed of Cambrian quartzites and sandstones, whilst valleys are eroded into shales and mudstones. There is evidence that, having undergone uplift during the Pliocene, the area was subject to processes of erosion related to former sea levels, and glacial erosion (Fjellanger & Sørbel 2007). The lake's bedrock is composed of sandstone and mudstone (The Geological survey of Norway; www.ngu.no) and the Komagdalen valley was probably deglaciated by 15.4–14.2 ka BP; the peninsula was certainly free of glacial ice by 13 000–12 000 cal. a BP (Stokes *et al.* 2014; Hughes *et al.* 2016; Stroeven *et al.* 2016). The current climate is low Arctic with a mean summer temperature of 8.7 °C (Norwegian Meteorological Institute; www.met.no).

Material and methods

Polar Urals lake sediment

Lake Bolshoye Shchuchye was cored during several expeditions between 2007 and 2009. The 24-m-long core 506-48 that was sampled for metabarcoding was obtained in July 2009 from the southern part of the lake (67°51′22.2″N, 66°21′30.1″E). Coring was conducted with a UWITEC Piston Corer (http://www.uwitec.at) using 2-m-long by 10-cm-diameter PVC or 2-m-long by 9-cm-diameter steel tubes. The full core was obtained by taking consecutive segments from the same hole. All sections were stored and transported at above 0 °C to avoid freezing of the material. The core was subsampled within the Centre for Geobiology and Microbiology (University of Bergen) in a laminar flow cabinet and in the presence of subsampling controls (open water samples) in order to detect laboratory contamination. Due to deformation near the top of each core segment, the samples form a non-continuous record and a second core was taken parallel to the first core at a 35 cm offset to account for the deformations (Svendsen *et al.* 2019) but was not sampled for this study. Age determination was based on 26 AMS radiocarbon dates from plant macrofossils provided by the Poznań Radiocarbon Laboratory. Dates were calibrated using INTCAL13 (Reimer *et al.* 2013) and the online Calib program (Stuiver *et al.* 2018). A full chronology and sedimentology of this core is described by Svendsen *et al.* (2019).

Varanger Peninsula lake sediment

The Uhca Rohči lake was cored in February 2016 with a modified Nesje piston-corer (Nesje 1992), using a 4-m-long and 10-cm-diameter ABS polymer pipe. A 2.5-m-core was retrieved and cut in the field into 1-m sections, which were sealed to reduce the risk of contaminating the sediments. The core sections were kept at above 0 °C conditions in the field and during transport to avoid freezing of the sediments and were stored in a 4 °C cold room at the Tromsø University Museum (TMU). Sampling of the core took place in a dedicated ancient DNA laboratory. The age of the core was determined based on seven AMS radiocarbon

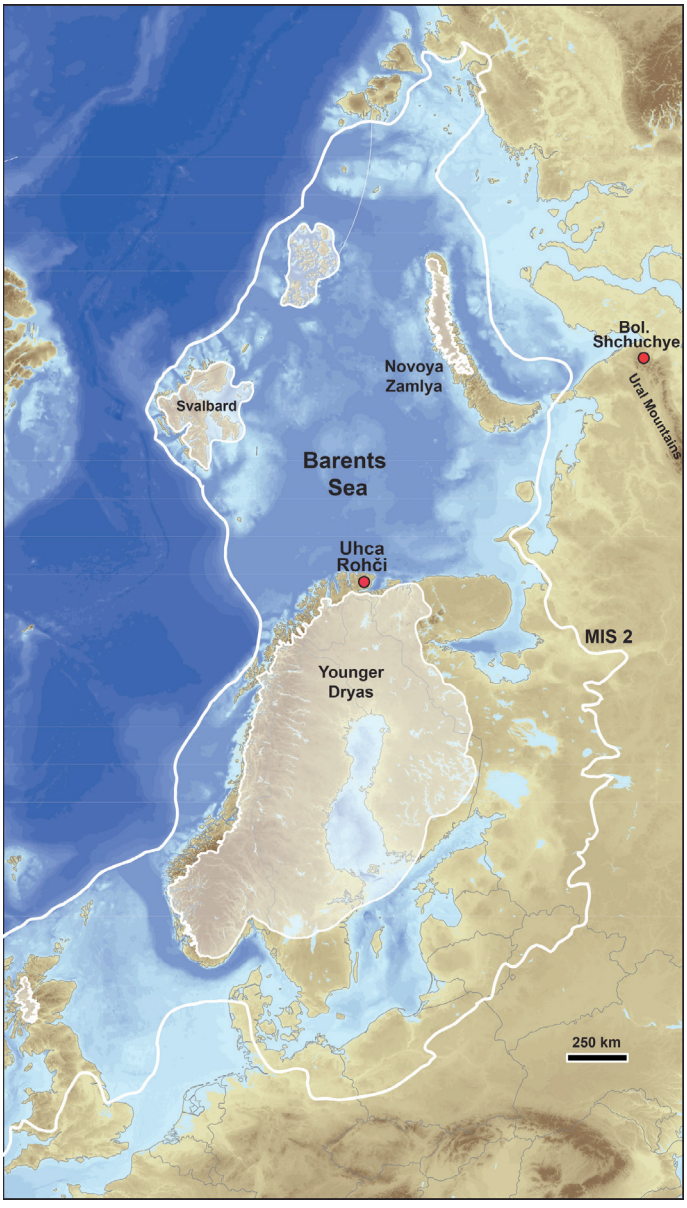


Fig. 1. The location of Lake Bolshoye Shchuchye in the Polar Urals of Arctic Russia and Lake Uhca Rohči on the Varanger Peninsula, northeast Finnmark, Norway. The outer white line represents the extension of the Eurasian ice sheet during Marine Isotope Stage 2 (20 000–15 000 a BP). The inner white shaded area represents the ice sheet during the Younger Dryas (12 800–11 400 a BP). [Colour figure can be viewed at www.boreas.dk]

dates on terrestrial plant macrofossils provided by the Poznań Radiocarbon Laboratory. Dates were calibrated using the terrestrial INTCAL13 curve (Reimer *et al.* 2013), and the age model was constructed using the Bayesian framework calibration software 'Bacon' (v2.2; Blaauw & Christen 2011), which was implemented in R (v3.2.4; R Core Team 2017). A full sedimentology and chronology of this core is described by Clarke *et al.* (2019).

DNA extraction

For the Polar Urals site, 153 lake sediment samples, 17 extraction controls and three subsampling controls underwent DNA extraction. DNA from the Varanger site was extracted from 77 sediment samples and nine extraction controls. All extractions were done at the Tromsø University Museum ancient DNA lab, using PowerMax soil DNA isolation kit (MOBIO Laboratories, Carlsbad, CA, USA), following the manufacturer's protocol with minor modifications by Alsos *et al.* (2016).

PCR amplification and sequencing

PCR reactions were carried out in a dedicated PCR room for ancient DNA at the Laboratoire d'Ecologie Alpine (Université Grenoble Alpes, France), using the MamP007F and MamP007R primers that target a ~70-bp-long part of the mammalian mitochondrial 16S rDNA (Giguet-Covex *et al.* 2014). Both forward and reverse primers had the same unique 8-bp tag on the 5′ end to allow sample multiplexing (Binladen *et al.* 2007; Valentini *et al.* 2009). In addition to the forward and reverse primers, the human blocking primer MamP007_B_Hum1 was added to suppress the amplification of human material (Giguet-Covex *et al.* 2014). The PCR reactions for each lake were carried out at different times to avoid cross-contamination of material. The Polar Urals samples included additional nine PCR negatives (excluding template DNA) and four PCR positives (including the marsupial *Didelphis marsupialis*, not found in Europe). The Varanger samples included six PCR negatives. For each sample, eight PCR repeats were carried out following a previously described PCR protocol (Giguet-Covex *et al.* 2014). PCR products were cleaned and pooled following the methods described by Alsos *et al.* (2016). Libraries (four for the Polar Urals core and two for the Varanger core) were prepared using the PCR free 'MetaFAST' library preparation protocol at Fasteris SA, Switzerland, and sequenced on an Illumina HiSeq 2500 at 2 × 125 bp paired-end sequencing.

DNA sequence analysis

The sequence data were analysed with the OBITOOLS software package (Boyer *et al.* 2016), using default settings unless otherwise specified. Paired-end data were merged with the *illuminapairedend* function and alignments with a score lower than 40 were removed. Data were demultiplexed with *ngsfilter* based on the known PCR tags. Identical sequences were merged with *obiuniq* and singleton sequences and those shorter than 10 bp were removed. Sequences were corrected for PCR and sequencing errors with *obiclean* with a 'head' to 'internal' ratio of 0.05. The remaining sequences were identified by comparing them to the EMBL nucleotide database (r133) with *ecotag*.

The identified sequences were further filtered in R (v3.4.2; R Core Team 2017) with a custom R script. Sequence occurrences that had fewer than 10 reads for a repeat were removed, to account for low-level sequence errors that survived the *obiclean* step and tag switching (Schnell *et al.* 2015). Only sequences that had a 100% match to reference data were kept. Furthermore, sequences had to be present in at least one sediment sample with two or more repeats. If that condition was met, single occurrences for other sediment samples were kept in. Finally, a sequence could only be present in the control samples with at most one repeat; if a sequence was found in a control sample with two or more repeats it was removed from the total data set. Common laboratory contaminants, such as human, *Homo sapiens*, pig, *Sus scrofa*, and chicken, *Gallus gallus* (Leonard *et al.* 2007), were manually removed from the list of sequences that survived filtering.

In silico primer analysis

The *ecoPCR* program (Ficetola *et al.* 2010) was used to calculate the mismatches between clitellate (=oligochaete) worms and the MamP007F – MamP007R primers. The target taxonomic group was set to NCBI TAXID 6381 (referred to as subclass Oligochaeta), the maximum number of mismatches in the primer to five, the amplicon size range to 10–100 bp and the EMBL r133 nucleotide release as database. For each clitellate family and species with available data in the EMBL release, the following were calculated: mean length of the amplicon, mean number of mismatches in each primer and the presence of mismatches in the last three bases of the primer 3′ end, which can hinder amplification (Kwok *et al.* 1990; Wu *et al.* 2009).

The same procedure was repeated for the following families that could be observed in the metabarcoding results: Cervidae (TAXID 9850), Hominidae (TAXID 9604), Phasianidae (TAXID 9005), Suidae (TAXID 9821) and Cercopagididae (TAXID 77756), with the exception that an amplicon size range of 25–150 bp was used to account for the longer expected fragment length.

Results

Polar Urals samples

A total of 80 983 160 raw reads was obtained for the four Polar Urals sequence libraries, which could be assigned to 68 521 unique sequences. Post-identification filtering reduced the number of sequences to 17, representing 1 123 241 reads. The sequences belonged to reindeer (*Rangifer tarandus* – two occurrences in the core at 23 000 and 14 000 cal. a BP, with a total of 27 133 reads) and eight clitellate taxa: two Enchytraeidae (*Enchytraeus norvegicus*, *Henlea perpusilla*), one Glossoscolecidae (*Pontoscolex corethrurus*) and five Lumbricidae (*Aporrectodea rosea*, *Dendrobaena octaedra*, *Bimastos norvegicus*, *Octolasion cyaneum* and *Octolasion tyrtaeum*) (Fig. 2, Table S2). The results also included seven

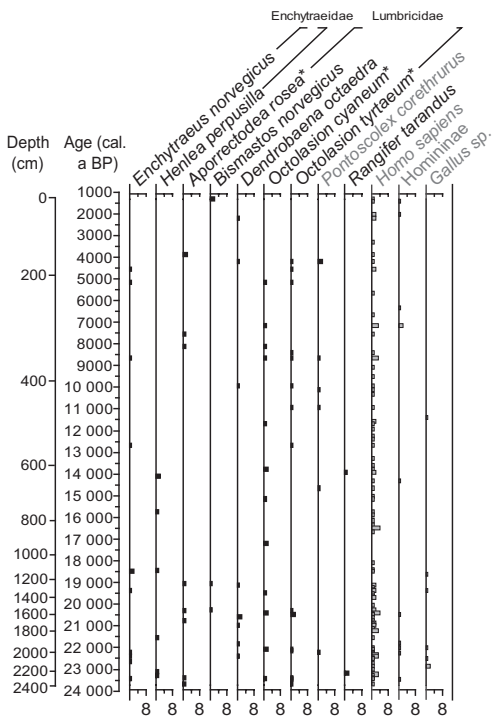


Fig. 2. Metabarcoding results for the Polar Urals core. The width of the bars indicates the number of PCR repeats. The grey taxa were assumed to be laboratory contaminants and were manually removed from the results. Non-cold tolerant taxa or taxa not found in the Polar Urals are indicated with an asterisk. The result for *Homo sapiens* is a combination of seven different *H. sapiens* sequences for which the maximum number of repeats is plotted.

Hominidae sequences (six assigned to *Homo sapiens* and one to Hominidae) and one *Gallus* sequence that survived the filtering criteria and were manually removed.

Species that did not survive filtering include steppe bison (*Bison priscus*, 100% match), Arctic lemming (*Dicrostonyx torquatus,* 98% match), rock ptarmigan (*Lagopus muta*, 99% match) and mountain hare (*Lepus timidus*, 98% match); these species are expected in the region, but none of them occurred in more than one sample and one repeat and thus did not survive our filtering criteria.

Varanger Peninsula samples

We obtained 52 562 858 raw reads for the two Varanger Peninsula (Uhca Rohči) libraries that represented 22 461 unique sequences. After R filtering, 18 sequences remained, representing 877 555 reads, which belong to: *Rangifer tarandus* (four occurrences at 10 800 cal. a BP and three between 3300 and 3600 cal. a BP, with 44 979 reads), the spiny water flea (cercopagidid cladoceran) *Bythotrephes longimanus* (six occurrences at 4900, 5600, 5700, 6300, 6500 and 9100 cal. a BP, sum 38 085 reads) and the oligochaete *Lumbriculus variegatus* (one occurrence at 10 800 cal. a BP with 227 reads) (Fig. 3, Table S3). A total of 12 *Homo sapiens*, one *Sus* and one *Gallus* sequences survived filtering and were manually removed.

Several worm taxa did not survive filtering, including *Dendrobaena octaedra*, *Tubifex tubifex* and a *Limnodrilus* sequence that could not be identified to species level. None of these taxa were detected in multiple repeats for a sample, but they are taxa that can be expected to occur in the Varanger area today.

In silico primer analysis

Primer matches between the mammal primer and annelid sequences could be calculated for 22 clitellate families and 1756 species (mean 175 sequences per family, SD = 317.7) out of the 28 families listed in the NCBI taxonomy database. The weighted average numbers of mismatches in the forward and reverse primer were 2.07 (SD = 0.05) and 2.04 (SD = 0.24), respectively, with an average estimated amplicon length of 35.7 bp (SD = 0.65; excluding primers).

The results for the clitellate families and the species that were detected in the metabarcoding results are displayed in Table 1, along with the mammalian and avian results. A full account of all clitellate families and species is provided in Table S1. The mismatch overview here is limited by the available clitellate data on EMBL, and some mismatch numbers might be over- or underestimated for some families depending on sampling and sequencing biases or depth.

Discussion

Mammal records

Rangifer tarandus was the only mammal in the Polar Urals and Varanger lake sediments that was detected in several PCR replicates (one Polar Urals sample with two repeats, Fig. 2, and three Varanger samples with two, three and four repeats, Fig. 3). *R. tarandus* was detected in a limited number of samples, furthermore, replicability was poor, with at most four out of eight PCR repeats. The limited presence is surprising, as *R. tarandus* has a circumpolar Eurasian distribution. It is known from western Norway at 13 500 cal. a BP from the village Blomvåg 30 km northwest of Bergen (Lie 1986; Mangerud *et al.* 2017) and it would be expected that *R. tarandus* was one of the first species immigrating north and west into Varanger after the ice receded after the LGM. Likewise, it is not surprising that *R. tarandus* was present in the Urals to the northeast of the Eurasian–Fennoscandian ice sheet during the Late Weichselian (24 000–15 000 a BP) as this area was probably its main glacial refugium based on genetic data (Flagstad & Roed 2003; Yannic *et al.* 2014; Kvie *et al.* 2016).

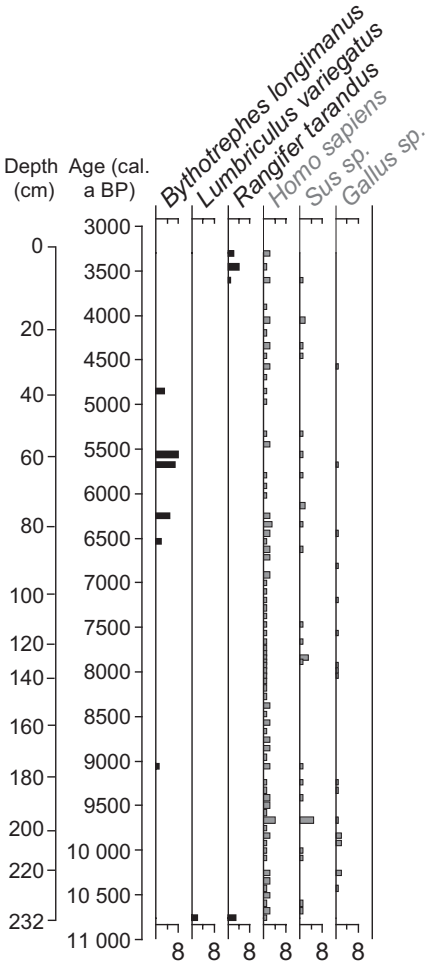


Fig. 3. Metabarcoding results for the Varanger core. The width of the bars indicates the number of PCR repeats. The grey taxa were assumed to be laboratory contaminants and were manually removed from the results. The result for *Homo sapiens* is a combination of 12 different *H. sapiens* sequences for which the maximum number of repeats is plotted.

The other mammals detected, *Bison priscus*, *Dicrostonyx torquatus* and *Lepus timidus*, are all likely for the sites in the period studied but were filtered out because they could only be observed in one sample and with one PCR repeat out of eight. There is always a trade-off between losing assumed true positives and keeping false negatives when setting a cut-off level for filtering (Ficetola *et al.* 2015). Lowering the cut-off level to include these taxa would increase our data set with many records that we suspected to be false positives. While probability statistics may be used to inform the likelihood of a record to represent a true positive, they require an independent record for calibration (Alsos *et al.* 2018). Thus, without records of bones, detection when there are low read numbers and few PCR repeats should be interpreted with caution. Furthermore, even if the filtered taxa were included, the limited occurrences in the records (only a single sample) suggest that the approach used here lacks the capability to reliably detect taxon presence, and thus is not appropriate for palaeoecological reconstructions.

The poor detection of mammals may either be explained by low DNA concentrations in extracts due to lack of template material, potentially caused by the low amounts of mammalian DNA deposited in the lakes, the age of the sediments or the size of the target amplicon. The amount of DNA deposited in the lake might be limited by accessibility for mammals, such as the steep slopes surrounding Lake Bolshoye Shchuchye. Alternatively, the plentiful water sources in the Komagdalen valley on Varanger Peninsula could have resulted in deposition of mammalian DNA over a large region, effectively diluting it in the process. Ancient DNA fragments found in lake sediments are of a relatively small length (Pedersen *et al.* 2015) and it is possible that the longer fragment required for the amplification of mammal material (*R. tarandus* requires a fragment of 111 bp, including primers) is too restricted in older sediments, especially considering the low biomass of mammals compared to other groups such as plants or invertebrates. Metabarcoding studies that successfully targeted ancient mammal DNA either worked with frozen material from localities affected by permafrost (Willerslev *et al.* 2003; Haile *et al.* 2009; Boessenkool *et al.* 2012) where conditions possibly preserved longer fragments (Pääbo *et al.* 2004; Willerslev *et al.* 2004), or with lake sediments from locations that had high mammalian concentrations, either due to migration routes (Pedersen *et al.* 2016), due to a waterhole (Graham *et al.* 2016) or due to human influence (Giguet-Covex *et al.* 2014). Thus, a combination of low mammal DNA concentration and long target fragment length may have caused the poor detection of mammals.

It is unlikely that failed DNA extractions are responsible for the poor mammal results, as the same DNA extracts were used for the metabarcoding of plants with the *g-h* universal plant primers (Taberlet *et al.* 2007) and produced successful results for both the Varanger (Clarke *et al.* 2019) and Polar Urals sites (C. L. Clarke, pers. comm. 2018). The success for plants could be explained by the obvious higher biomass and thus DNA contribution to the sediments and a potentially lower average fragment length required for amplification; for example the plant data from the Varanger site had an average length of 44.3 bp (±15.6) (Clarke *et al.* 2019) compared to the 73 bp of *Rangifer tarandus*.

The limited amount of mammal template material in the sediment extracts may have led to the amplification of laboratory contaminants and off-target species. *Homo sapiens* was by far the most dominant species in the filtered results for both the Polar Urals and Varanger

Table 1. The amplicon lengths and mismatches between the taxa and families that were detected in the metabarcoding results and the MamP007F – MamP007R mammal primers. * = this taxon is in reality a species complex.

Family	Species	Number of sequences	Average amplicon length (bp)	Forward primer		Reverse primer	
				Average mismatches	% 3′-end mismatches	Average mismatches	% 3′-end mismatches
Clitellate							
Acanthodrilidae		71	35.52 (±1.06)	2.06 (±0.29)	0	1.89 (±0.36)	0
Almidae		36	36.94 (±0.74)	2	0	2	0
Enchytraeidae		239	33.61 (±0.97)	2.05 (±0.25)	1.67	1.82 (±0.96)	1.67
	Enchytraeus norvegicus	1	34	2	0	1	0
	Henlea perpusilla	1	34	2	0	2	0
Eudrilidae		6	34.67 (±1.7)	2	0	2.67 (±0.94)	33.33
Glossoscolecidae		16	34.00 (±1.8)	2.38 (±0.48)	0	2.75 (±1.03)	12.5
	*Pontoscolex corethrurus**	7	33.86 (±1.64)	2	0	3.29 (±0.7)	14.29
Hormogastridae		585	35.38 (±1.34)	2.03 (±0.2)	0	1.98 (±0.19)	0.17
Lumbricidae		1037	35.74 (±1.22)	2.07 (±0.26)	2.41	1.97 (±0.2)	0.19
	*Aporrectodea rosea**	143	35.76 (±0.56)	2.23 (±0.42)	0.7	2.01 (±0.08)	0
	*Bimastos norvegicus**	9	37	2.22 (±0.63)	0	2	0
	*Dendrobaena octaedra**	15	35.73 (±0.77)	2	0	2	0
	Octolasion cyaneum	1	36	2	0	2	0
	*Octolasion tyrtaeum**	3	36.66 (±0.94)	2	0	2	0
Lumbriculidae		73	36.55 (±1.15)	2.03 (±0.16)	0	1.67 (±0.52)	0
	*Lumbriculus variegatus**	30	36.87 (±0.34)	2	0	2	0
Megascolecidae		763	35.86 (±1.21)	2.14 (±0.59)	0.66	1.86 (±0.4)	0.26
Moniligastridae		77	35.78 (±1.3)	2.13 (±0.41)	0	1.84 (±0.58)	2.6
Sparganophilidae		21	34.86 (±0.35)	2	0	2	0
Tubificidae		907	36.21 (±1.98)	2.06 (±0.29)	1.87	2.42 (±0.96)	20.84
Mammalia							
Cervidae		302	73.07 (±0.51)	0.03 (±0.27)	0.7	0.03 (±0.21)	0.7
	Rangifer tarandus	9	73	0	0	0	0
Hominidae		39 080	72.13 (±4.61)	0.07 (±0.58)	0.8	0.07 (±0.57)	0.9
	Homo sapiens	38 492	72.1 (±4.26)	0.06 (±0.53)	0.6	0.6 (±0.53)	0.8
Suidae		428	76.93 (±12.95)	0.66 (±1.65)	0.72	0.65 (±1.64)	10
	Sus sp.	400	76.99 (±13.4)	0.7 (±1.7)	0.78	0.69 (±1.68)	10.8
Aves							
Phasianidae		236	76.09 (±10.08)	2.23 (±0.81)	9.75	1.34 (±1.06)	5.5
	Gallus gallus	126	77.59 (±11.23)	2.33 (±0.98)	9.52	1.44 (±1.24)	8.7

samples before human DNA sequences were manually removed (767 186 out of 1 123 241 reads and 706 027 out of 877 555 reads for the Polar Urals and Varanger cores, respectively). Chicken, *Gallus gallus* (both sites), and pig, *Sus scrofa* (Varanger only), made up the remaining contaminants. Chicken could be amplified in both samples due to the limited differences between the mammalian primers used and the binding sites for chicken (Table 1). The amplification of *H. sapiens* was possible even in the presence of a human blocking primer, which is further indication that there was a limited amount of non-human template material available in the DNA extracts (Boessenkool *et al.* 2012).

The problems with the mammal primer presented here support the case for the exploration of alternative primers or methods for the detection of mammals in ancient sediments, especially where template material is probably low. Several metabarcoding primer sets have been suggested for mammals, with the shortest sets amplifying a mitochondrial 16S fragment of 68–71 bp (Rasmussen *et al.* 2009) or 60–84 bp (Giguet-Covex *et al.* 2014), both of which might be too long for reliable amplification of low concentration mammal material in ancient lake sediments. Alternative primer sets might yield better results if they target a shorter fragment or do not amplify common laboratory contaminants by targeting a narrower taxonomic group. Other alternatives are to bypass the usage of primers altogether by either shotgun sequencing sediment extracts (Pedersen *et al.* 2016; Seersholm *et al.* 2016) or by using DNA target capture methods (Slon *et al.* 2017).

Presence of worms

Off-target amplification of earthworms and other clitellates was observed in both the Polar Urals and the Varanger samples. Such amplification can be expected when there is limited target template available in the DNA extracts (Sipos *et al.* 2007; Schloss *et al.* 2011; Brown *et al.* 2015). The *in silico* amplification of clitellates with MamP007F and MamP007R primers revealed that 17 families and 849 species have a low number of mismatches (two or fewer outside the primer 3′ end) and that these

could potentially be amplified if there is limited competing mammal template available.

Metabarcoding potential of the mitochondrial 16S region targeted in this study has previously been demonstrated for earthworms with specific primers (Bienert *et al.* 2012). A comparison between the mammalian primers used in this study and the earthworm primers developed by Bienert *et al.* (2012) is given in Table 2. The forward primers are highly similar, with only a two base pair difference to account for the mismatches between the mammalian and earthworm primer binding sites; the reverse primer is shifted by four bases, but is otherwise comparable, once more indicating that the used mammal primers can amplify worms.

Another factor is the potential amount of DNA present in the sediment for various groups of organisms. Enchytraeidae biomass in Svalbard is estimated to be 1160 kg km^{-2} (Byzova *et al.* 1995) and Lumbricidae biomass in the northern Ural mountains is calculated to be 24 000 kg km^{-2} (Ermakov & Golovanova 2010). Thus, the clitellate numbers are far higher than common herbivorous mammals such as the North American brown lemming (*Lemmus trimucronatus*) at 30 kg km^{-2} in the Canadian Arctic (Fauteux *et al.* 2015) or *R. tarandus* in central Norway at 165 kg km^{-2} (Finstad & Prichard 2000; Vistnes *et al.* 2001). These rough biomass numbers give an indication that worms can produce vastly more DNA than the relatively sparse mammals, meaning that the clitellate DNA has a higher chance of being captured in the sediments. The difference in DNA production and contribution to the sediments, along with the additional problems of mammalian DNA described above, make worms more likely to be detected via metabarcoding.

Additionally, the clitellate amplicon length is considerably shorter than that of the mammalian taxa. The amplicon (excluding primer binding sites) for the mammals detected in the Polar Urals and Varanger cores is 74 bp on average (Table 1) and the average amplicon length for all clitellate families is 35 bp (Table S1). The shorter clitellate amplicon length increases the potential amount of template material in highly fragmented *sedaDNA* compared to the longer, and thus rarer mammalian target material. The downside of a shorter amplicon is the potential loss of taxonomic resolution, a problem that is difficult to estimate given the limited reference material available for clitellates.

Four worm species that are reported to be cold tolerant were recorded in the Polar Urals samples, and these could be expected to survive in the region. The enchytraeid *Henlea perpusilla* (six samples, one sample with two repeats) is found throughout Europe and is capable of surviving in the Arctic (Birkemoe *et al.* 2000). *Enchytraeus norvegicus* (10 samples, one sample with two repeats) is also known to have a broad range, extending from sea level in the Mediterranean (Rota *et al.* 2014) to colder temperate zones (Rota 1995) and at high (>1400 m a.s.l.) elevations in southern Norway (C. Erséus, unpublished data). The cosmopolitan lumbricid *Dendrobaena octaedra* (one sample with two repeats) has frost-tolerant populations in Finland, Greenland and Magadan Oblast, eastern Russia (Rasmussen & Holmstrup 2002). *Bimastos norvegicus* (three samples, one sample with two repeats) is part of the taxonomically difficult *Bimastos rubidus* (syn. *Dendrodrilus rubidus*) species complex, which is abundant in Scandinavia and European Russia, and is reported as freeze resistant. However, the known distribution today does not extend to the Ural region (Berman *et al.* 2010).

The remaining three lumbricid earthworms are less likely to be present in the northern Polar Urals, although they all show wide altitudinal ranges at lower latitudes. *Octolasion cyaneum* (12 samples, one sample with two repeats) is native to central and western Europe, and current records extend up to southern Finland and northern Sweden (Terhivuo & Saura 2006). In Norway it can be found to elevations of around 1000 m a.s.l. in the south, and in lowland localities north of the Arctic Circle (C. Erséus, unpublished data), but it is most often associated with human habitats. *Octolasion tyrtaeum* (also referred to as *Octolasion lacteum* (Shekhovtsov *et al.* 2014); 16 samples, one sample with two repeats) is a species complex with two cryptic lineages (Heethoff *et al.* 2004); it occurs in Europe, with populations extending to central Finland (Terhivuo & Saura 2006) and the taiga forests of European Russia (Perel 1979). *Aporrectodea rosea* (eight samples, one sample with two repeats) is also a species complex with a range that extends northwards from the Mediterranean towards central Finland (Terhivuo & Saura 2006) and the Middle Urals (Perel 1979). Tiunov *et al.* (2006) associate its occurrences in the northern part of the European Russian plain with cultivated soil (e.g. vegetable gardens), secondary deciduous forests and river valleys. The species found in the Urals is the one referred to as *A. rosea* L1 in the BOLD database, and this also occurs north of the Arctic Circle in Norway (C. Erséus, unpublished data). Although none of these lumbricids is recorded in the Polar Urals today, it is not unlikely that they were there during the Holocene Hypsithermal or other warmer periods.

Table 2. Comparison of the mammalian MamP007F – MamP007R primers with the ewB – ewC earthworm primers developed by Bienert *et al.* (2012).

	Forward	Reverse
Mammalian	5′–CGAGAAGACCCTATGGAGCT–3′	5′–CCGAGGTCRCCCCAACC–3′
Earthworm	5′–CAAGAAGACCCTATAGAGCTT–3′	5′–GGTCGCCCCAACCGAAT–3′

The glossoscolecid earthworm *Pontoscolex corethrurus* (six samples, one with two repeats) is a species complex with a circum-tropical distribution, native to South and Central America, but introduced in tropical and subtropical regions worldwide (Taheri *et al.* 2018). The family has no relatives in temperate environments and the genetic distance to other clitellate families rules out misidentification due to amplification or sequencing errors. The closest annelid sequence in GenBank belongs to the Asian *Amynthas glabrus* (Megascolecidae) recorded in China (Sun *et al.* 2017) and Japan (Blakemore 2003), at 77% sequence identity and with an edit distance of 8. The closest species in the results presented here is *Octolasion tyrtaeum* (Lumbricidae) at 63% sequence identity and an edit distance of 15. The most likely explanation for the detection of *Pontoscolex corethrurus* is contamination in the laboratory, possibly due to the reagents used.

The expected clitellate diversity in the Polar Urals is high based on previous lake diversity assessments. Baturina *et al.* (2014) recorded 30 aquatic species in the region. Unfortunately, little is known about the terrestrial clitellate diversity in the Polar Urals, making it difficult to assess how much of the diversity is captured in this study and what potential improvements can be made.

Only one annelid sequence was recorded at the Varanger site, representing a species in the *Lumbriculus variegatus* (one sample with two repeats) species complex. This complex has a current cosmopolitan distribution, but the particular lineage found on Varanger is an unidentified, probably undescribed, species; in its short (36 bp) 16S barcode it is 100% identical to the form of *L. variegatus* referred to as clade III by Gustafsson *et al.* (2009). Elsewhere, it has been recorded from Greenland, high-elevation sites (1000–1400 m a.s.l.) on the Scandinavian Peninsula (S. Martinsson & C. Erséus, unpublished data) and a lake at >3000 m a.s.l. in California (Gustafsson *et al.* 2009). This suggests that the complex is at least partially cold-adapted and could occur in northern Norway. In addition to the *Lumbriculus* species, the cercopagidid cladoceran *Bythotrephes longimanus* (six samples, five with two or more repeats) was detected, a species that is native to northern Europe and previously recorded on the Varanger Peninsula (Hessen *et al.* 2011).

Previous metabarcoding efforts of modern sediments on the Varanger Peninsula with enchytraeid specific primers targeting the mitochondrial 12S region resulted in identifications of *Cognettia sphagnetorum* and *Mesenchytraeus armatus* (Epp *et al.* 2012). Neither of these species could be detected in the results presented here. The discrepancy can be explained by the different primers used (the enchytraeid-specific primers can be expected to perform better than mammal primers used in this study), the age of the sediments (modern sediments compared to 3304–10 759 cal. a BP sediments), the type of sediment and how it retains DNA (heath and meadow plots compared to lake sediments) and local variation in clitellate diversity. In-lake sampling of northern Norwegian lakes (C. Erséus & M. Klinth, unpublished data) indicates a high clitellate diversity (20–30 species). Both the previous metabarcoding study and in-lake sampling indicate that the results obtained here are an underestimation of the true diversity.

Although not reported before, after re-analysing the data presented by Giguet-Covex *et al.* (2014) we noted that clitellate sequences were also recovered. However, the eight species that could be identified (*Aporrectodea caligninosa*, *Chamaedrilus sphagnetorum*, *C. glandulosus*, *Dendrodrilus rubidus*, *Eiseniella tetraedra*, *Henlea perpusilla*, *Lumbricus meliboeus* and *Tubifex tubifex*) neither survived the filtering criteria applied by the authors (amplicon length shorter than 50 bp or identified as non-mammalian) nor the criteria used in this study (each taxon was only detected in a single repeat). Their annelid results are probably worse than the results presented in this study, due to the overall higher quality and success rate for mammalian DNA, but confirm that the annelid bycatch in this study is not a fluke.

The overall scattered detections of worm sequences in the Polar Urals and Varanger samples are most likely due to the non 'worm-specific' primers used, hindering, but not completely preventing, the amplification of the material. Furthermore, the detection of the four unexpected earthworm species warrants an explanation. These species might represent true positives, which have not been recorded in the region and represent past distributions during warmer periods. Alternatively, they could be artefacts of limited DNA reference material and might be misidentified to the wrong species or a consequence of amplification or sequencing errors. Finally, the observed worm sequences could be the results of contamination, either in the field or during sampling, extracting and amplification of the DNA. The laboratory standards used along with the negative controls give some confidence that these results are true detections, but contamination cannot be fully ruled out and is a likely explanation for the tropical *Pontoscolex corethrurus*.

Palaeoenvironmental implications of the worm detections

The sediments of Bolshoye Shchuchye (Polar Urals) are low in organic matter (1–5% LOI, see Fig. 4, based on Svendsen *et al.* 2019) and are essentially silt and clay. Given the thermal sensitivity of worms and the long record at this site (0–24 000 cal. a BP), we might expect a temporal pattern in the worm occurrence. At the species level this is not the case with the two Enchytraeidae (*Enchytraeus norvegicus* and *Henlea perpusilla*) occurring in both warm periods, such as the Holocene, and cold periods, including Heinrich Stadial 2 (24 000–22 000 cal. a BP). This is also true for the lumbricid earthworms (*Aporrectodea rosea*, *Bimastos norvegicus*, *Dendrobaena octaedra*, *Octolasion cyaneum* and *O. tyrtaeum*), which occur in the Holocene and the Lateglacial. When aggregated, the DNA shows distinctly greater and more continuous values for the Lumbricidae in the Holocene but no trend in the Enchy-

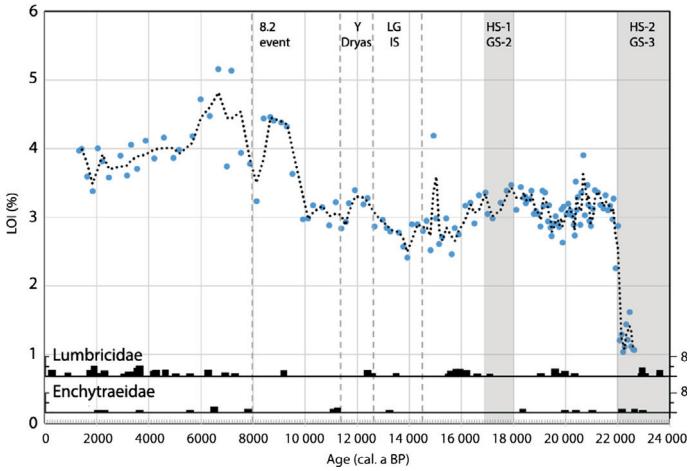


Fig. 4. Aggregated worm data from Lake Bolshoye Shchuchye (Polar Urals) with the core LOI and a two-period moving average (dotted line). Y Dryas is the Younger Dryas and LG IS is the Lateglacial interstadial. HS and GS are the Heinrich Stadial events and the Greenland Stadials from Rasmussen *et al.* (2014), respectively. [Colour figure can be viewed at www.boreas.dk]

traeidae (Fig. 4). These records suggest that both the soils and the lake sediments remained biologically active over the last 24 000 years, and that soils were almost certainly not set to zero biologically during the LGM or the Lateglacial stadials, when cold, dry conditions prevailed and the vegetation was predominantly tundra-steppe (Svendsen *et al.* 2014). However, the results do suggest higher rates of worm activity, and thus more soil formation, during the Holocene than during the Late Weichselian.

Potential for annelids in ancient DNA

Although the results for the worms detected in this study are not optimal, due to mismatched primers, overall poor metabarcoding results, low number of 'bycatch' taxa and perhaps some contamination of the samples, they indicate that earthworms and other clitellates can be identified in ancient sediments up to 24 000 years old. The use of a more optimized primer targeting short barcode regions in annelids, as has been done for the mitochondrial 16S and 12S regions (Bienert *et al.* 2012; Epp *et al.* 2012; Pansu *et al.* 2015), should increase clitellate diversity and detection reliability. Once detection methods have been optimized, tracking clitellate communities through time in ancient sediments may yield valuable information and proxies for various environmental conditions, such as temperature, soil moisture and acidity (Edwards & Lofty 1977; Beylich & Graefe 2009).

Conclusions

The results presented in this study show that the detection of mammalian material in ancient lake sediments in the Sub-Arctic via 16S metabarcoding is possible, but not without problems. The highly fragmented nature of *sedaDNA* means that amplification of long fragments of low biomass taxa is problematic and might benefit from alternative identification methods. By contrast, clitellate worms look like a more promising group for metabarcoding in older late-Quaternary sediments. Although a previous attempt to retrieve enchytraeid material from permafrost sediments failed (Epp *et al.* 2012), the combination of suitable primers for targeting short fragments, high biomass (for earthworms in particular) and DNA contribution to the sediments warrants further investigation in the group and the possible effects of age and sediment types on metabarcoding success.

Acknowledgements. – This work was supported by the Norwegian Research Council (grant: 255415, Climate History along the Arctic Seaboard of Eurasia (CHASE); grant: 213692, Ancient DNA of NW Europe reveals responses of climate change; grant: 250963, ECOGEN – Ecosystems change and species persistence over time: a genome-based approach). Y. Lammers was financed by an internal PhD position at the Tromsø University Museum. C. Erséus was supported by the Norwegian and Swedish Taxonomy Initiatives (Artsdatabanken and Artdatabanken) and the Adlerbert Foundation. We thank Francisco Javier Ancin-Murguzur and John-Arne Kristiansen for field assistance in Varanger, Dmitry Nazarov and Aleksei Matioushkov for field assistance in the Polar Urals, Marie Kristine Føreid Merkel for lab assistance, Frédéric Boyer for raw sequence data handling, Dorothee Ehrich and Peter Heintzman for fruitful discussions. Lastly, we thank Editor-in-Chief Jan A. Piotrowski, Guest Editor Martin Melles and two anonymous reviewers for their valuable and constructive feedback. Author contributions: YL analysed the data and wrote the first version of the manuscript; CLC, AGB and IGA carried out fieldwork on Varanger and JIS, HH and JM in the Polar Urals; CLC extracted the DNA; LG amplified the DNA and ran OBITOOLS; CE and ER provided taxonomic and ecological data; the project was devised by IGA / MEE and JIS / JM / HH. All authors contributed in various ways to the final version of the manuscript.

Data availability

The merged forward and reverse reads for both the Varanger and Polar Urals core, the used primer and tag sequences per sample, R script for the OBITOOLS filtering and the filtered OBITOOLS output are available on Dryad: https://doi.org/10.5061/dryad.g0f4hv0.

References

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L. & Edwards, M. E. 2018: Metabarcoding lake sediments: taphonomy and representation of contemporary vegetation in environmental DNA (eDNA) records. *PLoS ONE 13*, e0195403, https://doi.org/10.1371/journal.pone.0195403.

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

Barnes, M. A. & Turner, C. R. 2016: The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics 17*, 1–17.

Baturina, M., Timm, T. & Loskutova, O. 2014: Oligochaete (Annelida, Clitellata) communities in lakes of the Ural Mountains (Russia). *Zoosymposia 9*, 77–94.

Berman, D. I., Meshcheryakova, E. N. & Leirikh, A. N. 2010: Egg cocoons of the earthworm *Dendrodrilus rubidus tenuis* (Lumbricidae, Oligochaeta) withstand the temperature of liquid nitrogen. *Doklady Biological Sciences 434*, 347–350.

Beylich, A. & Graefe, U. 2009: Investigations of annelids at soil monitoring sites in Northern Germany: reference ranges and time-series data. *Soil Organisms 81*, 175–196.

Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J. J. & Taberlet, P. 2012: Tracking earthworm communities from soil DNA. *Molecular Ecology 21*, 2017–2030.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. 2007: The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE 2*, e197, https://doi.org/10.1371/journal.pone.0000197.

Birkemoe, T., Coulson, S. J. & Sømme, L. 2000: Life cycles and population dynamics of enchytraeids (Oligochaeta) from the High Arctic. *Canadian Journal of Zoology 78*, 2079–2086.

Blaauw, M. & Christen, J. A. 2011: Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis 6*, 457–474.

Blakemore, R. J. 2003: Japanese earthworms (Annelida: Oligochaeta): a review and checklist of species. *Organisms Diversity & Evolution 3*, 241–244.

Boessenkool, S., Epp, L. S., Haile, J., Bellemain, E., Edwards, M., Coissac, E., Willerslev, E. & Brochmann, C. 2012: Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology 21*, 1806–1815.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. 2016: OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources 16*, 176–182.

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., Oliver, A. K. & Jumpponen, A. 2015: Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecology 13*, 221–225.

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

Byzova, J. B., Uvarov, A. V. & Petrova, A. D. 1995: Seasonal changes in communities of soil invertebrates in tundra ecosystems of Hornsund, Spitsbergen. *Polish Polar Research 16*, 245–266.

Canti, M. G. 2003: Earthworm activity and archaeological stratigraphy: a review of products and processes. *Journal of Archaeological Science 30*, 135–148.

Clarke, C., Edwards, M. E., Brown, A. G., Gielly, L., Lammers, Y., Heintzman, P. D., Ancin-Murguzur, F. J., Bråthen, K. A., Goslar, T. & Alsos, I. G. 2019: Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (sedaDNA) and pollen. *Boreas 48*, 299–316.

Darwin, C. R. 1881: *The Formation of Vegetable Mould, Through the Action of Worms, with Observations on their Habits*. 326 pp. John Murray, London.

Domaizon, I., Winegardner, A., Capo, E., Gauthier, J. & Gregory-Eaves, I. 2017: DNA-based methods in paleolimnology: new opportunities for investigating long-term dynamics of lacustrine biodiversity. *Journal of Paleolimnology 58*, 1–21.

Dushin, V. A., Serdyukova, O. P., Malyugin, A. A., Nikulina, I. A., Kozmin, V. S., Burmako, P. L., Abaturova, I. V. & Kozmina, L. I. 2009: *State Geological Map of the Russian Federation 1:200000. Sheet Q-42-I*. VSEGEI, St. Petersburg.

Edwards, C. A. & Lofty, J. R. 1977: *Biology of Earthworms*. 333 pp. Chapman and Hall, London.

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

Ermakov, A. I. & Golovanova, E. V. 2010: Species composition and abundance of earthworms in the tundra biocenoses of Denezhkin Kamen' Mountain (Northern Urals). *Contemporary Problems of Ecology 3*, 10–14.

Fauteux, D., Gauthier, G. & Berteaux, D. 2015: Seasonal demography of a cyclic lemming population in the Canadian Arctic. *Journal of Animal Ecology 84*, 1412–1422.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P. & Pompanon, F. 2010: An in silico approach for the evaluation of DNA barcodes. *BMC Genomics 11*, 434, https://doi.org/10.1186/1471-2164-11-434.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G. & Taberlet, P. 2015: Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources 15*, 543–556.

Finstad, G. L. & Prichard, A. K. 2000: Growth and body weight of free-range reindeer in western Alaska. *Rangifer 20*, 221–227.

Fjellanger, J. & Sørbel, L. 2007: Origin of the palaeic landforms and glacial impact on the Varanger Peninsula, northern Norway. *Norwegian Journal of Geology/Norsk Geologisk Forening 87*, 223–238.

Flagstad, O. & Roed, K. H. 2003: Refugial origins of reindeer (*Rangifer tarandus* L.) inferred from mitochondrial DNA sequences. *Evolution 57*, 658–670.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J. & Taberlet, P. 2014: Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications 5*, 3211, https://doi.org/10.1038/ncomms4211.

Graham, R. W., Belmecheri, S., Choy, K., Culleton, B. J., Davies, L. J., Froese, D., Heintzman, P. D., Hritz, C., Kapp, J. D., Newsom, L. A., Rawcliffe, R., Saulnier-Talbot, É., Shapiro, B., Wang, Y., Williams, J. W. & Wooller, M. J. 2016: Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proceedings of the National Academy of Sciences of the United States of America 113*, 9310–9314.

Gustafsson, D. R., Price, D. A. & Erséus, C. 2009: Genetic variation in the popular lab worm *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation. *Molecular Phylogenetics and Evolution 51*, 182–189.

Haflidason, H., Lundekvam, J., Gyllencreutz, R., Svendsen, J. I., Gladysh, S. & Elizaveta, L. 2019: The Last Glacial and Holocene

seismostratigraphy and sediment distribution of the Lake Bolshoye Shchuchye, Polar Ural, Arctic Russia. *Boreas 48*, 452–469.

Haile, J., Froese, D. G., MacPhee, R. D. E., Roberts, R. G., Arnold, L. J., Reyes, A. V., Rasmussen, M., Nielsen, R., Brook, B. W., Robinson, S., Demuro, M., Gilbert, M. T. P., Munch, K., Austin, J. J., Cooper, A., Barnes, I., Moller, P. & 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*, 22352–22357.

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

Heethoff, M., Etzold, K. & Scheu, S. 2004: Mitochondrial COII sequences indicate that the parthenogenetic earthworm *Octolasion tyrtaeum* (Savigny 1826) constitutes of two lineages differing in body size and genotype. *Pedobiologia 48*, 9–13.

Hessen, D. O., Bakkestuen, V. & Walseng, B. 2011: The ecological niches of *Bythotrephes* and *Leptodora*: lessons for predicting long-term effects of invasion. *Biological Invasions 13*, 2561–2572.

Hughes, A. L. C., Gyllencreutz, R., Lohne, Ø. S., Mangerud, J. & Svendsen, J. I. 2016: The last Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1. *Boreas 45*, 1–45.

Karaca, A., Kizilkaya, R., Turgay, O. C. & Cetin, S. C. 2010: Effects of earthworms on the availability and removal of heavy metals in soil. *In* Sherameti, I. & Varma, A. (eds.): *Soil Heavy Metals*, 369–388. Springer, Berlin.

Kvie, K. S., Heggenes, J., Anderson, D. G., Kholodova, M. V., Sipko, T., Mizin, I. & Røed, K. H. 2016: Colonizing the High Arctic: mitochondrial DNA reveals common origin of Eurasian archipelagic reindeer (*Rangifer tarandus*). *PLoS ONE 11*, e0165237, https://doi.org/10.1371/journal.pone.0165237.

Kwok, S., Kellogg, D., McKinney, N. & Spasic, D. 1990: Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research 18*, 999–1005.

Leonard, J. A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R. K. & Fleischer, R. C. 2007: Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science 34*, 1361–1366.

Lie, R. 1986: Animal bones from the Late Weichselian in Norway. *Fauna Norwegica, Serie A 7*, 41–46.

Mangerud, J., Briner, J. P., Goslar, T. & Svendsen, J. I. 2017: The Bølling-age Blomvåg Beds, western Norway: implications for the Older Dryas glacial re-advance and the age of the deglaciation. *Boreas 46*, 162–184.

Nesje, A. 1992: A piston corer for lacustrine and marine sediments. *Arctic and Alpine Research 24*, 257–259.

Pääbo, S., Poinar, H., Serre, D., Svante, P., Jaenicke-despr, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. & Hofreiter, M. 2004: Genetic analyses from ancient DNA. *Annual Review of Genetics 38*, 645–679.

Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J.-M., Gielly, L., Cordonnier, T., Zinger, L., Brun, J.-J., Choler, P., Taberlet, P. & Cécillon, L. 2015: Landscape-scale distribution patterns of earthworms inferred from soil DNA. *Soil Biology and Biochemistry 83*, 100–105.

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*, 924–942.

Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., Bennett, K. D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M. E., Andersen, K., Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P., Houmark-Nielsen, M., Larsen, N. K., Orlando, L., Gilbert, M. T. P., Kjær, K. H., Alsos, I. G. & Willerslev, E. 2012: Glacial survival of boreal trees in northern Scandinavia. *Science 335*, 1083–1086.

Parducci, L., Väliranta, M., Salonen, J. S., Ronkainen, T., Matetovici, I., Fontana, S. L., Eskola, T., Sarala, P. & Suyama, Y. 2015: Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 370*, 20130382, https://doi.org/10.1098/rstb.2013.0382.

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

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., Spens, J., Thomsen, P. F., Bohmann, K., Cappellini, E., Schnell, I. B., Wales, N. A., Carøe, C., Campos, F., Schmidt, A. M. Z., Gilbert, M. T. P., Hansen, A. J., Orlando, L. & Willerslev, E. 2015: Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences 370*, 20130383, https://doi.org/10.1098/rstb.2013.0383.

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

Perel, T. S. 1979: *Range and Regularities in the Distribution of Earthworms of the USSR Fauna*. 268 pp. Nauka, Moscow.

Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T. O., Tucker, A. E., Sung, W. & Thomas, W. K. 2009: Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources 9*, 1439–1450.

R Core Team. 2017: *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

Rasmussen, L. M. & Holmstrup, M. 2002: Geographic variation of freeze-tolerance in the earthworm *Dendrobaena octaedra*. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology 172*, 691–698.

Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A. M., Vallelonga, P., Vinther, B. M., Walker, M. J. C., Wheatley, J. J. & Winstrup, M. 2014: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. *Quaternary Science Reviews 106*, 14–28.

Rasmussen, M., Cummings, L. S., Gilbert, M. T. P., Bryant, V., Smith, C., Jenkins, D. L. & Willerslev, E. 2009: Response to comment by Goldberg et al. on "DNA from Pre-Clovis Human Coprolites in Oregon, North America". *Science 325*, 148 LP-148, https://doi.org/10.1126/science.1167672.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M. & van der Plicht, J. 2013: IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon 55*, 1869–1887.

Rota, E. 1995: Italian Enchytraeidae (Oligochaeta). I. *Italian Journal of Zoology 62*, 183–231.

Rota, E., Caruso, T. & Bargagli, R. 2014: Community structure, diversity and spatial organization of enchytraeids in Mediterranean urban holm oak stands. *European Journal of Soil Biology 62*, 83–91.

Schloss, P. D., Gevers, D. & Westcott, S. L. 2011: Reducing the effects of PCR amplification and sequencing artifacts on 16s rRNA-based studies. *PLoS ONE 6*, e27310, https://doi.org/10.1371/journal.pone.0027310.

Schnell, I. B., Bohmann, K. & Gilbert, M. T. P. 2015: Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources 15*, 1289–1303.

Seersholm, F. V., Pedersen, M. W., Søe, M. J., Shokry, H., Mak, S. S. T., Ruter, A., Raghavan, M., Fitzhugh, W., Kjær, K. H., Willerslev, E., Meldgaard, M., Kapel, C. M. O. & Hansen, A. J. 2016: DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. *Nature Communications 7*, https://doi.org/10.1038/ncomms13389.

Shekhovtsov, S. V., Golovanova, E. V. & Peltek, S. E. 2014: Genetic diversity of the earthworm *Octolasion tyrtaeum* (Lumbricidae, Annelida). *Pedobiologia 57*, 245–250.

Siedlecka, A. & Roberts, D. 1992: *The Bedrock Geology of Varanger Peninsula, Finnmark, North Norway: An Excursion Guide*. 45 pp. Norges Geologiske Undersøkelse, Trondheim.

Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K. & Nikolausz, M. 2007: Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology 60*, 341–350.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., De La Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., De Lumley, H., Perrenoud, C., Gušić, I., Kućan, Ž., Rudan, P., Aximu-Petri, A., Essel, E., Nagel, S., Nickel, B., Schmidt, A., Prüfer, K., Kelso, J., Burbano, H. A., Pääbo, S. & Meyer, M. 2017: Neandertal and Denisovan DNA from Pleistocene sediments. *Science 356*, 605–608.

Solomina, O., Ivanov, M. & Bradwell, T. 2010: Lichenometric studies on moraines in the Polar Urals. *Geografiska Annaler: Series A, Physical Geography 92*, 81–99.

Stokes, C. R., Corner, G. D., Winsborrow, M. C. M., Husum, K. & Andreassen, K. 2014: Asynchronous response of marine-terminating outlet glaciers during deglaciation of the Fennoscandian Ice Sheet. *Geology 42*, 455–458.

Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B. W., Harbor, J. M., Jansen, J. D., Olsen, L., Caffee, M. W., Fink, D., Lundqvist, J., Rosqvist, G. C., Strömberg, B. & Jansson, K. N. 2016: Deglaciation of Fennoscandia. *Quaternary Science Reviews 147*, 91–121.

Stuiver, M., Reimer, P. J. & Reimer, R. W. 2018: *CALIB 7.1*. Available at: http://calib.org/.

Sun, J., James, S. W., Jiang, J., Yao, B., Zhang, L., Liu, M., Qiu, J. & Hu, F. 2017: Phylogenetic evaluation of *Amynthas* earthworms from South China reveals the initial ancestral state of spermathecae. *Molecular Phylogenetics and Evolution 115*, 106–114.

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M., Hubberten, H. W., Ingólfsson, Ó., Jakobsson, M., Kjær, K. H., Larsen, E., Lokrantz, H., Lunkka, J. P., Lyså, A., Mangerud, J., Matiouchkov, A., Murray, A., Möller, P., Niessen, F., Nikolskaya, O., Polyak, L., Saarnisto, M., Siegert, C., Siegert, M. J., Spielhagen, R. F. & Stein, R. 2004: Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews 23*, 1229–1271.

Svendsen, J. I., Færseth, L. M. B., Gyllencreutz, R., Haflidason, H., Henriksen, M., Hovland, M. N., Lohne, Ø. S., Mangerud, J., Nazarov, D., Regnéll, C. & Schaefer, J. M. 2019: Glacial and environmental changes over the last 60,000 years in the Polar Ural Mountains, Arctic Russia. *Boreas 48*, 407–431.

Svendsen, J. I., Krüger, L. C., Mangerud, J., Astakhov, V. I., Paus, A., Nazarov, D. & Murray, A. 2014: Glacial and vegetation history of the Polar Ural Mountains in northern Russia during the Last Ice Age, Marine Isotope Stages 5–2. *Quaternary Science Reviews 92*, 409–428.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. 2012: Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology 21*, 2045–2050.

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

Taheri, S., Pelosi, C. & Dupont, L. 2018: Harmful or useful? A case study of the exotic peregrine earthworm morphospecies *Pontoscolex corethrurus*. *Soil Biology and Biochemistry 116*, 277–289.

Terhivuo, J. & Saura, A. 2006: Dispersal and clonal diversity of North-European parthenogenetic earthworms. *In* Hendrit, P. F. (ed.): *Biological Invasions Belowground: Earthworms as Invasive Species*, 5–18. Springer, Dordrecht.

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

Tiunov, A. V., Hale, C. M., Holdsworth, A. R. & Vsevolodova-Perel, T. S. 2006: Invasion patterns of Lumbricidae into the previously earthworm-free areas of northeastern Europe and the western Great Lakes region of North America. *Biological Invasions 8*, 1223–1234.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., Swenson, J. E. & Taberlet, P. 2009: New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources 9*, 51–60.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., Peroux, T., Crivelli, A. J., Olivier, A., Acqueberge, M., Le Brun, M., Møller, P. R., Willerslev, E. & Dejean, T. 2015: Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology 25*, 929–942.

Vistnes, I., Nellemann, C., Jordhøy, P. & Strand, O. 2001: Wild reindeer: impacts of progressive infrastructure development on distribution and range use. *Polar Biology 24*, 531–537.

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

Willerslev, E., Hansen, A. J. & Poinar, H. N. 2004: Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution 19*, 141–147.

Wu, J. H., Hong, P. Y. & Liu, W. T. 2009: Quantitative effects of position and type of single mismatch on single base primer extension. *Journal of Microbiological Methods 77*, 267–275.

Yannic, G., Pellissier, L., Ortego, J., Lecomte, N., Couturier, S., Cuyler, C., Dussault, C., Hundertmark, K. J., Irvine, R. J., Jenkins, D. A., Kolpashikov, L., Mager, K., Musiani, M., Parker, K. L., Røed, K. H., Sipko, T., Þórisson, S. G., Weckworth, B. V., Guisan, A., Bernatchez, L. & Côté, S. D. 2014: Genetic diversity in caribou linked to past and future climate change. *Nature Climate Change 4*, 132–137.

Zimmermann, H., Raschke, E., Epp, L., Stoof-Leichsenring, K., 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*, 273, https://doi.org/10.3390/genes8100273.

Supporting Information

Additional Supporting Information may be found in the online version of this article at http://www.boreas.dk.

Table S1. Average mismatches between all clitellate families, species and the MamP007F – MamP007R mammal primers.

Table S2. Metabarcoding results for the Polar Urals core, including the repeats, read abundances and sequence information.

Table S3. Metabarcoding results for the Varanger core, including the repeats, read abundances and sequence information.

Paper IV

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 20 June 2019
doi: 10.3389/fevo.2019.00189

Shotgun Environmental DNA, Pollen, and Macrofossil Analysis of Lateglacial Lake Sediments From Southern Sweden

Laura Parducci[1], Inger Greve Alsos[2], Per Unneberg[3], Mikkel W. Pedersen[4], Lu Han[1,5], Youri Lammers[2], J. Sakari Salonen[6], Minna M. Väliranta[7], Tanja Slotte[8,9] and Barbara Wohlfarth[10]**

[1] Department of Ecology and Genetics, The Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden, [2] Tromsø Museum, University of Tromsø – The Arctic University of Norway, Tromsø, Norway, [3] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, [4] Department of Zoology, University of Cambridge, Cambridge, United Kingdom, [5] Ancient DNA Laboratory, College of Life Sciences, Jilin University, Changchun, China, [6] Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland, [7] Environmental Change Research Unit (ECRU), Ecosystems and Environment Research Programme, University of Helsinki, Helsinki, Finland, [8] Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden, [9] Science for Life Laboratory, Solna, Sweden, [10] Department of Geological Sciences and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

The lake sediments of Hässeldala Port in south-east Sweden provide an archive of local and regional environmental conditions ~14.5–9.5 ka BP (thousand years before present) and allow testing DNA sequencing techniques to reconstruct past vegetation changes. We combined shotgun sequencing with plant micro- and macrofossil analyses to investigate sediments dating to the Allerød (14.1–12.7 ka BP), Younger Dryas (12.7–11.7 ka BP), and Preboreal (<11.7 ka BP). Number of reads and taxa were not associated with sample age or organic content. This suggests that, beyond the initial rapid degradation, DNA is still present. The proportion of recovered plant DNA was low, but allowed identifying an important number of plant taxa, thus adding valid information on the composition of the local vegetation. Importantly, DNA provides a stronger signal of plant community changes than plant micro- and plant macrofossil analyses alone, since a larger number of new taxa were recorded in Younger Dryas samples. A comparison between the three proxies highlights differences and similarities and supports earlier findings that plants growing close to or within a lake are recorded by DNA. Plant macrofossil remains moreover show that tree birch was present close to the ancient lake since the Allerød; together with the DNA results, this indicates that boreal to subarctic climatic conditions also prevailed during the cold Younger Dryas interval. Increasing DNA reference libraries and enrichment strategies prior to sequencing are necessary to improve the potential and accuracy of plant identification using the shotgun metagenomic approach.

Keywords: environmental DNA, ancient DNA, shotgun sequencing (metagenomics), pollen, macrofossils remains, lake sediments

OPEN ACCESS

Edited by:
Gael Le Roux, UMR5245 Laboratoire Ecologie Fonctionnelle et Environnement (ECOLAB), France

Reviewed by:
Patrick Rioual, Institute of Geology and Geophysics (CAS), China
Charline Giguet-Covex, UMR5204 Environnements, DYnamiques et TErritoires de la Montagne (EDYTEM), France

****Correspondence:***
Laura Parducci
laura.parducci@ebc.uu.se
Barbara Wohlfarth
barbara.wohlfarth@geo.su.se

Specialty section:
This article was submitted to Paleoecology, a section of the journal Frontiers in Ecology and Evolution

Received: *04 December 2018*
Accepted: *08 May 2019*
Published: *20 June 2019*

Citation:
Parducci L, Alsos IG, Unneberg P, Pedersen MW, Han L, Lammers Y, Salonen JS, Väliranta MM, Slotte T and Wohlfarth B (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

INTRODUCTION

The final stages of the last glacial period were, in the North Atlantic region, characterized by distinct and alternating warmer/colder and wetter/drier climate states before interglacial temperatures were attained (Bjorck et al., 1996; Lowe et al., 2008; Steffensen et al., 2008; Schenk et al., 2018). These marked climatic shifts caused a series of environmental changes that are registered in various geological archives (Blockley et al., 2012; Rasmussen et al., 2014). The sediments of the ancient lake at Hässeldala Port in southern Sweden (**Figures 1A–D**) provide a detailed record for this time interval and suggest that local and regional responses to climate variability between 14.5 and 9.5 ka BP (thousand years before present) are recorded in multiple types of environmental proxy data (Wohlfarth et al., 2017 and references therein). Hässeldala Port's sediments thus offer an excellent testing ground for DNA analysis, as recently demonstrated in Ahmed et al. (2018), who investigated Archaeal composition changes in these sediments.

Previous pollen stratigraphic investigations, combined with a high-resolution chronology, identified several climatic shifts between warmer and colder time intervals at Hässeldala Port: the transition from a cold Older Dryas into a warmer Allerød at 14.1 ka BP, the transition from the Allerød into the cold Younger Dryas at 12.7 ka BP, and the transition between the Younger Dryas and the early Holocene (Preboreal) at 11.8–11.7 ka BP (Wohlfarth et al., 2017). The area around Hässeldala Port had become free of stagnant ice some time before 14.5 ka BP. Run off from unstable catchment soils was considerable during the early lake stage and continued during the early Allerød. These unstable soils likely prevented the initial establishment of denser local vegetation. With the start of the Holocene and due to gradual infilling, the lake changed into a peatbog with greatly altered soil conditions. This again had an impact on the type of plants on and around the site.

Over the last decade, environmental DNA (eDNA) studies of lake sediments have added a new dimension to traditional biological proxies (e.g., pollen and plant macrofossil analyses) often used to investigate and reconstruct past vegetation changes and palaeoenvironments (Parducci et al., 2017). We know that in sediments eDNA binds to mineral and organic components (extracellular DNA) and that it is also present within cells in plant and animal remains that become embedded in the sediments (for a review see Nagler et al., 2018). Indeed, PCR (polymerase chain reaction) amplification of short fragments of chloroplast DNA from sediments (metabarcoding) has been successfully used to demonstrate the presence of animals and plants in different palaeoenvironmental settings (e.g. Willerslev et al., 2003, 2014; Parducci et al., 2012; Epp et al., 2015; Alsos et al., 2016; Pedersen et al., 2016).

More recently, eDNA extracted from lake sediments has been directly converted to DNA libraries in order to "shotgun" sequence the entire metagenome present in the sample and to investigate the entire diversity of the taxonomic groups present (Pedersen et al., 2016). Providing that a good fraction of the analyzed taxa is represented in a reference library, metagenomics is a more powerful approach compared to metabarcoding for investigating biodiversity of ancient environments, as it allows all of the fragmented DNA molecules to be sequenced and determined, from microorganisms like bacteria and Archaea (Ahmed et al., 2018) to plants and animals (Pedersen et al., 2016), including humans (Slon et al., 2017). In addition, this approach permits statistical data analyses to detect specific substitutions (C to T and G to A) that are normally present at the ends of ancient DNA fragments, and which indicate their ancient origin (Briggs et al., 2007; Jónsson et al., 2013). Metagenomic data of lake sediment samples, in combination with pollen and plant macrofossil data, resulted in for example detailed palaeoenvironmental reconstructions of the ice-free corridor in North America (Pedersen et al., 2016) providing evidence for new postglacial colonization routes of humans.

Here, we employ low-coverage shotgun sequencing of eDNA and compare the full metagenome data set preserved in the Hässeldala Port record with plant micro- and macrofossil analyses conducted on the same sediment samples. We (1) examine whether sufficient and useful plant DNA information can be retrieved from the lake sediments using the shotgun sequencing technique, (2) discuss the differences between the genomic and micro/macrofossil data sets, (3) explore whether genomic eDNA add valid information on past local and regional vegetation, and (4) assess the new eDNA data set in relation to the marked climatic shifts of the last glacial/interglacial transition.

MATERIALS AND METHODS

Site Description

Hässeldala Port (56°16′N, 15°01′E; 63 m a.s.l.) in Blekinge province, southeast Sweden is today a peat bog underlain by a distinct Lateglacial lake sediment sequence (**Figures 1A–D**) (Wohlfarth et al., 2017). Multiple sediment cores have been obtained from this site over the years and studied using a variety of palaeo-environmental and palaeo-climatic proxies (lithostratigraphy, inorganic and organic geochemistry, pollen, diatoms, chironomids, biomarkers, hydrogen isotopes, charcoal) (Wohlfarth et al., 2017 and references therein). These allowed reconstructing temporal changes in lake status, evaporation, local and regional vegetation, and summer temperature, which occurred in response to large-scale hemispheric climatic shifts (Bjorck et al., 1996; Lowe et al., 2008; Steffensen et al., 2008). Hässeldala Port's multi-proxy data set shows that the small lake basin formed >14.5 ka BP and pollen assemblages indicate that the early catchment vegetation was dominated by herbs, shrubs and dwarf-shrubs. This type of vegetation also persisted during the Allerød (14.1–12.7 ka BP) and Younger Dryas (12.7–11.7 ka BP) pollen zones. However, *Betula pubescens* (tree birch) remains have been found at other sites in Blekinge demonstrating that tree birch has been present regionally before the Holocene (Berglund, 1966; Wohlfarth et al., 2017). With the start of the Holocene, *B. pubescens* and later also *Pinus sylvestris* (Scots pine) plant macro remains appear in Hässeldala Port's sediments, testifying to the presence of these trees in the immediate surroundings of the ancient lake. Around 11.8 ka BP, the shallow lake started to

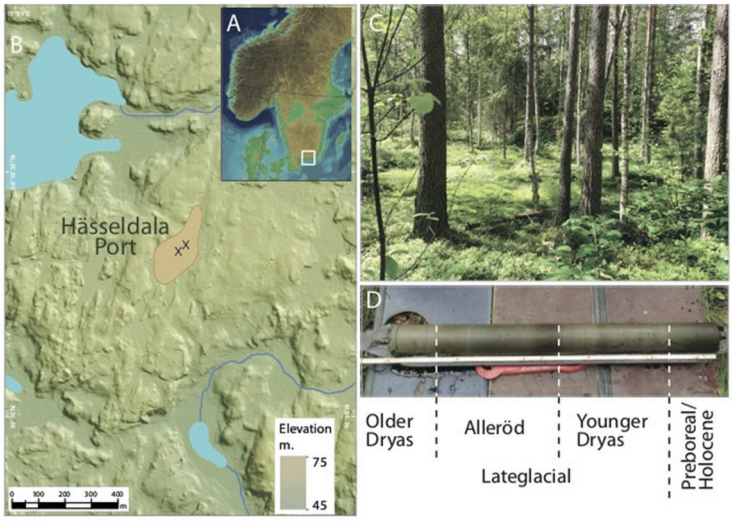


FIGURE 1 | (A) Location of Hässeldala Port in southernmost Sweden. **(B)** Close-up of the topography around Hässeldala Port. X marks the coring locations; blue = lakes; brown = peatbogs. **(C)** View over the peatbog at Hässeldala Port. **(D)** Hässeldala Port's typical Lateglacial/early Holocene sediment record showing the sediments assigned to the Older Dryas (>14.1 ka BP), Allerød (14.1–12.7 ka BP), Younger Dryas (12.7–11.7 ka BP), and Preboreal (<11.7 ka BP) regional pollen zones.

transform into minerotrophic mire and a few hundred years later turned into an ombrotrophic bog (Wohlfarth et al., 2017).

Coring and Sub-sampling

Two new sediment cores (#7.4 and #8) were obtained at a distance of ca. 5 m in June 2015 at Hässeldala Port using a Russian corer with a chamber length of 1 m and a diameter of 10 cm (**Figures 1B–D**). The two cores were taken from the deepest part of the basin in the south-western corner of the bog where several parallel cores (#1-#6) had previously been obtained and analyzed (see Wohlfarth et al., 2017 for a description of all previous studies on these cores). Core #7.4 extends between 340 and 440 cm depth and core #8 between 290 and 430 cm depth (see **Table S1** for details). The sediments of cores #7.4 and #8 were described and sub-sampled in August 2015, first for DNA and subsequently for loss-on-ignition (LOI), plant micro and macrofossil analyses.

Samples for DNA Analyses

To minimize the risks for contamination, we took a number of precautions during coring, sampling and DNA analysis. Immediately after collection, the cores were wrapped in plastic and as quickly as possible transported to the cold room at the Department of Geological Sciences (IGV) at Stockholm University where they were stored at 5°C until sub-sampling. The cold storage is in a part of the building where no DNA analyses are being performed. Sub-sampling was conducted in September 2015 in a clean laboratory at IGV using gloves and wearing lab coats and masks. Eight samples and corresponding replicates (16 samples) were taken from each core (total of 32 samples) as representative of the warm (Allerød, Preboreal) and cold (Younger Dryas) time intervals previously described for Hässeldala Port (Wohlfarth et al., 2017). For each sample, we removed the top 1.5 cm of the outer sediment with sterile scalpels, sampled uncontaminated material from the innermost part of the core with a new set of sterile scalpels and placed the sample in sterile tubes that were immediately closed. This procedure was repeated two times for each sample. For an overview of the 16 samples collected and analyzed in cores #7.4 and #8 see **Figure 2** and **Table 1**. The 32 samples were frozen within 2 h after sub-sampling and shipped frozen to the Center for GeoGenetics, University of Copenhagen for molecular analyses. Here, DNA extractions and library constructions were performed in ancient DNA facilities specifically dedicated to eDNA analyses following established criteria and rules during all the steps.

DNA Extraction, Libraries Construction, and High-Throughput Sequencing

One gram of wet sediment was added to 3 ml of lysis buffer [68 mM N-lauroylsarcosine sodium salt, 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 20 mM EDTA (pH 8.0) with addition of 150 ul 2-mercaptoethanol and 0.3 mM DTT], together with 170 µg of proteinase K, and vortexed vigorously for 2 × 20 s using a FastPrep-24 at speed 4.0 ms^{-1}. An additional 170 µg of proteinase K was added to each sample and incubated, then gently rotated overnight at 37°C. The MOBIO C2 and C3 buffers (MOBIO Laboratories, Carlsbad, CA) were used for removal of inhibitors, following the manufacturer's protocol. The

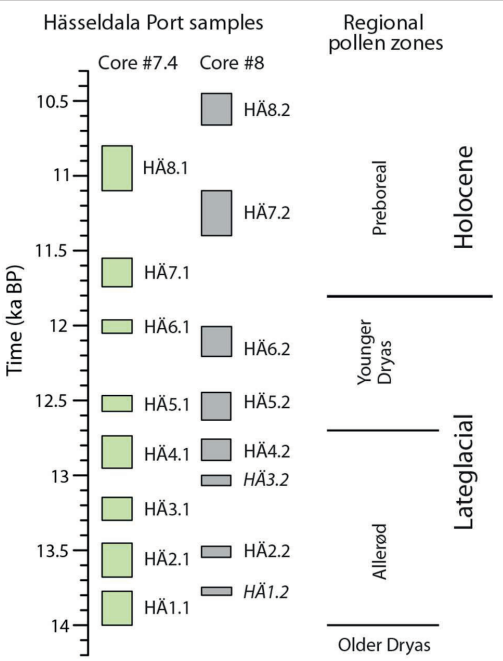


FIGURE 2 | The 16 samples analyzed for eDNA, pollen and plant macrofossil remains in cores #7.4 (green) and #8 (gray) from Hässeldala Port shown on an age scale and in relation to the regional pollen zones. Sample IDs shown in italics represent DNA samples of poor quality, which were excluded from DNA assignments.

TABLE 1 | The 16 samples analyzed for eDNA, pollen, and plant macrofossil remain in cores #7.4 and #8, their age interval, organic matter content, and their assignment to the regional pollen zones.

Core ID	Sample ID	Age interval (yr BP)	Organic matter content (%)	Regional pollen zone
#8	HÄ 8.2	10,440–10,680	76.6	Preboreal
#7.4	HÄ 8.1	10,800–11,110	55.7	Preboreal
#8	HÄ 7.2	11,110–11,400	67.5	Preboreal
#7.4	HÄ 7.1	11,540–11,750	25.7	Preboreal
#7.4	HÄ 6.1	11,930–12,020	18.8	Younger Dryas
#8	HÄ 6.2	12,000–12,200	20.3	Younger Dryas
#8	HÄ 5.2	12,430–12,610	18.4	Younger Dryas
#7.4	HÄ 5.1	12,400–12,590	17.8	Younger Dryas
#7.4	HÄ 4.1	12,740–12,940	29.2	Alleröd
#8	HÄ 4.2	12,790–12,900	27.8	Alleröd
#8	HÄ 3.2	13,000–13,050	17.4	Alleröd
#7.4	HÄ 3.1	13,150–13,300	23.1	Alleröd
#8	HÄ 2.2	13,460–13,540	16.8	Alleröd
#7.4	HÄ 2.1	13,440–13,660	18.0	Alleröd
#8	HÄ 1.2	13,730–13,800	9.0	Alleröd
#7.4	HÄ 1.1	13,770–14,000	8.6	Alleröd

extracts were further purified using phenol:chloroform and up-concentrated using 30 kDa Amicon Ultra-4 centrifugal filters and this step was repeated if more purification was needed. Lastly, all filters were washed twice with 500 ul Qiagen EB buffer and transferred to clean 1.5 ml Eppendorf tube. The 32 samples were extracted in four separate batches with two extraction controls each (40 samples in total). The extracted DNA was quantified using a Quant-iT dsDNA HS assay kit (Invitrogen) on the Qubit 2.0 Fluorometer according to the manufacturer's manual. Aliquots from the DNA extracts were subsequently converted into Illumina double stranded libraries using a NEBNext® DNA Library Prep Master Mix Set for 454 (New England BioLabs) following the manufacturer's protocol with the modifications described in Pedersen et al. (2016). Metagenome libraries were amplified using AmpliTaq Gold (Applied Biosystems) (14–20 cycles), purified using the MinElute PCR Purification kit (Qiagen), quantified on the 2100 BioAnalyzer and then pooled equimolarly. All pooled libraries were then sequenced using Illumina HiSeq 2500 platform (single-end reads). The eight extraction negative controls and the six library negative controls (one for each batch) were also prepared and sequenced as controls for contaminants (46 samples in total).

Bioinformatic Analyses

Sequence Quality Control

We sequenced a total of 1,340,975,627 DNA reads distributed across Hässeldala Port samples and controls. All reads were subjected to the following quality criteria. First, all reads were trimmed for the Illumina 3′ sequencing adapter (parameter setting -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC) using cutadapt v. 1.11 (Martin, 2011). In addition, we used the parameter settings *-m 30–trim-n -q 10,10* which were employed to discard reads >30 base pairs (bp), trim trailing N's, and to trim low-quality bases (phred score <= 10) from both ends of the reads. **Table S2** summarizes the total number of raw reads and the total number of reads passing quality control filters (48% for samples and 19% for controls). Second, we compared the quality of the raw and trimmed quality filtered reads using FASTQC (Andrews, 2010) version 0.11.5 and found that there were marked difference between the controls and the samples (**Figure S1**). Eventually, samples HÄ3.1_2, HÄ5.1_2, HÄ6.2_2, HÄ1.2, extraction blank E7 and library blank Blank4 failed to provide sufficient number of sequences. All library and extraction negative controls were markedly different from the "true" samples in terms of read lengths, phred scores and duplicate numbers (**Figure S2**). The bases sequenced for raw reads and for reads passing quality control filters (QC reads), showed two characteristic features of ancient DNA reads. First, the frequency of C to T substitutions is expected to be elevated at the 5′-end (Briggs et al., 2007); the Hässeldala Port samples display an increase in thymines at the 5′-end in both raw and trimmed reads, whereas all other bases display lower frequencies compared to later positions. Second, the frequency of G to A substitutions is expected to increase starting circa 10

bases from the 3′-end (Briggs et al., 2007); the Hässeldala Port samples display a longer increase in adenines toward the 3′-end. Intermediate positions show less variation than for raw reads (**Figure S3**). These features indicate the presence of ancient DNA reads in the Hässeldala Port samples, however we cannot conclude on the basis of this information alone that the samples contain DNA damage. We therefore used kPAL (Anvar et al., 2014) to perform a simple reference-free k-mer analysis of the data to identify issues with contamination and poor-quality samples. kPAL counts the number of k-mers of a given length and creates distance matrices based on the differences between k-mer profiles which we then use as input to a PCA using the R function prcomp.

Assessment of Ancient DNA Characteristics

QC reads were mapped with Bowtie2 version 2.2.8 (Langmead and Salzberg, 2012). The reference database consisted of 1,060,223 sequences of unicellular (prokaryote, archaeal, and eukaryote) origin designed to cover as many taxa as possible that are relevant for soil and environmental samples (Ahmed et al., 2018). Bam files were sorted with samtools version 1.3.1 (Li et al., 2009) and duplicate reads were removed with Picard version 2. (http://broadinstitute.github.io/picard) and finally mapped reads with mapping quality $<=25$ were filtered out with bamtools version 2.4.0 (Barnett et al., 2011). **Table S3** summarizes mapping and filtering statistics for the Hässeldala Port samples. On average 2.3% of the total sample reads mapped against the reference database, of which circa 3.4% remained after removing duplicate reads and mapping quality filtering and were used for taxonomic assignments. **Figure S2b** shows read length distributions per sample for QC-passed reads and suggests that a third characteristic feature to ancient DNA is visible in the Hässeldala Port samples. In ancient reads, the read length is normally very low and, in our case, it was concentrated to between 30 and 50 bp, all other lengths displayed lower frequencies.

Mapping, Filtering, and Taxonomic Assignment

We determined the taxonomic profiles of all plant DNA in our samples, by constructing a custom reference dataset containing all plastid sequences from the NCBI RefSeq project (2,477 sequences, 29th January 2018) and the plastid and ribosomal sequences from the PhyloNorway project (259 plant species from Norway included; **Table S4**). The sequence data was mapped against the custom reference dataset using BWA v0.7.5a aln algorithm (Li and Durbin, 2009) and all alignments with a mapping score of 0 (to account for multi-mapped reads) or $>=30$ were parsed for downstream quality control. Duplicate reads were removed using samtools v0.1.19 (Li et al., 2009) and low complexity reads (DUST score >1) were removed using PRINSEQ v0.20.4 (Schmieder and Edwards, 2011). In order to remove low-quality hits, the remaining reads were aligned to both the custom reference and the NCBI NT dataset (15th December 2017) with the NCBI-BLAST+ blastn algorithm v2.7.1 (Camacho et al., 2009). Due to a low number of plant sequences and high DNA damage, reads were kept if: (1) they had a similarity score $>=90\%$ and an alignment length $>=90\%$ and (2) they had a better alignment against the custom reference dataset than the NCBI NT dataset (bit score), or, in case of equal bit scores, (3) if the NCBI NT alignment was against a Viridiplantae. The up to 90% similarity level allowed for a maximum of three mismatches on the shortest possible reads (30 bp) and it was used to account for DNA damage, which leads to higher mismatches than expected. The double check against both the custom reference and GenBank nucleotide was needed to remove erroneous hits when using a lower threshold. A custom script was used to determine the last common ancestor for the remaining alignments. In addition, for each mapped read we double-checked that: (1) the assigned taxon was native in the region (Virtuella flora, http://linnaeus.nrm.se/flora/; Mossberg and Stenberg, 2010), (2) that it was likely to be present in the region during the last glacial/interglacial transition based on its current distribution (Hultén and Fries, 1986), and (3) that the reads could possibly be assigned to taxa available in the reference dataset. For each assigned taxon, we also checked for closely related native taxa present in the reference dataset and adjusted the taxonomic identification to a level that we could be certain of. For example, reads assigned to *Juncus filiformis* are presented as *Juncus* cf. *filiformis* because we are certain about the genus-level identification. However, since a full chloroplast reference genome is available only for four of the 25 *Juncus* species currently occurring in Sweden, we cannot exclude that the same read was shared among all *Juncus* species. It is also possible that, due to sequence similarity and absence of correct species in the reference dataset, reads were assigned to correct genus but wrong species. This is the case for example for *Diplazium dushanense*, which is presented as *Diplazium* sp. because the only species occurring in Sweden, but missing in the reference dataset, is *Diplazium sibiricum*. Finally, we only retained taxa that were represented by a minimum of two reads and used the maximum read count found in any control (extraction and/or library) as a removal threshold (i.e., we removed taxa that showed higher number of reads in the controls than in the samples). If a read was assigned to taxa that could not have been credibly present at our site based on the modern or ancient flora, we assigned the reads to a higher taxonomical level (either from species to genus or from genus to family), assuming that the first assigned taxon was absent in the reference database. If an assignment at a higher taxonomical level was still not plausible, the taxon was considered a false positive and removed from the dataset. Filtering data is always a trade-off between loosing true positive and keeping false positives (Ficetola et al., 2015; Alsos et al., 2018). We think that our filtering settings removed several true positives as e.g., *Ranunculus*, although the majority of taxa removed are not native to Scandinavia today and are assumed to be due to artifacts as e.g., sequence similarity, lack of native taxa in database, PCR or sequencing errors or contamination. All data was then parsed to MEGAN (Huson et al., 2007), R and excel for further downstream presentation and statistical analysis.

Loss-on-Ignition (LOI) and Chronology

Contiguous 1-cm thick samples were taken from both cores for LOI analyses. The samples were dried overnight at 105°C and subsequently ignited at 550°C following the procedures described

in Heiri et al. (2001) (**Table S1**). To obtain a timescale for the two new cores, we adopted the core-to-core alignment of Muschitiello et al. (2015) and described in Wohlfarth et al. (2017). This approach is based on a statistical correlation of the LOI% curves of #7.4 and #8 to that of master core #5 (**Table S1**), which serves as a chronological template for all of Hässeldala's sediment cores (Wohlfarth et al., 2017).

Plant Macro- and Micro-Fossil Analysis

For plant macro- and microfossil analyses 16 sub-samples were taken from the two cores at exactly the same levels as the samples for DNA analysis (**Table S1**). Our purpose was not to establish a new pollen stratigraphy for cores #7.4 and #8, as these are already available from previous cores (Wohlfarth et al., 2017), but to compare the pollen spectra in the respective samples to the plant macrofossil assemblages and to the DNA results. Prior to sieving for plant macro remains, the sediment samples were soaked in a sodium pyrophosphate [$Na_4P_2O_7 \times 10H_2O$] solution over night to disaggregate the clay minerals (Birks, 2002). Samples were sieved using a 250-μm mesh and remains retained on the sieve were identified using a binocular microscope. For microfossil analyses we subsampled 0.5 cm^3 of sediment from each of the 16 samples and prepared the samples using standard treatments (HCl, KOH, HF, acetolysis; Moore et al., 1991). The slides were mounted in glycerol, and pollen, spores and coenobia were counted using a light microscope at 400× magnification. Between 300 and 750 (mean = 575) pollen grains were counted in each sample. Microfossil identification followed Moore et al. (1991) and Reille (1992).

RESULTS

Lithology and Chronology

The lithostratigraphy of sediment cores #7.4 and #8 compares well to that of all previous sediment cores taken from Hässeldala Port (Wohlfarth et al., 2017): fine sands, silt and silty clay in the bottom part; clayey algae gyttja and algae gyttja in the middle part and gyttja and coarse detritus gyttja in the uppermost part (**Table S1**). The LOI% values of the two cores mirror the changes in sediment organic matter content and are almost identical to those obtained earlier for Hässeldala Port cores (Wohlfarth et al., 2017). The common features of the LOI% curves therefore allow precise statistical alignments to master core #5 (**Table S1**) for which an excellent chronological framework has been established (Muschitiello et al., 2015; Wohlfarth et al., 2017). Using the statistical alignments described in Muschitiello et al. (2015) and Wohlfarth et al. (2017), we transferred the chronology of core #5 to cores #7.4 and 8 to obtain an age assignment for each of our samples (**Table 1**). As seen in **Figure 2**, the ages of several samples in the two cores partly overlap (e.g., HÄ4.1 and HÄ4.2; HÄ5.1 and HÄ5.2), while other samples allow filling the respective age gaps. The age model for both cores combined with the almost identical lithostratigraphy permits combining the data sets of the two cores and presenting the samples in a temporal order (**Figure 2** and **Table 1**). We however clearly differentiate samples from core #7.4 from those of core #8 in all figures.

The statistical alignment of the LOI% data and transferring the age model from core #5 to cores #7.4 and 8, moreover allows us to assign regional pollen zone boundaries to our new cores by adopting the framework presented in Wohlfarth et al. (2017) (**Figure 2**). Accordingly, none of the samples compares to the regional Older Dryas pollen zone, while samples HÄ1.1–HÄ4.2 compare to the regional Allerød pollen zone, samples HÄ5.1–HÄ6.2 to the regional Younger Dryas pollen zone and samples HÄ7.1–HÄ8.2 to the Preboreal.

Plant Micro and Macrofossils

Microfossil analysis of the 16 samples allowed identifying more taxa than plant macrofossil analysis (45 vs. 22; **Table S1**). Spores and coenobia of ferns, mosses, algae, and fungi that were counted on the pollen slides, suggest a regional and local presence of these taxa. The two data sets are presented together with the eDNA data in **Figures 3A–D**. For comparison and convenience, we grouped the identified families and taxa as follows: (1) trees, shrubs and dwarf shrubs; (2) herbs, grasses and aquatics; and (3) ferns, mosses, algae, and fungi. A total of 58 taxa are represented in the plant micro- and macrofossil data sets; of these 36 taxa were identified as pollen, 13 taxa as macrofossils, and 9 taxa were present in both proxies (**Figure 4**).

Pollen of *Betula* sp. (undifferentiated) and *Pinus* were found in all analyzed samples. Pollen of these taxa are transported over long and short distances, which makes it difficult to determine whether they are of local or regional origin. Macroscopic remains of tree *Betula* (*B. pubescens*) however suggest that these trees grew close to the ancient lake during the Allerød, Younger Dryas, and Preboreal pollen zones (**Table S1**). The presence of tree *Betula* during the Allerød pollen zone has been reported earlier for Blekinge (Berglund, 1966; Wohlfarth et al., 2017), but its occurrence at Hässeldala Port during the Allerød and Younger Dryas pollen zones is novel. *Pinus* needles and bud scales are however only present in the uppermost samples, which represent the early Holocene. Pollen of shrubs and dwarf shrubs (e.g., Betulaceae, Ericaceae, *Juniperus*, *Salix*) occur throughout the record, while macroscopic remains only comprise frequent *Betula nana* and scarce *Empetrum nigrum* seeds and *Salix* sp. bark. Although the pollen record suggests local and regional presence of diverse herb/grass communities, herb/grass macrofossil finds are scarce and only include Cyperaceae/*Carex, Luzula, Dryas octopetala,* and possibly *Fragaria vesca* (**Table S1**). The micro and macrofossil samples analyzed here suggest that the local terrestrial vegetation was dominated by shrubs, dwarf shrubs, herbs and grasses and that tree *Betula* became established already during the Allerød pollen zone, while *Pinus* appeared first during the early Holocene.

Aquatic plants are well represented among macrofossils with taxa such as *Rancunculus* sect. *Batrachium*, various *Potamogeton* species, *Myriophyllum* cf. *alterniflorum*, *Nymphaea,* and *Isoëtes lacustris* (**Table S1**). *Nitella* oospores, spores of *Botryococcus* and *Pediastrum* coenobia are present in all samples, whereas spores of the fungi *Tilletia sphagnii* are only present in three samples. Moreover, macroscopic remains of *Sphagnum* and *Polytrichum strictum* also occur in several samples, as well as spores of

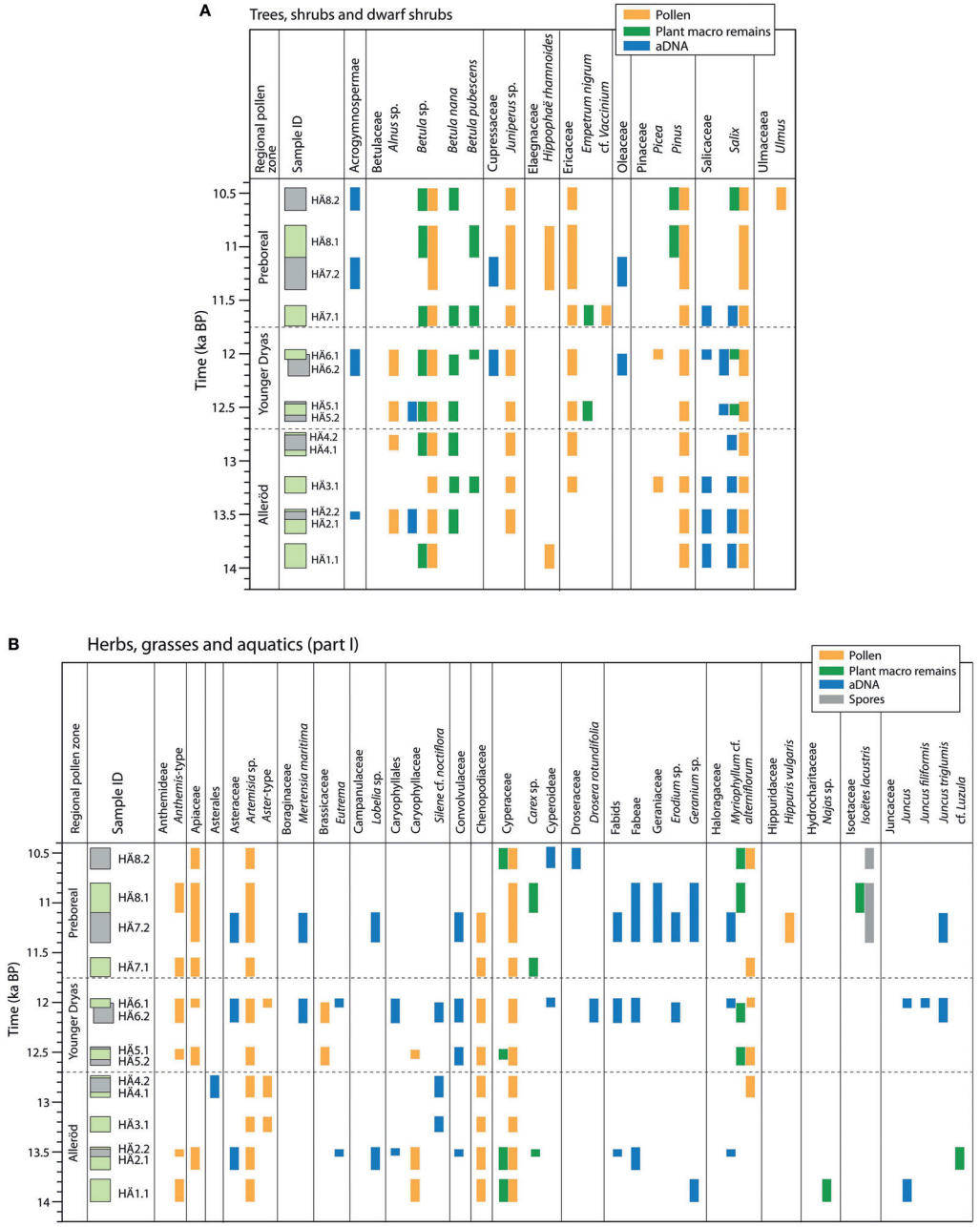


FIGURE 3 | Continued

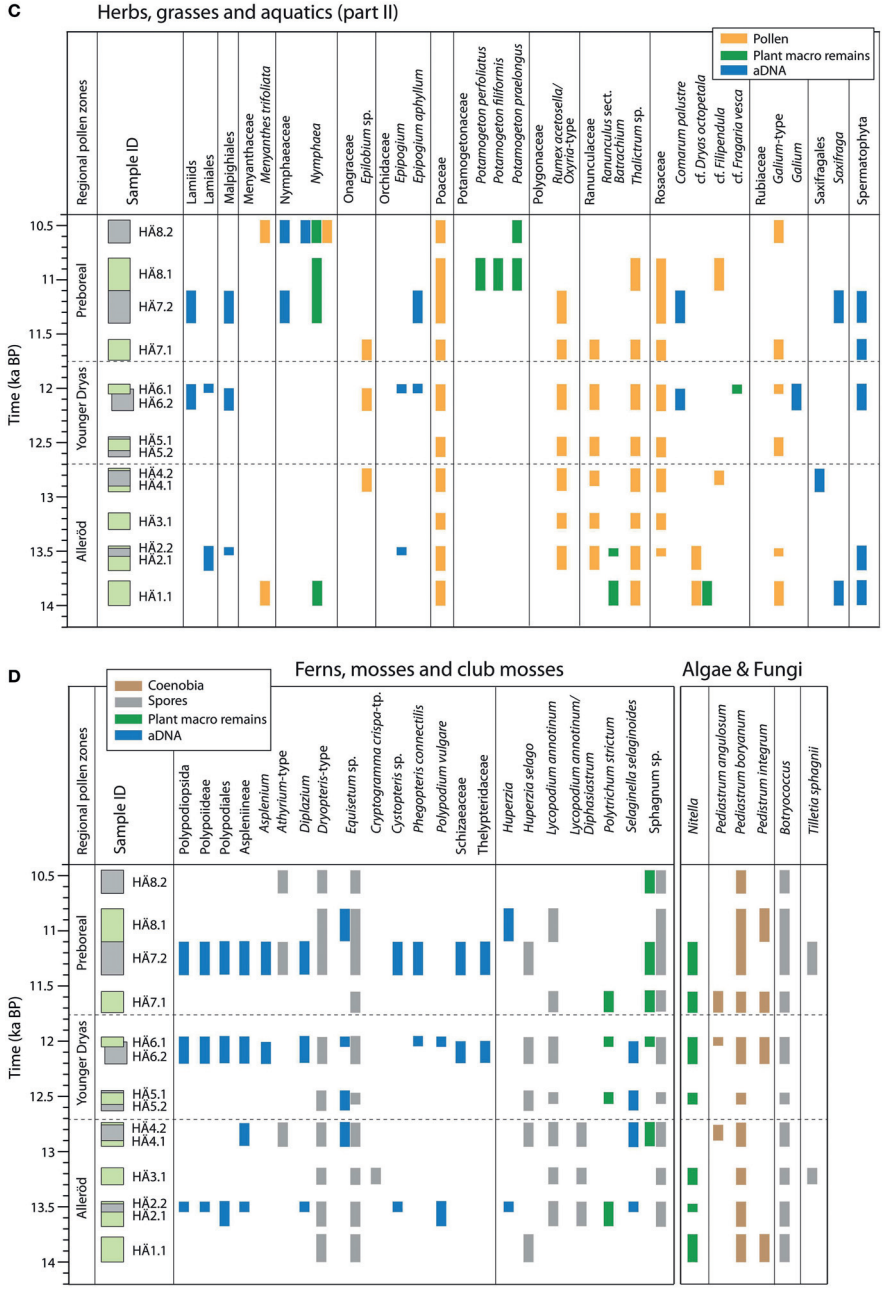


FIGURE 3 | Comparison between the results of micro- and macrofossil analyses and eDNA in Hässeldala Port's sediments shown according to age and in relation to the regional pollen zones. Samples from core #7.4 are shown in green and samples from core #8 in gray. **(A)** Trees, shrubs, and dwarf shrubs. **(B,C)** Herbs, grasses, and aquatic plants. **(D)** Ferns, mosses, algae, and fungi.

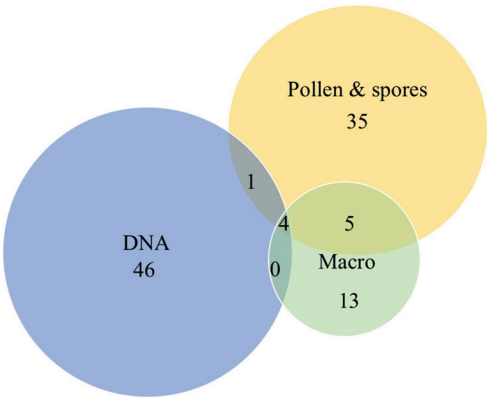


FIGURE 4 | Venn diagram showing the number of plant taxa detected from Hässeldala Port's sediments using micro- and macrofossil and eDNA analyses. See **Figure 3** for details on taxa.

for example *Dryopteris*-type, *Equisetum* sp., *Huperzia selago*, *Lycopodium,* and *Sphagnum.*

eDNA

Negative Controls

After DNA extraction none of the eight controls contained measurable DNA amounts. However, after library preparation and sequencing we found that 43,453,148 and 23,290,426 reads passed QC filters in the extraction and library controls, respectively (except for Blank4 that provided no sequences and E7 that had no reads passing quality filters). These reads however, show low quality scores (**Figure S1**) and do not show the damage patterns typical of ancient reads (**Figure S3**). In addition, K-mer distribution analyses (PCA of the distance matrix for k = 6, raw data) indicate sequence anomalies, suggesting low data quality due to errors and sequencing biases (**Figure S4**). Despite our efforts to prevent any contamination, the taxonomic alignment reveals taxa in the controls that originate from commonly known contaminant DNA like human and bacteria. A visual inspection preformed with MEGAN showed that the three most common plant contaminants were rice (*Oryza*), wheat (*Triticum*), and maize (*Zea*), while the rest of the plant reads were assigned to tropical or exotic species (with the exception of only one read assigned to *Pinus*). These were therefore not considered further in the analyses.

Samples

We generated a total of 1,014,951,5 reads for the Hässeldala Port samples, of which the per-sample read counts lie in the range of 12–62 million (**Table S2**), except for a few samples that failed to yield sufficient number of high-quality reads (HÄ1.2 and HÄ6.2_2) and those devoid of reads after filtering (HÄ3.1_2 and HÄ5.1_2). Read length distributions per sample for QC-passed reads was concentrated to between 30 and 50 bp and showed no clear association of post-mortem DNA decay (fragmentation and damage) with age (**Figures S2, S3**). This suggests that, beyond the initial rapid degradation, eDNA is well-preserved at Hässeldala Port.

On average 2.3% of the total sample reads mapped against the reference database, and of these only 3.4% remained after cleaning and filtering and were used for taxonomic assignments. A general overview of mapped reads passing filters that were assigned to taxa from the three domains of cellular life (Bacteria, Archaea, and Eukaryotes) can be found in **Table S5**. Of all mapped reads the majority were assigned to Bacteria (79% on average), while 15% was assigned to Eukaryota and <6% to Archaea. These proportions do not necessarily reflect the true abundance of each domain in our samples due to bias and limitations existing in the database. However, given that the alignment with the plant custom reference dataset eventually detected only 1,634 plant reads (see below), we assume the plant percentage to be very low and in the range of 0.1%.

The 1,634 reads obtained matched 122 unique plant taxa, of which 69 were excluded by the following criteria: (1) taxa identified by only one read, (2) taxa present in the controls, (3) taxa exotic to Sweden during the studied time period (**Table S6**). The excluded taxa are to a large part assigned to tropical plant species (e.g. *Musa*, *Aloe*, *Bomarea, Cytinus, Ephedra,* Cycadales). Out of the remaining 53 taxa, two pairs of reads were assigned to the same taxa resulting in 51 single taxa. Seven of these taxa showed reads in the controls that were however below the removal threshold and were therefore kept. We also excluded sample HÄ3.2 where only one read could be assigned. This resulted in a total of 14 samples containing 51 unique plant taxa assumed to represent true positives. Of these ∼22% were assigned at the family level, 31% at the genus level and 22% at the species level. The remaining 25% of the taxa were identified at higher taxonomical levels (order and clade). The high percentage of taxa identified at or below the genus level (53%) is likely due to the inclusion of the local PhyloNorway data set in our reference database for taxonomical assignment.

The largest number of reads (166) was found for the fern *Diplazium.* The 16 other taxa recorded with 10 or more reads were trees and shrubs such as Salicaceae (27 reads) and *Salix* (43), the large clade of Acrogymnospermae (31), herbs of Convolvulaceae (54), and Asteraceae (11), aquatic or semi-aquatic plant species e.g., *Juncus triglumis* (31) and the clubmoss *Selaginella selaginoides* (10).

Comparison of eDNA, Micro- and Macro-Fossil Data Sets

Overall, we recovered more taxa by DNA (51) than singularly from the micro- (44) and the macro-fossil (22) record (**Figure 4**). The number of taxa identified at the species level varied in the three data sets between 14 for microfossils (32%), 11 for macrofossils (45%), and 11 for DNA (21%). The three proxies show important differences in respect to identified major plant taxa and only four taxa were detected by all three proxies (*Betula* sp., *Salix*, *Myriophyllum* cf. *alterniflorum,* and *Nymphaea*).

Among the trees, shrubs, and dwarf shrubs (**Figure 3A**), *Betula* and Salicaceae were recorded by all three proxies, although

the taxonomic level of identification varied. *Pinus* was recorded in the pollen (all periods) and macrofossil (Preboreal) data sets, whereas the one-read DNA record did not pass filtering criteria. Other taxa detected as pollen (e.g., *Ulmus, Alnus* sp., *Hippophaë rhamnoides, Picea,* Chenopodiaceae, Poaceae, Apiaceae) were not recorded as macrofossil or DNA. DNA of Cupressaceae was detected and could potentially represent *Juniperus*, which was identified by pollen. Ericaceae were present in the pollen spectra (some of it was tentatively assigned to *Vaccinium*) whereas *E. nigrum* was identified by macrofossils. Oleaceae (family including *Fraxinus*) and Acrogymnospermae (including *Picea, Pinus, and Juniperus*) were only recorded by DNA.

Among herbs, grasses and aquatics (**Figures 3B–D**), Cyperaceae (including *Carex* sp. and Cyperoideae identified by macroscopic remains and eDNA, respectively), *Myriophyllum* cf. *alterniflorum,* and *Nymphaea* were present in all three proxies. Rosaceae could also be identified by all three proxies, but with assumed different taxa. Ranunculaceae were identified by pollen and macroscopic remains but as different taxa. *Isoëtes lacustris* was identified both as spores and macroscopic remains. Other herbs are only recorded as pollen, such as e.g., *Anthemis*-type, Apiaceae, Chenopodiaceae, *Hippuris vulgaris*, *Menyanthes trifoliata*, *Epilobium* sp., Poaceae, and *Rumex acetosella. Two* taxa appear in the macrofossil record only: *Najas* sp. and various *Potamogeton* species. However, eDNA detected families (Asteraceae) that include taxa found as pollen (*Artemisia*); species or genera (*Eutrema, Silene noctiflora, Comarum palustre, Galium*) that belong to families identified in the pollen record; and several taxa that show no signal in the micro- and macrofossil records: Boraginaceae (with the species *Mertensia maritima),* Campanulaceae (with the genus *Lobelia),* Convolvulaceae, Droseraceae (with the genus *Drosera rotundifolia),* Geraniaceae (with *Erodium* and *Geranium),* Juncaceae (with *J. filiformis* and *J. triglumis*), Orchidaceae (with *Epipogium* and *Epipogium aphyllum*) the genus *Saxifraga*, and larger clades like Fabids, Fabeae, Lamiids, and Lamiales. A large number of ferns without identifiable spores were also identified exclusively by eDNA (Schizaeaceae, Thelypteridaceae, and several taxa within the Polypodiaceae family). Moreover, eDNA could detect a few club mosses (*Huperzia*, *Selaginella selaginoides*), but none of the algae species *Pediastrum* and *Botryococcus* that were identified by microfossil analysis (**Figure 3D**), although these were present in the database.

Except for a few taxa—trees/shrubs growing close to the lake (*Betula* and *Salix*), some aquatic plants (*Myriophyllum alterniflorum, Nymphaea*) and some ferns (*Equisetum)*—we found limited overlap between the three proxies (**Figure 4**). However, several of the taxa that were recorded by eDNA now allow complementing information regarding the local vegetation, especially in respect to herbs/grasses/aquatics (*M. maritima*, *Eutrema, Lobelia, S. noctiflora, D. rotundifolia, Erodium, Geranium, J. filiformis*, *J. triglumis, E. aphyllum, C. palustre, Saxifraga)* and a variety of ferns and mosses (**Figures 3B–D**). Taken together, the eDNA results allow us to increase the total number of identified taxa from 67 (pollen and plant macro remains) to more than 100.

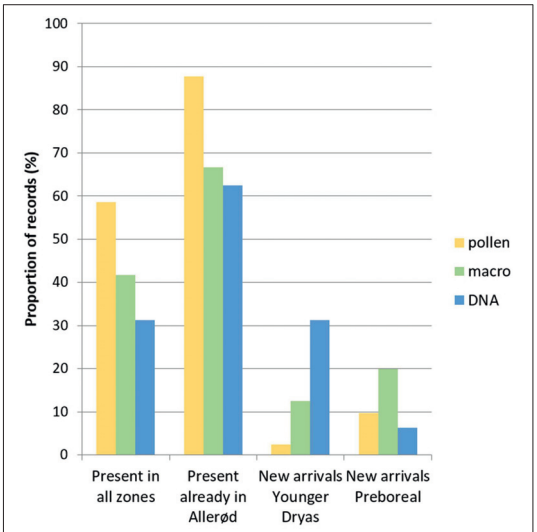


FIGURE 5 | Presences and arrival of taxa (in percentage) detected in pollen, macrofossils and ancient DNA analyses.

Taxa Representation During the Regional Pollen Zones

The diversity of plant DNA taxa was compared to the regional pollen zones (Allerød, Younger Dryas, and Preboreal) earlier identified at Hässeldala Port, which correspond to warmer/colder/warmer time intervals (Wohlfarth et al., 2017). The eDNA of some tree and shrub species, such as *Betula* sp. and Salicaceae/*Salix,* was present in Allerød and Younger Dryas samples, in accordance with the pollen and macrofossil data (**Figure 3A**), which suggests that boreal to subarctic climate conditions also prevailed during the Younger Dryas pollen zone.

The proportion of taxa detected in all pollen zones was generally lower for eDNA (31%) than for pollen (59%) or macrofossils (42%) (**Figure 5**). The eDNA data set however suggests that a higher number of new taxa became established during Younger Dryas, while the pollen and macrofossil data indicate that new plants became established only during the Preboreal. We also found a higher number of taxa and reads per sample in Younger Dryas samples (average 17.3 and 103.8, respectively), than in Allerød (average 7.8 and 34, respectively) and Preboreal (11.5 and 49.3, respectively) samples (**Figure 5**). This may be a mapping artifact because our reference library is biased toward cold tolerant species. It is however notable that all taxa, except for two of the new species that appeared during Younger Dryas, persisted into the Preboreal.

DISCUSSION

At Hässeldala Port, the metagenomic data allowed to identify a limited number of plant reads, but sufficient to identify a larger

number of taxa than either micro- or macro-fossils alone and therefore increased the overall information about past floras.

The eDNA Record

eDNA signals were mainly derived from terrestrial taxa. These likely represent common and dominant taxa that were growing close to the lake as they have the strongest signal (Alsos et al., 2018). Aquatic taxa, such as Juncaceae, Nymphaceae, *Myriophyllum alternifolium* may have been less common or had a scattered presence at Hässeldala, since aquatic taxa have usually a higher chance to be detected due to less taphonomical restrictions (Alsos et al., 2018).

The Shotgun Sequencing Approach

Starting from more than a billion reads only very few reads (<2,000) could eventually be assigned to the likely local plant taxa. Thus, the majority of reads did not match any plants in our reference library and could not be used in our study. Similarly, in two recent genomic studies on sediments only a small fraction of the reads that aligned to the database could be eventually assigned to plants using the shotgun sequencing approach. Slon et al. (2017) shotgun sequenced sediments from seven caves and aligned between 4.9 and 21% of the reads to the database and only a small proportion of these (1–2%) aligned to plant sequences (Viridiplantae). Pedersen et al. (2016) aligned and assigned to plant taxa only 2,596 out of 1 billion aligned reads. In our study, more than 78% of the mapped reads (2.3% of the total) were assigned to Bacteria, circa 15% to Eukaryotes and more than 5% to Archaea (see also Ahmed et al., 2018).

Overall these results suggest that the genomic data is complementary to micro and macrofossil records and potentially holds many more taxa than those detected in this study. Thus, if whole community reconstructions are the goal of a study, genomics is a practical approach for exploring a large number of taxa, from large animals like mammoth, to bacteria, algae, and archaea. However, the number of plant taxa we found in each sample was less than seven for the majority of the samples, which is lower than the 20–60 taxa detected with PCR-based metabarcoding approaches that are generally used to target plants (Pansu et al., 2015; Parducci et al., 2015; Alsos et al., 2016, 2018; Sjögren et al., 2016; Niemeyer et al., 2017; Zimmermann et al., 2017). The PCR-based method however, may be biased due to differences in the amplification of different taxa. For example, Cyperaceae is normally underrepresented in metabarcoding studies probably due to a longer than average fragment size (Alsos et al., 2018). Our hope using the shotgun sequencing approach was to overcome this bias and provide data that also allow for more quantitative interpretations. However, we note that Cyperaceae pollen or macro remains were found in many samples, but Cyperaceae DNA was only found in two samples. As metabarcoding is cheaper and requires less bioinformatic work than the shotgun sequencing approach, we conclude that, if vegetation reconstruction alone is the goal of a study, metabarcoding is currently a more useful and cost-efficient method.

We think that several issues currently limit the shotgun sequencing approach. The first is the lack of sufficiently curated plant reference genomes. For mapping we here used plastid DNA sequences present at NCBI in combination with plastid DNA and ribosomal DNA derived from the PhyloNorway data set. However, we did not check the nuclear genomes of the plants present in NCBI, which could have contributed with more information. At the moment there is little suitable reference material available in the NCBI database for plant nuclear genomes and the majority of the sequences come from economically important crop species like maize, rice, soya and potato. The lack of reference genome sequence data for non-commercial plant species as those analyzed in our study, substantially reduces the potential for success of the bioinformatic analyses with metagenomic data. In addition, metagenomic data can vary in content across samples from the same or similar environments (Quince et al., 2017), as it was also seen in our two cores which were located only few meters from each other, but reported in several instances a different taxa composition (e.g., Hä2.1/Hä2.2, Hä7.1/Hä7.2, or Hä8.1/Hä8.2 in **Table S6**). This complicates the detection of biologically meaningful differences between samples of different ages.

The second issue is methodological and may be related to the protocols used for extraction and library construction. These often allow only a small percentage of eukaryote reads (and in particular plant reads) to be extracted and sequenced as compared to archaeal and bacterial reads. The problem can be partially circumvented by using new developed DNA extraction methods for sediments like the phosphate-based protocols developed by Taberlet et al. (2012); Zinger et al. (2016); a disadvantage is however a strong detriment of the final DNA amount obtained. More recent advances in high-throughput DNA sequencing technologies allow circumventing the problem more efficiently. Today it is possible to target-enrich specific genomic DNA regions via hybridization capture prior to sequencing (Schmid et al., 2017; Slon et al., 2017). This allows generating larger data sets for multiple DNA target loci in parallel with high efficiency and improving substantially the quality of the taxonomical assignments after sequencing. The target capture sequencing technology coupled with shotgun sequencing seems a good alternative and a powerful strategy for molecular ecologists who want to work with taxa for which few genomic resources are available.

A third issue to consider is the properties of the sediments in relation to DNA preservation. It is very difficult to assess the time frame for which DNA molecules are preserved in lake sediments. We know, however that local environmental conditions may be more important than other factors, including time (Allentoft et al., 2012). Cold, dry and anoxic environments and a neutral or slightly alkaline pH represent optimal conditions for DNA preservation. The degree of preservation may also be a function of water depth, with the poorest preservation resulting from decay in shallow and relatively turbulent water and/or high algae content (Rich, 1989; Alsos et al., 2018). Even if current studies seem to show little success with eDNA extracted from below bogs (Pedersen personal communication), at Hässeldala Port we did not find a clear correlation between sample age and post-mortem decay (damage and fragmentation), suggesting that, beyond the initial rapid degradation, DNA molecules were well-preserved. We found on the other hand, a high variation among samples both in number of reads and in number of taxa detected. This

observation compares to plant macrofossil studies, which can show large variations in number and type of taxa recovered (Birks, 2013).

Finally, because of contamination risks, we employed a very cautious approach during DNA read assignment; this further limited the power of the shotgun sequencing approach. To avoid misidentifications, we used 90% of the sequences matching to at least two different reference database entries occurring at least two times in a sample. We therefore excluded several taxa that were assigned by only one read. The reason for using a lower percentage threshold (90%) was that we did not have access to a complete reference database. Indeed, a shotgun read can originate from any part of the genome, and since it is a whole molecule and not an amplified fragment as in the metabarcoding approach, the DNA damage patterns become a more serious issue for identifications. By limiting to a 90% match we included also some damaged sequences. The successive GenBank comparison was done to make sure to exclude sequences originating from non-plant sources (most likely bacteria).

The fact that our extraction controls did not contain any measurable DNA but still yielded reads after sequencing analyses highlights the challenges we are still facing for decontaminating laboratory reagents and consumables (Champlot et al., 2010). Mostly, background contamination comes from reagents used in the laboratory since the controls contained mainly exotic plant species and not much of the north European taxa found in the samples. We therefore find our samples to contain a mixture of authentic ancient DNA as well as low-background contamination of modern human, bacterial and plant taxa from the reagents which were eventually removed from the dataset.

Proxy Comparison

We found limited overlap between the three proxies (**Figure 4**). This is not surprising since pollen reflect the local and regional vegetation as well as long-distance transportation, while the plant macrofossil and eDNA records register the local vegetation in and around the site. It is also a positive result since a larger number of taxa could be identified using the three proxies. Some major taxa detected by pollen (e.g., *Pinus*, *Ulmus,* Chenopodiaceae, Poaceae, Apiaceae), some assigned to long-distance transport, were not recorded either by the macrofossil or eDNA data sets, confirming that pollen does not contribute to eDNA in lake sediments (Parducci et al., 2015; Alsos et al., 2016; Pedersen et al., 2016; Sjögren et al., 2016; Clarke et al., 2018), likely because only few pollen grains per species are present in each sediment sample (in our case 623 on average **Table S1**) and each pollen grain contains only two or three cells.

Surprisingly, some taxa as the algae *Pediastrum* and *Botryococcus*, the clubmosse *Lycopodium*, the moss *Sphagnum,* the dwarf shrub *Empetrum* and some aquatic plant species (*Potamogeton*) were identified by both micro- and macrofossil analyses, but not by eDNA even though they were present in our reference database. This may relate to differences in preservation conditions for different taxa and at different locations and may also explain the variations we observed across samples in our two cores located only a few meters apart. There is still limited knowledge regarding the taphonomy of DNA in sediments (Yoccoz et al., 2012; Alsos et al., 2018; Edwards et al., 2018). Importantly, all of the three proxies only detect a fraction of the total diversity of the past vegetation (Allen and Huntley, 1999; Drake and Burrows, 2012; Birks and Birks, 2016; Alsos et al., 2018). Thus, a complete overlap can never be expected.

One strange observation was *Juniperus*, as this genus is present in our database by several *Cupressaceae* species, including *Juniperus communis*. However, 19 reads were assigned to *Cupressus chengiana* (a tree endemic to China) and were therefore presented as *Cupressaceae*. We have no explanation why it was not assigned to *J. communis* and we suggest therefore to consider this record with caution.

Climate and Plant Diversity

All three proxies show a subset of species that are only recorded during warmer periods, as has also been observed in previous studies from the same site using pollen and macrofossils (Wohlfarth et al., 2006, 2017). Surprisingly, eDNA is the only proxy to show a high number of new taxa appearing during the Younger Dryas pollen zone (**Figure 5**). Even more puzzling is that most of these new taxa persist into the Preboreal period. There is only one new taxon detected by pollen appearing in Younger Dryas (Brassicaceae), and this does not persist into the Preboreal. Macrofossils detect far less species, but notably, the three new species appearing in Younger Dryas (*E. nigrum*, *Salix,* and *Myriophyllum* cf. *alterniflorum*) persist also into the Preboreal period. Thus, cold adapted species may indeed have survived for some time after the start of the Holocene warming due to a lag response to climate change (Alexander et al., 2018). These results confirm previous studies suggesting that in lake sediments eDNA is more efficient in recording persistence of species as compared to macrofossils (Alsos et al., 2016), and may help to better understand how climate change affected not only colonization but also the survival and persistence of plant species.

The appearance of new DNA taxa during a cold period may be related to an over representation of cold-tolerant species in the reference library and/or better preservation of DNA in sediments under cold conditions, or also to the presence of clays in the sediments to which DNA can more easily bind and remain protected by microbial enzymatic activity. During the Younger Dryas the ancient lake was likely stratified, ice-covered and anoxic large parts of the year, which may have also favored DNA preservation in the sediments. The area around Hässeldala Port became ice free some time before 14.5 ka BP. Sediment lithology, geochemistry and the low organic matter content of the basal sediments (**Table 1**) suggest unstable catchment soils and higher run-off during the early lake stage (Kylander et al., 2013). These unstable soils likely prevented the initial establishment of denser local vegetation. Catchment erosion gradually decreased during the Alleröd and was low during the Younger Dryas (Kylander et al., 2013). Lower organic matter content during the Younger Dryas is therefore related to lower lake organic productivity and not to catchment run-off. With the start of the Holocene and due to infilling, the lake gradually transformed into dry mire and subsequently into a dry shrub heathland as shown by diatom and insect studies, respectively, and this change led to a shift in soil conditions from alkaline to more acid (Wohlfarth et al., 2017).

This again had an impact on the type of plants on and around the site. Furthermore, water depth and lake status changes may have played an important role for plant establishment and survival, or the fact that smaller plants did not have to compete with taller plants during the cold Younger Dryas and could expand more easily. An alternative hypothesis is that boreal-subarctic conditions also prevailed during the Younger Dryas, as also recently proposed by Schenk et al. (2018).

Our knowledge regarding the local impact of the distinct climatic shifts that characterized the end of the last glacial period is still limited and depends largely on the choice and interpretation of available proxies. Wohlfarth et al. (2017) for example had suggested, based on their multi-proxy study of Hässeldala Port's sedimentary record, that climatic shifts had a major impact on the site's environment affecting plant, diatom, and chironomid communities. In contrast, Schenk et al. (2018) recently noted that summer temperatures did not change markedly between Allerød and Younger Dryas, but that Younger Dryas summers were distinctly drier and winter temperatures significantly lower as compared to the Allerød. A re-evaluation of Hässeldala Port's chironomid-based summer temperature record (Schenk et al., personal communication) moreover shows that Younger Dryas summer temperatures were not significantly different from Allerød summer temperatures. Our new plant macrofossil and metagenomic data sets confirm these observations, as they show no change in plant community response or taxa composition between warmer (Allerød) and colder (Younger Dryas) time intervals, but the presence of a boreal-subarctic vegetation.

CONCLUSIONS

We present the full metagenome data set preserved in the Hässeldala Port lake sediment record and compare it to the plant micro- and macrofossil datasets analyzed in the same sediment samples. Our hypothesis was that the shotgun metagenomic approach allows retrieving complementary information to the micro- and macro-fossil record, thus providing details about plant community changes in relation to the climatic shifts that occurred at the transition from a glacial to an interglacial climate state. The genomic approach however, permitted only a small percentage of reads to map against our database (2.3%) and of these, plant reads were a small percentage (ca 0.1% of the total). We did not find clear associations between the number of reads/taxa, sample age and sediment organic content, suggesting that, beyond the initial rapid degradation, eDNA in the Hässeldala Port sediments was present and overall well-preserved.

Overall, we found: (1) limited overlap between proxies (but more taxa could be in common at different taxonomic levels), (2) highest diversity in number of reads and number of read/sample in the colder period (Younger Dryas) with no direct association with climate, (3) no association between DNA post-mortem decay (damage and fragmentation) and sample age or changes in organic matter content in the sediments, (4) several taxa detected only by eDNA.

Although the proportion of plant DNA was very low in the analyzed samples, the shotgun metagenomic approach added a considerable number of new taxa to those identified as pollen or plant macrofossils and thus increased the overall information about past floral assemblages (from 67 taxa identified by pollen and plant macro remains to more than 100).

The combined eDNA and plant macrofossil data sets indicate that tree *Betula* (*B. pubescens*) grew close to the ancient lake during the Allerød, Younger Dryas, and Preboreal pollen zones and that boreal to subarctic climate conditions and not arctic conditions prevailed during the Younger Dryas pollen zone. Comparisons between the pollen and eDNA data sets also confirm earlier observations that eDNA in lake sediments reflects local plant occurrences only. Moreover, the eDNA data set suggests that a higher number of new taxa became already established during Younger Dryas while the pollen and macrofossil data indicate that new plants became established only during the Preboreal.

Shotgun sequencing, though less time-consuming is still quite expensive, and for the moment, metabarcoding seems to be a more cost-efficient approach for the detection of plants from eDNA samples extracted from lake sediments. In the future, coupling target capture technology with shotgun metagenomics could significantly enhance the ability to investigate past genomic diversity in lake sediments, particularly as full-genome reference databases are being built up. So far, the performance of these combined methodologies has seen limited investigations with ancient lake sediments and we think it has a potential for becoming an important tool for future investigations of plant biodiversity changes in lake sediments, in combination with complementary pollen and plant macrofossil analyses.

AUTHOR CONTRIBUTIONS

LP, TS, and BW designed the research. LH, BW, and LP organized and performed the coring. LH, PU, MP, MV, and JS performed the lab research. LP, TS, BW, PU, MP, YL, IA, MV, and JS analyzed the data. LP, BW, IA, and MP wrote the paper with final contributions from MV, JS, PU, YL, LH, and TS.

ACKNOWLEDGMENTS

We would like to thank Frederik Schenk for help during field work, Eske Willerslev (University of Copenhagen, Denmark) for being able to use the facilities of the Center of Geo-Genetics during DNA extraction and library preparation; Engy Ahmed for help during coring, for loss-on-ignition analyses of cores #7.4 and #8, and for preparing the DNA samples at the Center of Geo-Genetics; Francesco Muschitiello for creating the age models for cores #7.4 and #8. The present study was supported by SFO grants from the Faculty of Science, Stockholm University (#5.1.2-3222-14) to BW, LP, and TS; Bioinformatics Long-term Support (WABI) to BW, LP, and TS; the Swedish Research Council (grant no. 2013- D0568401), Formas (Dnr 2016-01236), and the Carl Triggers' Foundation (grant no. 14:371) to LP. PU was supported by a grant from the Knut

and Alice Wallenberg Foundation to the Wallenberg Advanced Bioinformatics Infrastructure. Bioinformatics Infrastructure, IA by grants from the the Research council of Norway (226134/F50), and JS by grants from the Academy of Finland (projects 1278692 and 1310649).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2019.00189/full#supplementary-material

REFERENCES

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., et al. (2018). Archaeal community changes in Lateglacial lake sediments: evidence from ancient DNA. *Quat. Sci. Rev.* 181, 19–29. doi: 10.1016/j.quascirev.2017.11.037

Alexander, J. M., Chalmandrier, L., Lenoir, J., Burgess, T. I., Essl, F., Haider, S., et al. (2018). Lags in the response of mountain plant communities to climate change. *Glob. Chang. Biol.* 24, 563–579. doi: 10.1111/gcb.13976

Allen, J. R. M., and Huntley, B (1999). Estimating past floristic diversity in montane regions from macrofossil assemblages. *J. Biogeogr.* 26, 55–73. doi: 10.1046/j.1365-2699.1999.00284.x

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., et al. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B Biol. Sci.* 279, 4724–4733. doi: 10.1098/rspb.2012.1745

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

Alsos, I. G., Sjögren, P., Edwards, M. E., Landvik, J. Y., Gielly, L., Forwick, M., et al. (2016). Sedimentary ancient DNA from Lake Skartjørna, Svalbard: assessing the resilience of arctic flora to Holocene climate change. *Holocene* 26, 1–16. doi: 10.1177/0959683615612563

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed May 5, 2017).

Anvar, S. Y., Khachatryan, L., Vermaat, M., van Galen, M., Pulyakhina, I., Ariyurek, Y., et al. (2014). Determining the quality and complexity of next-generation sequencing data without a reference genome. *Genome Biol. Evol.* 15:555. doi: 10.1186/s13059-014-0555-3

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., and Marth, GT (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27,1691–1692. doi: 10.1093/bioinformatics/btr174

Berglund, B. E. (1966). Late-Quaternary vegetation in eastern Blekinge, south-eastern Sweden. *Opera Botanica*, 12, 5–180.

Birks, H. H. (2002). "Plant macrofossils," in *Tracking Environmental Change Using Lake Sediments, Vol. 3*, ed J. P. Smol, H. J. B. Birks, W. M. Last, R. S. Bradley, and K. Alverson (Dordrecht: Springer) 49–74.

Birks, H. J., and Birks, H. H. (2016). How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? *New Phytol.* 209, 499–506. doi: 10.1111/nph.13657

Birks, H. J. B. (2013). "Plant macrofossil introduction," in *Encyclopedia of Quaternary Science,* eds H. P. Mock and S. A. Elias (Amsterdam: Elsevier), 593–612. doi: 10.1016/B978-0-444-53643-3.00203-X

Bjorck, S., Kromer, B., Johnsen, S., Bennike, O., Hammarlund, D., Lemdahl, G., et al. (1996). Synchronised terrestrial-atmospheric deglacial records around the North Atlantic. *Science* 274, 1155–1160. doi: 10.1126/science.274.5290.1155

Blockley, S. P. E., Lane, C. S., Hardiman, M., Rasmussen, S. O., Seierstad, I. K., Steffensen, J. P., et al. (2012). Synchronisation of palaeoenvironmental records over the last 60,000 years, and an extended INTIMATE1 event stratigraphy to 48,000 b2k. *Quat. Sci. Rev.* 36, 2–10. doi: 10.1016/j.quascirev.2011.09.017

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14616–14621. doi: 10.1073/pnas.0704665104

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

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

Clarke, C. L., Edwards, M. E., Brown, A. G., Gielly, L., Lammers, Y., Heintzman, P. D., et al. (2018). Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (sedaDNA) and pollen. *Boreas* 12:299. doi: 10.1111/bor.12357

Drake, H., and Burrows, C. J. (2012). The influx of potential macrofossils into Lady Lake, north Westland, New Zealand. *N.Z. J. Bot.* 18, 257–274. doi: 10.1080/0028825X.1980.10426924

Edwards, M. E., Alsos, I. G., Yoccoz, N. G., 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. doi: 10.1177/0959683618798095

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

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* 15, 543–556. doi: 10.1111/1755-0998.12338

Heiri, O., Lotter, A. F., and Lemcke, G. (2001). Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *J. Paleolimnol.* 25, 101–110. doi: 10.1023/A:1008119611481

Hultén, E., and Fries, M. (1986). *Atlas of North European Vascular Plants North of the Tropic of Cancer*. Königstein: Koeltz Scientific Books.

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

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

Kylander, M. E., Klaminder, J., Wohlfarth, B., and Löwemark, L. (2013). Geochemical responses to paleoclimatic changes in southern Sweden since the late glacial: The Hässeldala Port lake sediment record. *J. Paleolimnol.* 50, 57–70. doi: 10.1007/s10933-013-9704-z

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lowe, J. J., Rasmussen, S. O., Björck, S., Hoek, W. Z., Steffensen, J. P., Walker, M. J. C., et al. (2008). Synchronisation of palaeoenvironmental events in the North Atlantic region during the Last Termination: a revised protocol recommended by the INTIMATE group. *Quat. Sci. Rev.* 27, 6–17. doi: 10.1016/j.quascirev.2007.09.016

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200

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

Mossberg, B., and Stenberg, L. (2010). *Den nya Nordiska Flora (The new Nordens Flora)*. Stockholm: Wahlström & Widstrand.

Muschitiello, F., Pausata, F. S. R., Watson, J. E., Smittenberg, R. H., Salih, A. A. M., Brooks, S. J., et al. (2015). Fennoscandian freshwater control on Greenland

hydroclimate shifts at the onset of the Younger Dryas. *Nat. Commun.* 6:8939. doi: 10.1038/ncomms9939

Nagler, M., Insam, H., Pietramellara, G., and Ascher-Jenull, J. (2018). Extracellular DNA in natural environments: features, relevance and applications. *Appl. Microbiol. Biotechnol.* 116, 67–14. doi: 10.1007/s00253-018-9120-4

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

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

Parducci, L., 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

Parducci, L., Jørgensen, T., Tollefsrud, M. M., Elverland, E., Alm, T., Fontana, S. L., et al. (2012). Glacial survival of boreal trees in northern Scandinavia. *Science* 335, 1083–1086. doi: 10.1126/science.1216043

Parducci, L., Väliranta, M., Salonen, J. S., Ronkainen, T., Matetovici, I., Fontana, S. L., et al. (2015). Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20130382. doi: 10.1098/rstb.2013.0382

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., et al. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature* 53, 45–49. doi: 10.1038/nature19085

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844. doi: 10.1038/nbt.3935

Rasmussen, S. O., Birks, H. H., Blockley, S. P. E., Brauer, A., Hajdas, I., Hoek, W. Z., et al. (2014). Dating, synthesis, and interpretation of palaeoclimatic records of the Last Glacial cycle and model-data integration: advances by the INTIMATE (INTegration of Ice-core, MArine and TErrestrial records) COST Action ES0907. *Quat. Sci. Rev.* 106, 1–13. doi: 10.1016/j.quascirev.2014.10.031

Reille, M (1992). *Pollen et Spores d'Europe et d'Afrique du Nord*. Marseille: URA CNRS.

Rich, F. J. (1989). A review of the taphonomy of plant remains in lacustrine sediments. *Rev. Palaeobot. Palynol.* 58, 33–46. doi: 10.1016/0034-6667(89)90055-9

Schenk, F., Väliranta, M., Muschitiello, F., Tarasov, L., Heikkilä M., Björck, S., et al. (2018). Warm summers during the Younger Dryas cold reversal. *Nat. Commun.* 9:1634. doi: 10.1038/s41467-018-04071-5

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., et al. (2017). HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods Ecol. Evol.* 8, 1374–1388. doi: 10.1111/2041-210X.12785

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27, 863–864. doi: 10.1093/bioinformatics/btr026

Sjögren, P., Edwards, M. E., Gielly, L., Langdon, C. T., Croudace, I. W., Merkel, M. K., 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

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., et al. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science* 53:eaam9695. doi: 10.1126/science.aam9695

Steffensen, J. P., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Fischer, H., et al. (2008). High-resolution Greenland ice core data show abrupt climate change happens in few years. *Science* 321, 680–684. doi: 10.1126/science.1157707

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., et al. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* 21, 1816–1820. doi: 10.1111/j.1365-294X.2011.05317.x

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., 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

Wohlfarth, B., Blaauw, M., Davies, S. M., Andersson, M., Wastegård, S., Hormes, A. et al. (2006). Constraining the age of lateglacial and early Holocene pollen zones and tephra horizons in southern Sweden with bayesian probability methods. *J. Quaternary Sci.* 21, 321–334. doi: 10.1002/jqs.996

Wohlfarth, B., Muschitiello, F. L., Greenwood, S., Andersson, A., Kylander, M., Smittenberg, R. H., et al. (2017). Hässeldala - a key site for Last Termination climate events in northern Europe. *Boreas* 46, 143–161. doi: 10.1111/bor.12207

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., et al. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* 21, 3647–3655. doi: 10.1111/j.1365-294X.2012.05545.x

Zimmermann, H. H., Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schirrmeister, L., Schwamborn, G., et al. (2017). The history of tree and shrub taxa on Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial uncovered by sedimentary ancient DNA and pollen data. *Genes (Basel)* 8:273. doi: 10.3390/genes8100273

Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., et al. (2016). Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. *Soil Biol. Biochem.* 96, 16–19. doi: 10.1016/j.soilbio.2016.01.008

Conflict of Interest Statement: 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.

Paper V

Last Glacial Maximum environmental condition of Andøya, a northern ice-edge ecological "hotspot"

Inger G. Alsos[1], Per Sjögren[1], Antony G. Brown[1,2], Ludovic Gielly[3], Marie Kristine Føreid Merkel[1], Aage Paus[4,5] , Youri Lammers[1], Mary E. Edwards[1,2], Torbjørn Alm[1], Melanie Leng[6], Tomasz Goslar[7], Catherine T. Langdon[2], Jostein Bakke[5,8], Willem G.M. van der Bilt[5,8]

1) UiT – The Arctic University of Norway, Tromsø Museum, N-9037 Tromsø, Norway
2) University of Southampton, Department of Geography and Environment, SO17 1BJ Southampton, UK
3) Laboratoire d'Ecologie Alpine (LECA), Université Grenoble Alpes, C2 40700 38058, Grenoble, Cedex 9, France
4) University of Bergen, Department of Biological Science, N-5020 Bergen Norway
5) The Bjerknes Centre for Climate Research, Bergen, Norway
6) British Geological Survey, Centre for Environmental Geochemistry, NG12 5GG Nottingham, UK
7) Faculty of Physics, Adam Mickiewicz University, Poznan, Poland
8) University of Bergen, Department of Earth Science, N-5020 Bergen, Norway

Highlights

- Unusually high organic content of LGM core from northern refugium
- Nitrogen and carbon isotopes together with bird bones indicate local bird cliff
- DNA of pine and spruce found in all three cores, but signal not consistent
- Review of previous Andøya LGM studies shows a mixture of boreal and arctic taxa
- Environmental reconstruction does not preclude growth conditions for trees
- Currently a non-analogue situation but could presage future conditions further north

Abstract

Andøya on the NW coast of Norway is a key locality for understanding Last Glacial Maximum (LGM) in northern Europe. Controversy has arisen concerning the local conditions, especially about the timing and extent of local glacial cover, maximum July temperatures and whether pine and/or spruce could have grown there. We reviewed all existing data and analysed ancient sedimentary DNA, pollen, macrofossils, geochemistry and stable isotopes from three new lake sediment cores from Øvre Æråsvatnet. The longest core covers the period 23.4-8k cal. BP (Bayesian model) and possibly 26.7 k cal. BP. A total of 23 new dates and age-depth modelling suggests the lake has been ice-free since GI2 (<22.8 k cal. BP) and possibly GS3 (<27.4 k cal. BP). *Pinus* and *Picea* DNA was found in all three cores but at such low frequencies that it could not be distinguished from background contamination. LGM samples had unusually high organic matter content (Loss On Ignition 60-70%), with isotopic values indicating that they are derived from a marine source. Along with finds of bones of the little auk (*Alle alle*), this indicates that the lake received guano from an adjacent bird colony. DNA, pollen and macrofossil assemblages were dominated by Poaceae, Brassicaceae and *Papaver*, but scattered occurrence of species currently restricted to the Low Arctic Tundra Zone (July temperature of 8-9°C) such as Apiaceae (DNA, 8-9°C), *Betula* tree-type (pollen, 10-12°C), and *Alchemilla alpina* (macrofossil, 8-9°C) were also recorded. The review showed that 94 vascular plant taxa had been recorded, of which 38% have a northern limit in Shrub Tundra further south. This unusual assemblage likely stems from a combination of proximity to ice-free water in summer, geographical isolation linked with stochastic long-distance dispersal events, and the presence of bird-fertilized habitats. The environmental reconstruction based on all records from the area does not preclude local growth of tree species as the local climate combined with high nutrient input, may have led to periodically suitable environmental 'hotspot' conditions.

Key words: ancient DNA (aDNA), Andøya, climate variability, environmental conditions, glacial survival, last glacial maximum, late Weichselian, MIS2, refugia, sedimentary DNA (sedDNA)

1. Introduction

Evidence for cryptic glacial-age refugia in northern hemisphere has long been sought but remains elusive (Birks and Willis, 2008; Brochmann et al., 2003; Stewart and Lister, 2001; Tzedakis et al., 2013), despite the fact that some phylogenetic data indicate that they must have existed (Anderson et al., 2006; Fedorov and Stenseth, 2001; Westergaard et al., 2019). While it seems clear that late-glacial tundra zones supported small populations of boreal trees in Alaska (Brubaker et al., 2005), Yukon (Zazula et al., 2006), Siberia (Binney et al., 2009; Tarasov et al., 2009), and Estonia (Heikkilä et al., 2009), it is debated if pollen records at Bodmin Moor (SW England) represent local growth of trees (Kelly et al., 2010; Scourse, 2010). Similarly, it is questioned if macrocharcoal of oak, beech and pine indicate survival in the Hartz Mt., central Germany (Giesecke, 2016; Robin et al., 2016). It remains even more controversial, however, if tree taxa grew within the maximum limits of the Scandinavian ice sheet (Birks et al., 2005; Kullman, 2005) as indicated by megafossils of spruce and pine in the Scandinavian mountains (Kullman, 2002), and DNA in lake sediments from a glacial refugium at Andøya (Birks et al., 2012; Parducci et al., 2012a; Parducci et al., 2012b). As all proxies for reconstructing past flora and environmental conditions have some uncertainties, a multi-proxy study may provide better answers. The focus on environmental reconstruction is often on temperature (Birks and Birks, 2014; Trondman et al., 2015). However, the effect of temperature on vegetation can be indirect, as via nutrient cycling, and high nutrient levels may compensate for low temperature, as seen, for example, at high-latitude bird cliffs (González-Bergonzoni et al., 2017). Thus, estimation of nutrient availability and isotope analyses to assess trophic status may further elucidate the environmental conditions in refugia.

The northern Norwegian island of Andøya has long been a key locality for understanding LGM environments (here and throughout considered to be diachronous within the time window ca. 26-18 k cal. BP) on the North Atlantic margin (Vorren, 1978), and it has been studied on many occasions. Andøya is situated where the Norwegian continental shelf is at its narrowest. The result of this geomorphology is that this area is a key fixed point for the maximum extent of the Eurasian Ice Sheet (Hughes et al., 2016; Patton et al., 2017). While it is clear that higher elevations on northern parts of Andøya remained ice-free throughout the LGM, it has been less clear whether the lowland was continuously ice free from ca. 26 k ca. BP (Alm, 1993; Vorren and Plassen, 2002; Vorren et al., 2015).

Palaeobotanical investigations have been carried out on three lakes on the northern ice-free tip of Andøya (Fig. 1): Endletvatn (Alm and Elverland, 2012; Parducci et al., 2012b; Vorren, 1978; Vorren et al., 2013), Nedre Æråsvatnet (Alm and Birks, 1991; Vorren et al., 1988) and Øvre Æråsvatnet (Alm, 1993). The recorded Late-Glacial vegetation, combined with low, minerogenic sedimentation, has been interpreted as being grass-dominated and typifying cold and dry polar desert conditions (*sensu lato*). There may have been interruptions by periods of warmer climate when mean July temperatures reached up to 10°C, as indicated by features such as high concentrations/accumulation rates of pollen and/or macrofossils and the occasional presence of more thermophilous plant taxa (Alm, 1993; Alm and Birks, 1991; Elverland and Alm, 2012; Vorren, 1978). Recently, sedimentary ancient DNA (sedaDNA) of pine and spruce of LGM age has been found (Parducci et al., 2012b). The discovery of conifer aDNA on Andøya was unexpected, and it was debated as to whether the origin was due to contamination, long-distance pollen, driftwood, re-sedimentation, or possibly in-situ growth (Birks et al., 2012; Parducci et al., 2012a; Parducci et al., 2012b). The issue has been further polarized by Vorren *et al.* (2013) who, based on both new and re-interpreted past data, concluded that LGM mean July temperatures never exceeded 3°C. However, this was based mainly on the inference that the combination of the dominant moss species *Syntrichia ruralis* and *Aulacomnium turgidum* represent Polar Desert vegetation, and it contradicts previous palaeoecological interpretations of past climatic conditions (Alm, 1993; Alm and Birks, 1991; Elverland and Alm, 2012; Parducci et al., 2012b; Vorren, 1978). Such a radical interpretation would preclude any tree growth during the LGM on Andøya. It follows that the chronology, environmental conditions and palaeoecology of N Andøya warrant further clarification.

This study seeks to capitalize on recent advances in both the methodology and understanding of sedimentary sedaDNA to: **1**) determine the duration of the ice-free period, **2**) evaluate the sedaDNA results from Parducci et al. (2012b) by investigating a second lake on Andøya using improved methods; **3**) assess the local LGM palaeoenvironment based on additional proxy records, including stable isotopes, and **4**) review previous investigations, with special emphasis on environmental conditions and the potential for tree growth on Andøya during LGM.

2. Regional setting

At Andøya (Fig. 1), the proximity of the continental shelf to abyssal depths (6.1 km coast to the top of the Andøya Canyon (Laberg et al., 2000) limited the extent of ground-based glaciers during the low sea levels of the Weichselian glaciation, and it was suggested that the island became deglaciated early (Vorren et al., 2015). This is in line with recent work on continental glaciation that suggests that topography/trough geometry had an overriding effect on glacial extent and recession rates (Small et al., 2018). The tip of Andøya is also bisected by moraines that have been correlated with the earliest moraine (Egga I), which marks the shelf edge (Vorren and Plassen, 2002), and Egga II, which has been dated to 23-22.2 k cal. BP (Vorren et al., 2015).

The presence of lakes that may have received sediment input during this time has made the northern tip of the island an important site for palaeoenvironmental studies of the LGM. One of these sites is Øvre Æråsvatnet (69°15'22''N; 16°02'03''E), the lake targeted for this study. The basin sits at 43 m a.s.l, just above the local marine limit. At present, the lake receives water from an inlet to the SW and is drained by an outlet to the NE (Fig. 1). Mean temperature in July is 11°C, mean temperature in February -2.2°C, annual mean temperature 3.6°C and annual precipitation 1060 mm (Norwegian Metrological Institute; eKlima 2016; 1961–1990, station 87110, 10 m a.s.l., Andøya). The lake area is 20.4 ha, an area similar to that of two adjacent lakes, Nedre Æråsvatnet (20.6 ha, 34 m a.s.l.) and Endletvatn (28.6 ha, 35 m a.s.l.; Fig. 1). The bedrock is entirely non-calcareous and the catchment size is 3.6 km^2.

3. Material and methods

3.1. Field work

Lake Øvre Ærsåsvatnet was chosen as, in contrast to Endletvatn (where the previous conifer DNA was detected) it was never below the marine limit, and thus the chance of driftwood having entered the lake catchment is low. Fieldwork was conducted early in March 2014. A north-south transect across the lake was test-cored using a Hiller sampler (And1-And7, Fig. 2). The laminated gyttja of expected LGM age (Alm, 1993) was only found in samples from the shallower, south-central part of the lake. To further assess sediment distribution and water depth, we surveyed the basin with Ground Penetrating Radar (GPR). For this purpose, we used a Malå ProEx unit fitted with a shielded 50 MHz antenna. The system was fitted in PVC tube and dragged behind a snow scooter driving at constant speed and traversing the lake in

semi-regular grids. Following acquisition, all data were processed in version 1.4 of the RadExplorer software package with a set of prescribed bandpass filtering, DC removal and time-zero adjustment routines. We then traced the interfaces between water, sediment and bedrock based on these optimized GPR reflections. Finally, all data were exported to ArcMap 10.4 to construct maps and models. Coring was conducted with a Geonor piston corer (110 mm diameter) and a modified Nesje corer (110 mm; (Paus et al., 2015). Only the deepest part of the lake-sediment package was collected. A 3.46 m continuous core was retrieved with the modified Nesje-corer (And-11, 69.25579°N, 16.03517°E; 3.15 m water, total depth 11.88 m; subsequently divided in three parts). Two shorter cores taken with the Geonor corer (And-8, 3.15 m water, total depth 11.98 m, core length 1.14 m; And-10, 69.25552°N, 16.03516°E, 2.9 m water, total depth 10.16 m, core length 1.16 m) were collected 5 m west and 30 m south, respectively, from And-11 (Fig. 2).

3.2. Radiocarbon dating

The radiocarbon ages of 23 identified macroscopic plant remains were determined using accelerator mass spectrometry at Poznan Radiocarbon Laboratory (Table 1). The bulk of the dated material consisted of bryophytes (moss stems), but seeds and leaf fragments of vascular plants were included when available. Low mass (TOC $\leq$ 0.2 mg) of many samples led to relative large error intervals, but this was preferred to combining material. The ^{14}C ages were calibrated using with OxCal 4.2 (Bronk Ramsey, 2009) using IntCal13 (Reimer et al., 2013), and an age-depth modelling was undertaken using Bacon v. 2.3.9.1 (Blaauw and Christen, 2011).

3.3. Geochemical analyses

Colour line-scan images with a resolution of approximately 70 µm were acquired using a Jai L-107CC 3 CCD RGB Line Scan Camera installed on an Avaatech XRF core scanner (Fig. 2). Qualitative element-geochemical analyses were carried out with the Avaatech XRF core scanner. The measurements were carried out at continuous 10-mm steps. Instrument settings were 10 kV, 1000 µA, 10 seconds count time, and no filter. Data processing was performed using WinAxil version 4.5.6. The results are presented as ratios of selected elements divided by the sum of the 7 most abundant elements (Ca, Cl, Fe, K, S, Si and Ti; Rhodium (Rh) not included as it is induced by the equipment) to minimize the influence of water and matrix effects (Tjallingii et al., 2007; Weltje and Tjallingii, 2008). Loss-on-ignition (LOI) was analysed every 4 cm. About 10 g of sediment was dried overnight at 105°C, weighed, and

then burned for four hours at 550°C. LOI was calculated as the percent dry-weight loss after burning.

Thirty samples were selected from the And-11 core for $\delta^{13}C$ and $\delta^{15}N$ analysis and determination of %C and %N. The isotope analyses were conducted in the Stable Isotope Facility at the British Geological Survey, UK. Samples for carbon isotopes were decarbonated in 5% HCl prior to analysis (to remove inorganic carbon) while a separate aliquot for nitrogen isotopes was run with no pre-treatment. $\delta^{13}C$ analyses were performed by combustion in a Costech ECS4010 Elemental Analyser (EA) on-line to a VG TripleTrap (plus secondary cryogenic trap) and Optima dual-inlet mass spectrometer, with $\delta^{13}C$ values calculated to the VPDB scale using a within-run laboratory standard (BROC2) calibrated against NBS-19 and NBS-22. Replicate analysis of well-mixed samples indicated a precision of $\pm$ <0.1‰ (1 SD). Percent C and N analyses were run at the same time, calibrated against an Acetanilide standard. $\delta^{15}N$ analyses were performed by combustion in a Thermo Finnigan Flash EA (1112 series) on-line to a Delta Plus XL mass spectrometer. $\delta^{15}N$ was calculated to the $\delta^{15}N$ value of air using the BROC2 standard as a within-run laboratory standard calibrated against UGS40 and UGS41. Replicate analysis of well mixed samples indicated a precision of ±<0.2‰ (1 SD).

3.4 sedaDNA analysis

The DNA analyses of these sediments proved challenging, and we repeated the whole process three times. For the first extraction, we followed the phosphate buffer extraction protocol of (Taberlet et al., 2012). While this works well for modern soils samples, we had rather poor results, as also experienced for other ancient samples (Alsos pers. obs.), so this method might be less suited for extracting DNA from ancient sediments. We then tried the PowerMax extraction kit (MO BIO Laboratories, Carlsbad, CA, USA), a method that has worked well for other sediments (Alsos et al., 2016; Clarke et al., 2019). Here also the results were poor. We suspect that the main problem was the high organic content of the lower part of core And-11 and all of cores And-8 and And-10, as we have experienced similar problems with other sediments with high organic content (Clarke et al., 2018); further, the top, less organic part of core And-11 gave reasonable results in all three analyses. During this period, the Tromsø Museum dedicated ancient DNA laboratory has been moved twice and between buildings and all reagents have been replaced, so the three runs exhibited different background contamination levels. Here we only present data from the third analyses, but we note that we

did detect pine (*Pinus*) and spruce (*Picea*) in all three runs, both in samples and in negative controls, but not always in the same samples.

For the third extraction, we used sterile plastic tools for taking 74 samples from the three cores. We extracted DNA using DNeasy PowerSoil Kit (Qiagen) with a minor modification (Heintzman et al., in prep.). We included negative controls during sampling (n=2), extraction (n=4), transferring extract from tubes to plates (n=2), PCR setup (n=2), and post-PCR (n=2), as well as a synthetic positive control (n=2), in total 14 controls.

For all three runs, the short and variable P6 loop region of the chloroplast trnL (UAA) intron (Taberlet et al., 2007) was used as diagnostic marker, following the same analysis protocol (Alsos et al., 2016; Sjögren et al., 2017), running 8 PCR replicates on each DNA extract. The PCR replicates were pooled, cleaned and quantified with Qubit (Invitrogen™ Quant-iT™ and Qubit™ dsDNA HS Assay Kit, Thermofisher). The pool was converted into a DNA libraries using a Truseq DNA PCR-free low throughout library prep kit (Illumina). The library was quantified by qPCR using the KAPA Library Quantification Kit for Illumina sequencing platforms (Roche) and a Prism 7500 Real Time PCR System (Life Technologies, Fisheries faculty, UiT). The library was normalised to a working concentration of 10 nM using the molarity calculated from qPCR adjusted for fragment size, and were sequenced on an Illumina HiSeq 2000 platform (2x150 bp, mid-output mode, dual indexing) at Genomics Support Centre Tromsø (UiT).

All next-generation sequence data were aligned, filtered and trimmed using the OBITools software package (Boyer et al., 2016); http://metabarcoding.org/obitools/doc/index.html) using similar criteria to Alsos et al. (2016) and Sjögren et al. (2017). Resulting barcodes were assigned to taxa using the *ecotag* program (Yoccoz, 2012) and two independent reference datasets. One reference contained regional arctic and boreal sequences (Soininen et al., 2015; Sønstebø et al., 2010; Willerslev et al., 2014) and the other the NCBI nucleotide database (January 2018 release). The resulting identifications were merged and filtered, retaining barcode sequences if 1) they were identified to 100% in either reference set, 2) they were present in at least 3 PCR repeats and 3) had at least 10 reads across the entire dataset. Furthermore, each observation required 3 or more reads in order to be kept in. We removed the likely false positives relating to common PCR errors and food contaminants, based on experience from the analyses of 15 other sediment cores in Tromsø Museum, as well as taxa

identified above family level (Supplementary Table 1). For the last step of filtering, we looked at frequency of PCR repeats in samples compared to negative controls. There is no clear way to set the cut-off (Alsos et al., 2018; Sjögren et al., 2017), and we chose a rather conservative value of only keeping sequences with an overall frequency of PCR repeats in samples at least twice as high as in negative controls. We present the data semi-quantitatively as proportion of PCR replicates where the taxa occur.

3.5. Pollen and macrofossils

Pollen was analysed for 19 samples from the And-11 core in the depth range 910-1182 cm. Pollen samples (1 cm^3) were prepared using the acetolysis method (Berglund and Ralska-Jasiewiczowa, 1986) and mounted in silicon oil for analysis at the Palaeoecological laboratory, University of Southampton. Left-over material from the DNA extraction 1 was used. As we resampled the core for DNA extraction 3, there might be minor stratigraphic difference between pollen and DNA samples. Approximately every second or third level analysed for DNA in extraction 3 corresponded to a pollen sample, except for 1038 cm, which was only analysed for pollen. Identifications were based on (Fægri and Iversen, 1989) and (Moore et al., 1991), in combination with a reference collection of modern material at the University of Bergen. A target sum of 350 terrestrial grains was only achieved in the two uppermost levels (910, 918cm) below that a sum of exotics was used, but in two levels (1122, 1142 cm) the concentrations were so low (<400 grains 1 cm^3) that only partial counts were possible.

Macrofossils were collected from 44 levels from core And-11 across the depth range 884-1181 cm. Approximately 2-cm thick slices were sampled at about every 8 cm (from half the core width, ca. 50 ml volume), thus often between samples for DNA/pollen. If necessary, the samples were soaked in 10% sodium hydroxide (NaOH) to disperse organic material and/or sodium pyrophosphate ($Na_4P_2O_7 * 10H_2O$) to disaggregate clay particles. The macrofossils were retrieved by gently sieving the sample using a 250-µm mesh. The herbarium and the macrofossil reference collection at Tromsø Museum were used to aid the identification.

3.6 DNA, pollen and macrofossil data analyses

Initial diagrams were plotted in R studio version 1.2.5001 using rioja, vegan and ggplot2 package. We explored zonation within each of the three proxies using the Constrained

incremental sum of squares (CONISS) as implemented in ggplot2 v3.2.1. Final diagrams were constructed using Tilia v.2.6.1 (TiliaIT.com).

3.7 Review of botanical records and reconstruction of minimum July temperature

All palaeo-records of species from Andøya for the period 26.7-14.7 k cal. BP were compiled from publications. The northern limits of vascular plants and vegetation types are closely linked to summer temperature (Karlsen and Elvebakk, 2003; Karlsen et al., 2005), and therefore the Arctic can be divided into bioclimatic zones representing mean July temperatures (Walker et al., 2005). We used the Pan-Arctic flora checklist (Elven et al., 2011) to assign species to the northernmost bioclimatic zones where they were (1) present or (2) frequent. If the palaeorecords were not determined to species level, the northernmost potential species in that taxon was used. Some taxa only identified to a higher taxonomic level were not classified, as ranges can be global (for example, Poaceae). The choice of classification inevitably introduces bias, and our choices here lead to opposing biases regarding the kind of environment we reconstruct. First, at their northern limits most species have small population sizes and pollen production is typically relatively low (Lamb and Edwards, 1988). Thus, a rare taxon has a low chance of occurring in palaeorecords, whereas frequent species are more likely to be recorded. Thus, alternative (1) represents a conservative (cold) estimate of the minimum temperature (results given in Supplementary Table S2), whereas alternative (2) represents a more mid-range alternative (results in main text). In contrast, the choice of the northernmost potential species in a group causes a strong bias towards more northerly (hence, cold) zones (for example, *Puccinellia* and *Ranunculus* occur in the polar desert zone, although the majority of species in these genera do not reach the High Arctic; (Elven et al., 2011). Thus, even alternative 2 is biased towards colder environment.

4. Results

4.1. Bathymetry, chronology, lithostratigraphy and geochemical analysis

The new bathymetry (Fig. 2) shows a irregular basin morphology with shallow sediments across the middle of the lake, and a bedrock or boulder mound. Given that the deepest sediments occur in a generally shallow area, an inversion of normal sedimentation patterns, it is unlikely that sediments have been rearranged by erosion. Rather, there may be a bedrock bar, or more probably a coarse boulder-dominated moraine. Whether a moraine or not, the

altitude of the basin, and its basal irregularity suggests disturbance of the basin in its early history. This is also suggested by the disturbed/tilted laminations in the basal sediments of Alm's (1993) core C.

The 23 ^{14}C dates, which are all on identified plant macrofossils, ranged from ca. 8 k cal. BP to 26.7 k cal. BP (Table 1). Several of the dates were based on small sample sizes (<0.1 mg), and some of these have dates older (Poz-77610) or younger (several) than expected. There was a clear hiatus in And-11 at 1089; the same hiatus was also observed in And-8 at 1130 cm (Fig. 2). We constructed an age-depth model for And-11 with and without specifying this hiatus (And-8 and And-10 have too few dates to allow a meaningful model for comparison). When including a hiatus at 1089 cm depth, 88% of the dates overlapped with the age-depth model at 95% ranges, compared to 75% of the dates without the hiatus. The addition of another hiatus at 994 cm was found not to improve the model or significantly alter the model. Thus, the model including a single hiatus at 1089 cm was preferred (Fig. 3). The modelled basal range and median was similar for the model with one hiatus (21,880-26,780, median 23,446 cal. BP) and without (21,828-26,732, median 23,366 cal. BP). However, there are good reasons to assume that the basal date (Poz-77656, median 26.7 k cal. BP) is accurate as it is in accordance with two bulk dates obtained by Alm (1993; T-8029A and T-8029B) date to 27,068-25,282 cal. BP and 26,069-25,541 cal. BP (at 2σ)). This date on moss suggests that, contrary to Vorren et al. (2015), Alm´s pre-22.0 k cal. BP dates cannot be dismissed just because they were bulk dates on gyttja. This strongly suggests that a lake existed, and the basin was deglaciated, at least in part of the Øvre Ærsåsvatnet basin, in the later part of GS-3 from ca. 26.7 k cal. BP; the oldest sediments, however, have been disturbed, probably during or very soon after deposition.

The scatter of ages in the lowermost part of the core requires explanation. Since samples were all individual terrestrial macrofossils and their ages are old, recent contamination is highly unlikely, but they could reflect re-worked material. If this is the case, it requires a local terrestrial source during sediment deposition, which implies a lack of glacial cover somewhere in the basin. The most parsimonious explanation is that the dates, the basal bathymetry, sediment depths and stratigraphic disturbance indicate the partial presence of ice in the lake during GS3-GS2 (27.5-17.2 k cal. BP), a situation that is highly likely given the location of the site adjacent to the postulated ice marginal limits (Vorren et al., 2015).

The new complete core (And-11, Fig. 4) is divided into 5 units, labelled U1 to U5, based on lithology. In the most parsimonious correlation, And-8 covers only part of U2, while And-10 covers U1-U3. Unless otherwise stated, the results are coherent for the three cores. Based on the combined dates of And-11 and And-8 the age interval of unit U2 can be estimated as ca. 23.2-17.2 k cal. BP. The hiatus occurs between units U2 and U3, and the sediments from the period ca. 17.2-16.2 k cal. BP are missing. The most important aspect of the lithology is the occurrence at 1178-1017 cm depth (And-11) of a laminated gyttja that is markedly different from typical glacial clays but was deposited during the LGM.

Geochemical analyses (C, N, ^{15}N, ^{13}C and XRF) were done on the And-11 core whereas LOI was done on all three cores. The most striking feature in the geochemical analyses is the exceptionally high organic content in these MIS2 sediments from ca. 23.2 k cal. BP onwards (Figs. 2 and 4).

The results from the geochemical analyses reveal three major trends (Fig. 4): Firstly the organic content and associated elements (LOI, C%, N%, C/N, ^{13}C, ^{15}N, S, Cl and Ca) reach high values in unit U2 and U3 (Fig. 4). All values, with exception of C/N and ^{13}C, show a distinct drop in the first half of U3, i.e. U3a. The LOI (60-70%, Figs. 2, 4) and C (~50%, Fig. 4) reflect rapid deposition of organic material, which is exceptionally high for MIS2. S, Cl and Ca co-vary with the LOI and C%, with exception of an increase in Ca in unit U5b. Secondly, a contrasting trend is seen in K, Ti and Fe. In U1 to U4 these elements are negatively correlated with the organic content, and they are interpreted as representing material eroded from the catchment. The third trend is in Si. This element is also negatively correlated with the organic content, but in contrast to K, Ti and Fe, it increases markedly in U5. This likely signifies erosion of base depleted soils (Boyle, 2007). Low LOI combined with high K, Ti and Fe in U3–U4 indicate mineral soil depletion, suggesting at least seasonally, temperatures above 0°C, limited soil formation and erosion.

The unusually high values of $\delta^{13}C$ (-16 to -11) and $\delta\ ^{15}N$ (18 to 22) in units U2 and U3b (Fig. 5) indicate that the organic material is derived from a high trophic level and it is well outside the normal values for temperate, boreal or arctic lakes (Gąsiorowski and Sienkiewicz, 2013; Osburn et al., 2019; Thompson et al., 2018). Combined with the other elements this confirms that most of the organic material is derived from sea-bird faeces (guano) a suggestion originally made by Alm (1993). The C/N ratio is interpreted as reflecting this preservation of

organic matter with moderately high N prior a reduction of both elements in the Late Glacial and into the early Holocene.

4.3 Ancient DNA record

From 74 samples we obtained in total 22,888,821 raw reads, of which 1,707,668 reads of 45 sequences passed the filtering criteria of our pipeline (Supplementary Table S3). Two sequences matching *Vaccinium myrtillus/vitis-idaea* and three matching *Ranuculus reptans* were assumed homopolymer variants, and only the most frequent sequence was kept. Also two sequences matching pine (*Pinus*) were found. *Pinus1* was found in 19 samples (3.1% of the repeats of all samples) whereas *Pinus2* was only found in six PCR repeats at 894 cm depth in core And-11. The following taxa were found in negative controls after filtering pipeline: spruce (*Picea,* one repeat in each of three negative controls from sampling and extraction), and one PCR repeat of each of *Betula, Pinus1,* Poaceae and Brassicaceae. The frequency of PCR repeats of these taxa was lower in samples than in negative controls for *Picea* (0.83) and Brassicaceae (0.26), and these taxa were therefore excluded, whereas the frequency of *Pinus1* and *Betula* were 2.7 and 4.6 times higher in samples than in negative controls, respectively, and therefore kept in the dataset (Table S4). All 39 taxa, including negative controls, are presented in Supplementary Table S4, whereas the 37 assumed true positive taxa are included in Figs. 7-8. All taxa were found in core And-11. Cores And-8 and And-10 contained each eight taxa (Supplementary Table S4). The majority of the taxa were identified to a taxonomic level that allowed classification according to bioclimatic zones (Supplementary Table S5).

The diversity was generally low in samples older than 12.7 k cal. BP, with 0-5 taxa per sample, but it increased to 13-22 taxa in more recent samples. The highest read abundances were found for *Pinus, Ceratophyllum demersum, Myriophyllum alterniflorum,* Poaceae*,* and the algae *Nannochloropsis* spp (all > 90,000 reads). However, a more conservative estimate of DNA quantity is the number of PCR repeats, and here *Nannochloropsis gaditana* and *Nannochloropsis* sp. were by far the most dominant with 307 and 301 repeats, respectively, compared to 41 of *Caltha palustris,* 40 *Myriophyllum alterniflorum*, 39 Poaceae, and 37 PCR repeats of *Betula*.

4.4 Pollen record

In total 60 pollen and spore types were identified in the 19 samples (see raw pollen and spore count in Supplementary Table S6). As is typical with cold-period spectra, the concentrations

were low - ranging from 12,500 to 67 grains cm^3. No pollen was found in the basal diamicton. Also, three samples contained only 1-3 pollen grains and only in the three uppermost samples, more than 75 pollen grains were identified. There were less than 5 taxa in each sample older than 14.2 k cal. BP and up to 29 in each sample from 12.7 k cal. BP and more recent. The most common taxa were Pteropsida (496), Poaceae (303), *Betula*-tree type (165), and *Gymnocarpium dryopteris* (120). We tentatively classified 46 of them to bioclimatic zones (Supplementary Table S5). The conventional pollen diagram shows a typical increase in richness and thermophilous types from U2-3 into U4-5 (Supplementary Fig. S1). This is also reflected in the pollen concentration rates.

4.5 Macrofossils record

The 44 macrofossil samples included 503 records of 19 taxa/types of vascular plants, bryophytes, insect fragments, *Daphnia* ephippia and *Chara* oospores, with taxa mostly identified to species or genus level (raw counts in Supplementary Table S7). For the majority of samples, 0-3 taxa of vascular plants were found, with 4-6 taxa for the three topmost samples. Bryophytes were found in all samples (mostly with <50 fragments) and insect fragments in all except the lowermost (>1000 in many samples). Other abundant types of macrofossil were Poaceae (148 seeds), *Papaver* (110 seeds), and Brassicaceae *Draba*-type (15 seeds). There was a clear turnover in the macrofossil record from a dominance of *Papaver,* Poaceae and Brassicaceae until 14.2 k cal. BP, to *Salix* and *Saxifraga* to around 10.5 k cal. BP when *Betula* and aquatic dominated (Supplementary Fig. S2).

4.6 Combined vegetation zones

The CONISS analyses suggested five periods for each of DNA, pollen and macrofossils, but zone boundaries differed in age/depth (Fig. 6). The only boundary identified in all three proxies, and also in the lithology, was at ca. 1018 cm depth (14.2 k cal. BP, range 13.9-14.6 k cal. BP). This is close to what is generally seen as the end of GS-2.1a (14.7 k cal. BP, Rasmussen et al. 2014), so we use this as a major boundary. Zonation before 14.2 cal. BP is based on few taxa in each of the records and thus not robust. Therefore, we keep this as one zone. After 14.2 k cal. BP, there is a step-wise zonation with first a boundary in macrofossils around 12.8, pollen at 11.7 and 10.7, and then DNA at 9.7 k cal. BP (Fig. 6). Below, we discuss the two major zones and their minor zonation. Cores And-8 and And-10 are both within zone 1.

4.7 Zone 1 DNA, pollen and macrofossils 24.0-14.2 cal. BP (1182-1018 cm)

The number of taxa found per sample was low for DNA (1-4), pollen (1-4), and macrofossil (2-6, Fig. 7). Poaceae is a common taxon in all three proxies with concentration of 67-1900 and 1-22 in the pollen and macrofossil records, respectively. *Papaver* was recorded in all three proxies, but mostly as macrofossils where it was found in all samples with concentration of 1-11 seeds per cm^3. Poaceae and *Papaver* were also recorded in And-8 and And-10 (Supplementary Tables S4). Brassicaceae (*Draba* type) was found in about half of the pollen and macrofossil samples, whereas the one record in DNA was filtered out (see above and Supplementary Table S4). Bryophytes were found in all macrofossil samples. Although they are regularly detected in aDNA studies and were found in four younger samples at this site, they were not detected in aDNA samples <14.2 k cal. BP. This may be because the DNA record was "swamped" by algae (*Nannochloropsis* sp., *N. gaditana*, and for one sample in And-8 also *N. granulata*).

Pinus DNA was found in three sample of core And-11 (Fig. 7) and one and two samples in And-8 and And-10, respectively (Supplementary Table S4). It was also found as single grains in each of two pollen samples (Supplementary Table S6), but not the same as the DNA record (Fig. 7). *Salix* was found in a single DNA sample (And-10, Supplementary Table S4) and as one pollen grain (Fig. 7). Other woody taxa recorded were *Betula*-tree type (single grain), *Quercus* (single grain in two samples), and *Sorbus* (uppermost DNA samples of this zone). DNA of Apiaceae (most likely *Angelica archangelica*) was found in all three cores (few PCR repeats). Single records were also found of *Aster* sect. *Aster* (pollen), *Potentilla* (pollen), *Potamogeton* cf. *grammineus/alpinus* (DNA) and a macrofossil of cf. *Alchemilla alpina* (Fig. 7, Supplementary Tables S4, S6, and S7). The number of *Nannochloropsis* repeats drops between 16.2-15.2 k cal. BP, which is also when the sediments show a drop in LOI, C and N isotopes (Fig. 4-5).

Insect fragments are found in all except the lowermost samples and increase from around 15.5 k cal. BP. Also from ca. 15.0 k cal. BP, *Daphnia* ephippia rapidly become abundant (Fig. 7). Although generally eurythermal, most *Daphnia* species require a minimum water temperature of 10ºC (Clare 2001). A bone identified to little auk (*Alle alle*) was found at 1178 cm in And-8 (ca. 22.2-21.0 k cal. BP), and a similar bird bone was found at 1004 cm in And-10.

4.10 Zone 2 DNA, pollen and macrofossils 14.2-8.2 k cal. BP (1018- 850 cm)

Pinus is still scattered in both DNA and pollen record, and sometimes in the same samples (Fig. 8). *Salix* becomes common, first in macrofossils from around 14.2 k cal. BP, and soon after in pollen and DNA (Fig. 8a). Note that there are no samples analysed for pollen and macros in the top sediments (Fig. 6), so *Salix* probably remains common throughout the period as seen in the DNA record. Poaceae is still common in DNA and pollen, but almost disappears in macrofossils (only 2 single seeds after 14.7 k cal. BP). Also, *Papaver* and Brassicaceae (*Draba* type) almost disappear. On the contrary, some new forbs appear, e.g. *Oxyria* (pollen), *Ranuculus glacialis* (pollen), and *Saxifraga* spp. (pollen and macrofossils), suggesting an arctic tundra flora (Fig. 8b). *Salix* was not identified to species, but given the other species in the assemblage, it likely represents dwarf shrubs such as *S. herbacea, S. polaris*, and/or *S. reticulata*. *Salix* is rare in bioclimatic zone A, the Polar Desert Zone (Walker et al., 2005), so conditions must have been warmer than that. The pollen assemblage also includes some more "southern" forbs such as *Artemisia* (common in low-arctic tundra but potentially long-distant dispersed) and the boreal taxa *Rumex* and *Ranuculus acris* from around 13.2 k cal. BP, suggesting an increase in temperature around that time.

Nannochloropsis taxa show a clear drop from 14.2 k cal. BP, with only scattered occurrence after this time, whereas *Chara* oospores occur in every macrofossil sample from 12.2 k cal. BP upward 9.5 k cal. BP. This suggests a limnological change.

From 12.9-12.8 k cal. BP, there is a short-lived appearance of several bryophyte taxa in the DNA, reflected also in higher abundance of bryophyte macrofossils (Fig. 8c), and a band of bryophytes in the lithology (U4b; Figs. 4 and 6). The single pollen sample from this interval shows a drop in pollen concentration (Table S6). Soon after, Apiaceae re-appears, as well as new taxa with current distribution north to the Low Arctic Tundra Zone such as *Thalictrum,* Caryophyllaceae (*Arenaria* type), *Betula nana* and *Gentianella,* suggesting a transition to Low Arctic or Shrub Tundra.

From 11.7 k cal. BP the onset of the Holocene, the first ferns occur, whereas a larger change is from around 10.7 k cal. BP, when *Betula* is common in all three proxies and identified to *Betula* tree type in pollen. At the same time, *Filipendula ulmaria* appears in all three proxies, along with many ferns suggesting a tall shrub birch forest. Also the aquatic flora is rich, as seen by the abrupt appearance of a range of aquatic taxa: *Calta palustris, Isoetes, Menyanthes*

trifoliata, Myriophyllum alterniflorum, Potamogeton spp*., Sparganium* and *Subularia aquatica* (Fig. 8c).

Pinus is found scattered in both pollen and DNA, with the highest concentration around 12.1 to 10.0 k cal. BP (Fig. 8a). *Picea* is also found in a total of 7 of the 23 DNA samples covering the period after 14.2 k cal. BP (Supplementary Table S4), but note that we suspect these might be false positives.

4.13 Synthesis of Andøya plant and animal record >14.7 cal. BP

When taken together the DNA, macrofossils and pollen show that the environment was botanically rich. In total 94 vascular plant taxa have been recorded from studies of pollen (77), macrofossils (19) and aDNA (9) and a megafossil of *Betula pubescens* from nearby Stavdalen (ca. 20.4 k cal. BP, Kullman 2006) (Supplementary Table S2, S8). The largest quantities both in this and previous records are of Poaceae, followed by Brassicaceae and *Papaver*. Poaceae may potentially include *Bromus, Festuca, Phippsia algida* and *Puccinellia.* Brassicaceae may include *Braya*-type, *Cardamine nymannii, Cochlearia,* and *Draba*-type. Vorren (1978) notes that the pollen indicates two different taxa of *Papaver*, whereas Alm and Birks (1991) note that the variation within *Papaver* seeds falls within *P. radicatum* s.lat. At least 38% of all recorded taxa have a current northern limit in the shrub tundra and southwards (July temperatures 10-12°C, Supplementary Table S2). However, we note that the majority of these are only found as occasional pollen types and may represent long-distance pollen. However, also some taxa found in the macrofossil and/or DNA record have their currently northern limit in the Shrub Tundra or Boreal Zone (Table 3, Supplementary Table S2).

Further taxa identified during the LGM and early late glacial include caddisflies and chironomids (23.5 k cal. BP, Øvre Æråsvatnet, (Solem and Alm, 1994); Nedre Æråsvatn – mainly 16.9 k cal. BP onwards (Alm and Willassen, 1993), Endletvatn, 22-14.7 k cal. BP (Elverland and Alm 2012), and the beetle *Diernerella filum* (ca. 18-17 and 15.5 k cal. BP, Endletvatn, Elverland and Alm 2012). The records of taxa indicating warmer conditions largely coincide with periods of higher pollen and macrofossil concentrations (Alm, 1993; Parducci et al., 2012a). The caddisfly *Apatania zonella* is a continental species with a distribution extending east of the LGM limit to the Urals (Fauna Europaea) although it could have survived the LGM at Andøya (Solem and Alm, 1994). *Dienerella* is cosmopolitan genus,

with its current Norwegian distribution restricted to a few sites in the south and one in the north. It is associated with rotting wood and musty fruit bodies of soil fungi, but it is also found in arctic tundra (Elverland and Alm 2012, https://www.artsdatabanken.no/).

Furthermore, there are records of little auk dated ca. 20-15 k cal. BP (Alm and Elverland, 2012), eider duck (*Somateria* sp.) ca. 17.3 k cal. BP (Vorren et al., 1988), and stoat (*Mustela ermine*) in the interval 20.1-21.4 k cal. BP (Endletvatn, (Fjellberg, 1978), re-dated by (Vorren et al., 2013)).

5. Discussion

5.1 The LGM and Glaciation of Andøya

Our new dates confirm that Øvre Æråsvatnet contains sediments older than 22 k cal. BP and add another date that is over 26 k cal. BP. The earliest period of sedimentation, which corresponds to the later part of GS-3, is disturbed, probably due to ice and the lake's location at the glacial margin. However, these new dates confirm that the north tip of Andøya cannot have been glaciated after ca. 26 k cal. BP. This is also in agreement with three cosmogenic dates from Store Æråsen (105 m a.s.l.; (Nesje et al., 2007)) which is adjacent to the lake (Fig. 1). Although the cosmogenic data are cited in Vorren et al. (2013), the authors suggested that the dates do not exclude non-erosive, cold-based glacial ice. However, our data showing an open lake with surrounding vegetation and a bird cliff does. Vorren et al. (2013) also regard the earlier pre-20 k cal. BP dates of Alm (1993) as potential due to reworking. Given the radiocarbon dates provided here, 7 of which pre-date 20 k cal. BP, all based on macrofossils, reworking seems even more unlikely as does the presence of cold-based glacial ice after 26 k cal. BP. The complex bathymetry and disturbance of the basal sediments in Øvre Æråsvatnet prior to 16 k cal. BP suggest ice melt within the lake, as has been observed on the floor of glaciated lakes in southern Norway (Eilertsen et al., 2016), or paraglacial disturbance (Ballantyne, 2002). These dates therefore constrain the ice at the LGM on this northern tip of Andøya to a brief period after the Ålesund Interstadial (38-35 k cal. BP; (Mangarud et al., 1981)) to 26 k cal. BP, or, as suggested by (Mangerud, 2003), the existence throughout the last glacial cycle of an unglaciated refugium that included Røyken and adjacent peaks on northern Andøya. This also has implications for the local glacial sequence. Either the outermost Egga I moraine is earlier than ca. 26 k cal. BP, as originally argued by Vorren and

Plassen (2002) or it is possible that it represents a terminal moraine of a glacier in Andfjorden, the surface of which was just below Øvre Æråsvatnet. This is possible, as the Egga I moraine lies at -240 to -250 m b.s.l. and cannot be linked to an outer moraine further to the south, due to the presence of the Andøya canyon. However, it also follows that ice depositing Egga 2 cannot have covered the lake, and the only correlative moraines on the tip of Andøya (Kjølhaug, Endleten, and off shore, Bleik) are all to the east and below Øvre Æråsvatnet (Fig. 1). It therefore appears that during the LGM a very small area of Andøya, including Øvre Æråsvatnet, was an ice-free island bounded to the north by ocean (ice or water depending on season), to the east by the edge of the ice sheet, and protected to the south and west by nunataks.

Given the finds in the Sunnmøre caves of little auk, other seabirds, fox and reindeer (Larsen et al., 1987), the data now available, suggest that an ice-free corridor existed along the outer islands of Norway, with most areas being overrun during the LGM after 26 k cal. BP, except a small part of northern Andøya. In addition to the proximity of the continental shelf edge it is likely that a polynya existed here, as has been suggested for an apparent MIS 2 ice-free area off Svalbard (van der Bilt and Lane, 2019); this would be a critical asset for the large sea bird population suggested by our data.

5.2 Does sedaDNA of pine and spruce derive from locally growing trees?

Just as in the study of the nearby lake Endletvatn (Parducci et al., 2012b), DNA of pine and spruce was recorded. For the conifer DNA in Endletvatn, local growth was suggested, but alternative sources such as driftwood, reworked older material, DNA leaching, or pollen origin were discussed as well (Birks et al., 2012; Parducci et al., 2012a; Parducci et al., 2012b). Studies since then indicate that DNA leaching is not a problem (Clarke et al., 2019; Sjögren et al., 2017), as also confirmed by the high number of taxa observed in the Holocene layers, of which very few were recorded in samples older than 14.2 k cal. BP (Fig. 8). Similarly, pollen is not a likely source of chloroplast sedaDNA (Niemeyer et al., 2017; Parducci et al., 2017; Sjögren et al., 2017), but we are aware that there is less empirical evidence showing this for gymnosperms, where chloroplast is paternally inherited. However, as only few pine pollen grains were found, and these occurred in DNA samples with no pine, pollen is not a likely source of our pine and spruce DNA. The likelihood of driftwood is low at Øvre Æråsvatnet, which lies at a higher elevation and in which only lacustrine sediments have been recorded. Furthermore, if driftwood were the source, we would expect to see a

higher frequency of conifer DNA around 20.5 k cal. BP, when the nearby Endletvatn was transgressed by sea for a short period (Vorren et al., 2013; Vorren et al., 1988). In line with previous interpretation (Alm, 1993), our new dates and palaeorecord do not indicate any signs of reworked material. Thus, of all the suggested sources, we think we can rule out driftwood, pollen, leaching and reworked material. This leaves contamination or local growth as explanations.

In the previous record, contamination was not thought to be a source of the DNA as the result had been repeated in two sediment cores and three independent laboratories (Parducci et al., 2012a; Parducci et al., 2012b). Compared to that study, we have improved all parts of the methods (extraction protocol, negative controls at all steps, running eight rather than one PCR, unique tagging to minimize tag jumps, sequencing, and bioinformatics). Nevertheless, for low-frequency sequences, it is often impossible to distinguish between true and false positives (Alsos et al., 2018; Ficetola et al., 2015; Zinger et al., 2019). When comparing sedaDNA with local vegetation, a trade-off exists between retaining the true positive (true, according to the vegetation) and removal of the false positives. In a recent study of modern vegetation-DNA relationships, strict cut-off levels that removed the majority of false positives also removed 33% of the true positives (Alsos et al., 2018).

For ancient DNA studies, the issue of true and false positives is more challenging, due to difficulties in finding independent proxies that allow their identification. Authentication of ancient DNA via ancient damage patterns is one possibility, and this was tried. However, shotgun sequencing of two samples from Lake Øvre Æråsvatnet did not identify sufficient pine and spruce for aDNA damage pattern analysis (Lammers et al. In prep.). In other studies, scattered DNA records have been confirmed by macrofossils, for example, *Arabis alpina* in Svalbard (Alsos et al., 2016). Pine and spruce DNA are especially challenging, as they, like common food plants, can be part of background contamination potential due to wood or paper labels where the reagent are produced (Boessenkool et al., 2014; Clarke et al., 2018). In our laboratory, pine and spruce were detected in 1.62% (SD=1.08, range 0-2.50%) and 1.99% (SD=2.52, range 0-6.25), respectively, of the negative control PCR repeats (n=1360 PCR repeats of 170 negative control samples), for samples processed in the same period as this study. Thus, the frequency of pine (3.3%) is marginally above the range of background contamination in our lab, whereas the frequency of spruce (3.1%) in the three Øvre Æråsvannet samples was within that range, meaning that the DNA record is not robust in the

face of contamination. We note, however, that even low frequencies of spruce DNA may represent true positives, as a recent discovery of sedimentary ancient DNA in samples from the Polar Urals dated to ca. 21.0 and 18.0 k cal. BP (Clarke et al., 2019) has been confirmed by finds of two *Picea abies* stomata dated to ca. 20.4 and 18.8 k cal. BP (Anne Bjune pers. comm. 2019.). Clearly, there is more to learn about glacial survival of tree species.

As of the spruce record dated 10.3-6.3 k cal. BP found at Lake Rundtjørna, about 700 km further south (Parducci et al., 2012b), no new samples have been analysed from that region. However, as that study was based on a different DNA method that specifically targeted conifers and was focused on sediments that were considerably younger and closer in age to finds of megafossils in that region, there is no strong reason to reject the Rundtjørna spruce record.

5.3 The LGM and Late Glacial environment: a productivity hotspot?

The stable $\delta^{13}C$ /$\delta^{15}N$ values from the lower units of And-11 (U2-U3b) are highly unusual for lake sediments and indicate a marine nutrient source and high trophic level. An explanation for these anomalous values comes from the discovery of bones of little auk at this point in the core and at ca. 21.6 k cal. BP, which is within the period where numerous little auk bones were found at Endletvatn (ca. 20-15 k cal. BP, Elverland and Alm 2012), as well as bones of eider duck (*Somateria* sp.) at ca. 17.3 k cal. BP (Vorren et al., 1988). The occurrence of ground-nesting birds is further supported by the discovery of bones of stoat (*Mustela erminea*) at Endletvatn in the interval 20.1-21.4 k cal. BP (Fjellberg, 1978, re-dated by Vorren et al., 2013). Profundal sediments from high-arctic lakes in NW Greenland affected by marine-derived nutrients from little auk colonies show a 10-fold increase in $\delta^{15}N$ values over other sites ($\delta^{15}N$ 20.7 ± 2.4 SD), together with values for aquatic moss of 17.3 ± 5.8 and benthic algae of 17.9 ± 8.8 (González-Bergonzoni et al., 2017); these are similar to the $\delta^{15}N$ values of unit U2 (23.2-17.2 k cal. BP.) and unit U3b (15.1-14.2 ka cal. BP) at Øvre Æråsvatnet. The isotope data and the finds of little auk bones both in this study and by Alm and Elverland (2012) strongly suggest a substantial sea-bird colony existed in the area, and it is likely the lake and its surroundings were heavily manured. Furthermore, the bryophyte *Syntrichia ruralis* prefers alkaline substrates and high nitrogen (Vorren et al., 2013, http://www.arcticatlas.org). It has been estimated that marine derived nutrients from little auk colonies underpins more than 85% of terrestrial and aquatic biomass in affected areas (González-Bergonzoni et al., 2017). Such an increase in nutrient input would also greatly

700 increase primary production, which fits with the unusually high organic content of the
701 sediments for LGM and the dominance of the algae *Nannochloropsis* in the sedaDNA record.
702
703 Our DNA, pollen and macrofossil record for the period 24-14.2 k cal. BP was relatively
704 species-poor and dominated by very few taxa (Poaceae, Brassicaceae, and *Papaver*). Neither
705 fossil pollen assemblages nor modern plant communities provide close analogues. The
706 modern Arctic vegetation unit B1 (cryptogam-herb barren) of the Circumpolar Arctic
707 vegetation map (Walker et al., 2005) is most similar, but fossil pollen, sedaDNA and
708 macrofossil records from the northernmost bioclimatic zones show a diversity in the flora
709 rather than dominance by these three taxa (e.g. (Alsos et al., 2016; Bennike, 1999; Bennike
710 and Hedenäs, 1995; Birks, 1991; Hyvärinen, 1970). In our records, the high abundance of
711 *Papaver* probably reflects a relative lack of competition (Modin, 2016) combined with the
712 availability of favourable habitats, such as screes and gravel bed channesl. In the Arctic, high
713 values of Brassicaceae pollen are found in association with bird-manured soils (Rozema et al.,
714 2006; van der Knaap, 1988). Bird–manured arctic vegetation, combined with disturbed
715 ground habitats in other parts of the catchment, may thus be the closest modern analogue to
716 the LGM vegetation on Andøya.
717
718 The four phases of warmer temperature that have been observed in previous studies (Table 3,
719 (Alm, 1993; Alm and Birks, 1991; Elverland and Alm, 2012; Vorren et al., 2013; Vorren et
720 al., 1988)), are not pronounced in the current record (Fig. 7). The high percentages of algae
721 may have masked the occurrence of terrestrial taxa in the DNA record, as has also been
722 inferred based on modern studies (Alsos et al., 2018). The species-poor pollen and
723 macrofossil record in the current study may be due to less extensive analyses (fewer samples,
724 less pollen counted) than in previous studies, although counting more would not have
725 increased the pollen or macrofossil concentrations. If we assume that the DNA records of
726 *Picea* and *Pinus* are due to contamination and that pollen of *Betula* and *Quercus* is exotic, the
727 only thermophilic species prior to 16 k cal. BP in the current record is Apiaceae. The
728 sequence has 100% match to several species of *Angelica* and *Heracleum*, as well as
729 *Conioselinum tataricum* and *Podistera macounii.* Based on current biogeography and
730 northern limits of these taxa, the sequence most likely represents *Angelica archangelica*, a
731 species that can reach high abundances along bird cliffs today (Grønlie, 1948) and that has its
732 current northernmost limit in the Low Arctic Tundra (8-9°C July temperature) were it is
733 frequent (Elven et al., 2011). We note that the occurrence of Apiaceae is within the Andøya

734 thermomer 2 (Table 3), concurrent with Endletvatn DNA record (Parducci et al., 2012b), and
735 within the period of high nitrogen and carbon isotope values (Figs. 4-5) and high insolation
736 values (Berger and Loutre, 1991; Hughes et al., 2013). This is the period with strongest
737 evidence of climate amelioration, with the occurrence of several macrofossils with thermal
738 limit at around 8°C (Elverland and Alm, 2012; Kullman, 2006; Vorren, 1978) and *Urtica*
739 *dioica* which is rare in the Shrub Tundra (10-12°C) and common in the Boreal Zone (Elven et
740 al., 2011; Parducci et al., 2012b).
741
742 A few more thermophilic taxa occur in our record from 15.8-14.2 k cal. BP. We note that
743 these are species that are either common beneath bird cliffs (cf. *Alchemilla alpine*, *Angelica*
744 *archangelica*) and/or easily dispersed by birds (*Potamogeton*, *Sorbus aucuparia*), and their
745 occurrence is within lithological unit U3b that is rich in nitrogen and carbon (Figs. 4-5). Both
746 *Sorbus aucuparia* and *Potamogeton* are common north to the boreal zone (>12°C), but with
747 scattered occurrence in Shrub Tundra (10-12°C). For the same period, Vorren (1978) inferred
748 an oceanic climate with a temperature above 10° based on macrofossils of *Sphagnum*
749 *papillosum* as well as pollen of Apiaceae and cf. *Melampyrium*, an observation he later
750 disregarded as likely contamination or reworked material (Vorren et al., 2013). We note that
751 the period Vorren et al. (2013) recorded high frequency of the arctic/alpine bryophyte
752 *Aulacomnium turgidum*, ca 16.6-17.5 k cal. BP, falls within the hiatus in our core. Our core
753 may therefore not cover the coldest period. However, the earlier record of Apiaceae pollen
754 (Vorren, 1978) was confirmed by our Apiaceae DNA. This, as well as our other new record
755 and review of previous records, strongly suggest at least shorter periods with July
756 temperatures up to 10°C.
757
758 High nutrient input from birds may compensate for low temperatures (González-Bergonzoni
759 et al., 2017). In addition, the southeast-facing slope of Æråsen, which is the closest likely site
760 for such a bird cliff (Fig. 2), could have provided favourable microclimate as south-facing
761 slopes maximize sun exposure in the Arctic (Armbruster et al., 2007). Many bird cliffs in the
762 Arctic today are south-facing, highly productive environments with rare species
763 (svalbardflora.no). In addition, birds may also facilitate long-distance dispersal (Alsos et al.,
764 2015). Thus, the combination of south-facing slope, nutrient input, and bird dispersal may
765 have facilitated the presence of unusual (azonal) plant assemblages and allowed plants to
766 grow beyond their normal temperature limit.
767

Picea and *Pinus* are currently scattered and rare, respectively, in Shrub Tundra (10-12°C) (Elven et al., 2011). *Picea abies* has been reported to survive and occasionally even to produce viable seeds at mean July-August temperatures down to 5°C (Kullman, 2002) and seeds of both *Pinus sylvestris* and *Picea abies* germinate and may even have increased seedling survival above current treeline (Bougnounou et al., 2018). Nutrient-rich, south-facing slopes, cliffs providing wind protection, and enhanced summer degree-day sums related to local topography, could have been relatively favourable localities for tree establishment, at least during the warmest phases of LGM. Thus, the environmental conditions based on the entire available evidence does not exclude local growth of *Picea* and *Pinus*, at least during short warm phases, whereas *in situ* survival during the entire LGM seems unlikely.

5.4 Enrichment of the flora from 14.2 k cal. BP

The stepwise increase in floristic richness from 14.2 k cal. BP has also been observed in other records from the same lake and neighbouring lakes (Fig. 1), as well as other lakes studied in the region (Aarnes et al., 2012; Birks et al., 2014), and it represents the transition first to shrub tundra and then birch forest. The DNA record is especially rich in aquatic macrophytes and spore plants, including taxa not recorded in previous pollen or macrofossil records from the site, showing that this proxy may contribute to new knowledge even at one of the most studied palaeosites in NW Europe.

5.5 Glacial survival of plant taxa on Andøya

The presence of a *Papaver* seed in the lowermost sampled diamicton (ca. 26.7 k cal. BP) indicates that *Papaver* may have survived the last glaciation *in situ*, supporting the conclusion of Alm and Birks (1991) for *Papaver radicatum*. This is a genetically diverse genus, and the existence of a separate genetic group of *P. radicatum* in northern Norway (Solstad, 2009) also supports the glacial survival hypothesis (Brochmann et al., 2003). That hardy taxa such as *Papaver* could have persisted through both cold and warmer phases of the last glacial period on Andøya is not particularly controversial. Other arctic or low arctic species may have survived dormant as seeds or other propagules frozen in the ground, which is also a form of survival. The scattered record of more thermophilous plant species both in this and in previous records, may indicate short-term presence, which seems more likely now that we understand the high nutrient conditions supplied by the bird cliff.

The youngest record we have of *Papaver* is 12.0 k cal. BP, and it is not found in the region today https://www.artsdatabanken.no/, thus this taxon did not survive the Holocene warming at the site. In contrast to a site in the Polar Urals, where the arctic-alpine species were reduced but mainly appeared resilient to expansion of shrubs and trees (Clarke et al., 2019), Andøya does not provide a warm-stage refugia for high-arctic species.

6. Conclusions

Our new records from three cores in Øvre Æråsvatnet confirm that northern Andøya was ice-free from at least 23.4 k cal. BP and probably earlier (26.7 k cal. BP), as previously suggested by Alm (1993). When combined with higher Northern Hemisphere insolation after 24 k cal. BP, local conditions were ideal for populations of cliff-nesting seabirds. This is reflected in the Øvre Æråsvatnet stratigraphy and most clearly in stable isotope values, along with further discoveries of auk bones. The presence of thermophilous taxa in sedaDNA and macrofossils indicate at least short periods of Low Arctic Tundra condition (July mean up to 8-9°C) and possibly Shrub Tundra conditions (July mean 10-12°C) in the period 24-14.2 k cal. BP. We did record *Pinus* and *Picea* DNA, but the frequency was so low that it could not be distinguished from background contamination. Several recorded species have climate limits similar to those of *Pinus* and *Picea*. Based on this, and the local high-nutrient input, we conclude that environmental conditions, at least temporarily, would not exclude growth of pine and spruce, but we provide no firm evidence for this. It is clear that in northern Andøya a combination of proximity to warm oceanic water, coastal nunataks and sea-bird colony produced an environmental 'hotspot' at the edge of the Eurasian Ice sheet, which is unlikely to have been unique. While such environmental analogues do not occur today near the edge of ice sheets, these conditions may be realized in the near future with rapid ice-sheet retreat, ecological range changes and human-aided plant dispersal.

Acknowledgements

We thank Sandra Garces Pastor for help with the DNA analyses of run 3, Arve Elvebakk for identification of a macrofossil as *Warnstorfia fluitans*, and Charlotte Clarke and Lyn Aspden for help with figures. The work was supported by the Research Council of Norway (grant nos. 213692/F20, 230617/E10 and 250963/F20) to Alsos.

Author contributions

IGA, MEE and TA planned, designed and organised the research, IGA, AP, LG and PS carried out the coring, LG, IGA, MKFM and PS performed the DNA analysis, YL performed the bioinformatics analyses, AP and CTL performed the pollen analysis, PS performed the macrofossil analysis, PS, ML, AGB and MF performed the geochemical analyses, IGA. PS and TA carried out the review, IGA, AGB and PS organized the data and wrote the manuscript with input from all co-authors.

Conflict of Interest

Ludovic Gielly is one of the co-inventors of patents related to g-h primers and the subsequent use of the P6 loop of the chloroplast *trn*L (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers

References

Aarnes, I., Bjune, A., Birks, H., Balascio, N., Bakke, J., Blaauw, M., 2012. Vegetation responses to rapid climatic changes during the last deglaciation 13,500–8,000 years ago on southwest Andøya, arctic Norway. Veg. Hist. Archaeobot. 21, 17-35.

Alm, T., 1993. Øvre Æråsvatn – palynostratigraphy of a 22,000 to 10,000 B.P. lacustrine record on Andøya, Northern Norway. Boreas 22, 171-188.

Alm, T., Birks, H.H., 1991. Late Weichselian flora and vegetation of Andøya, Northern Norway - macrofossil (seed and fruit) evidence from Nedre Æråsvatn. Nord. J. Bot. - Section of geobotany 11, 465-476.

Alm, T., Elverland, E., 2012. A Late Weichselian *Alle alle* colony on Andøya, northern Norway - a contribution to the history of an important Arctic environment, in: Elverland, E. (Ed.), Late Weichselian to early Holocene vegetation and bird activity on Andøya, Nordland County. As evident primarily by macrofossils. Univerity of Tromsø, PhD thesis.

Alm, T., Willassen, E., 1993. Late Weichselian chironomidae stratigraphy of Nedre Æråsvatn, Andøya, northern Norway. Hydrobiologia 254, 21-32.

Alsos, I.G., Ehrich, D., Eidesen, P.B., Solstad, H., Westergaard, K.B., Schönswetter, P., Tribsch, A., Birkeland, S., Elven, R., Brochmann, C., 2015. Long-distance plant dispersal to North Atlantic islands: colonization routes and founder effect. AoB Plants 7.

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.

Alsos, I.G., Sjögren, P., Edwards, M.E., Landvik, J.Y., Gielly, L., Forwick, M., Coissac, E., Brown, A.G., Jakobsen, L.V., Føreid, M.K., Pedersen, M.W., 2016. Sedimentary

ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene 26, 627-642.
Anderson, L.L., Hu, F.S., Nelson, D.M., Petit, R.J., Paige, K.N., 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proceedings of the National Academy of Sciences 103, 12447-12450.
Armbruster, W.S., Rae, D.A., Edwards, M.E., 2007. Topographic complexity and terrestrial biotic response to high-latitude climate change: variance is as important as the mean, in: Ørbæk, J.B., Kallenborn, R., Tombre, I., Hegseth, E.N., Falk-Petersen, S., Hoel, A.H. (Eds.), Arctic Alpine Ecosystems and People in a Changing Environment. Springer Verlag, Berlin.
Ballantyne, C.K., 2002. A general model of paraglacial landscape response. The Holocene 12, 371-376.
Bennike, O., 1999. Colonisation of Greenland by plants and animals after the last ice age: a review. Polar Rec. 35, 323-336.
Bennike, O., Hedenäs, L., 1995. Early Holocene land floras and faunas from Edgeøya, Eastern Svalbard. Polar Res. 14, 205-214.
Berger, A., Loutre, M.F., 1991. Insolation values for the climate of the last 10 million years. Quat. Sci. Rev. 10, 297-317.
Berglund, B.E., Ralska-Jasiewiczowa, M., 1986. Handbook of Holocene palaeoecology and palaeohydrology. John Wiley & Sons Ltd., Chichester
Binney, H.A., Willis, K.J., Edwards, M.E., Bhagwat, S.A., Anderson, P.M., Andreev, A.A., Blaauw, M., Damblon, F., Haesaerts, P., Kienast, F., Kremenetski, K.V., Krivonogov, S.K., Lozhkin, A.V., MacDonald, G.M., Novenko, E.Y., Oksanen, P., Sapelko, T.V., Väliranta, M., Vazhenina, L., 2009. The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. Quat. Sci. Rev. 28, 2445-2464.
Birks, H.H., 1991. Holocene vegetational history and climatic changes in west Spitsbergen - plant macrofossils from Skardtjørna, an Arctic lake. The Holocene 1, 209-218.
Birks, H.H., Aarnes, I., Bjune, A.E., Brooks, S.J., Bakke, J., Kühl, N., Birks, H.J.B., 2014. Lateglacial and early-Holocene climate variability reconstructed from multi-proxy records on Andøya, northern Norway. Quat. Sci. Rev. 89, 108-122.
Birks, H.H., Birks, H.J.B., 2014. To what extent did changes in July temperature influence Lateglacial vegetation patterns in NW Europe? Quat. Sci. Rev. 106, 262-277.
Birks, H.H., Giesecke, T., Hewitt, G.M., Tzedakis, P.C., Bakke, J., Birks, H.J.B., 2012. Comment on “Glacial survival of boreal trees in northern Scandinavia”. Science 338, 742.
Birks, H.H., Larsen, E., Birks, H.J.B., 2005. Did tree-*Betula*, *Pinus* and *Picea* survive the last glaciation along the west coast of Norway? A review of the evidence, in light of Kullman (2002). J. Biogeogr. 32, 1461-1471.
Birks, H.J.B., Willis, K.J., 2008. Alpines, trees, and refugia in Europe. Plant Ecology & Diversity 1, 147-160.
Blaauw, M., Christen, J., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Analysis 6, 457-474.

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

Bougnounou, F., Hulme, P.E., Oksanen, L., Suominen, O., Olofsson, J., 2018. Role of climate and herbivory on native and alien conifer seedling recruitment at and above the Fennoscandian treeline. J. Veg. Sci.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. OBITOOLS: a unix-inspired software package for DNA metabarcoding. Mol. Ecol. Res. 16, 176-182.

Boyle, J.F., 2007. Loss of apatite caused irreversible early-Holocene lake acidification. The Holocene 17, 543-547.

Brochmann, C., Gabrielsen, T.M., Nordal, I., Landvik, J.Y., Elven, R., 2003. Glacial survival or *tabula rasa*? The history of North Atlantic biota revisited. Taxon 52, 417-450.

Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. Radiocarbon 51, 337-360.

Brubaker, L.B., Anderson, P.M., Edwards, M.E., 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.

Clarke, C.L., Edwards, M.E., Brown, A.G., Gielly, L., Lammers, Y., Heintzman, P.D., Ancin-Murguzur, F.J., Bråthen, K.-A., Goslar, T., Alsos, I.G., 2018. Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (sedaDNA) and pollen. Boreas 0.

Clarke, C.L., Edwards, M.E., Gielly, L., Ehrich, D., Hughes, P.D.M., Morozova, L.M., Haflidason, H., Mangerud, J., Svendsen, J.I., Alsos, I.G., 2019. Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. Scientific Reports 9, 19613.

Eilertsen, R.S., Bøe, R., Hermanns, R., Longva, O., Dahlgren, S., 2016. Kettle holes, 'dead-ice' topography and eskers on a lake floor in Telemark, southern Norway, in: Dowdswell, J.A., Canals, M., Jakobsson, M., Todd, B.J., Dowdeswell, E.K., Hogan, K.A. (Eds.), Atlas of submarine glacial landforms: modern, quaternary and ancient. Geological Society of London Memoirs, pp. 113-114.

Elven, R., Murray, D.F., Razzhivin, V.Y., Yurtsev, B.A., 2011. Annotated checklist of the Panarctic Flora (PAF). Vascular plants. CAFF/University of Oslo, Natural History Museum, University of Oslo.

Elverland, E., Alm, T., 2012. High resolution macrofossil analyses of Late Weichselian Arctic lacustrine sediments on Andøya, northern Norway, in: Elverland, E. (Ed.), Late Weichselian to early Holocene vegetation and bird activity on Andøya, Nordland County. As evident primarily by macrofossils. Univerity of Tromsø, PhD thesis.

Fedorov, V.B., Stenseth, N.C., 2001. Glacial survival of the Norwegian lemming (*Lemmus lemmus*) in Scandinavia: inference from mitochondrial DNA variation. Proc. R. Soc. Lond. B 268, 809-814.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Rayé, G., Taberlet, P., 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol. Ecol. Res. 15, 543-556.

959 Fjellberg, A., 1978. Fragments of a Middle Weichselian fauna on Andøya, north Norway.
960 Boreas 7, 39-39.
961 Fægri, K., Iversen, J., 1989. in: Fægri, K., Kaland, P.E., Krzywinski, K. (Eds.), Textbook of
962 pollen analysis 4. Revised edition. Wiley, Chichester, p. 314.
963 Gąsiorowski, M., Sienkiewicz, E., 2013. The sources of carbon and nitrogen in mountain
964 lakes and the role of human activity in their modification determined by tracking stable
965 isotope composition. Water, Air, & Soil Pollution 224, 1498.
966 Giesecke, T., 2016. Did thermophilous trees spread into central Europe during the Late
967 Glacial? New Phytol. 212, 15-18.
968 González-Bergonzoni, I., Johansen, K.L., Mosbech, A., Landkildehus, F., Jeppesen, E.,
969 Davidson, T.A., 2017. Small birds, big effects: the little auk (*Alle alle*) transforms high
970 Arctic ecosystems. Proceedings of the Royal Society B: Biological Sciences 284.
971 Grønlie, A.M., 1948. The ornithocoprophilous vegetation of the bird-cliffs of Røst in the
972 Lofoten islands, northern Norway. Nyt magazin for naturvidenskaberne 86, 117-243.
973 Heikkilä, M., Fontana, S.L., Seppä, H., 2009. Rapid Lateglacial tree population dynamics and
974 ecosystem changes in the eastern Baltic region. J. Quat. Sci. 24, 802-815.
975 Hughes, A.L.C., Gyllencreutz, R., Lohne, Ø.S., Mangerud, J., Svendsen, J.I., 2016. The last
976 Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1.
977 Boreas 45, 1-45.
978 Hughes, P.D., Gibbard, P.L., Ehlers, J., 2013. Timing of glaciation during the last glacial
979 cycle: evaluating the concept of a global 'Last Glacial Maximum' (LGM). Earth-
980 Science Reviews 125, 171-198.
981 Hyvärinen, H., 1970. Flandrian pollen diagrams from Svalbard. Geogr. Ann. 52 A, 213-222.
982 Karlsen, S.R., Elvebakk, A., 2003. A method using indicator plants to map local climatic
983 variation in the Kangerlussuaq/Scoresby Sund area, East Greenland. J. Biogeogr. 30,
984 1469-1491.
985 Karlsen, S.R., Elvebakk, A., Johansen, B., 2005. A vegetation-based method to map climatic
986 variation in the arctic-boreal transition area of Finnmark, north-easternmost Norway. J.
987 Biogeogr. 32, 1161-1186.
988 Kelly, A., Charman, D.J., Newnham, R.M., 2010. A Last Glacial Maximum pollen record
989 from Bodmin Moor showing a possible cryptic northern refugium in southwest
990 England. J. Quat. Sci. 25, 296-308.
991 Kullman, L., 2002. Boreal tree taxa in the central Scandes during the Late-Glacial:
992 implications for Late-Quaternary forest history. J. Biogeogr. 29, 1117-1124.
993 Kullman, L., 2005. On the ocurrence of late-glacial trees in the Scandes. J. Biogeogr. 32,
994 1499-1500.
995 Kullman, L., 2006. Late-glacial trees from arctic coast to alpine tundra: response to Birks et
996 al. 2005 and 2006. J. Biogeogr. 33, 377-378.
997 Laberg, J.S., Vorren, T.O., Dowdeswell, J.A., Kenyon, N.H., Taylor, J., 2000. The Andøya
998 Slide and the Andøya Canyon, north-eastern Norwegian–Greenland Sea. Marine
999 Geology 162, 259-275.
1000 Lamb, H., Edwards, M.E., 1988. The Arctic, in: Huntley, B., Webb, T.I. (Eds.), Vegetation
1001 history. Handbook of vegetation science 7 Kluwer Academic Publishers, Dordrecht, pp.
1002 519-555.

Larsen, E., Gulliksen, S., Lauritzen, S.-E., Lie, R., Løvlie, R., Mangerud, J., 1987. Cave stratigraphy in western Norway; multiple Weichselian glaciations and interstadial vertebrate fauna. Boreas 16, 267-292.

Mangarud, J., Gulliksen, S., Larsen, E., Longva, O., Miller, G.H., Sejrup, H.-P., Sønstegaard, E., 1981. A Middle Weichselain ice-free period in Western Norway: the Ålesund Interstadial. Boreas 10, 447-462.

Mangerud, J., 2003. Ice sheet limits in Norway and on the Norwegian continental shelf, in: Ehlers, J. (Ed.), Glacial deposits in North-West Europe. A.A. Balkema, Rotterdam, pp. 61-73.

Modin, H., 2016. Higher temperatures increase nutrient availability in the High Arctic, causing elevated competitive pressure and a decline in Papaver radicatum, Dept of Physical Geography and Ecosystem Science.

Moore, P.D., Webb, J.A., Collinson, M.E., 1991. Pollen analysis. Blackwell Scientific Publications, Oxford.

Nesje, A., Dahl, S.O., Linge, H., Ballantyne, C.K., McCarroll, D., Brook, E.J., Raisbeck, G.M., Yiou, F., 2007. The surface geometry of the Last Glacial Maximum ice sheet in the Andøya-Skånland region, northern Norway, constrained by surface exposure dating and clay mineralogy. Boreas 36, 227-239.

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

Osburn, C.L., Anderson, N.J., Leng, M.J., Barry, C.D., Whiteford, E.J., 2019. Stable isotopes reveal independent carbon pools across an Arctic hydro-climatic gradient: Implications for the fate of carbon in warmer and drier conditions. Limnology and Oceanography Letters 4, 205-213.

Parducci, L., Bennett, K.D., Ficetola, G.F., Alsos, I.G., Suyama, Y., Wood, J.R., Pedersen, M.W., 2017. Transley Reviews: Ancient plant DNA from lake sediments. New Phytol. 214, 924-942.

Parducci, L., Edwards, M.E., Bennett, K.D., Alm, T., Elverland, E., Tollefsrud, M.M., Jørgensen, T., Houmark-Nielsen, M., Larsen, N.K., Kjær, K.H., Fontana, S.L., Alsos, I.G., Willerslev, E., 2012a. Response to Comment on "Glacial Survival of Boreal Trees in Northern Scandinavia". Science 338, 742.

Parducci, L., Jørgensen, T., Tollefsrud, M.M., Elverland, E., Alm, T., Fontana, S.L., Bennett, K.D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M.E., Andersen, K., Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P., Houmark-Nielsen, M., Larsen, N.K., Orlando, L., Gilbert, M.T.P., Kjær, K.H., Alsos, I.G., Willerslev, E., 2012b. Glacial survival of boreal trees in northern Scandinavia. Science 335, 1083-1086.

Patton, H., Hubbard, A., Andreassen, K., Auriac, A., Whitehouse, P.L., Stroeven, A.P., Shackleton, C., Winsborrow, M., Heyman, J., Hall, A.M., 2017. Deglaciation of the Eurasian ice sheet complex. Quat. Sci. Rev. 169, 148-172.

Paus, A., Boessenkool, S., Brochmann, C., Epp, L.S., Fabel, D., Haflidason, H., Linge, H., 2015. Lake Store Finnsjøen – a key for understanding Lateglacial/early Holocene

vegetation and ice sheet dynamics in the central Scandes Mountains. Quat. Sci. Rev. 121, 36-51.
Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T.J., Hoffmann, D.L., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., Manning, S.W., Niu, M., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Staff, R.A., Turney, C.S.M., van der Plicht, J., 2013. IntCal13 and marine13 radiocarbon age calibration curves 0–50,000 years cal BP.
Robin, V., Nadeau, M.-J., Grootes, P.M., Bork, H.-R., Nelle, O., 2016. Too early and too northerly: evidence of temperate trees in northern Central Europe during the Younger Dryas. New Phytol. 212, 259-268.
Rozema, J., Boelen, P., Doorenbosch, M., Bohncke, S., Blokker, P., Boekel, C., Broekman, R., Konert, M., 2006. A vegetation, climate and environment reconstruction based on palynological analyses of high arctic tundra peat cores (5000–6000 years BP) from Svalbard. Plants and Climate Change 41, 155-174.
Scourse, J., 2010. Comment: A Last Glacial Maximum pollen record from Bodmin Moor showing a possible cryptic northern refugium in southwest England. (Kelly et al., 2010). J. Quat. Sci. 25, 826-827.
Sjögren, P., Edwards, M.E., Gielly, L., Langdon, C.T., Croudace, I.W., Merkel, M.K.F., Fonville, T., Alsos, I.G., 2017. Lake sedimentary DNA accurately records 20th Century introductions of exotic conifers in Scotland. New Phytol. 213, 929-941.
Small, D., Smedley, R.K., Chiverrell, R.C., Scourse, J.D., Cofaigh, C.Ó., Duller, G.A.T., McCarron, S., Burke, M.J., Evans, D.J.A., Fabel, D., Gheorghiu, D.M., Thomas, G.S.P., Xu, S., Clark, C.D., 2018. Trough geometry was a greater influence than climate-ocean forcing in regulating retreat of the marine-based Irish-Sea Ice Stream. GSA Bulletin 130, 1981-1999.
Soininen, E.M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., Gussarova, G., Bellemain, E., Hassel, K., Stenøien, H.K., Epp, L., Schrøder-Nielsen, A., Brochmann, C., Yoccoz, N.G., 2015. Highly overlapping diet in two sympatric lemming species during winter revealed by DNA metabarcoding. Plos One 10, e0115335.
Solem, J.O., Alm, T., 1994. Southwards migration of freshwater invertebrates from northern Norway. Fauna norv. Ser. A 15, 9-18.
Solstad, H., 2009. Taxonomy and evolution of the diploid and polyploid *Papaver* sect. *Meconella* (Papaveraceae), National Scentre for Biosystematics, Faculty of Mathematics and Natural Science. University of Oslo, Oslo.
Stewart, J.R., Lister, A.M., 2001. Cryptic northern refugia and the origins of the modern biota. Trends Ecol. Evol. 16, 608-613.
Sønstebø, J.H., Gielly, L., Brysting, A.K., Elven, R., Edwards, M., Haile, J., Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P., Brochmann, C., 2010. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Mol. Ecol. Res. 10, 1009-1018.

1088 Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T.,
1089 Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the
1090 chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 35, e14.
1091 Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L.,
1092 Rioux, D., Choler, P., ClÉMent, J.-C., Melodelima, C., Pompanon, F., Coissac, E.,
1093 2012. Soil sampling and isolation of extracellular DNA from large amount of starting
1094 material suitable for metabarcoding studies. Mol. Ecol. 21, 1816-1820.
1095 Tarasov, P., Müller, S., A. Andreev, Werner, K., Diekmann, B., 2009. Younger Dryas *Larix*
1096 in eastern Siberia: A migrant or survivor? PAGES news 17, 122-124.
1097 Thompson, H.A., White, J.R., Pratt, L.M., 2018. Spatial variation in stable isotopic
1098 composition of organic matter of macrophytes and sediments from a small Arctic lake
1099 in west Greenland. Arct. Antarct. Alp. Res. 50, S100017.
1100 Tjallingii, R., Röhl, U., Kölling, M., Bickert, T., 2007. Influence of the water content on X-
1101 ray fluorescence core-scanning measurements in soft marine sediments. Geochemistry,
1102 Geophysics, Geosystems 8, Q02004.
1103 Trondman, A.K., Gaillard, M.J., Mazier, F., Sugita, S., Fyfe, R., Nielsen, A.B., Twiddle, C.,
1104 Barratt, P., Birks, H.J.B., Bjune, A.E., Björkman, L., Broström, A., Caseldine, C.,
1105 David, R., Dodson, J., Dörfler, W., Fischer, E., Geel, B., Giesecke, T., Hultberg, T.,
1106 Kalnina, L., Kangur, M., Knaap, P., Koff, T., Kuneš, P., Lagerås, P., Latałowa, M.,
1107 Lechterbeck, J., Leroyer, C., Leydet, M., Lindbladh, M., Marquer, L., Mitchell, F.J.G.,
1108 Odgaard, B.V., Peglar, S.M., Persson, T., Poska, A., Rösch, M., Seppä, H., Veski, S.,
1109 Wick, L., 2015. Pollen-based quantitative reconstructions of Holocene regional
1110 vegetation cover (plant-functional types and land-cover types) in Europe suitable for
1111 climate modelling. Global Change Biol. 21, 676-697.
1112 Tzedakis, P.C., Emerson, B.C., Hewitt, G.M., 2013. Cryptic or mystic? Glacial tree refugia in
1113 northern Europe. Trends in Ecology & Evolution 28, 696-704.
1114 van der Bilt, W.G.M., Lane, C.S., 2019. Lake sediments with Azorean tephra reveal ice-free
1115 conditions on coastal northwest Spitsbergen during the Last Glacial Maximum. Science
1116 Advances 5, eaaw5980.
1117 van der Knaap, W.O., 1988. A pollen diagram from Brøggerhalvøya, Spitsbergen: changes in
1118 vegetation and environment from ca. 4400 to ca. 800 B.P. Arct. Alp. Res. 20, 106-116.
1119 Vorren, K.D., 1978. Late and Middle Weichselian stratigraphyof Andøya, north Norway.
1120 Boreas 7, 19-38.
1121 Vorren, T.O., Plassen, L., 2002. Deglaciation and palaeoclimate of the Andfjord-Vågsfjord
1122 area, North Norway. Boreas 31, 97-125.
1123 Vorren, T.O., Rydningen, T.A., Baeten, N.J., Laberg, J.S., 2015. Chronology and extent of the
1124 Lofoten–Vesterålen sector of the Scandinavian Ice Sheet from 26 to 16 cal. ka BP.
1125 Boreas 44, 445-458.
1126 Vorren, T.O., Vorren, K.-D., Aasheim, O., Dahlgren, K.I.T., Forwick, M., Hassel, K., 2013.
1127 Palaeoenvironment in northern Norway between 22.2 and 14.5 cal. ka BP. Boreas, 876–
1128 895
1129 Vorren, T.O., Vorren, K.-D., Alm, T., Gulliksen, S., Løvlie, R., 1988. The last deglaciation
1130 (20,000 - 11,000 B.P.) on Andøya, Northern Norway. Boreas 17, 41-77.

Walker, D.A., Raynolds, M.K., Daniels, F.J.A., Einarsson, E., Elvebakk, A., Gould, W.A., Katenin, A.E., Kholod, S.S., Markon, C.J., Melnikov, E.S., Moskalenko, N.G., Talbot, S.S., Yurtsev, B.A., 2005. The circumpolar arctic vegetation map. J. Veg. Sci. 16, 267-282.

Weltje, G.J., 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.

Westergaard, K.B., Zemp, N., Bruederle, L.P., Stenøien, H.K., Widmer, A., Fior, S., 2019. Population genomic evidence for plant glacial survival in Scandinavia. Mol. Ecol. 28, 818-832.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergard, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Bråthen, K.A., Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I.G., Bellemain, E., Brysting, A.K., Elven, R., Sonstebo, J.H., Murton, J., Sher, A., Rasmussen, M., Ronn, R., Mourier, T., Cooper, A., Austin, J., Moller, P., Froese, D., Zazula, G., Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R.G., MacPhee, R.D.E., Gilbert, M.T.P., Kjaer, K.H., Orlando, L., Brochmann, C., Taberlet, P., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature 506, 47-51.

Yoccoz, N.G., 2012. The future of environmental DNA in ecology. Mol. Ecol. 21, 2031-2038.

Zazula, G.D., Telka, A.M., Harington, C.R., Schweger, C.E., Mathewes, R.W., 2006. New Spruce (Picea spp.) macrofossils from Yukon Territory: implications for Late Pleistocene refugia in eastern Beringia. Arctic 59, 391-400.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., De Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Mol. Ecol. 28, 1857-1862.

Figures

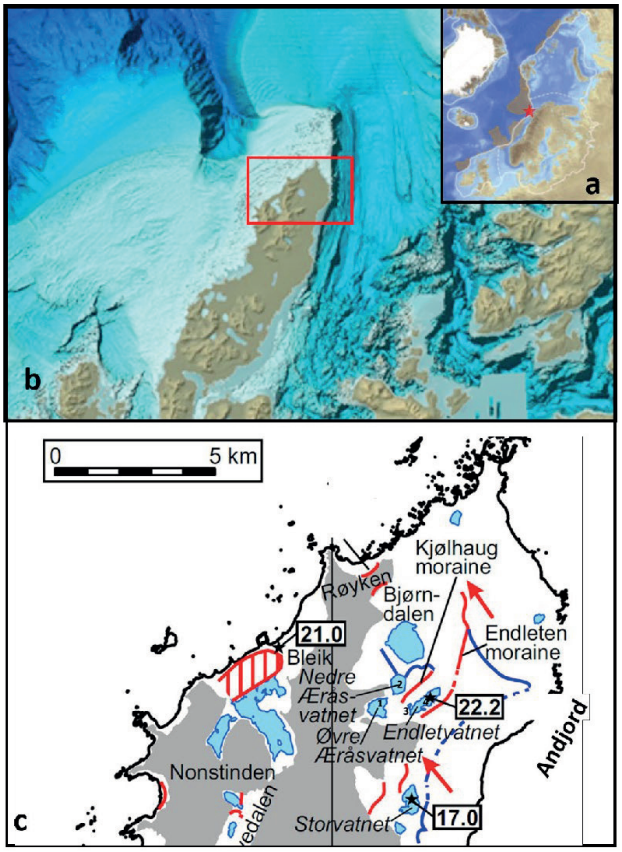


Fig. 1. The island of Andøya showing the coring site at lake Øvre Æråsvatnet. The ice limit is drawn according to Hughes et al. 2016. Coring sites are marked with stars. Numbers mark previous investigations: 1) Alm, 1993; 2) Vorren et al., 1988; Birks and Alm, 1991; 3) Vorren, 1978; Elverland and Alm, 2012; Parducci et al., 2012b; Vorren et al., 2013; 4) Vorren and Alm 1999.

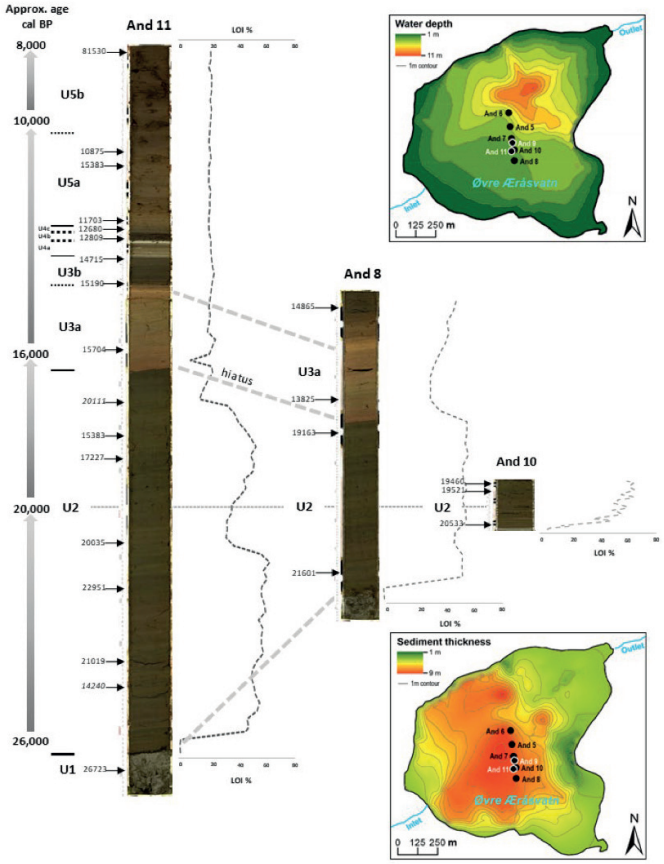


Fig. 2. Correlation among core And-11 (main core) and the shorter cores And-8 and And-10. The alignment is based on dates and allows a depth and stratigraphy match between And-8 vs. And-11. It is not possible to correlate And-10 precisely with the other two cores. Line graphs show Loss on Ignition (LOI).

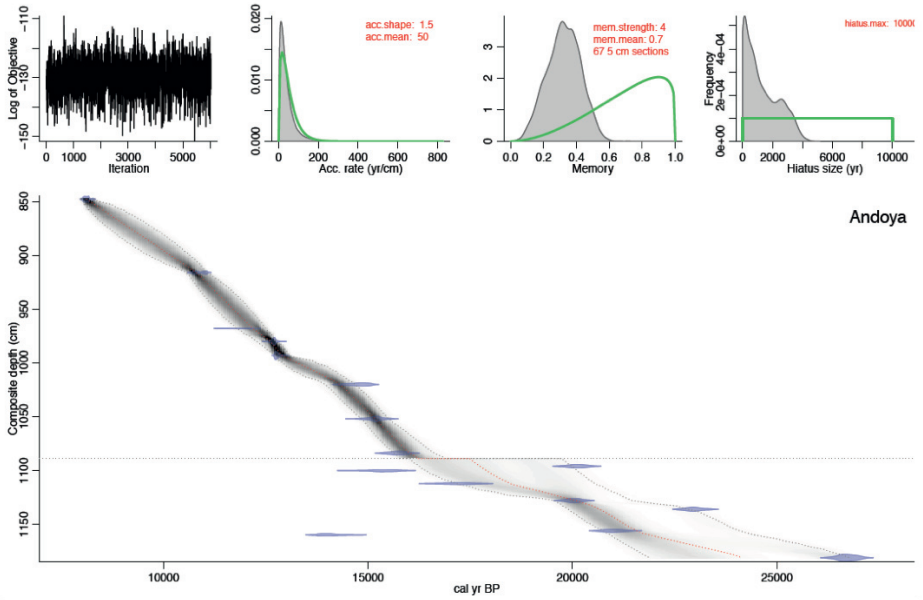


Fig. 3. Age-depth model for core And-11 from Øvre Æråscvatnet, Andøya, Norway. The calibrated 14C dates are shown in blue. The red lines show the best model based on average of the mean, and the stippled lines show the 95% confidence interval.

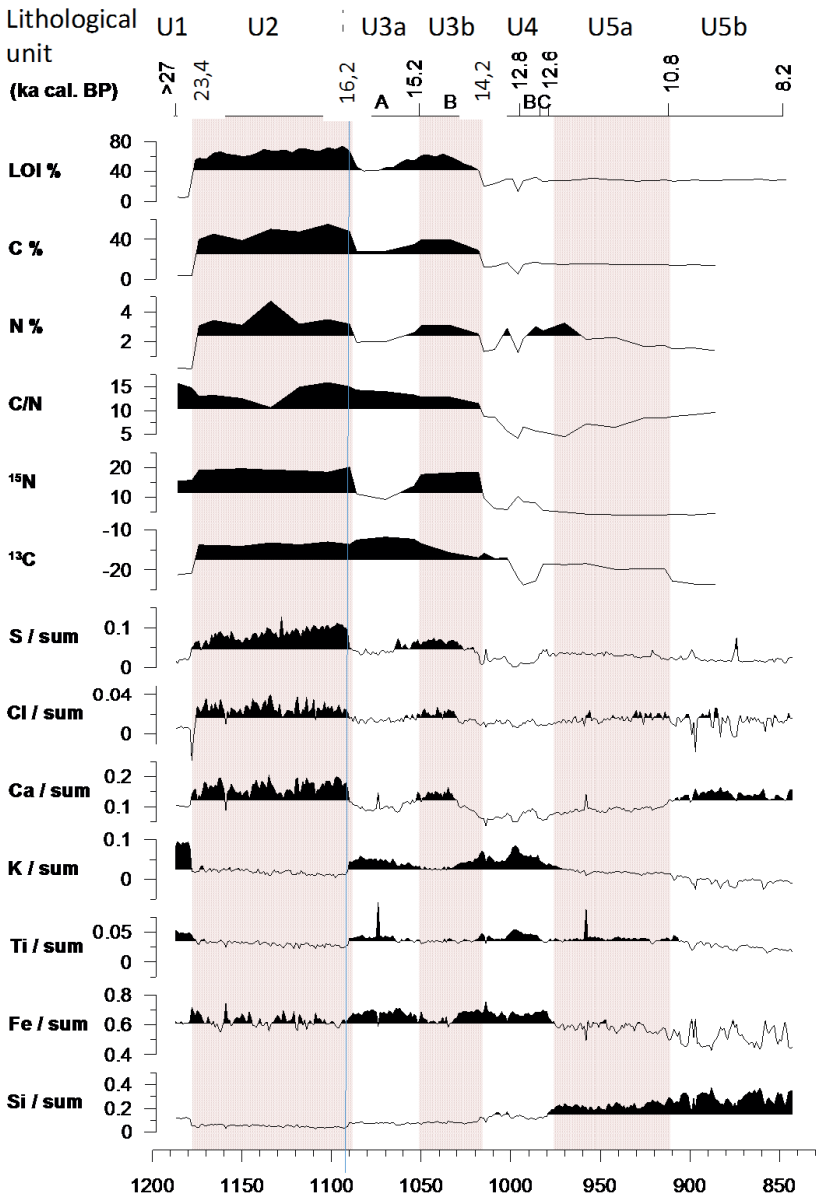


Fig. 4. Sediment properties of the core And-11 from Øvre Æråsvatnet, Andøya, Norway. Lithostratigraphic units U1-U5 are marked. The data are shown on a depth scale, with the age shown for unit boundaries. Selected elements of XRF analyses are shown as a ratio with Ti, and values above the mean are shown in black.

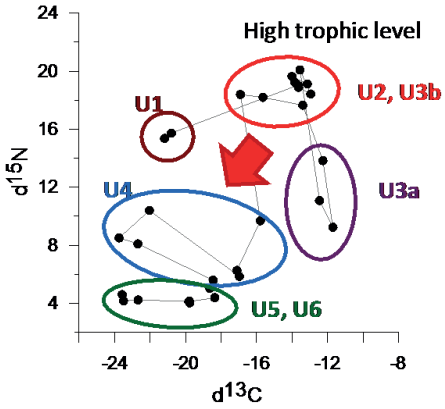


Fig. 5. A bi-plot of stable isotopes of nitrogen d15N and carbon d13C (N) in core And-11, Øvre Æråsvatnet, Andøya, Norway. The lithological units U1-U5 are marked. Note that the age of the units spans the period 24,000 cal. BP (U1) to 8000 (U5).

Fig. 6. Correlation between lithological units (Figs. 2 and 4) and zones in DNA, pollen and macrofossils according to CONISS analyses. Note that especially for the period 23,500-14,200 cal. BP, there are few proxy counts and therefore the ordination is not robust. The number of samples analysed in each zone are given in brackets and main taxa for each proxy and period are indicated.

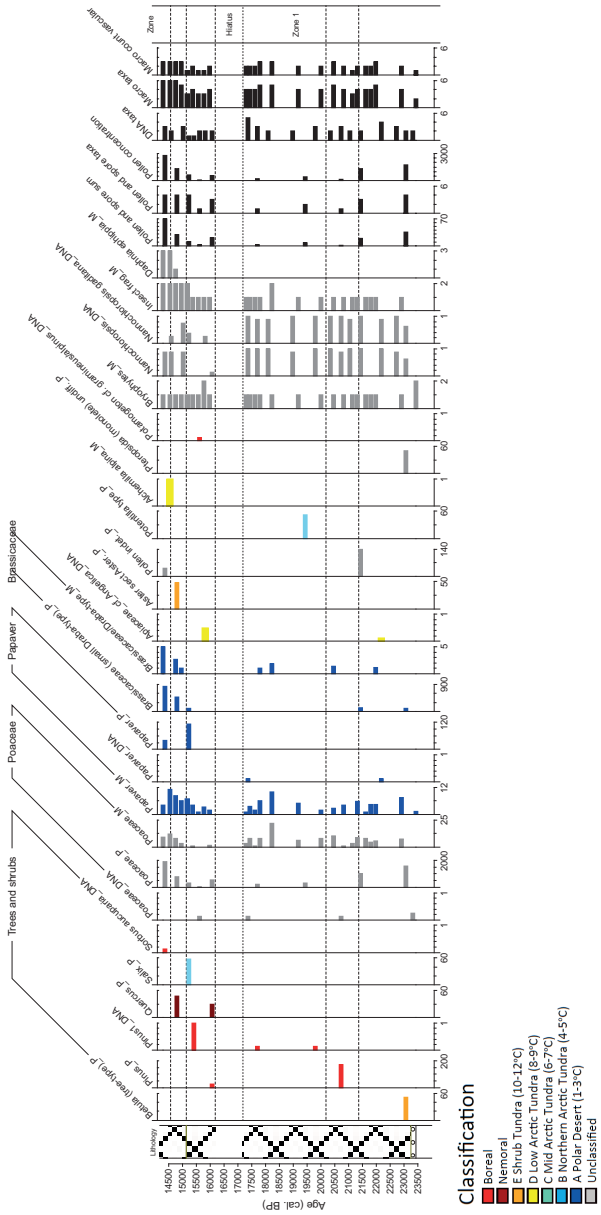
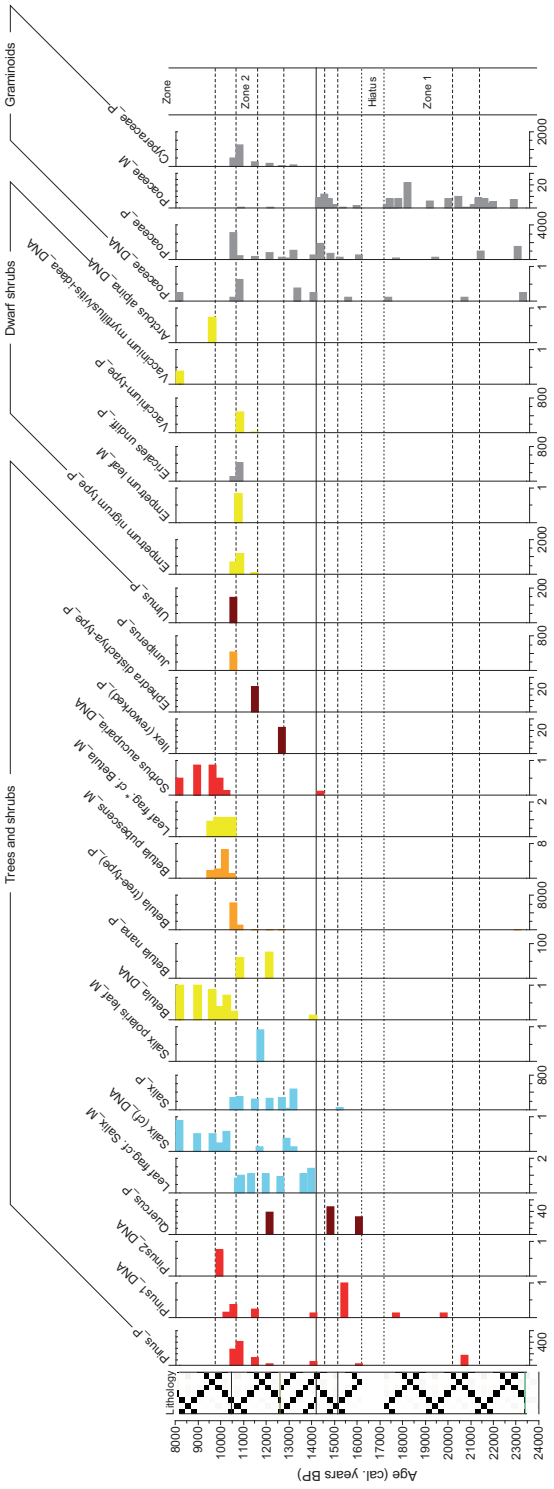
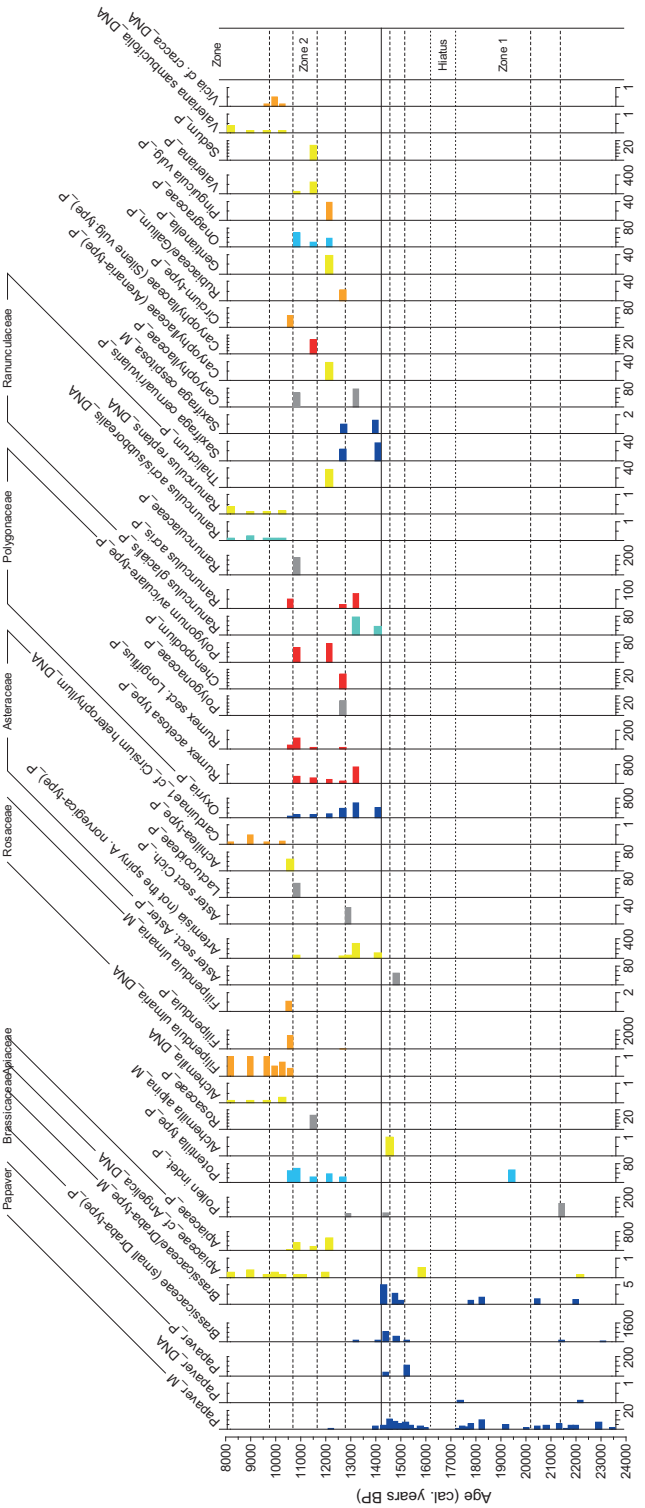


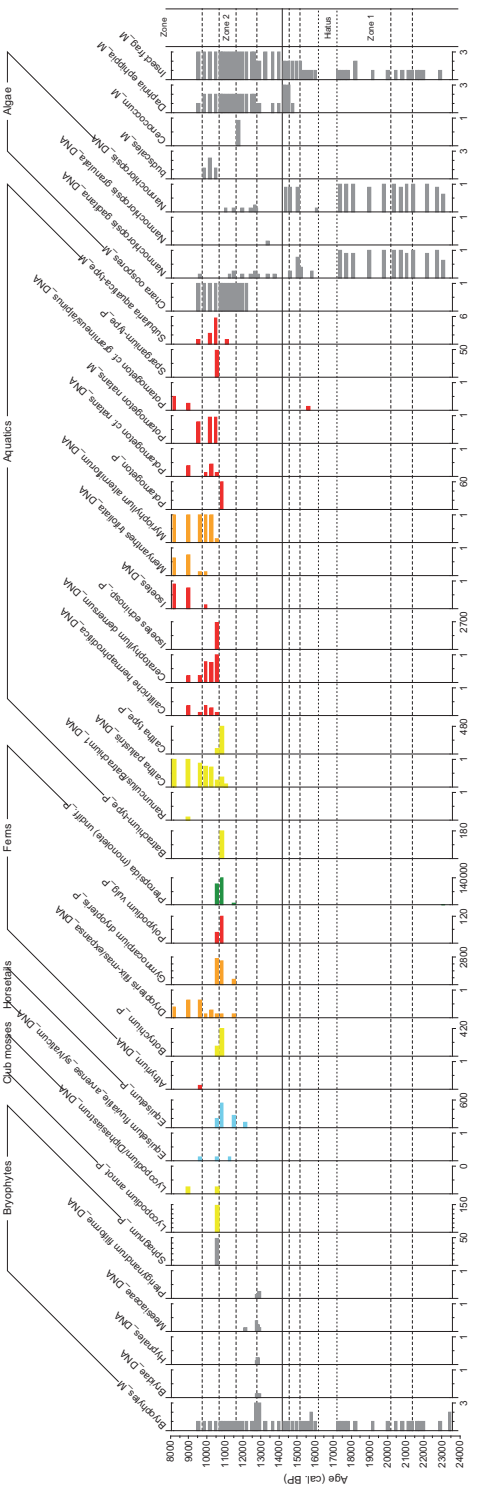
Fig. 7. Ancient sediment DNA, pollen and macrofossils recorded in core And-11 from Øvre Æråsvatnet (Andøya, Norway) from 14,200 cal. BP to 23,500 cal. BP. DNA data are presented as proportion out of 8 PCR repeats, whereas macrofossils and pollen as presented as concentrations per 50 cc, respectively. Note that the x-axes are scaled according to occurrence within each taxa and proxy; DNA data are all scaled to 1. Colour codes are according to northernmost bioclimatic subzone where the taxa is frequent (see methods).


Trees and shrubs
Dwarf shrubs
Graminoids
Pinus_P
Pinus1_DNA
Pinus2_DNA
Quercus_P
Leaf frag.cf Salix_M
Salix (cf)_DNA
Salix_P
Salix pollen leaf_M
Betula_DNA
Betula nana_P
Betula (tree-type)_P
Betula pubescens_M
Leaf frag.* cf Betula_M
Sorbus aucuparia_DNA
Ilex (rewort/red)_P
Ephedra distachya-type_P
Juniperus_P
Ulmus_P
Empetrum nigrum type_P
Empetrum nigrum leaf_M
Ericales undiff._P
Vaccinium-type_P
Arctous alpina_DNA
Poaceae_DNA
Poaceae_P
Poaceae_M
Cyperaceae_P
Zone
Zone 2
Hiatus
Zone 1
Lithology
Age (cal. years BP)
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
21000
22000
23000
24000

1207


Age (cal. years BP)
Zone
Zone 2
Hiatus
Zone 1
Papaver
Brassicaceae
Asteraceae
Rosaceae
Polygonaceae
Ranunculaceae
Papaver_M
Papaver_DNA
Papaver_P
Brassicaceae (small Draba-type)_P
Brassicaceae_cf_Arabis-type_M
Apiaceae_P
Apiaceae_cf_Daucus-type_P
Potentilla-type_P
Potentilla_undiff._P
Rosaceae_M
Aphanes_DNA
Filipendula_undiff._P
Filipendula ulmaria_DNA
Aster sect. Aster_M
Artemisia (not the spiny A. norvegica-type)_P
Aster sect. Cich._P
Lactucoideae_P
Achillea-type_P
Carduinae_P
Oxyria_P
Taraxacum_cf_Crisum heterophyllum_DNA
Rumex acetosa-type_P
Rumex sect. Longifolius_P
Chenopodiaceae_P
Polygonum_P
Polygonum aviculare-type_P
Ranunculus glacialis_P
Ranunculus acris_P
Ranunculaceae_P
Ranunculus aconitifolius-type_P
Thalictrum_P
Saxifraga cernua/hirculus_P
Saxifraga oppositifolia_M
Caryophyllaceae_P
Caryophyllaceae (Arenaria-type)_P
Cerastium-type_P
Rubiaceae/Galium_P
Gentianella_P
Onagraceae_P
Polygonum viviparum_P
Valeriana_P
Sedum_P
Valeriana sambucifolia_DNA
Vicia cf. cracca_DNA

1208


Age (cal. BP)
Zone
Zone 2
Hiatus
Zone 1
Algae
Aquatics
Ferns
Horsetails
Club mosses
Bryophytes
Botryhyle_M
Bryale_DNA
Hyphales_DNA
Meesiaceae_DNA
Pseudocalliergon trifarium_DNA
Sphagnum_P
Lycopodium annot._P
Lycopodium.Diphasiastrum_DNA
Equisetum fluviatile.arvense.sylvaticum_DNA
Equisetum_P
Athyrium_DNA
Botrychium_P
Dryopteris.fix.mas.expansa_DNA
Gymnocarpium dryopteris_P
Polypodium vulg._P
Pteropsida (monolete) undif._P
Baratchamn.type_P
Ranunculus.flammula_P
Calla palustris_DNA
Calla type_P
Calliergonella cuspidata_DNA
Ceratodon.purpureus_DNA
Isoetes echinosp._P
Isoetes_DNA
Menyanthes trifoliata_DNA
Myriophyllum alterniflorum_DNA
Potamogeton_P
Potamogeton cf. filiformis_DNA
Potamogeton natans_M
Spar.gramin.type_P
Subularia aquatica-type_P
Chara oospores_M
Nematoctenopsis gracilis_DNA
Nematoctenopsis spatiens_DNA
Nematoctenopsis_M
Botryococcus_M
Coelastrum_M
Cyclotella_M
Daphnia ephippia_M
Tissue frag_M

1209

Fig.8. Ancient sediment DNA, pollen and macrofossils recorded in core And-11 from Øvre Æråsvatnet (Andøya, Norway) from 8,000 cal. BP to 23,500 cal. BP. DNA data are presented as proportion out of 8 PCR repeats, whereas macrofossils and pollen as presented as seed and grains per 50 cm^3, respectively. Note that the x-axes are scaled according to occurrence within each taxa and proxy; DNA data are all scaled to 1. Colour codes indicate northernmost bioclimatic subzone where the taxa is frequent (see methods and Fig. 7).

Supplementary figures

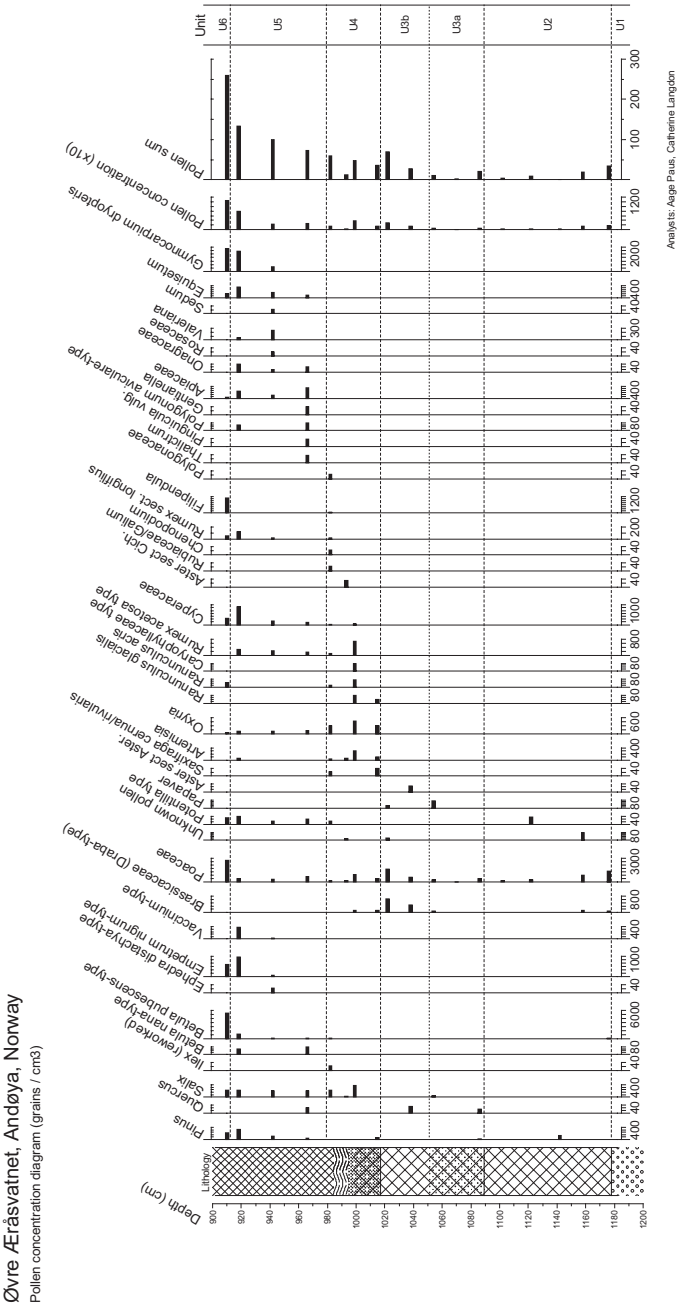


Fig. S1. Pollen diagram for And-11. Taxa only occurring in the top two samples are not included to improve readability. Pollen sum and pollen concentrations refers to total dry land pollen and spore sum and concentration per cm^3, respectively.

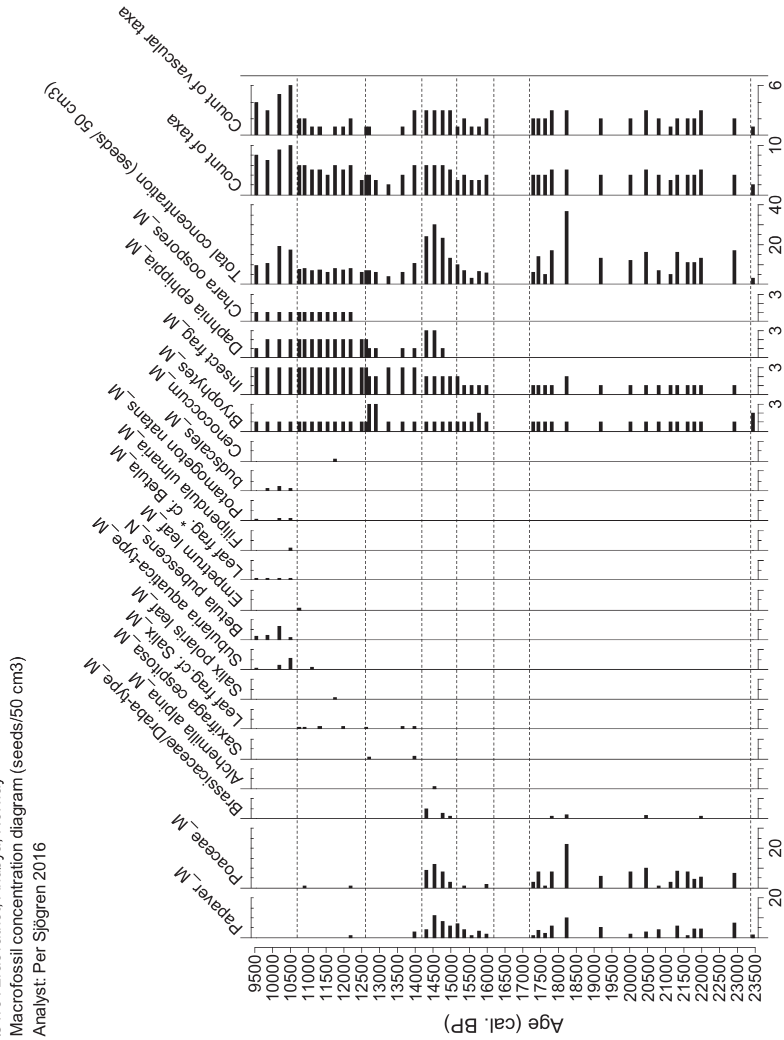


Fig. S2. Macrofossil diagram for And-11. Macrofossils are given as concentration if not otherwise noted. Macrofossil of leaf fragments of *Salix* and *Betula* which are given as presence/absence. Insects, *Daphnia, Chara* and bryophytes are given as 1=few (1-50 fragments), 2=common (50-1000 fragments) and 3=abundant (>1000 fragments).

Table 1. Radiocarbon dates. Weight of dated material is added when this is so low that it might influences the probability estimate.

Core	Depth	Lab #	^{14}C BP (1σ)	Cal. BP (2 σ)	Cal. BP (2 σ) median	Material
And-11	848 cm	Poz-72236	7320 ± 50	8300–8010	8155	Moss, plant mat., leaves (0.2 mg)
And-11	916 cm	Poz-72235	9520 ± 50	11090–10660	10875	Moss, plant mat., leaves (0.2 mg)
And-11	968 cm	Poz-77608	10120 ± 120	12150–11260	11705	Moss (0.06 mg)
And-11	980 cm	Poz-72234	10770 ± 80	12800–12570	12685	Moss, leaves (0.2 mg)
And-11	993 cm	Poz-72233	10900 ± 60	12930–12690	12730	Moss (*Warnstorfia fluitans*)
And-11	1020 cm	Poz-72231	12530 ± 80	15130–14300	14715	Moss, seeds
And-11	1052 cm	Poz-72230	12750 ± 90	15550–14830	15190	Moss, seeds (0.2 mg)
And-11	1084 cm	Poz-72229	13120 ± 110	16070–15340	15705	Moss, seeds, plant mat.
And-11	1096 cm	Poz-77610	16670 ± 150	20510–19710	20110	Moss, seeds (0.14 mg)
And-11	1100 cm	Poz-104680	12880 ± 190	16019–14673	15346	Moss stems (0.06mg)
And-11	1112 cm	Poz-104682	14180 ± 250	17891–16501	17196	Seed case (0.05mg)
And-11	1128 cm	Poz-72219	16610 ± 110	20360–19710	20030	Moss, seeds
And-11	1136 cm	Poz-104683	19060 ± 130	23364–22579	22972	Moss stems (0.18mg)
And-11	1156 cm	Poz-77611	17390 ± 150	21450–20590	21020	Moss, seeds (0.2 mg)
And-11	1160 cm	Poz-104684	12120 ±150	13760–11110	12435	Herb stem fragments (0.02mg)
And-11	1181 cm	Poz-77656	22410 ± 120	27120–26330	26725	Moss
And-8	1090 cm	Poz-72215	12630 ± 100	15310–14420	14865	Moss, seeds (0.2 mg)
And-8	1122 cm	Poz-77607	11930 ± 180	14280–13370	13825	Moss, seeds (0.02 mg)
And-8	1134 cm	Poz-72214	15840 ± 200	19610–18720	19165	Moss, seeds, plant mat. (0.14 mg)
And-8	1182 cm	Poz-72213	17840 ± 230	22230–20980	21605	Moss, seeds (0.11 mg)
And-10	908 cm	Poz-72217	16120 ± 190	19940–18980	19460	Moss, seeds (0.13 mg)
And-10	960 cm	Poz-77606	16180 ± 100	20040–19000	19520	Moss, seeds (0.06 mg)
And-10	1004 cm	Poz-72216	16990 ± 240	21150–19910	20530	Moss, seeds (0.08 mg)

Table 2. Lithological descriptions and modelled median age ranges. And-8 date in brackets inferred from And-11. Bacon dates in brackets extrapolated (1178 extrapolated from 1089 to 1156 using weighted mean).

Depth (cm)	Bacon k cal. BP	Unit	Description
And-11			
842–912	8.11–10.64	U5b	Dark olive brown gyttja
913–979	10.68–12.60	U5a	Olive brown silty gyttja
980–984	12.62–12.71	U4c	Dark brown silty gyttja
985–995	12.73–12.99	U4b	Moss layer (*Warnstorfia fluitans*).
996–1017	13.04–14.18	U4a	Bands of light olive grey to white silt. Lower boundary sharp
1018–1051	14.24–15.15	U3b	Laminated olive green gyttja. Lower boundary gradual.
1052–1089	15.17–16.18	U3a	Laminated rusty brown gyttja. Lower boundary sharp
	Hiatus (ca. 1 k years)		
1090–1178	17.23 – 23.25	U2	Laminated olive green gyttja. Lower boundary sharp
1179–1181*	23.31 – 23.45	U1	Silty-sandy diamicton. Sand layer at top (1 cm thick)
And-8			
1084–1093	*14.7-14.9*	U3b	Laminated olive green gyttja. Lower boundary gradual.
1093–1130	*14.9-(15.8)*	U3a	Laminated rusty brown gyttja. Lower boundary sharp
	Hiatus ?		
1130–1188	*19.0-21.9*	U2	Laminated olive green gyttja
1188–1191	*21.9-22.1*	U1/U2	Brown clay/gyttja
1191–1198	*22.1- ?*	U1	Silty diamicton, including pebbles
And-10			
900–954	*19.4-20.0*	U2	Coarse detritus gyttja
954–998	*20.0-20.5*	U2	Laminated gyttja
998–1010	*20.5-20.6*	U2	Laminated silty gyttja
1010–1016	*20.6- ?*	U1	Silt

Table 3. Thermophilous and other selected taxa of MIS2 age. July temperature of the bioclimatic zone where the taxon is found as frequent (Temp.1) or rare/scattered (Temp.2) today. *Indirectly dated to the period 21.4–20.1 ka cal. BP by Vorren et al., 2013). Uncertain temperatures due to uncertain taxon identifications are in brackets.

Taxa	Date	Temp.1	Temp2	Type	Site	Reference
Andøya thermomer 1, ca. 24–23 ka cal. BP (IS2, Andøya interstadial)						
Sphagnum leaves	ca. 25	8-9		macrofossil	Øvre Æråsvatnet	Alm 1993, unpub. data
Rumex acetosa	23.2	>12	10-12	macrofossil	Nedre Æråsvatnet	Alm and Birks, 1991
Saxifraga cespitosa	23.2	1-3	1-3	macrofossil	Nedre Æråsvatnet	Alm and Birks, 1991
Andøya thermomer 2, ca. 22–20 ka cal. BP (LGM)						
Chrysosplenium	22.0	6-7	4-5	macrofossil	Nedre Æråsvatnet	Alm and Birks, 1991
Pinus	22.0	(>12)	(10-12)	*sed*aDNA	Endletvatn	Parducci et al., 2012
Brya-type (*B. linearis*)	≤21.8	(8)	(6-7)	macrofossil	Endletvatn	Elverland and Alm, 2012
Apiaceae	≤21.8	8-9	8-9	*sed*aDNA	Øvre Æråsvatnet	Present
Apioideae (Apiaceae)	≤21.8	8-9	8-9	*sed*aDNA	Endletvatn	Parducci et al., 2012
Sphagnum papillosum	ca. 21	8		macrofossil	Endletvatn	Vorren, 1978
Mustela erminea	ca. 20.8*	3		vertebrae	Endletevatn	Fjellberg, 1978
Betula pubescens	20.4	10-12	10-12	megafossil	Stavedalen	Kullman, 2006
Urtica dioica	19.8	>12	10-12	macrofossil	Endletvatn	Parducci et al., 2012
Andøya thermomer 3, ca. 19–18 ka cal. BP (LGM)						
Pinus	19.2	(>12)	(10-12)	*sed*aDNA	Endletvatn	Parducci et al., 2012
Sphagnum papillosum	ca. 18	8		macrofossil	Endletvatn	Vorren, 1978
Picea abies	17.7	(>12)	(10-12)	*sed*aDNA	Endletvatn	Parducci et al., 2012
Andøya thermomer 4, ca. 15 ka cal. BP (pre-Bølling warming)						
Sphagnum papillosum	ca. 15	8		macrofossil	Endletvatn	Vorren, 1978
S. cf. p*latyphyllum*	ca. 15	8		macrofossil	Endletvatn	Vorren, 1978
Scorpidium scorpioides	ca. 15	3		macrofossil	Endletvatn	Vorren, 1978
cf. *Alchemilla alpina*	14.9	(8-9)	(8-9)	macrofossil	Øvre Æråsvatnet	Present

Paper VI

Environmental palaeogenomic reconstruction of an Ice Age algal population

Youri Lammers[1], Peter D. Heintzman[1], Inger Greve Alsos[1]

[1]The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Tromsø, Norway

Abstract

Despite being located far north of the Scandinavian ice sheet extent, the coastal Norwegian island of Andøya was partially unglaciated during the last glacial maximum (LGM). The potential of this locality as a cryptic northern refugium has therefore been a focus of previous palaeoecological studies based on analyses of lake sediment cores. Previous geochemical studies have suggested that these sediments contain an abundance of algae, while metabarcoding of sedimentary ancient DNA (*sed*aDNA) suggests that a major constituent of this LGM algal community was *Nannochloropsis*, although species-level resolution could not be determined. Given the inferred abundance of *Nannochloropsis* DNA in the sediment record, together with a panel of available reference genomes, we used shotgun sequencing to robustly identify which and how many *Nannochloropsis* species were present in two LGM layers of the Lake Øvre Æråsvatnet record. Both metagenomic analysis and mapping against the reference genome panel confirmed that *N.* cf. *limnetica* was the dominant *Nannochloropsis* taxon present. We then reconstructed full chloroplast (cp) and mitochondrial (mt) genomes from the *Nannochloropsis* sequences and examined the nucleotide variants present to examine within-lake population genetic variation within *N.* cf. *limnetica*. At least two major haplogroups were present in both organellar genomes and occur at similar ratios. These haplogroups are consistent, and their relative frequencies comparable, between the two broadly coexisting samples, suggesting that the results are replicable. The shotgun sequencing approach presented here demonstrates the potential of *sed*aDNA of ancient environmental samples for population genetic analysis.

Keywords

*sed*aDNA, *Nannochloropsis*, algae, Andøya, environmental palaeogenome reconstruction, phylogenetic analysis, population genetics

Introduction

Andøya, an island located in northwest Norway, was partially unglaciated during the last glacial maximum (LGM, Figure 1) and has therefore been the focus of palaeoecological study (Vorren *et al.* 2015), especially for its potential as a cryptic northern refugium (Parducci *et al.* 2012, Birks *et al*. 2012, Parducci *et al*. 2012). Lake sediment cores have been the main source of palaeoecological insights, including those taken from Lake Endletvatn (Vorren 1978, Elverland and Alm 2012, Vorren *et al*. 2013), Lake Nedre Æråsvatnet (Vorren *et al*. 1988, Alm and Birks 1991, Parducci *et al*. 2012), and Lake Øvre Æråsvatnet (Alm 1993, Vorren *et al*. 2013, Alsos *et al*. in prep.). These studies have reported the presence of an LGM Arctic flora containing taxa such as Poaceae, Brassicaceae, and *Papaver*, along with macrofossils of the Little Auk/dovekie (*Alle alle*). Furthermore, recent geochemical and DNA metabarcoding analyses by Alsos *et al*. (in prep.) indicate both the presence and an inferred high abundance of the algae *Nannochloropsis* in LGM layers. Together with the high organic content of the sediments, the presence and abundance of *Nannochloropsis* DNA suggests eutrophication of Lake Øvre Æråsvatnet during the LGM and possible long-lasting algal blooms.

Nannochloropsis is a genus of unicellular microalgae in Family Eustigmatophyceae. All species in the genus have relatively high lipid contents, making up to 50% of their cellular weight (Su *et al*. 2011, Radakovits *et al*. 2012), and are thus of interest as a potential source of biofuels. As a result of this economic interest, six species in the genus have had their organelle and nuclear genomes sequenced (Radakovits *et al*. 2012, Vieler *et al*. 2012, Wei *et al*. 2013, Schwartz *et al*. 2018). In total, eight species have been described (Table 1), where most species are known from marine environments, except *N. limnetica*, which is known from freshwater or brackish habitats (Krienitz *et al*. 2000, Fawley and Fawley 2007, Fietz *et al*. 2005). The genus *Nannochloropsis* is cosmopolitan in its distribution, with the marine species being reported from most oceans (Hibberd 1981, Karlson *et al*. 1996, Suda *et al*. 2002). The freshwater/brackish *N. limnetica* also has a broad distribution and is known from lakes from Central Europe (Krienitz *et al*. 2000), Lake Baikal (Fietz *et al*. 2005), North Dakota, Minnesota and North Carolina (Andersen *et al*. 1998, Fawley and Fawley 2007) and Antarctica (Karlov *et al*. 2017).

Species of *Nannochloropsis* range in size from 2 to 6 µm (Suda *et al*. 2002, Fietz *et al*. 2005) and are problematic to identify due to a lack of defined morphological structures (Gladu *et al*. 1995, Andersen *et al*. 1998), which is in stark contrast to diatoms (Smol *et al*. 2004). Though potential identification in palaeoecological records is possible due to their cell walls containing a family of highly resistant aliphatic macromolecules called algaenans (Gelin *et al*. 1997, Leeuw *et al*. 2005). Reliable identification to species level is possible with molecular markers such as *rbcL*, 18S, or ITS2 (Andersen *et al*. 1998, Fawley and Fawley 2007, Kryvenda *et al*. 2018). Consequently, this genus can be identified from analysis of sedimentary ancient DNA (*sed*aDNA) via metabarcoding of the 18S or *trn*L p6 loop region or through shotgun sequencing. Using these approaches, *Nannochloropsis* has been reported from additional regions, such as Greenland (Epp *et al*. 2015), British Columbia and Alberta (Pedersen *et al*. 2016), China (Li *et al*. 2016), and Svalbard (Voldstad *et al*. submitted).

*Sed*aDNA is increasingly used to infer the taxonomic composition of past ecosystems (Parducci *et al*. 2017). The most commonly applied method is DNA metabarcoding, which allows for the targeting of particular groups of organisms with generic primers (Taberlet *et al*. 2012, Taberlet *et al*. 2018). Although powerful, the ability to confidently identify sequences is related to the completeness of reference databases and the length and variability of the barcode targeted. For ancient DNA applications, short barcodes are necessarily targeted due to small DNA fragment lengths, which can impede identification and exploration of genetic variation (Prüfer *et al*. 2010). Additional shortcomings are that it is often difficult to distinguish rare ancient from contaminating amplicons (Parducci *et al*. 2017, Zinger *et al*. 2019) and that DNA damage patterns, used to authenticate ancient DNA (Briggs *et al*. 2007, Skoglund *et al*. 2014), are effectively removed by the PCR. An alternative approach is shotgun sequencing, which is non-targeting and therefore requires no prior knowledge of the DNA sequences present. Although unbiased, the lack of targeting can result in most sequence data being unidentified due to incomplete genomic reference databases (Pedersen *et al*. 2016, Stat *et al*. 2017). Shotgun sequenced data retain DNA damage patterns allowing for authentic ancient DNA to be distinguished from modern contaminating sequence. With deep enough sequencing, shotgun data can be used for genome reconstruction, which in turn allows for robust phylogenetic reconstruction (Seersholm *et al*. 2016, Slon *et al*. 2017), as well as the potential exploration of population genetic variation in an environmental sample (Søe *et al*. 2018).

In this study, we shotgun sequenced two LGM sediment layers from Lake Øvre Æråsvatnet that previous metabarcoding had shown to contain *Nannochloropsis* (Alsos *et al*. in prep.). Although confidently identified to the genus, metabarcoding did not allow for finer scale identification to species level, nor exploration of the potential within-lake *Nannochloropsis* community through time. The depth of our shotgun sequence data, together with the availability of a reference genome panel, allowed us to determine that *Nannochloropsis* dominates the identifiable taxonomic profile of the LGM sediments from Andøya and that the taxon likely to be present is *N.* cf. *limnetica*. We also show that it is possible to investigate past genetic diversity from lake core *sed*aDNA samples, which, in this case, show no major population genetic difference between the contemporaneous samples.

Material and methods

Site description, chronology, and sampling

A detailed description of the site, coring methods, age-depth model reconstruction, and sampling strategy can be found in Alsos *et al*. (in prep.). Briefly, Lake Øvre Æråsvatnet is located on Andøya, Northern Norway (69.25579°N, 16.03517°E) (Figure 1). In 2013, three cores were collected from the deepest sediments, AND8, AND10 and AND11. Macrofossil remains from all three were dated and for AND8 and AND10, all dates were within the LGM. For the longer core AND11, a Bayesian age-depth model allowed more precise estimate of age of each layer (Alsos *et al*. in prep.). In this study, we selected one sample from LGM sediments from two different cores. According to the Bayesian age-depth model, sample AND11_1102 was dated to a median age of 17,700 (range 20,200-16,500) cal a BP. Based on an interpolated median date of two adjacent macrofossils, AND10_938 was dated to 19,500 cal a BP. As the second sample falls within the age range of the first sample, we interpreted the samples as broadly contemporaneous.

Sampling, DNA extraction, library preparation, and sequencing

The two cores were subsampled at the selected layers for DNA extractions under clean conditions, in a dedicated ancient DNA laboratory at the The Arctic University Museum in Tromsø, Norway. We extracted DNA from 15 g of sediment following the Taberlet phosphate extraction protocol (Taberlet *et al*. 2012). We shipped a 210 µL aliquot of the DNA extract to the ancient DNA dedicated laboratories at the Centre for GeoGenetics,

Copenhagen, for double-stranded DNA library construction. After concentrating the DNA extract to 80 μL, half of the extract (40 μL, totalling between 31.7-36.0 ng of DNA) from each of the two selected samples was converted to Illumina-compatible libraries using established protocols (Pedersen *et al*. 2016). The number of indexing PCR cycles was determined using qPCR and each sample was dual indexed. The libraries were then purified using the AmpureBead protocol (Beckman Coulter, Indianapolis, IN, USA), adjusting the volume ratio to 1:1.8 library:AmpureBeads, and quantified using a BioAnalyzer (Agilent, Santa Clara, CA, USA). The indexed libraries were pooled equimolarly and sequenced on a lane of the Illumina HiSeq 2500 platform using 2x 80 bp paired-end chemistry.

Raw read pre-processing

For each sample, we merged and adapter-trimmed the paired-end reads with *SeqPrep* (https://github.com/jstjohn/SeqPrep/releases, v1.2) using default parameters. We only retained the resulting merged sequences, which were then filtered with the preprocess function of the *SGA toolkit* v0.10.15 (Simpson and Durbin 2012) by the removal of those shorter than 35 bp or with a DUST complexity score >1 from the dataset.

Metagenomic analysis of the sequence data

We first sought to obtain an overview of the taxonomic composition of the samples and therefore carried out a metagenomic analysis on the two filtered sequence datasets. To make the datasets more computationally manageable, we subsampled the first and last one million sequences from the filtered dataset of each sample and analysed each separately. The data subsets were each identified against the NCBI nucleotide database (release 223) using the blastn function from the *NCBI-BLAST+* suite v2.7.1+ (Camacho *et al*. 2009) under default settings. For each sample, the results from the two subsets were merged into one dataset and loaded into *MEGAN* v6.12.3 (Huson *et al*. 2016). Analysis and visualization of the Last Common Ancestor (LCA) was carried out for the taxonomic profile using the following settings: min score=35, max expected=1.0E-5, min percent identity=95%, top percent=10%, min support percentage=0.01, LCA=naive, min percent sequence to cover=95%. We define sequences as the reads with BLAST hits assigned to a taxa post filtering, thus ignoring "unassigned" and "no hit" categories.

Alignment to a reference genome panel

We mapped our filtered data against three different reference panels to help improve taxonomic identifications and provide insight into the abundance of the identified taxa. The first reference panel consisted of 37 nuclear genomes that included taxa expected from Northern Norway, exotic taxa, and five *Nannochloropsis* species. The inclusion of exotic taxa was to give an indication of the background random mapping rate, which can result from erroneous mappings to conserved parts of the genome and/or due to short, fragmented, and damaged ancient DNA molecules. The other two reference panels were based on either all mitochondrial or chloroplast genomes on NCBI GenBank (as of January 2018). The chloroplast data set was augmented with 247 partial or complete chloroplast genomes generated by the PhyloNorway project. Full details for the nuclear reference panel is given in Supplementary Table S1. The filtered data were mapped against each reference genome or organelle genome set individually using *bowtie2* v2.3.4.1 (Langmead and Salzberg 2012) under default settings. The resulting bam files were processed with *SAMtools* v0.1.19 (Li *et al*. 2009). We removed unmapped sequences with *SAMtools view* and collapsed PCR duplicate sequences with *SAMtools rmdup*.

For the nuclear reference panel, we reduced potential erroneous or nonspecific sequence mappings, by comparing the mapped sequences to both the aligned reference genome and the NCBI nucleotide database using *NCBI-BLAST+*, following the method used by Graham *et al*. (2016) and modified by Wang *et al*. (2017). The sequences were aligned using the following *NCBI-BLAST+* settings: num_alignments=100 and perc_identity=90. Sequences were retained if they had better alignments, based on bit score, to reference genomes as compared to the NCBI nucleotide database. If a sequence had a better or equal match against the NCBI nucleotide database, it was removed, unless the LCA of the highest NCBI nucleotide bit score was from the same genus as the reference genome (based on the NCBI taxonID). To standardize the relative mapping frequencies to genomes of different size, we calculated the number of retained mapped sequences per Mb of genome sequence.

The sequences mapped against the chloroplast and mitochondrial reference panels were filtered and reported in a different manner than the nuclear genomes. First, to exclude any non-eukaryote sequences, we used *NCBI-BLAST+* to search sequence taxonomies and retained sequences if the LCA was from Eukaryota. Secondly, for the sequences that were retained, the LCA was calculated and reported in order to summarize the mapping results

across the organelle datasets. LCAs were chosen as the reference sets are composed of multiple genera.

Within the *Nannochloropsis* nuclear reference alignments, the relative mapping frequency was highest for *Nannochloropsis limnetica*. In addition, the relative mapping frequency for other *Nannochloropsis* taxa was higher than those observed for the exotic taxa. This could represent the mapping of sequences that are conserved between *Nannochloropsis* taxa or suggest the presence of multiple *Nannochloropsis* taxa in a community sample. We therefore cross-compared mapped sequences to determine the number of uniquely mapped sequences per *Nannochloropsis* reference genome. First, we individually remapped the filtered data to five available nuclear genomes or the six available *Nannochloropsis* chloroplast reference genomes, the accession codes of which are provided in Supplementary Table S1. For each sample, we then calculated the number of sequences that uniquely mapped to or overlapped between each *Nannochloropsis* nuclear and chloroplast genome.

We observed a greater proportion of uniquely mapped sequences across all chloroplast genomes, as compared to nuclear genomes. To check whether the retained uniquely mapped sequences were the result of mismapped ultra-short (35-40 bp) molecules that may have been missed by our other filters, we estimated DNA fragment length distributions for all mapped and uniquely mapped sequences per sample in the chloroplast genome alignments.

Environmental palaeogenome reconstruction of the Andøya *Nannochloropsis* community chloroplast and mitochondria

To place the Andøya *Nannochloropsis* community taxon into a phylogenetic context, and provide suitable reference sequences for variant calling, we reconstructed environmental palaeogenomes for the *Nannochloropsis* mitochondria and chloroplast. First, the raw read data from both samples were combined into a single dataset and re-filtered with the *SGA toolkit* to remove sequences shorter than 35 bp but retain low complexity sequences to assist in the reconstruction of low complexity regions in the organelle genomes. This re-filtered sequence dataset was used throughout the various steps for the environmental palaeogenome reconstruction.

The re-filtered sequence data were mapped onto the *Nannochloropsis limnetica* reference chloroplast genome (NCBI GenBank accession: NC_022262.1) with *bowtie2* using default

232 settings. *SAMtools* was used to remove unmapped sequences and PCR duplicates, as above.
233 We generated an initial consensus genome from the resulting bam file with *BCFtools* v1.9 (Li
234 *et al*. 2009), using the *mpileup*, *call*, *filter*, and *consensus* functions. For variable sites, we
235 produced a majority-rule consensus using the --variants-only and --multiallelic-caller options,
236 and for uncovered sites the reference genome base was called. The above steps were repeated
237 until the consensus could no longer be improved. The re-filtered sequence data was then re-
238 mapped onto the initial consensus genome sequence with *bowtie2*, using the above settings.
239 The *genomecov* function from *BEDtools* v2.17.0 (Quinlan and Hall 2010) was used to
240 identify gaps in the resulting alignment.
241
242 We attempted to fill these gaps, which likely consisted of diverged or difficult-to-assemble
243 regions. For this, we assembled the re-filtered sequence dataset into *de novo* contigs with the
244 MEGAHIT pipeline v1.1.4 (Li *et al*. 2015), using a minimum *k*-mer length of 21, a maximum
245 *k*-mer length of 63, and *k*-mer length increments of six. The MEGAHIT contigs were then
246 mapped onto the initial consensus genome sequence with the *blastn* tool from the *NCBI-*
247 *BLAST+* toolkit. Contigs that covered the gaps identified by *BEDtools* were incorporated into
248 the initial consensus genome sequence, unless a *blast* comparison against the NCBI
249 nucleotide database suggested a closer match to non-*Nannochloropsis* taxa.
250
251 We repeated the *bowtie2* through gap-filling steps iteratively, using the previous consensus
252 sequence as reference, until a gap-free consensus was obtained. The re-filtered sequence data
253 were again mapped, the resulting final assembly was visually inspected and the consensus
254 was corrected where necessary. This was to ensure the fidelity of the consensus sequence,
255 which incorporated *de novo*-assembled contigs that could potentially be problematic, due to
256 the fragmented nature and deaminated sites of ancient DNA impeding accurate assembly
257 (Seitz and Nieselt 2017).
258
259 Annotation of the chloroplast genome was carried out with *GeSeq* (Tillich *et al*. 2017), using
260 the available *Nannochloropsis* chloroplast genomes for the annotation. (Accession codes
261 provided in Supplementary Table S2.) The resulting annotated chloroplast was visualised
262 with *OGDRAW* (Greiner *et al*. 2019).
263
264 The same assembly and annotation methods outlined above were used to reconstruct the
265 mitochondrial palaeogenome sequence, where the initial mapping assembly was based on the

N. limnetica mitochondrial sequence (NCBI GenBank accession: NC_022256.1). The final annotation was carried out by comparison against all available *Nannochloropsis* mitochondrial genomes. (Accession codes provided in Supplementary Table S2.)

If the *Nannochloropsis* sequences derived from more than one taxon, then alignment to the *N. limnetica* chloroplast genome could introduce reference bias, which would underestimate the diversity of the *Nannochloropsis* sequences present. We therefore reconstructed *Nannochloropsis* chloroplast genomes, but using the six available *Nannochloropsis* chloroplast genome sequences, including *N. limnetica*, as starting points. (Accession codes are in Supplementary Table S1.) The assembly of the consensus sequences followed the same method outlined above, but with two modifications to account for the mapping rate being too low for complete genome reconstruction based on alignment to the non-*N. limnetica* reference sequences. First, consensus sequences were called with *SAMtools*, which does not incorporate reference bases into the consensus at uncovered sites. Secondly, neither additional gap filling, nor manual curation was implemented.

Assembly of high and low frequency variant consensus sequences

The within-sample variants in each reconstructed organelle palaeogenome was explored by creating two consensus sequences, which included either high or low frequency variants at multiallelic sites. For each sample, the initial filtered sequence data were mapped onto the reconstructed *Nannochloropsis* chloroplast palaeogenome sequence with *bowtie2* using default settings. Unmapped and duplicate sequences were removed with *SAMtools*, as above. We used the *BCFtools mpileup*, *call*, and *normalize* functions to identify the variant sites in the mapped dataset, using the --skip-indels, --variants-only, and --multiallelic-caller options. The resulting SNPs were divided into two sets, based on either high or low frequency variants. High frequency variants were defined as those present in the reconstructed reference genome sequence. Both sets were further filtered to only include SNPs with a quality score of 30 or higher and a coverage of at least half the average coverage of the mapping assembly (minimum coverage: AND10_938=22x, AND11_1102=14x). We then generated the high and low frequency variant consensus sequences using the consensus function in *BCFTools*. The above method was repeated for the reconstructed *Nannochloropsis* mitochondrial genome sequence in order to generate comparable consensus sequences of high and low frequency variants (minimum coverage: AND10_938=16x, AND11_1102=10x).

Analysis of ancient DNA damage patterns

We checked for the presence of characteristic ancient DNA damage patterns for nuclear genome alignments consisting of >1000 mapped sequences aligned, which included human and the five *Nannochloropsis* nuclear genomes. We further analysed damage patterns for sequences aligned to both the *N. limnetica* reference and reconstructed organelle genomes, as well as those aligned to the high or low frequency variant data sets. Damage analysis was conducted with *mapDamage* v2.0.8 (Jónsson *et al*. 2013) using the following settings: --merge-reference-sequences and --length=160.

Phylogenetic analysis of the reconstructed environmental palaeogenomes

We determined the phylogenetic placement of our reconstructed organelle genomes, including those derived from high and low frequency variant calls, within *Nannochloropsis*. For this, we created five alignments using the nine *Nannochloropsis* strains that have both mitochondrial and chloroplast genome sequences available (Supplementary Table S2), with the addition of: (1+2) reconstructed organelle genome sequence, (3+4) organelle genome high/low frequency variant organelle consensus sequences, and (5) reconstructed chloroplast genome sequence and consensus sequences derived from the alternative *Nannochloropsis* genome starting points. We generated alignments using *MAFFT* v7.427 (Katoh and Standley 2013) with the maxiterate=1000 setting, which was used for the construction of a maximum likelihood tree in *RAxML* v8.1.12 (Stamatakis 2014) using the GTRGAMMA model and without outgroup specified. We assessed branch support using 1000 replicates of rapid bootstrapping. Due to the limited availability of *Nannochloropsis* reference organelle genomes, we also reconstructed comparable phylogenies using ~1100 bp of *rbcL*, a chloroplast barcode locus (Fawley *et al*. 2015). We extracted *rbcL* data from all *Nannochloropsis* species accessions on NCBI GenBank (Supplementary Table S2) and all reconstructed chloroplast genomes, to give three chloroplast alignments analogous to those listed above. The *rbcL* alignments were analysed using the same phylogenetic methods and settings.

***Nannochloropsis* variant proportions and haplogroup diversity estimation**

We calculated the proportions of high and low variants in the sequences aligned to our reconstructed *Nannochloropsis* mitochondrial and chloroplast genomes, in order to estimate the diversity of major haplogroups. For each sample, we first mapped the initial filtered sequence data onto the high and low frequency variant consensus sequences with *bowtie2*. To

avoid potential reference biases, and for each organelle genome, the sequence data were mapped separately against both frequency consensus sequences. The resulting bam files were then merged with *SAMtools* merge. We removed exact sequence duplicates from the merged bam file by randomly retaining one copy. This step was replicated five times to examine its impact on the estimated variant proportion. After filtering, remaining duplicate sequences - those with identical mapping coordinates - were removed with *SAMtools* rmdup. We then called SNPs from the duplicate-removed bam files using *BCFTools* under the same settings as used in the assembly of the high and low frequency variant consensus sequences. Depending on the variant, the resulting SNPs were included in three data sets: (1) all variants, (2) with C/T and G/A variants removed (to reduce the impact of deaminated sites), or (3) transversions only (to further reduce the impact of deaminated sites). For each SNP in these data sets, the proportion of reference and alternative alleles was calculated. We removed rare alleles occurring at a proportion of <0.1, as these could be due to noise.

To infer the minimum number of haplogroups in each reconstructed organelle genome sequence, we inspected the phasing of adjacent SNPs that were linked by the same read in the duplicate-removed bam files, akin to the method used by Søe *et al*. (2018). For this, we first identified all positions, from both samples, where two or more transversion-only SNPs occurred within 35 bp windows. We then examined the allelic state in mapped sequences that fully covered each of these linked positions. We recorded the combination of alleles to calculate the observed haplotype diversity at each of the linked positions. We removed low frequency haplotypes, which were defined as those with <3 sequences or <15% of all sequences that covered a linked position, and the remaining haplotypes were scored.

Reanalysis of published *sed*aDNA data and *Nannochloropsis* estimation

We performed a meta-analysis of the global prevalence of *Nannochloropsis* since the last ice age using published and available lake *sed*aDNA data sets. Three published shotgun datasets from Lake Hill, Alaska (Graham *et al*. 2016, Wang *et al*. 2017), Charlie Lake, BC and Spring Lake, AB (Pedersen *et al*. 2016), and Hässeldala Port, Sweden (Parducci *et al*. 2019) were reanalysed for the presence of *Nannochloropsis* using the same nuclear genome methods as used in this study. Furthermore, a metabarcode dataset was reanalysed from Skartjørna, Svalbard (Alsos *et al*. 2016), using the same methods for analysis as the original study, but lowering the minimum barcode length to 10 bp, in order to retain the shorter *Nannochloropsis* sequences. These data sets were supplemented with studies that reported

Nannochloropsis, including; Bliss Lake, Greenland (Epp *et al*. 2015), Qinghai Lake, China (Li *et al*. 2016), Lielais Svētiņu, Latvia (Stivrins *et al*. 2018), Lake Øvre Æråsvatnet (Alsos *et al.* in prep.), and Jodavannet, Svalbard (Voldstad *et al*. submitted). We estimated the occurrence and abundance of *Nannochloropsis* in 5000-year time windows for the above datasets. Abundance was coarsely divided into four categories: (1) dominant, scored when *Nannochloropsis* was the only taxon detected or most abundant of the taxa identified in the sequence data; (2) common, assigned when it was in the top 10 most abundant taxa identified; (3) rare, scored for any other detections, and (4) absent, assigned if *Nannochloropsis* was not detected.

Our meta-analysis showed that the data set from Spring Lake, AB (Pedersen *et al*. 2016), had the highest number of *Nannochloropsis* shotgun sequences, and so we subjected these data to further analysis. We mapped the sequence data from each sample onto the *N. limnetica* reference chloroplast genome with *bowtie2*, using the same settings as used for our Andøya samples. *N. limnetica* was chosen as the reference as it is the only known non-marine species and the continental location of Spring Lake precludes a marine influence. We used *SAMtools* to merge the resulting assemblies from each sample and called a single consensus genome sequence. We aligned this consensus genome sequence to all *Nannochloropsis* reference chloroplast genome sequences, as well as the Andøya high and low frequency variant chloroplast consensus sequences. The alignment was carried out using *MAFFT* and a maximum likelihood tree was constructed with *RAxML,* both tools used the same settings as described previously.

Results

Identification of *Nannochloropsis* taxa present

We shotgun sequenced two contemporaneous *sed*aDNA samples from Lake Øvre Æråsvatnet for identification of the present taxa. The results are reported for AND10_938 and AND11_1102 respectively throughout this section. We generated 223,999,748 and 133,265,048 paired-end reads, of which 127,429,489 and 53,000,389 reads were retained after filtering (Supplementary Table S3).

Based on comparison to the NCBI nucleotide database, we could identify 29,488 (1.47%) and 32,720 (1.64%) of the two million sequences used from each sample. The majority of the

identified taxa were bacterial, although *Nannochloropsis* constituted ~20% of the identified sequences (Table 2). A detailed breakdown of the metagenomic results is provided in Supplementary Figure S1 and Supplementary Table S4.

Identification of the sequence data via the nuclear genome panel indicated that *Nannochloropsis limnetica* was the most abundant taxa in the data, based on both the filtered counts as well as those corrected for genome size (Figure 2 and Supplementary Table S1). The *N. limnetica* genome resulted in 681,934 and 321,317 sequences mapped, 20,351 and 9320 sequences per 1 Mb and a coverage of 1.13x and 0.48x for the samples. All other *Nannochloropsis* nuclear genomes had lower rates than *N. limnetica.* The most sequence abundant non-*Nannochloropsis* taxon was human, with 2253 and 11,101 sequences, 0.69 and 3.41 sequences per 1 Mb. All alignments against the *Nannochloropsis* nuclear genomes indicated that the material was of ancient origin due to the presence of deaminated bases (Supplementary Figure S2 and S3), on the contrary, the human alignments lacked such damage profiles (Supplementary Figure S3). A comparison between the exotic and non-exotic taxa is given in Table 3.

The most abundant species in both organelle reference panels was *Nannochloropsis limnetica*, with 37,934 (30.7% of total) and 23,631 (33.1%) sequences for the chloroplast and 14,097 (11.4%) and 8609 (12.1%) for the mitochondria (Supplementary Table S5). The most abundant non-*Nannochloropsis* species was the algae *Choricystis parasitica* with 1784 (1.5%) and 255 (0.4%) sequences for the chloroplast and 276 (0.2%) and 38 (0.1%) sequences for the mitochondria.

When comparing the sequence data mapped against the *Nannochloropsis* nuclear genomes, the genome with most unique sequences was *N. limnetica*, with 580,720 and 254,874 sequences. Similar results are obtained for the chloroplast, where *N. limnetica* again has the most uniquely mapped data, with 11,291 and 7719 sequences. (The unique and shared sequence counts are provided in Supplementary Table S1.) When comparing the sequence lengths for the chloroplast data, the uniquely mapped data tends to be shorter, with 54 and 49.9 bp, compared to the shared 59.1 and 54.1 bp, possibly due to the longer reads being easier to align as they allow for more mismatches. A box plot of the lengths is available in Supplementary Figure S4.

Reconstruction of the *Nannochloropsis* palaeo organelle genomes and phylogenetic placement

We reconstructed the organelle genomes for the *Nannochloropsis* variant present in Andøya. The resulting chloroplast sequence was 117,734 bp in length and had a coverage of 64.3x. The mitochondria was 38,543 bp in length and had a coverage of 62.4x. (Full description of the features is provided in Supplementary Table S6.) Ancient DNA damage estimation for the reconstructed organelle genomes, as well as the high and low frequency variants, indicate that the reconstructed genomes are ancient (Supplementary Figure S6 and S7). An overview of reconstructed palaeo organelle genomes is given in Figure 3, with a full annotated map in Supplementary Figure S5.

In the *Nannochloropsis* organelle phylogenies, both the reconstructed palaeo organelle genomes, as well as the high and low frequency variant sequences, group closest with the *N. limnetica* reference organelle genomes (Figure 4 and Supplementary Figure S8). A similar observation is made based on the *rbcL* phylogeny (Supplementary Figure S10 and S11). However, all trees indicate some distance between the reconstructed sequences and the available references, we therefore refer to the *Nannochloropsis* present in Andøya as *N.* cf. *limnetica*. The chloroplast phylogeny that include the consensus sequences reconstructed via the alternative reference and method has more complex grouping. The new sequences fail to converge on the reconstructed *N.* cf. *limnetica* genome, indicating that the alternative approach was not capable of fully reconstructing the present variant (Supplementary Figure S9 and S12).

Nannochloropsis cf. *limnetica* variant and haplotype estimation

We explored the variation present in the reconstructed organelle sequences and used it to estimate the haplotypes present for both samples. We detected an average of 338 and 96 transversion only variants for the chloroplast and mitochondria, respectively. For each sample, the average proportion of the transversion only alternative allele is 0.39 and 0.42 for the chloroplast variants and 0.4 and 0.43 for the mitochondrial variants. The positions and proportions for the variants is spread across both reconstructed organelle genomes (Figure 5 and Supplementary Figure S13). The number of transversion only linked positions is 70 and 21 for the chloroplast and mitochondria respectively. For each sample, the average number of haplotypes observed for the chloroplast linked positions is 2.09 and 1.93. The average for the mitochondrial linked positions is 2.29 and 2.05. The distribution of the observed haplotypes

is given in Supplementary Figure S14. A table of all variant proportions counts and observed haplotypes is provided in Supplementary Table S7.

Reanalysis of published *sed*aDNA and *Nannochloropsis* estimation

Nannochloropsis was either reported or could be detected in eight out of the 10 previously published *sed*aDNA studies, whereas no *Nannochloropsis* was detected based on shotgun *sed*aDNA data from the Lake Hill, Alaska and Hässedala Port, Sweden sites (Figure 6). Sequences could be mapped against the *N. limnetica* chloroplast genome for eight out of the 19 Spring Lake samples, which had between 273 and 3,192 sequences mapped. The merged Spring Lake consensus sequences was placed in a phylogeny along with the Andøya and reference *Nannochloropsis* chloroplast sequences and is given in Supplementary Figure S15.

Discussion

All of our analyses identified *Nannochloropsis* as the most abundant eukaryotic taxon in the LGM lake sediments from Andøya, confirming earlier results based on plant DNA metabarcoding. We observed ancient DNA deamination patterns for all reference sequence combinations, which supports the ancient authenticity of our data. The phylogenetic placement of both the reconstructed chloroplast and mitochondrial environmental palaeogenomes as well as the *rbcL* marker, indicate that the *Nannochloropsis* detected in Andøya is evolutionarily closest to *N. limnetica*. Given the distance of the Andøya reconstructed sequences from available *N. limnetica* sequences however, as well as the fact that not all *Nannochloropsis* species and strains are represented by sequence data, we therefore conservatively refer to the Andøya *Nannochloropsis* taxon as *N.* cf. *limnetica*.

The sequence material from our Andøya samples could potentially contain more than one *Nannochloropsis* species. An indication for this is the number of uniquely mapped sequences to other *Nannochloropsis* reference sequences and the phylogeny containing the alternative reconstructed consensus sequences, which don't fully converge on the reconstructed *N.* cf. *limnetica* sequence. This could, however, have several technical causes. First, the sequence data is of ancient origin and is thus highly fragmented and damaged, which could result in mismapped sequences (Orlando *et al*. 2015). Secondly, the reconstructed palaeogenome organelle sequences, as well as the high and low frequency consensus sequences might not capture all the variation present in the samples and thus fail to map sequence data from rarer

variants. Thirdly, the sheer sequence depth for these samples means that the above problems are more pronounced and that random maps can be expected, as is also observed from sequences mapping to the exotic nuclear genomes. Finally, the method used to construct the alternative consensus sequences might not be able to correct the more diverged regions of the chloroplast (Wei *et al*. 2013), as is suggested by the alternative *N. limnetica* consensus chloroplast grouping closer with the reference upon which it was based, rather than the reconstructed *N.* cf. *limnetica* chloroplast. The reconstruction issues could be due to a lack of gap filling and post assembly curation that the reconstructed *N.* cf. *limnetica* organelle genomes were subjected to. Thus, the identification and detection of potential *Nannochloropsis* species other than *N. limnetica* remains uncertain.

The sequence depth, as well as the within sample variation allowed for the reconstruction of high and low frequency variant consensus sequences, which represent the two extreme ends of the variation detected in the samples. The high frequency variants for both samples grouped together in the phylogeny, as well as the low frequency variants, indicating the presence of similar variants between the samples. The *Nannochloropsis* cf. *limnetica* organelle variant analysis indicates that most rare alleles have a proportion that falls between 0.39-0.43. This suggests that there are at minimum two abundant variants present. Further exploration of the possible number of haplotypes by inspecting linked positions, again indicate that there are two main haplogroups present, with a few SNP windows containing three haplogroups. This method, given the filtering and limited window size, is quite conservative and likely underestimates the haplotypes present. Furthermore, the above methods deal with unknown phasing between the isolated variants and the linked sites. So, although a specific variant or linked window indicates the presence of two haplogroups the actual number of haplogroups present in the data could be far higher, when fully independent. Part of the issue is the limited number of reference organelle sequences for *N. limnetica*, with just one variant sequenced for both the organelle and nuclear genome. The only alternative *N. limnetica* data available is a consensus that we called from the Spring Lake samples (Pedersen *et al*. 2016), but it is grouped with the reference upon it was constructed, possibly due to an overall lower coverage. Additional reference material from distinct populations or strains will help out with the detection and estimation of the haplogroups present in *sed*aDNA. In addition, the methods employed here are still novel and potential methodological and statistical advances will help with estimation and quantification of the variation present. Regardless, the comparable variant proportions and haplogroups for the

SNP windows observed, in combination with similar high and low frequency variant consensus sequences, suggest that both samples are comparable, which is as expected from the contemporaneous samples.

Nannochloropsis limnetica has not been reported before in Norway according to either the available GBIF data or published literature, though the genus could be detected based on reanalyses (Lammers pers. obs.) of metabarcoding data from modern sites (Alsos *et al*. 2018). The lack of morphological observations from contemporary sources could be due to the general difficulty of observing and identifying the taxa (Gladu *et al*. 1995, Andersen *et al*. 1998). The genus has been reported from various locations and periods based on *sed*aDNA, including Lake Øvre Æråsvatnet (Alsos *et al.* in prep.). For most *sed*aDNA samples where *Nannochloropsis* is detected, it occurs with various other taxa, though in a number of samples it is more abundant or the dominant taxa, such as the oldest samples of Bliss Lake, Greenland from 10,800 cal a BP (Epp *et al*. 2015), Spring Lake, Alberta, Canada, 9100-9300 cal a BP (Pedersen *et al*. 2016) and Jodavannet, Svalbard, 9000-11,000 cal a BP (Voldstad *et al*. submitted). Furthermore, based on modern occurrence data, *N. limnetica* seems to thrive in cold water and can handle resources stresses (Fawley and Fawley 2007, Pal *et al*. 2011). Therefore, the presence of *N.* cf. *limnetica* on Andøya is not unexpected during the LGM.

The abundance of *Nannochloropsis* material suggests a high biomass or turnover in Lake Øvre Æråsvatnet. The genus is known to be capable of forming algal blooms, which can reach up to 10^{10} cells per litre (Zhang *et al*. 2015), and such blooming events have been reported for *N. gaditana* in the Comacchio Lagoons, Italy (Andreoli *et al*. 1999) and *N. granulata* in the Yellow Sea (Zhang *et al*. 2015). *Nannochloropsis limnetica* itself was first described from spring blooms in Germany, reaching concentrations up to 5.7×10^9 cells per litre (Krienitz *et al*. 2000). Such blooms could explain the reported sequence abundance. The Andøya geochemical analysis resulted in high LOI values and organic elemental proportions for the period that was sampled for shotgun sequencing (Alsos *et al*. in prep.). In addition, stable isotope data suggest that the nitrogen and carbon is of a high trophic origin, most likely bird guano from an adjacent bird cliff (Alsos *et al*. in prep.), which corresponds with the detection of bird bones, such as little auk (*Alle alle*), from the Andøya sites (Vorren *et al*. 1988, Elverland and Alm 2012, Alsos *et al*. in prep.). The high inflow of nutrients into the lake could have resulted in eutrophication of the lake ecosystem and thus support blooms of *Nannochloropsis*.

The metagenomic and mapping analysis indicated the presence of some additional taxa. Another algal species, *Choricystis parasitica* (Trebouxiophyceae), was identified based on the organelle mappings. The taxa, however, was not abundant, with at most 1.45% of the identified chloroplast sequences assigned to it. Furthermore, the identification of *C. parasitica* remains somewhat uncertain as there no alternative references to closely related taxa, thus the actual species present might be different. The most common group of non-eukaryotic species observed is *Mycobacterium*, which made up the majority of the sequences identified through the metagenomic analysis but could not be identified to a more specific group. The overall proportion of material identified by the metagenomic analysis is low with 1.47% and 1.64% for samples AND10_938 and AND11_1102 respectively. These percentages suggest that in addition to the reported *Nannochloropsis* and *Mycobacterium*, there are probably additional taxa in our data that make up a significant part of the remaining sequences. Unfortunately, at this moment the available reference material does not allow for the identification of these hidden taxa.

Human is the most abundant taxa after the *Nannochloropsis* taxa, based on the number of mapped sequences against the nuclear genomes. However, due to the lack of ancient DNA damage patterns, the material for both samples originated from modern contamination. In addition, preservation of the 10.4 bp helical pitch periodicity in the fragment length distribution suggests cellular DNA (Pedersen *et al*. 2014). Such periodicity is unexpected of true ancient environmental data, as additional fragmentation over time will mask these earlier patterns (Graham *et al*. 2016). Although sequences could be mapped against several reference genomes, when the mapping counts were corrected for the genome size, most were indistinguishable from the exotic controls. Random mappings to a reference genome, especially to conserved regions, can be expected given the fragmented and damaged nature of ancient DNA (Prüfer *et al*. 2010). This is particularly true for larger genomes, as the chance of random alignments increases with size. Stringent filtering, correcting for genome size and validation of the results through ancient damage pattern analysis is required to rule out misidentifications. This can be problematic for rarer taxa, such as *Selaginella kraussiana*, which although it has mapping rates higher than most exotic taxa, cannot be confirmed present due to a lack of sequence material for reliable ancient DNA damage profiles.

Andøya has received research interested due to it being partially unglaciated during the LGM and possibly acting as a refugia for various taxa (Parducci *et al*. 2012, Birks *et al*. 2012, Parducci *et al*. 2012). Most studies were carried out with pollen and macrofossils (Vorren *et al*. 1988, Alm and Birks 1991, Vorren *et al*. 2013), with some studies utilizing *sed*aDNA (Parducci *et al*. 2012, Alsos *et al*. in prep.). The data presented in this study differs from the conventional metabarcode studies in that the shotgun data allows for the exploration of different questions. The ability to detect and identify *Nannochloropsis* microalgae could potentially be an informative proxy for estimating the salinity from sedimentary records, as the different species occur in either marine or brackish and freshwater habitats. The detection of the fresh or brackish water *N. limnetica* in Lake Øvre Æråsvatnet matches earlier studies that indicated a lacustrine sediment record (Alm 1993, Vorren *et al*. 2013, Vorren *et al*. 2015). Though additional reference sequences for the various *Nannochloropsis* species and strains, along with habitat preferences, will help out with the fresh and brackish water distinction as well as give additional information for palaeoecological reconstruction.

Our reconstructed *N.* cf. *limnetica* chloroplast palaeogenome represent the first reconstructed chloroplast genome that was derived from sedimentary ancient DNA. Although mitochondrial palaeogenomes have previously been reconstructed from cave sediments (Slon *et al*. 2017), and archaeological middens and latrines (Seersholm *et al*. 2016, Søe *et al*. 2018), the mitochondrial genome obtained is the first derived from lake sediments. The high depth of coverage for our sample-combined palaeogenomes (chloroplast: 64x, mitochondria: 62x) allowed for both their complete reconstruction as well as exploration of allele frequencies and haplotype diversity using *sed*aDNA.

Conclusion

This study explores several ways to analyse shotgun *sed*aDNA data. First the identification of the taxa present, via metagenomic and mapping analysis, which resulted in an abundance of *Nannochloropsis* cf. *limnetica* material in Andøya, expanding upon the earlier metabarcode results. Due to the amount of *N.* cf. *limnetica* material present, additional methods for the analysis of shotgun *sed*aDNA could be explored. The full organelle genomes could be reconstructed for the *N.* cf. *limnetica* variant present in Andøya, which allowed for finer placement the *Nannochloropsis* phylogeny, though the lack of reference information means that exact strain cannot be identified. In addition, the sequence depth allowed for estimation

of the population variation, both within and between the contemporaneous samples. The general within sample variation and variation observed in linked sites suggest the presence of at least two different *N.* cf. *limnetica* populations for both samples. The methods used for the population estimates are simple and likely underestimate the true variation. Potential future modelling methods should help with a more quantitative measurement of the present populations. Still, this study demonstrates that there is more information to be extracted from shotgun *sed*aDNA beyond species identification, which opens the door to answering more complex palaeoecological questions.

Author contribution

YL analysed the data and wrote the first draft of the manuscript; IGA carried out the fieldwork and extracted the DNA; PDH provided guidance on the analysis of the sequence data; All authors were involved in various stages of the design of the research and contributed to the final version of the manuscript.

Acknowledgements

This paper is a part of a larger project on the past environment of Andøya and we thank the Andøya team (Alsos *et al.* in prep.) for the usage their age-depth model. We thank Per Sjögren, Aage Paus and Ludovic Gielly for help with the fieldwork, Mikkel W. Pedersen for the library preparation and sequencing, Edana Lord, Love Dalén and Vendela K. Lagerholm for the usage of their *Lemmus lemmus* genome and Sandra Garcés Pastor for informative discussions. This project was financed by the Research Council of Norway (grant: 213692, Ancient DNA of NW Europe reveals responses of climate change; grant: 250963, ECOGEN – Ecosystems change and species persistence over time: a genome-based approach). Youri Lammers was financed by an internal PhD position at the Tromsø University Museum.

References

Alm, T. (1993) Øvre Æråsvatn - palynostratigraphy of a 22,000 to 10,000 BP lacustrine record on Andøya, northern Norway. Boreas 22, 171–188.

Alm, T. & Birks, H.H. (1991) Late Weichselian flora and vegetation of Andøya, Northern Norway-macrofossil (seed and fruit) evidence from Nedre Æråsvatn. Nordic Journal of Botany 11, 465–476.

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.

Alsos, I.G., Sjögren, P., Brown, A.G., *et al*. (in prep.) Late Glacial Maximum environmental condition of Andøya, a northern ecological "hotspot."

Alsos, I.G., Sjögren, P., Edwards, M.E., *et al*. (2016) Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. The Holocene 26, 627–642.

Andersen, R.A., Brett, R.W., Potter, D. & Sexton, J.P. (1998) Phylogeny of the Eustigmatophyceae Based upon 18S rDNA, with Emphasis on *Nannochloropsis*. Protist 149, 61–74.

Andreoli, C., Bresciani, E., Moro, I., Scarabel, L., La Rocca, N., Dalla Valle, L. & Ghion, F. (1999) A Survey on a Persistent Greenish Bloom in the Comacchio Lagoons (Ferrara, Italy). Botanica Marina 42.

Birks, H.H., Giesecke, T., Hewitt, G.M., Tzedakis, P.C., Bakke, J. & Birks, H.J.B. (2012) Comment on "Glacial survival of boreal trees in northern Scandinavia." Science 338, 742; author reply 742.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America 104, 14616–14621.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009) BLAST : architecture and applications. BMC Bioinformatics 10, 421.

691 Elverland, E. & Alm, T. (2012) A Late Weichselian Alle alle colony on Andøya, northern
692 Norway - a contribution to the history of an important Arctic environment. PhD thesis.
693 UiT - The Arctic University of Norway.

694 Epp, L.S., Gussarova, G., Boessenkool, S., *et al*. (2015) Lake sediment multi-taxon DNA
695 from North Greenland records early post-glacial appearance of vascular plants and
696 accurately tracks environmental changes. Quaternary Science Reviews 117, 152–163.

697 Fawley, K.P. & Fawley, M.W. (2007) Observations on the diversity and ecology of
698 freshwater Nannochloropsis (Eustigmatophyceae), with descriptions of new taxa. Protist
699 158, 325–336.

700 Fawley, M.W., Jameson, I. & Fawley, K.P. (2015) The phylogeny of the genus
701 *Nannochloropsis* (Monodopsidaceae, Eustigmatophyceae), with descriptions of *N.*
702 *australis sp. nov.* and *Microchloropsis gen. nov.* Phycologia 54, 545–552.

703 Fietz, S., Bleiss, W., Hepperle, D., Koppitz, H., Krienitz, L. & Nicklisch, A. (2005) First
704 record of *Nannochloropsis limnetica* (Eustigmatophyceae) in the autotrophic
705 picoplankton from Lake Baikal. Journal of Phycology 41, 780–790.

706 Gelin, F., Boogers, I., Noordeloos, A.A., Damste, J.S.S. & De Leeuw JW, R.R.A. (1997)
707 Resistant biomacromolecules in marine microalgae of the classes Eustigmatophyceae
708 and Chlorophyceae: Geochemical implications. Organic geochemistry 26, 659–675.

709 Gladu, P.K., Patterson, G.W., Wikfors, G.H. & Smith, B.C. (1995) Sterol fatty acid and
710 pigment characteristics of UTEX 2341 a marine eustigmatophyte identified previously as
711 *Chlorella minutissima* (Chlorophyceae). Journal of Phycology 31, 774–777.

712 Graham, R.W., Belmecheri, S., Choy, K., *et al*. (2016) Timing and causes of mid-Holocene
713 mammoth extinction on St. Paul Island, Alaska. Proceedings of the National Academy of
714 Sciences of the United States of America 113, 9310–9314.

715 Greiner, S., Lehwark, P. & Bock, R. (2019) OrganellarGenomeDRAW (OGDRAW) version
716 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic
717 acids research 47, W59–W64.

718 Hibberd, D.J. (1981) Notes on the taxonomy and nomenclature of the algal classes
719 Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). Botanical Journal of

the Linnean Society 82, 93–119.

Hughes, A.L.C., Gyllencreutz, R., Lohne, Ø.S., Mangerud, J. & Svendsen, J.I. (2016) The last Eurasian ice sheets - a chronological database and time-slice reconstruction, DATED-1. Boreas 45, 1–45.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J. & Tappu, R. (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS computational biology 12, e1004957.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F. & Orlando, L. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684.

Karlov, D.S., Marie, D., Sumbatyan, D.A., Chuvochina, M.S., Kulichevskaya, I.S., Alekhina, I.A. & Bulat, S.A. (2017) Microbial communities within the water column of freshwater Lake Radok, East Antarctica: predominant 16S rDNA phylotypes and bacterial cultures. Polar Biology 40, 823–836.

Karlson, B., Potter, D., Kuylenstierna, M. & Andersen, R.A. (1996) Ultrastructure, pigment composition, and 18S rRNA gene sequence for *Nannochloropsis granulata sp. nov.* (Monodopsidaceae, Eustigmatophyceae), a marine ultraplankter isolated from the Skagerrak, northeast Atlantic Ocean. Phycologia 35, 253–260.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30, 772–780.

Krienitz, L., Hepperle, D., Stich, H.-B. & Weiler, W. (2000) *Nannochloropsis limnetica* (Eustigmatophyceae), a new species of picoplankton from freshwater. Phycologia 39, 219–227.

Kryvenda, A., Rybalka, N., Wolf, M. & Friedl, T. (2018) Species distinctions among closely related strains of Eustigmatophyceae (Stramenopiles) emphasizing ITS2 sequence-structure data: Eustigmatos and Vischeria. European Journal of Phycology 53, 471–491.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

749 Leeuw, J.W. de, de Leeuw, J.W., Versteegh, G.J.M. & van Bergen, P.F. (2005)
750 Biomacromolecules of algae and plants and their fossil analogues. Plants and Climate
751 Change, 209–233.

752 Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. (2015) MEGAHIT: an ultra-fast
753 single-node solution for large and complex metagenomics assembly via succinct de
754 Bruijn graph. Bioinformatics 31, 1674–1676.

755 Li, G., Dong, H., Hou, W., Wang, S., Jiang, H., Yang, J. & Wu, G. (2016) Temporal
756 Succession of Ancient Phytoplankton Community in Qinghai Lake and Implication for
757 Paleo-environmental Change. Scientific reports 6, 19769.

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

761 Orlando, L., Gilbert, M.T.P. & Willerslev, E. (2015) Reconstructing ancient genomes and
762 epigenomes. Nature reviews. Genetics 16, 395–408.

763 Pal, D., Khozin-Goldberg, I., Cohen, Z. & Boussiba, S. (2011) The effect of light, salinity,
764 and nitrogen availability on lipid production by *Nannochloropsis sp.* Applied
765 Microbiology and Biotechnology 90, 1429–1441.

766 Parducci, L., Alsos, I.G., Unneberg, P., Pedersen, M.W., Han, L., Lammers, Y., Sakari
767 Salonen, J., Väliranta, M.M., Slotte, T. & Wohlfarth, B. (2019) Shotgun Environmental
768 DNA, Pollen, and Macrofossil Analysis of Lateglacial Lake Sediments From Southern
769 Sweden. Frontiers in Ecology and Evolution 7.

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

772 Parducci, L., Edwards, M.E., Bennett, K.D., *et al*. (2012) Response to Comment on "Glacial
773 Survival of Boreal Trees in Northern Scandinavia." Science 338, 742–742.

774 Parducci, L., Jørgensen, T., Tollefsrud, M.M., *et al*. (2012) Glacial survival of boreal trees in
775 northern Scandinavia. Science 335, 1083–1086.

776 Pedersen, J.S., Valen, E., Velazquez, A.M.V., *et al*. (2014) Genome-wide nucleosome map

and cytosine methylation levels of an ancient human genome. Genome research 24, 454–466.

Pedersen, M.W., Ruter, A., Schweger, C., *et al*. (2016) Postglacial viability and colonization in North America's ice-free corridor. Nature 537, 45–49.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J. & Green, R.E. (2010) Computational challenges in the analysis of ancient DNA. Genome biology 11, R47.

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Radakovits, R., Jinkerson, R.E., Fuerstenberg, S.I., Tae, H., Settlage, R.E., Boore, J.L. & Posewitz, M.C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. Nature Communications 3.

Schwartz, A.S., Brown, R., Ajjawi, I., McCarren, J., Atilla, S., Bauman, N. & Richardson, T.H. (2018) Complete Genome Sequence of the Model Oleaginous Alga CCMP1894. Genome announcements 6.

Seersholm, F.V., Pedersen, M.W., Søe, M.J., *et al*. (2016) DNA evidence of bowhead whale exploitation by Greenlandic Paleo-Inuit 4,000 years ago. Nature communications 7, 13389.

Seitz, A. & Nieselt, K. (2017) Improving ancient DNA genome assembly. PeerJ 5, e3126.

Simpson, J.T. & Durbin, R. (2012) Efficient de novo assembly of large genomes using compressed data structures. Genome research 22, 549–556.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J. & Jakobsson, M. (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences 111, 2229–2234.

Slon, V., Hopfe, C., Weiß, C.L., *et al*. (2017) Neandertal and Denisovan DNA from Pleistocene sediments. Science 356, 605–608.

Smol, J.P., Birks, H.J. & Last, W.M. (2004) Tracking Environmental Change Using Lake Sediments: Volume 4: Zoological Indicators, Springer Science & Business Media.

805 Søe, M.J., Nejsum, P., Seersholm, F.V., *et al*. (2018) Ancient DNA from latrines in Northern
806 Europe and the Middle East (500 BC-1700 AD) reveals past parasites and diet. PloS one
807 13, e0195481.

808 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis
809 of large phylogenies. Bioinformatics 30, 1312–1313.

810 Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey,
811 E.S. & Bunce, M. (2017) Ecosystem biomonitoring with eDNA: metabarcoding across
812 the tree of life in a tropical marine environment. Scientific Reports 7.

813 Stivrins, N., Soininen, J., Tõnno, I., Freiberg, R., Veski, S. & Kisand, V. (2018) Towards
814 understanding the abundance of non-pollen palynomorphs: A comparison of fossil algae,
815 algal pigments and seda DNA from temperate lake sediments. Review of Palaeobotany
816 and Palynology 249, 9–15.

817 Su, C.-H., Chien, L.-J., Gomes, J., Lin, Y.-S., Yu, Y.-K., Liou, J.-S. & Syu, R.-J. (2011)
818 Factors affecting lipid accumulation by *Nannochloropsis oculata* in a two-stage
819 cultivation process. Journal of Applied Phycology 23, 903–908.

820 Suda, S., Atsumi, M. & Miyashita, H. (2002) Taxonomic characterization of a marine
821 *Nannochloropsis species*, *N. oceanica sp. nov.* (Eustigmatophyceae). Phycologia 41,
822 273–279.

823 Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. (2018) Environmental DNA for functional
824 diversity, Oxford: Oxford University Press.

825 Taberlet, P., Prud'Homme, S.M., Campione, E., *et al*. (2012) Soil sampling and isolation of
826 extracellular DNA from large amount of starting material suitable for metabarcoding
827 studies. Molecular ecology 21, 1816–1820.

828 Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R. & Greiner,
829 S. (2017) GeSeq - versatile and accurate annotation of organelle genomes. Nucleic acids
830 research 45, W6–W11.

831 Vieler, A., Wu, G., Tsai, C.-H., *et al*. (2012) Genome, functional gene annotation, and
832 nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica*
833 CCMP1779. PLoS genetics 8, e1003064.

Voldstad, L.H., Alsos, I.G., Farnsworth, W., Heintzman, P.D., Håkansson, L., Kjellman, S.E., Rouillard, A., Schomacker, A. & Eidesen, P.B. (submitted) A 12,000-year lake sediment ancient DNA record reveals a long standing high-Arctic plant diversity hotspot in northern Svalbard.

Vorren, K.-D. (1978) Late and Middle Weichselian stratigraphy of Andøya, north Norway. Boreas 7, 19–38.

Vorren, T.O., Rydningen, T.A., Baeten, N.J. & Laberg, J.S. (2015) Chronology and extent of the Lofoten-Vesterålen sector of the Scandinavian Ice Sheet from 26 to 16 cal. ka BP. Boreas 44, 445–458.

Vorren, T.O., Vorren, K.-D., Aasheim, O., Torbjørn Dahlgren, K.I., Forwick, M. & Hassel, K. (2013) Palaeoenvironment in northern Norway between 22.2 and 14.5 cal. ka BP. Boreas 42, 876–895.

Vorren, T.O., Vorren, K.-D., Torbjørn, A., Gulliksen, S. & Løvlie, R. (1988) The last deglaciation (20,000 to 11,000 B. P.) on Andoya, northern Norway. Boreas 17, 41–77.

Wang, Y., Heintzman, P.D., Newsom, L., Bigelow, N.H., Wooller, M.J., Shapiro, B. & Williams, J.W. (2017) The southern coastal Beringian land bridge: cryptic refugium or pseudorefugium for woody plants during the Last Glacial Maximum? Journal of Biogeography 44, 1559–1571.

Wei, L., Xin, Y., Wang, D., *et al*. (2013) Nannochloropsis plastid and mitochondrial phylogenomes reveal organelle diversification mechanism and intragenus phylotyping strategy in microalgae. BMC genomics 14, 534.

Zhang, X., Kan, J., Wang, J., Gu, H., Hu, J., Zhao, Y. & Sun, J. (2015) First record of a large-scale bloom-causing species *Nannochloropsis granulata* (Monodopsidaceae, Eustigmatophyceae) in China Sea waters. Ecotoxicology 24, 1430–1441.

Zinger, L., Bonin, A., Alsos, I.G., *et al*. (2019) DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology 28, 1857–1862.

Table and figures

Species	18S	*rbcL*	Chloroplast	Mitochondria	Nuclear genome
Nannochloropsis australis	Yes	Yes	No	No	No
Nannochloropsis gaditana	Yes	Yes	Yes	Yes	Yes
Nannochloropsis granulata	Yes	Yes	Yes	Yes	Yes
Nannochloropsis limnetica	Yes	Yes	Yes	Yes	Yes
Nannochloropsis maritima	Yes	Yes	No	No	No
Nannochloropsis oceania	Yes	Yes	Yes	Yes	Yes
Nannochloropsis oculata	Yes	Yes	Yes	Yes	No
Nannochloropsis salina	Yes	Yes	Yes	Yes	Yes

Table 1. Overview of the *Nannochloropsis* species and the available genomic data.

	AND10_938			AND11_1102		
	N	I	A	N	I	A
Bacteria	18,852	64.9	0.9	21,873	66.9	1.1
Mycobacterium	6268	21.3	0.3	8535	26.1	0.4
Pseudomonas	920	3.1	0.05	904	2.8	0.05
Eukaryota	9333	31.7	0.5	9563	29.2	0.5
Nannochloropsis	5913	20	0.3	6179	18.9	0.3

Table 2. Summary of the most common (>500 sequences) taxa detected in the metagenomic analysis. N=Number of identified sequences, I=Percentage of identified sequences, A=Percentage of all sequences subjected to metagenomic analysis.

Group	AND10_938		AND11_1102	
	Count	Mean/Mb	Count	Mean/Mb
Non-exotics	32,186	1004	15,432	484.5
Non-exotics, excluding *Nannochloropsis*	29.8	0.05	11.8	0.02
Exotics	368	0.14	1605	0.49
Exotics, excluding human	54.5	0.05	22.5	0.01

Table 3. Comparison of the number of sequences obtained between the exotic and non-exotic taxa. Count=Mean count of sequences, Mean/Mb=Mean count of sequences per Mb of reference genome.

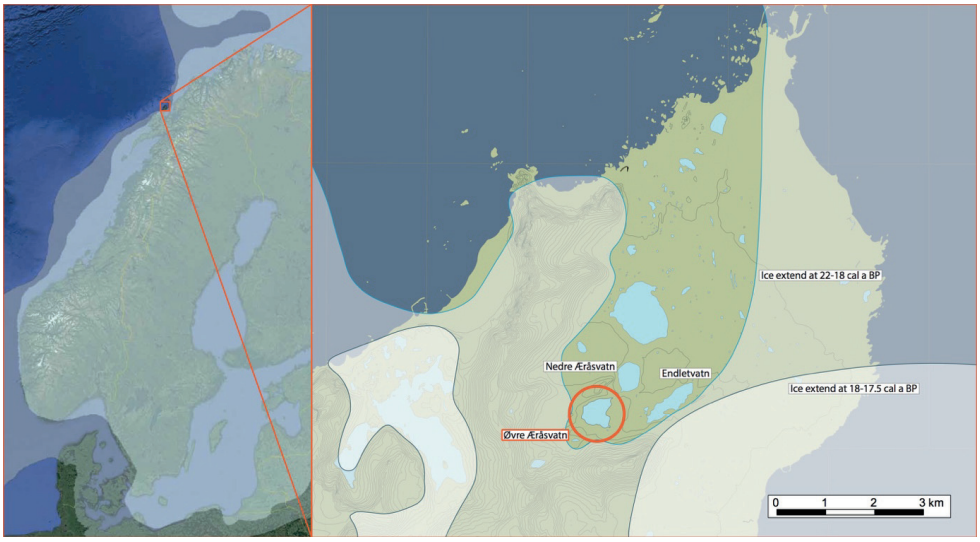


Figure 1. Location of Andøya in northwest Norway and Lake Øvre Æråsvatnet (circled in red). The regional ice extend for Scandinavia has been plotted for 22 (outer) and 17 (inner) cal a BP and is based on Hughes *et al*. (2016). The local ice extend is plotted for 22-18 and 18-17.5 cal a BP is plotted and based on Vorren *et al*. (2015).

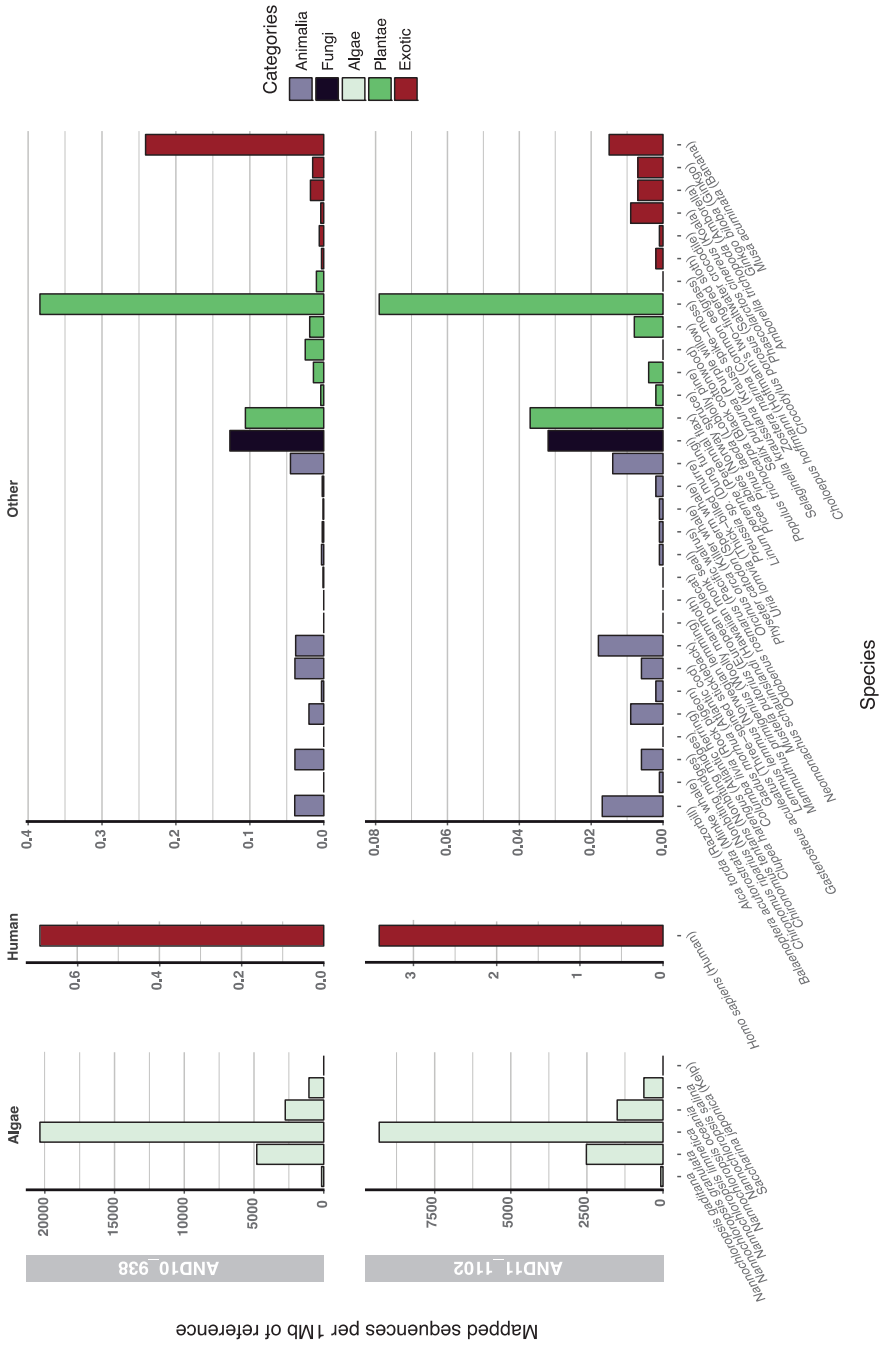


Figure 2. Taxonomic composition of the Andøya samples based on alignment to our nuclear genome reference panel. For readability, the algal and human results have been plotted separately.

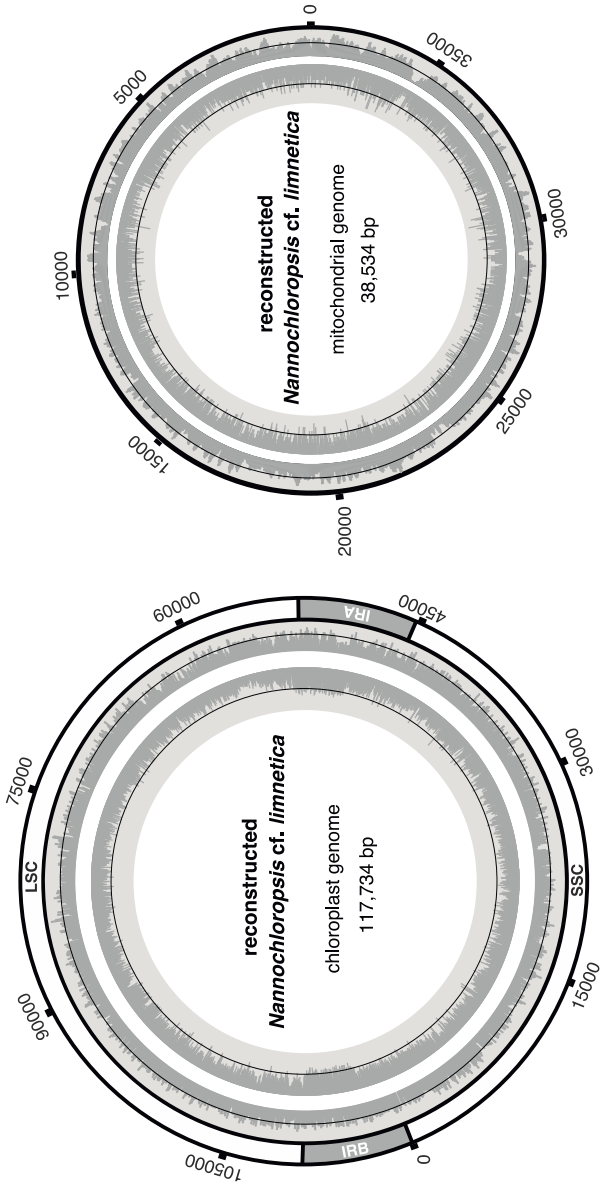


Figure 3. The reconstructed *Nannochloropsis* cf. *limnetica* chloroplast and mitochondrial genomes. The innermost circle contains a distribution of the GC content in dark grey, with the black bar representing the 50% mark. The outer distribution contains the coverage for the assembly in dark grey, with the black line representing the average coverage of 64.3x for the chloroplast and 64.9x for the mitochondria.

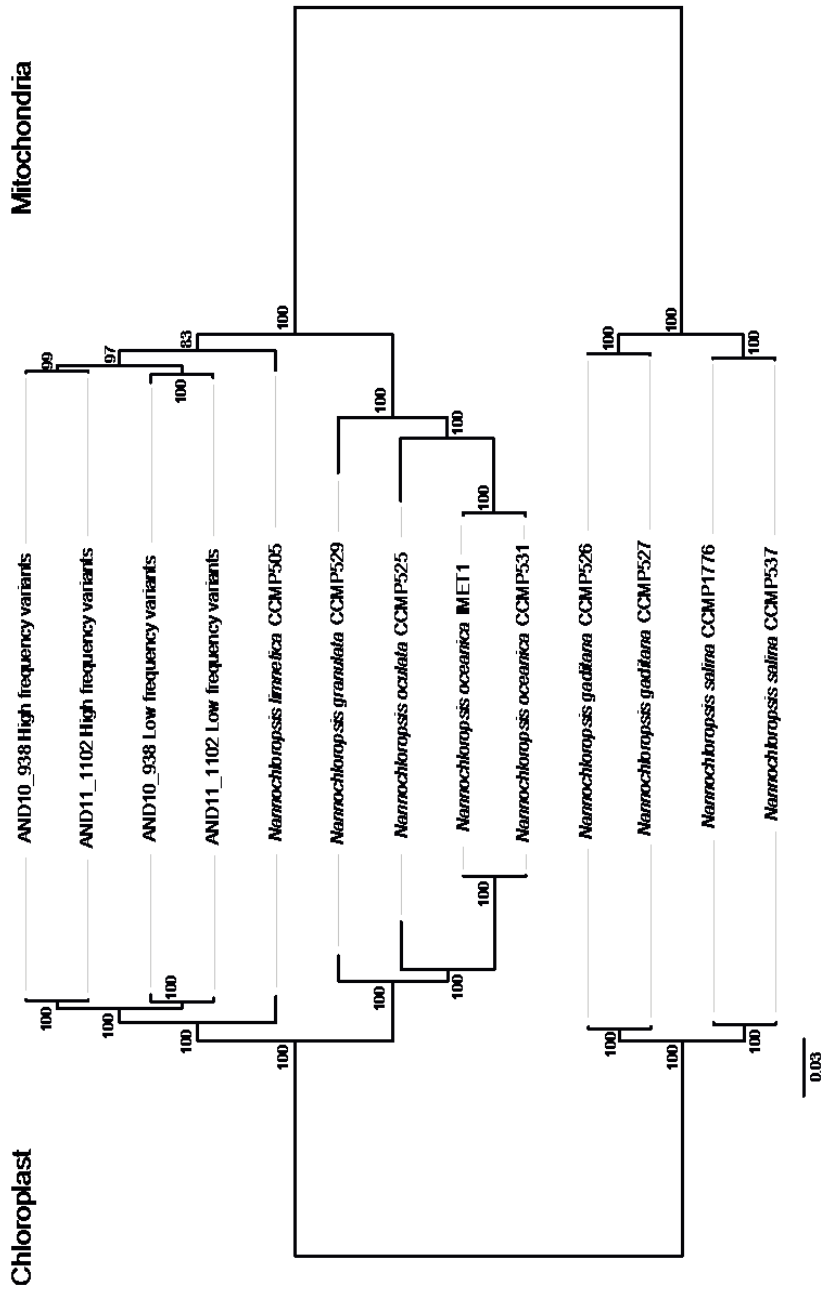


Figure 4. Phylogeny of the *Nannochloropsis* chloroplast (left) and mitochondrial (right) sequences, including the reconstructed *N.* cf. *limnetica* high and low frequency variant consensus sequences, representing the extreme end of the present variation.

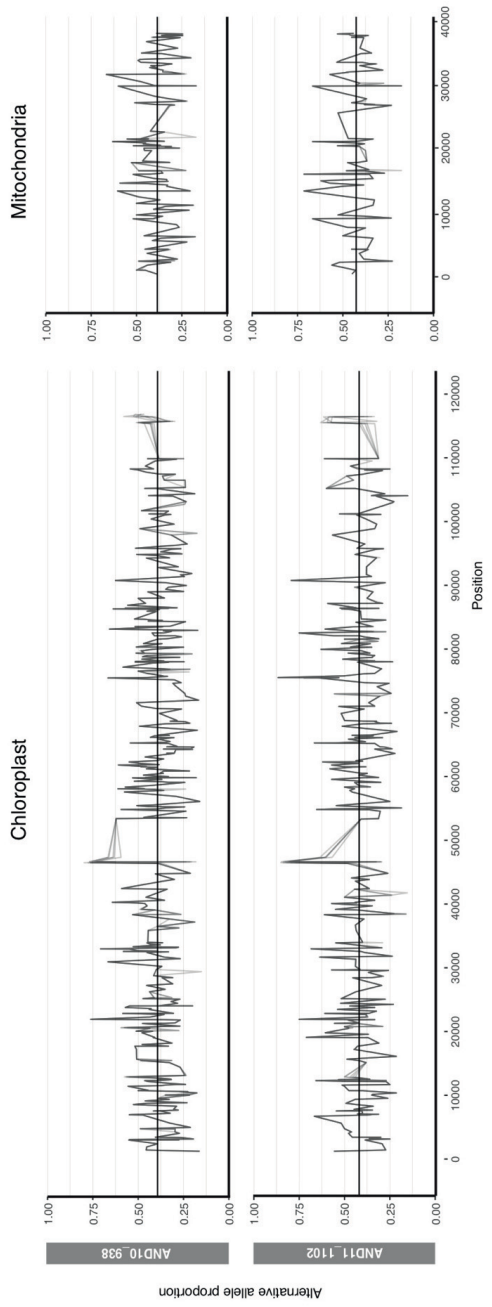


Figure 5: The proportion of the alternative alleles for the transversion only dataset across the organelle genomes. The horizontal black lines represent the averages: 0.39 and 0.42 for the chloroplast and 0.4 and 0.43 for the mitochondria, for samples AND10_938 and AND11_1102 respectively.

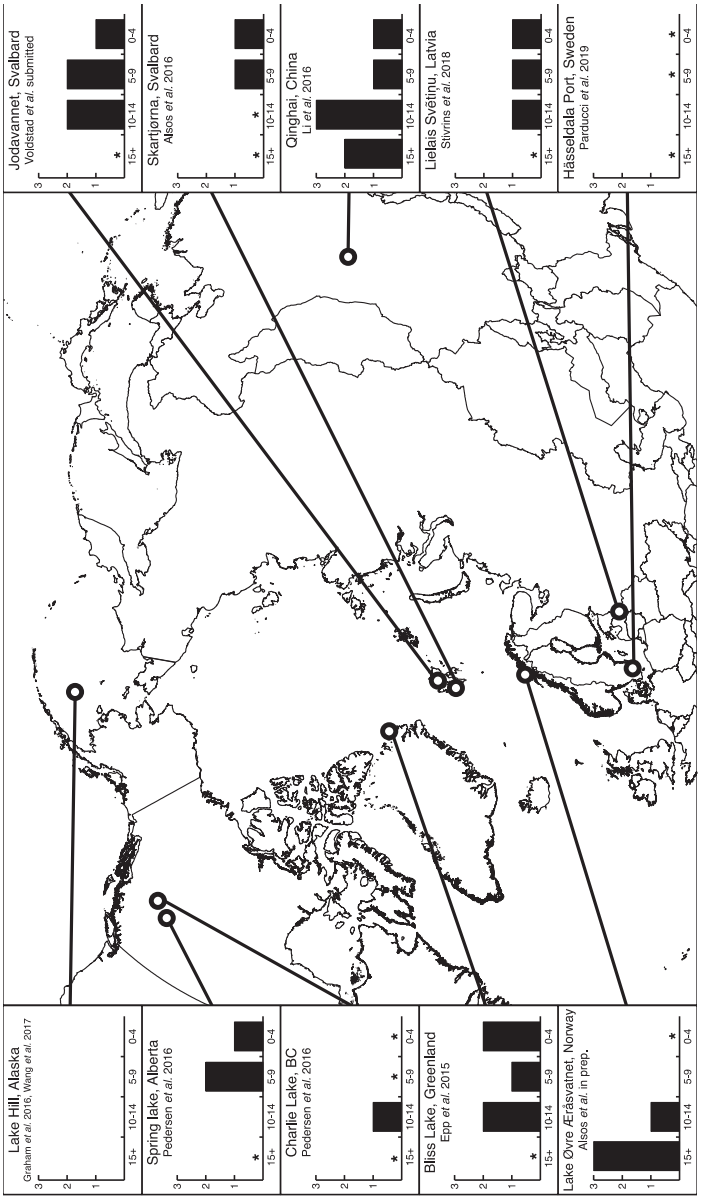


Figure 6. Overview of *Nannochloropsis* detections with *sed*aDNA, from both previously published results and reanalysed data. Plotted on the Y-axis is the inferred abundance of *Nannochloropsis*; 3: dominant in the period, 2: common, 1: rare and 0 is absent. Data was binned into 5000 years periods and might obscure finer details. No DNA was available for periods marked with an asterisk.

Supplementary Table S1: The nuclear and chloroplast sequences used for the mapping analysis. Includes the raw and filtered read counts, corrected read counts and coverage, as well as the number of shared and unique sequences between the *Nannochloropsis* nuclear and chloroplast sequences.

Supplementary Table S2: Lists of the *Nannochloropsis* chloroplast, mitochondria and *rbcL* sequences used for the phylogenies.

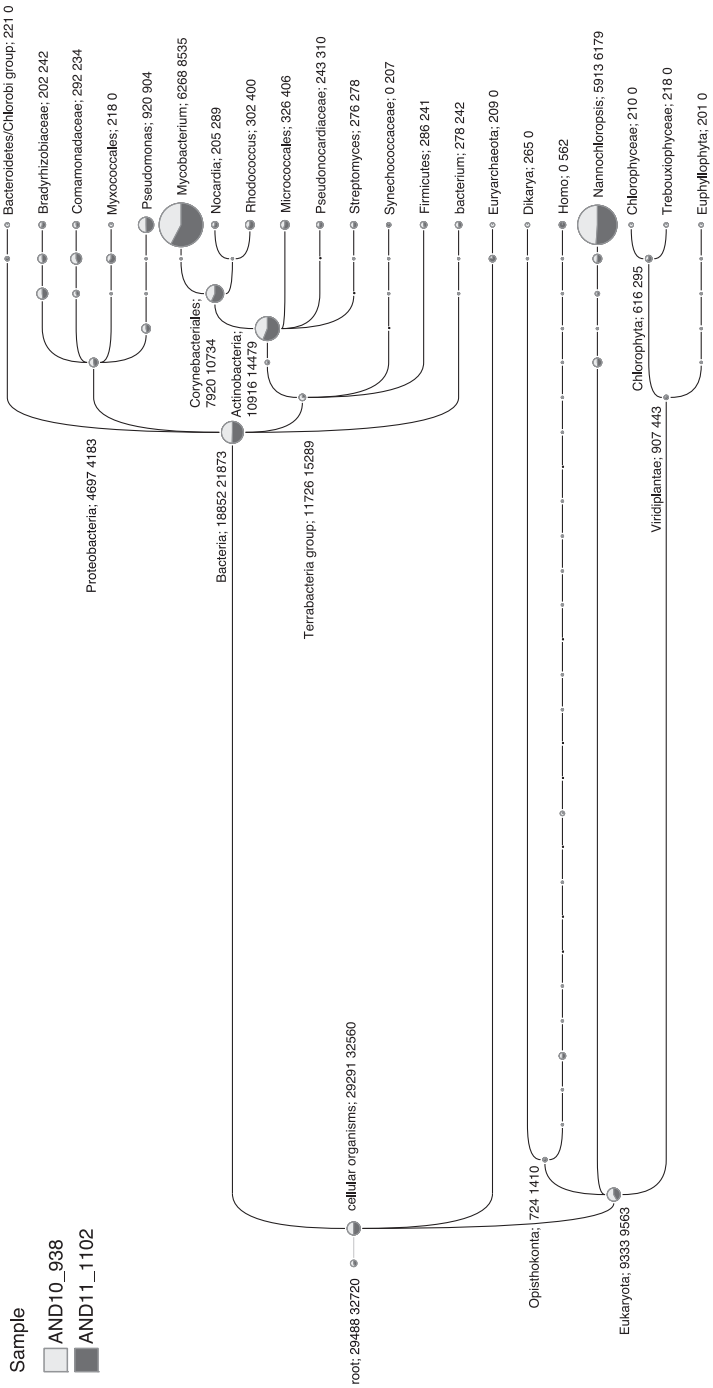
Supplementary Table S3: The raw and filtered read counts obtained via shotgun sequencing.

Supplementary Table S4: The MEGAN output in table form for both samples.

Supplementary Table S5: The number of sequences across the different LCA levels for the organelle mapping analysis.

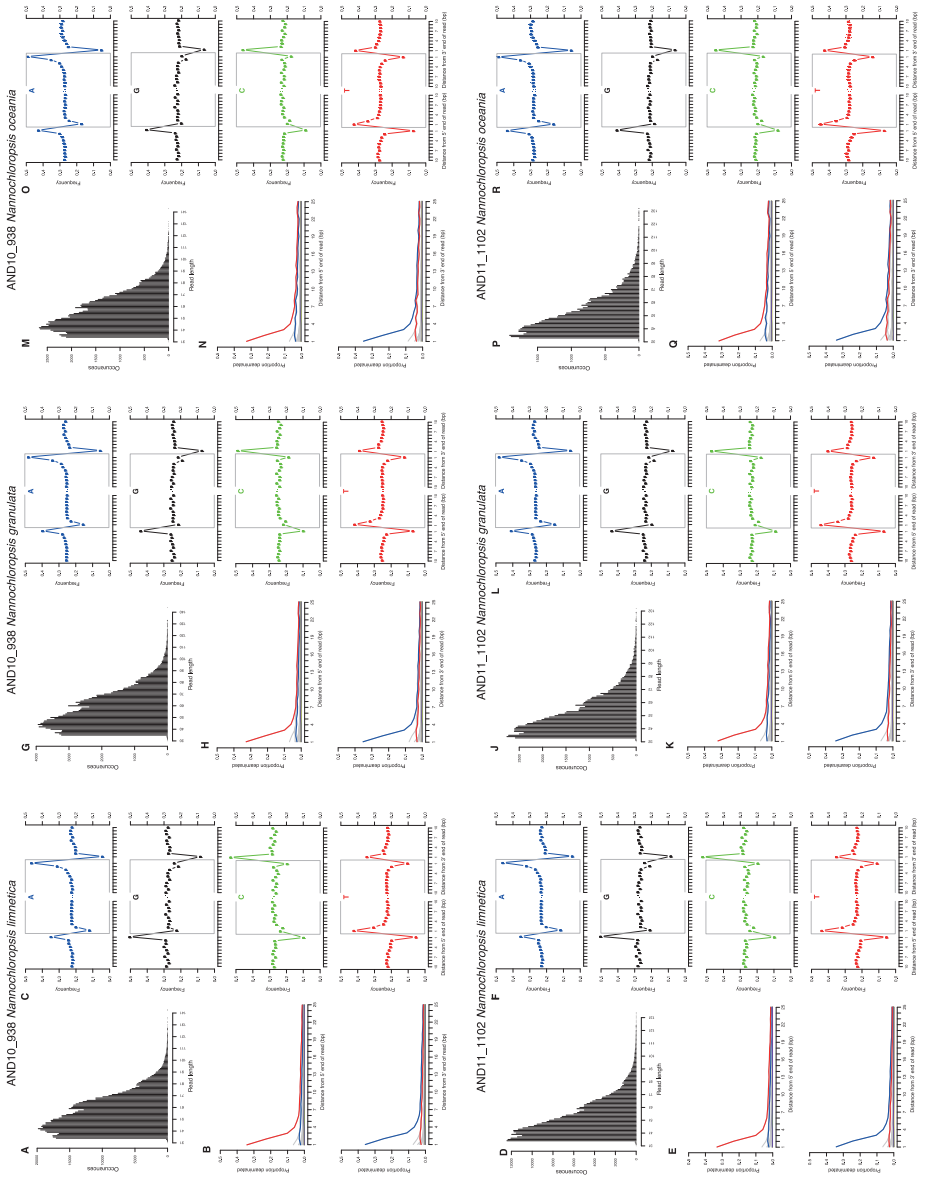
Supplementary Table S6: Summary of the reconstructed *Nannochloropsis* cf. *limnetica* organelle genomes, as well as the features on the *N. limnetica* reference.

Supplementary Table S7: Table containing the variant analysis output. Includes the average rare allele to reference allele proportions for both samples, reference genomes and filtering types. As well as the linked variant haplotype analysis.

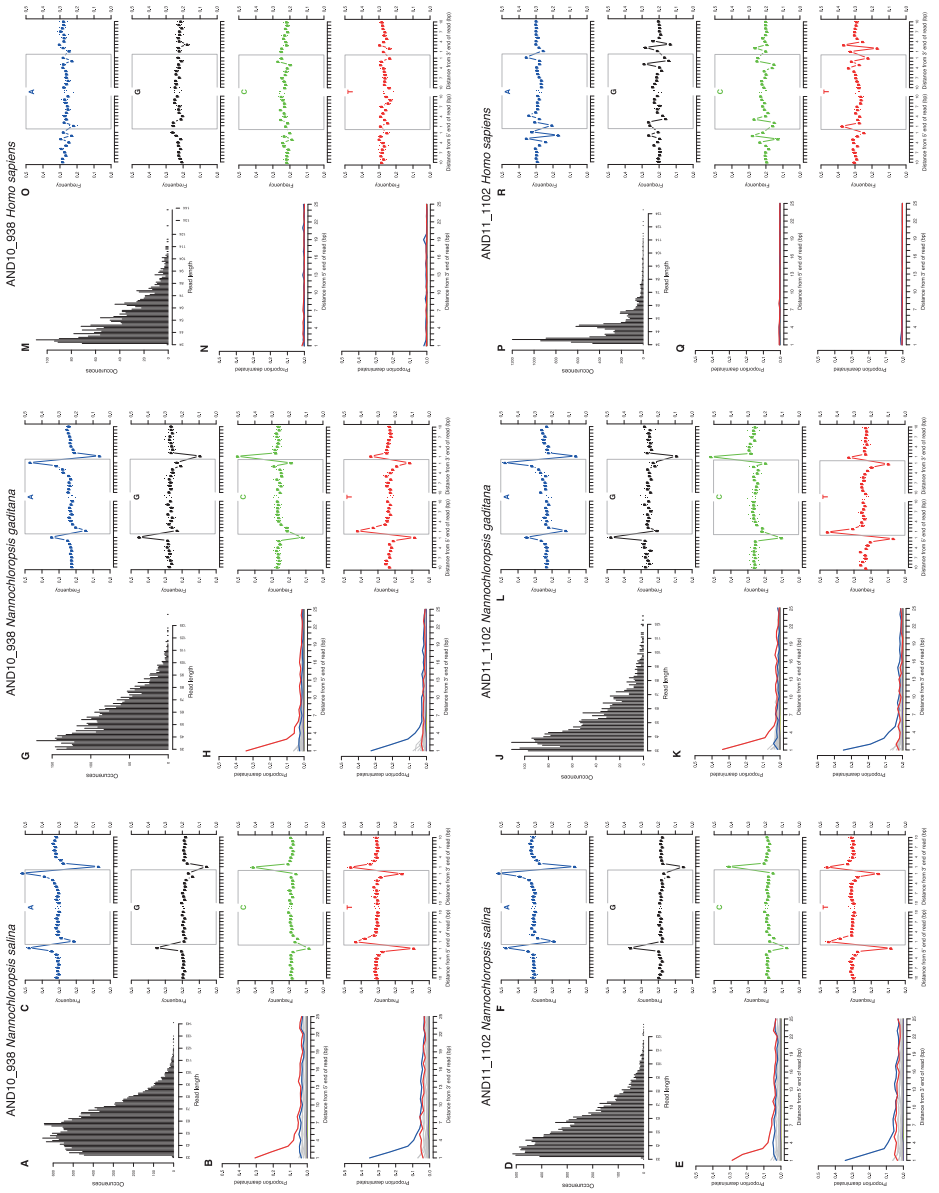


930

931 Supplementary Figure S1: Visualization of the MEGAN output.



Supplementary Figure S2: mapDamage output for the *Nannochloropsis limnetica*, *N. granulata* and *N. oceania* nuclear genomes.

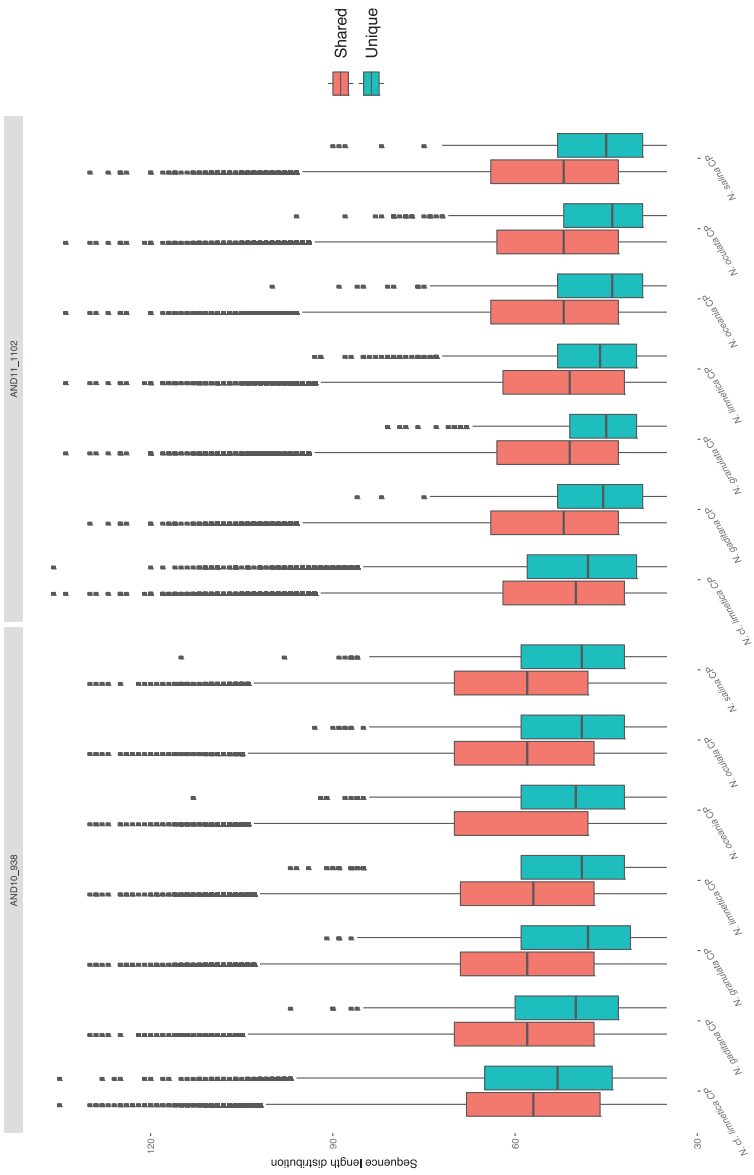


936

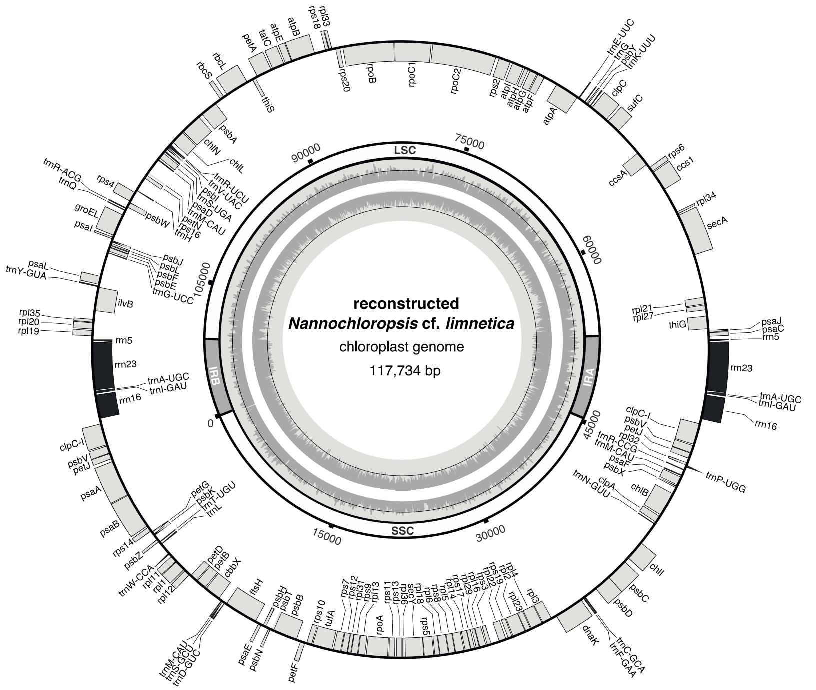
937 Supplementary Figure S3: mapDamage output for the *Nannochloropsis salina*, *N. gaditana*

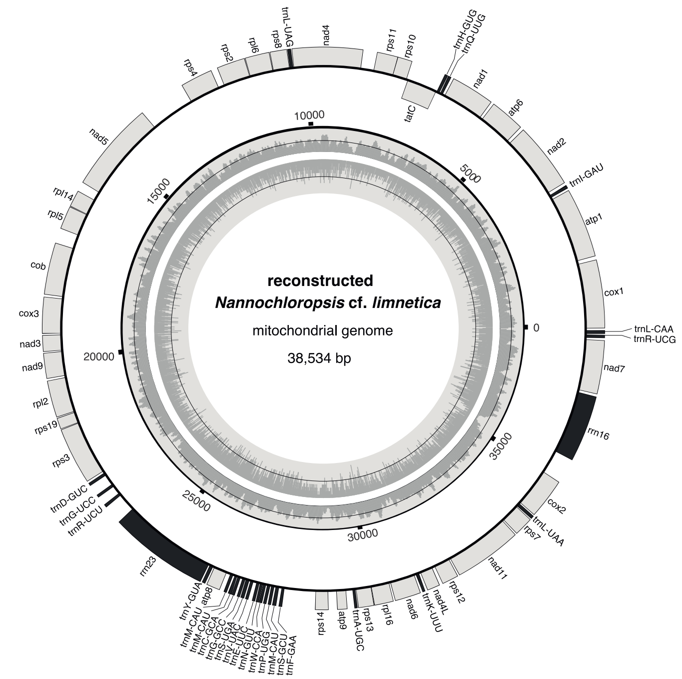
938 and *Homo sapiens* nuclear genomes.

939

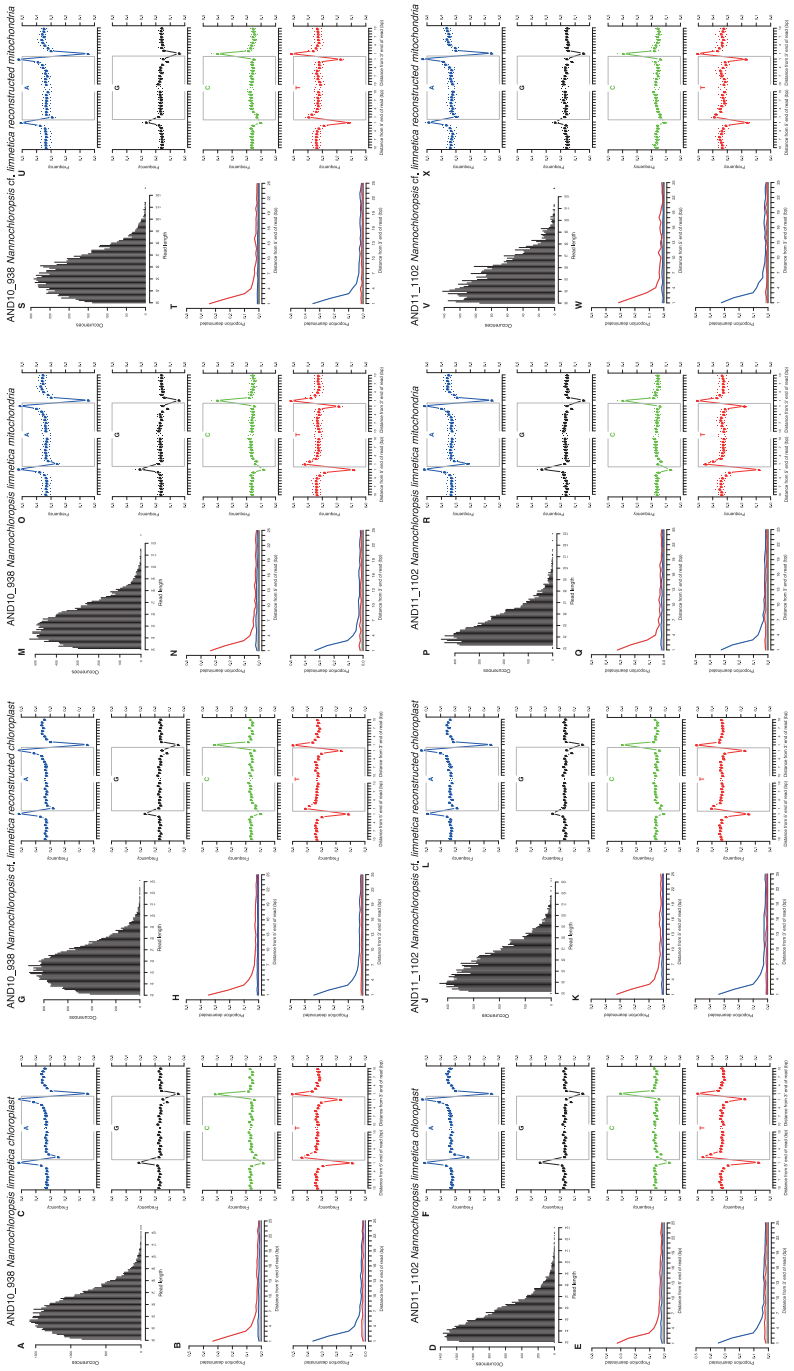


Supplementary Figure S4: Boxplots for the average fragment lengths for the shared and unique reads across the different *Nannochloropsis* chloroplast genomes.

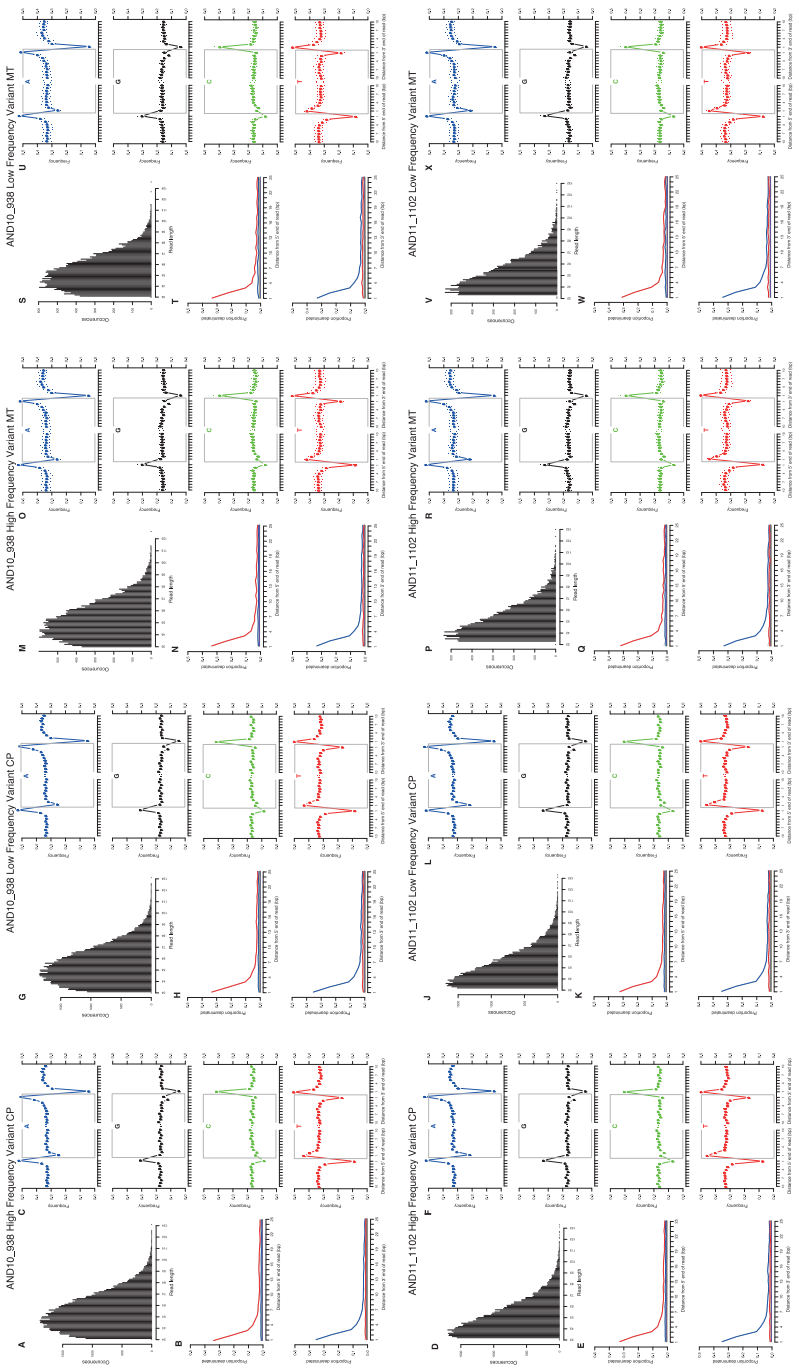

reconstructed
Nannochloropsis cf. limnetica
chloroplast genome
117,734 bp
LSC
SSC
IRA
IRB


reconstructed
Nannochloropsis cf. limnetica
mitochondrial genome
38,534 bp

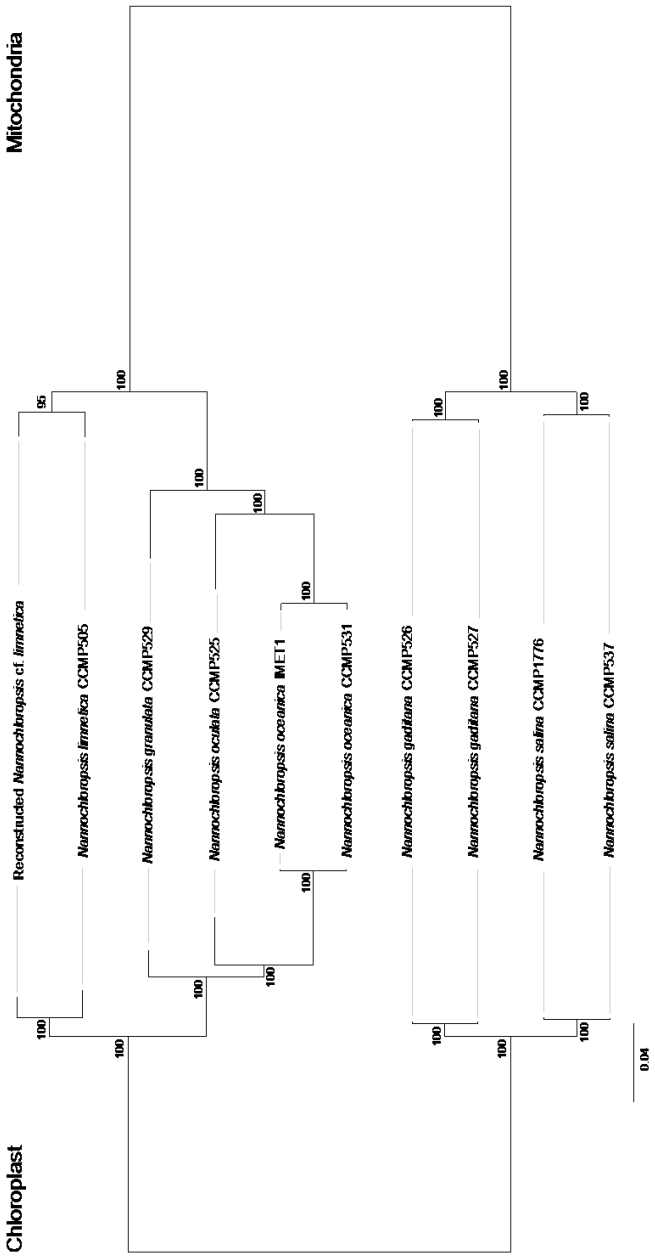
Supplementary Figure S5: Full annotated maps for the reconstructed *Nannochloropsis* cf. *limnetica* chloroplast and mitochondrial genomes. The innermost circle contains a distribution of the GC content in dark grey, with the black bar representing the 50% mark. The outer distribution contains the coverage for the assembly in dark grey, with the black line representing the average coverage of 64.3x for the chloroplast and 64.9x for the mitochondria. The genomic features are given on the outermost circle, where the coding genes are coloured grey and RNAs black. The features located on the inside are transcribed clockwise, those on the outside anticlockwise.



Supplementary Figure S6: mapDamage output for the *Nannochloropsis limnetica* organelle genomes and the reconstructed *N.* cf. *limnetica* organelle genomes.



Supplementary Figure S7: mapDamage output for the *Nannochloropsis* cf. *limnetica* high and low frequency variant consensus sequences.

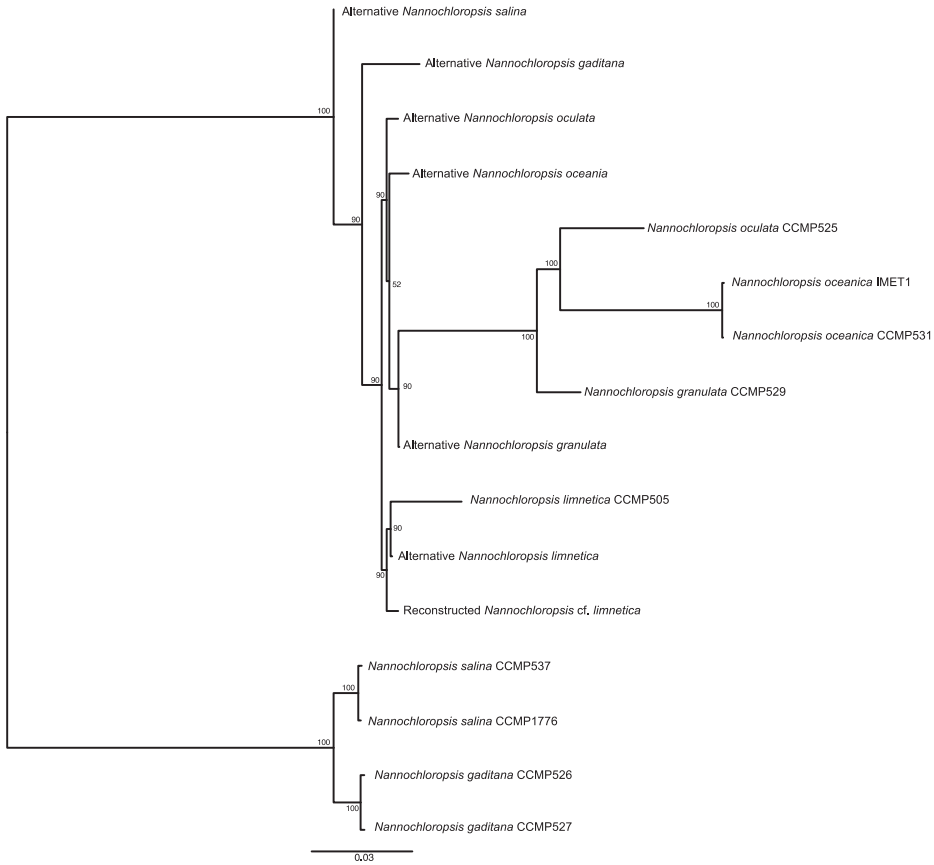


961

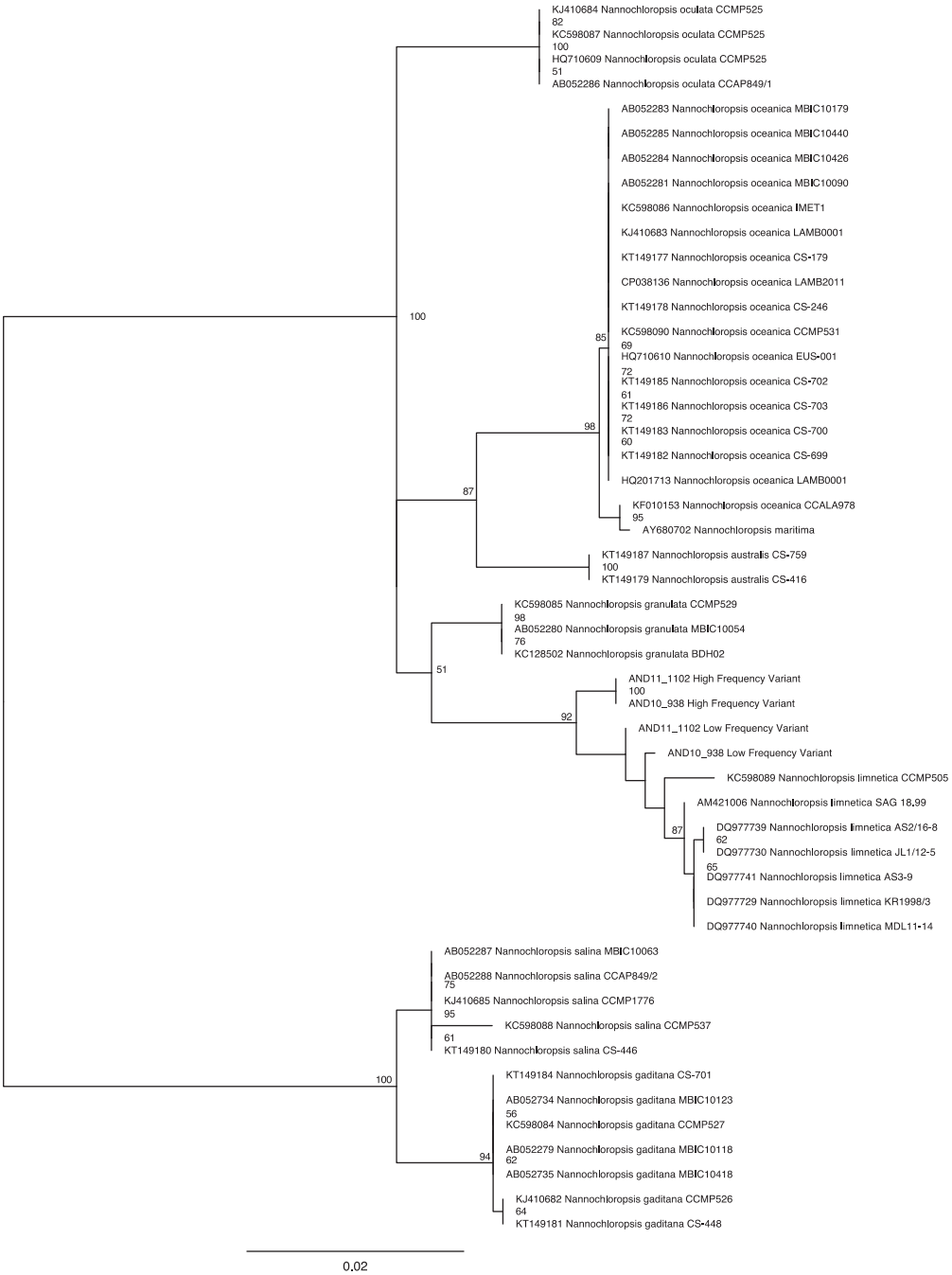
962 Supplementary Figure S8: Phylogeny of the *Nannochloropsis* chloroplast (left) and

963 mitochondrial (right) sequences, including the reconstructed *N.* cf. *limnetica* organelle

964 genomes.



Supplementary Figure S9: Phylogeny containing the *Nannochloropsis* reference chloroplast genomes, the reconstructed *N.* cf. *limnetica* chloroplast genomes and the chloroplast consensus sequences generated via the alternative chloroplast reconstructions.

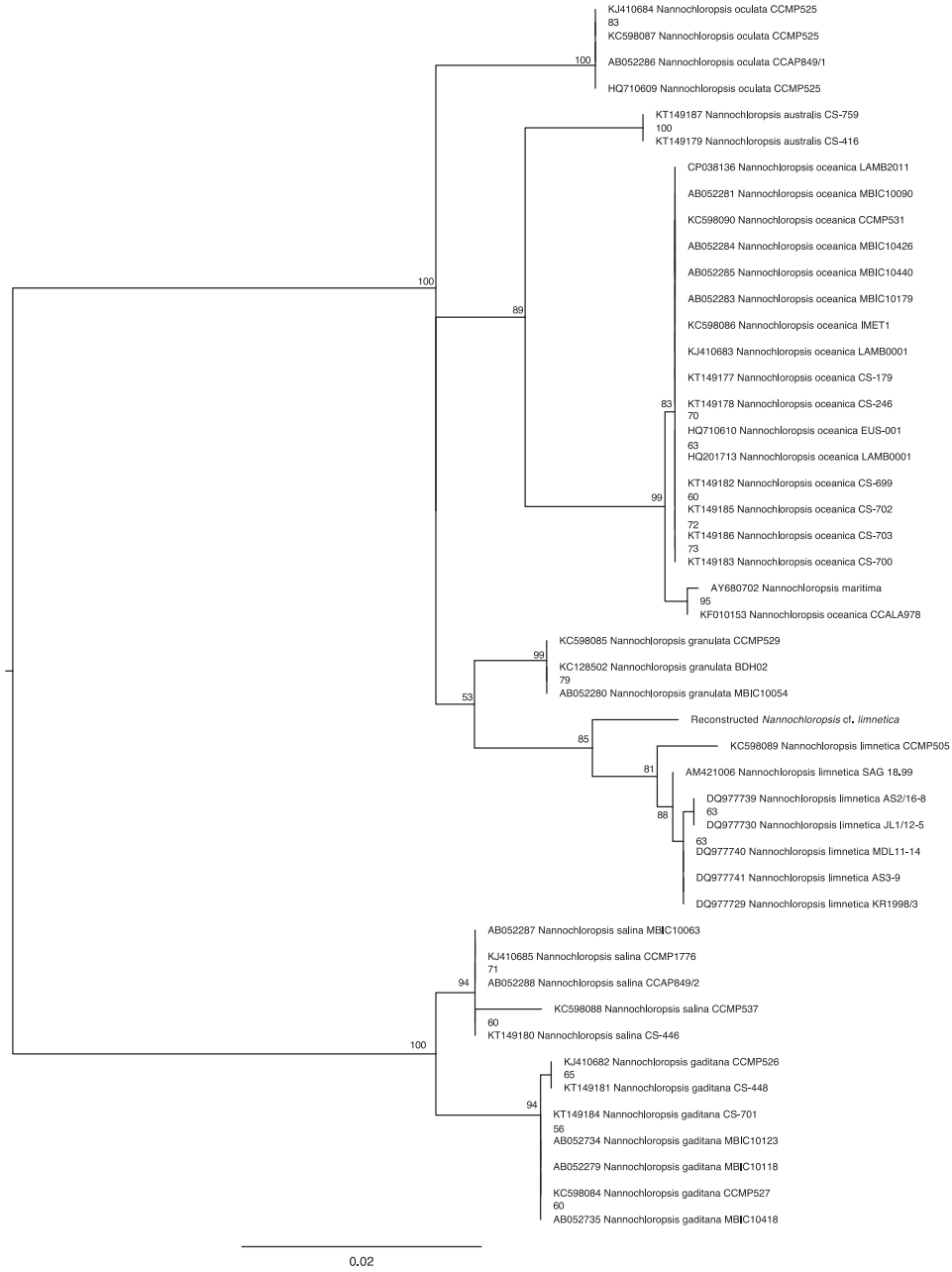


970

971 Supplementary Figure S10: Phylogeny of the *Nannochloropsis rbcL* sequences, including the

972 reconstructed *N.* cf. *limnetica* high and low frequency variant consensus sequences.

973

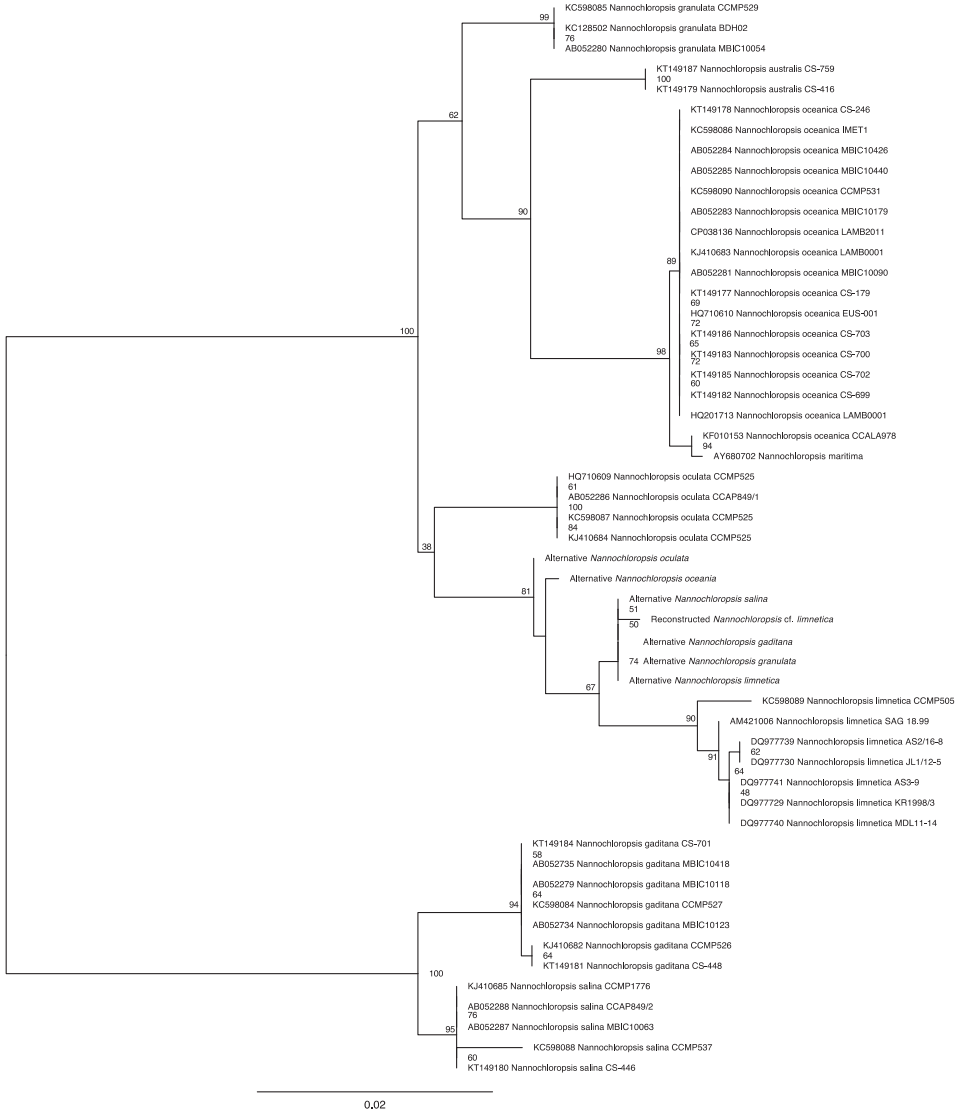


974

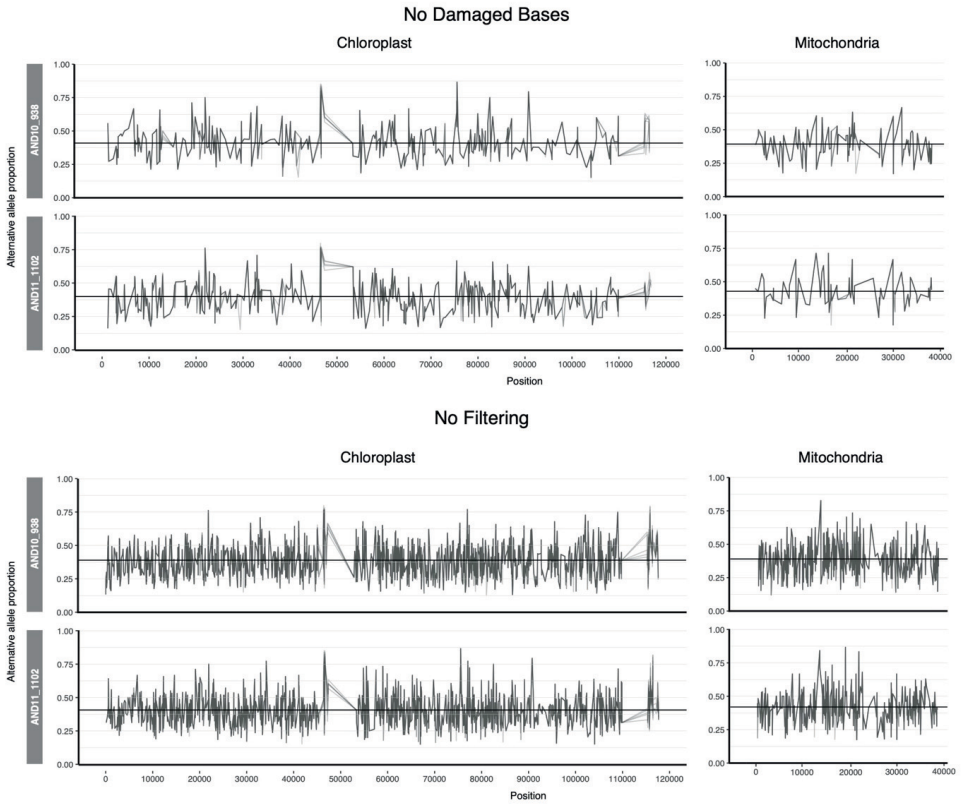
975 Supplementary Figure S11: Phylogeny of the *Nannochloropsis rbcL* sequences, including the

976 reconstructed *N.* cf. *limnetica rbcL* sequence.

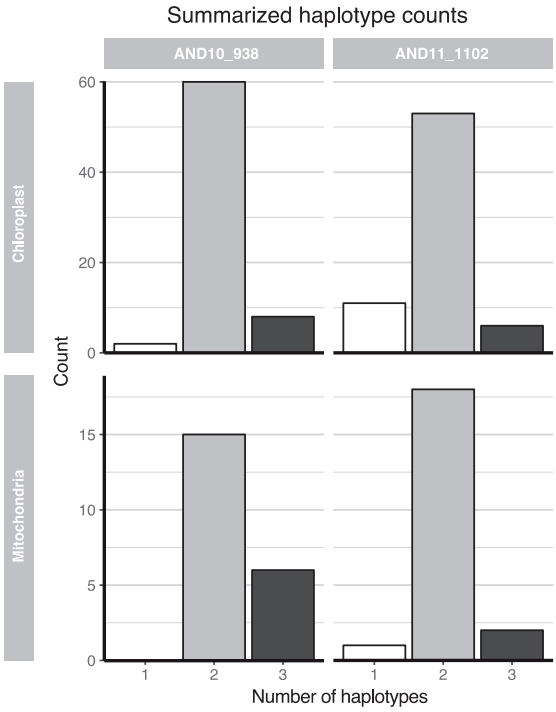
977



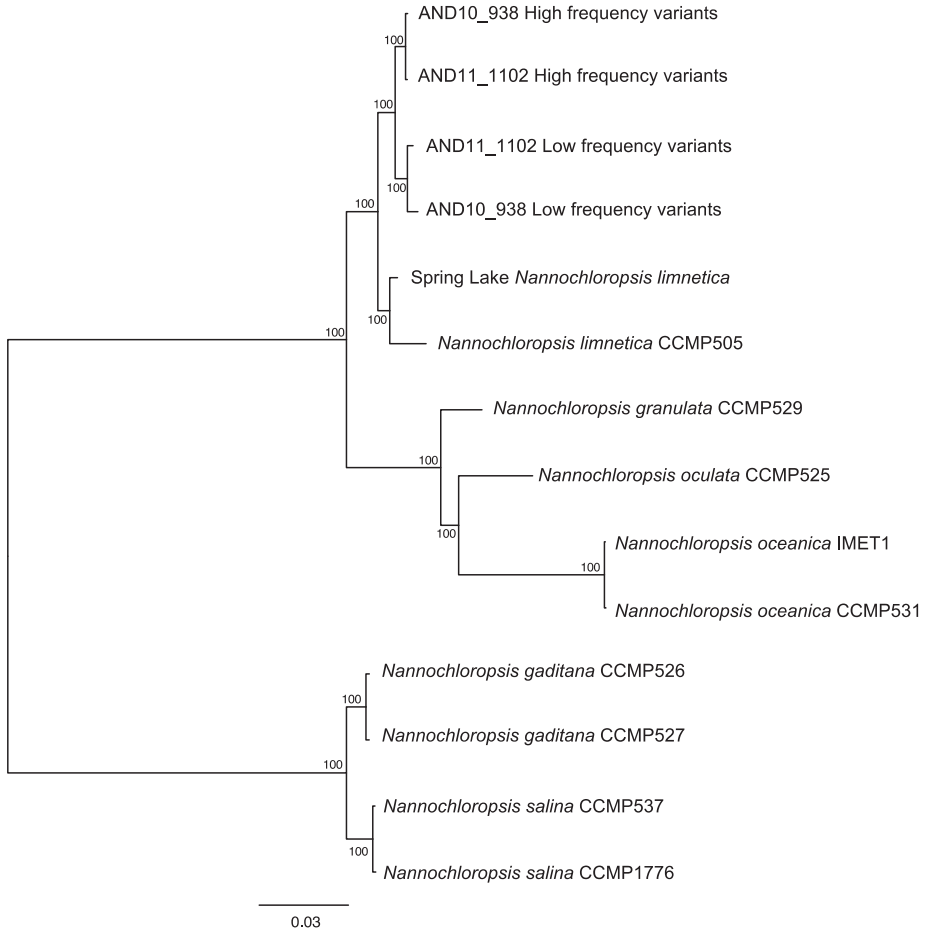
Supplementary Figure S12: Phylogeny of the *Nannochloropsis rbcL* sequences, including the reconstructed *N.* cf. *limnetica rbcL* sequence and the *rbcL* sequences extracted from the chloroplast consensus sequences generated via the alternative chloroplast reconstructions.



Supplementary Figure S13: The proportion of the alternative alleles for the no damage and unfiltered datasets across the organelle genomes. The horizontal black lines represent the averages and are for samples AND10_938 and AND11_1102 respectively: 0.4 and 0.41 for the chloroplast and 0.39 and 0.43 for the mitochondria of the no damaged bases dataset and 0.39 and 0.41 for the chloroplast and 0.39 and 0.42 for the mitochondria of the unfiltered dataset.



Supplementary Figure S14: Bar graph of the observed haplotype counts for both samples and the organelle genomes.



Supplementary Figure S15: Phylogeny containing the *Nannochloropsis* reference chloroplast genomes, the reconstructed *N.* cf. *limnetica* chloroplast genomes and the Spring Lake *N. limnetica* consensus sequence.