MOLECULAR ECOLOGY

Molecular Ecology (2013) doi: 10.1111/mec.12298

Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia

LAURA PARDUCCI,*† IRINA MATETOVICI,*‡ SONIA L. FONTANA,*[1] K. D. BENNETT,§¶ YOSHIHISA SUYAMA,** JAMES HAILE,†[2] KURT H. KJÆR,† NICOLAJ K. LARSEN,†† ANDREAS D. DROUZAS‡‡ and ESKE WILLERSLEV†

*Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden, †Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, ØsterVoldgade 5-7, DK-1350 Copenhagen, Denmark, ‡Molecular Biology Centre, Interdisciplinary Research Institute on Bio-Nano-Sciences, Babes-Bolyai-University ClujNapoca, 42 TreboniuLaurian Street, RO-400271 Cluj-Napoca, Romania, §School of Geography, Archaeology & Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, UK, ¶Department of Earth Sciences, Uppsala University, Villavägen 16, 75236 Uppsala, Sweden, **Graduate School of Agricultural Science, Tohoku University, 232-3 Yomogida, Naruko-onsen, Osaki, Miyagi 989-6711, Japan, ††Department of Geoscience, Aarhus University, Høegh Guldbergs Gade 2, DK-8000 Aarhus, Denmark, ‡‡School of Biology, Aristotle University of Thessaloniki, P.O. Box: 104, GR-54124, Thessaloniki, Greece

Abstract

Plant and animal biodiversity can be studied by obtaining DNA directly from the environment. This new approach in combination with the use of generic barcoding primers (metabarcoding) has been suggested as complementary or alternative to traditional biodiversity monitoring in ancient soil sediments. However, the extent to which metabarcoding truly reflects plant composition remains unclear, as does its power to identify species with no pollen or macrofossil evidence. Here, we compared pollen-based and metabarcoding approaches to explore the Holocene plant composition around two lakes in central Scandinavia. At one site, we also compared barcoding results with those obtained in earlier studies with species-specific primers. The pollen analyses revealed a larger number of taxa (46), of which the majority (78%) was not identified by metabarcoding. The metabarcoding identified 14 taxa (MTUs), but allowed identification to a lower taxonomical level. The combined analyses identified 52 taxa. The barcoding primers may favour amplification of certain taxa, as they did not detect taxa previously identified with species-specific primers. Taphonomy and selectiveness of the primers are likely the major factors influencing these results. We conclude that metabarcoding from lake sediments provides a complementary, but not an alternative, tool to pollen analysis for investigating past flora. In the absence of other fossil evidence, metabarcoding gives a local and important signal from the vegetation, but the resulting assemblages show limited capacity to detect all taxa, regardless of their abundance around the lake. We suggest that metabarcoding is followed by pollen analysis and the use of species-specific primers to provide the most comprehensive signal from the environment.

Keywords: ancient DNA, barcoding, environmental DNA, palaeoecology, pollen

Received 5 November 2012; revision received 18 January 2013; accepted 29 January 2013

Correspondence: Laura Parducci, Fax: +46184716484; E-mail: laura.parducci@ebc.uu.se

[1]Present address: Department of Palynology and Climate Dynamics, University of Göttingen, UntereKarspule 2, 37073, Göttingen, Germany

[2]Present address: Murdoch University Ancient DNA Laboratory, Murdoch University, Perth, Australia

Introduction

Historically, much of what we know about plant community distribution in the past comes from palaeoecological records such as fossil pollen and macrofossils accumulated in lake and peat sediments. These records have provided important information for reconstructing

past migration of plants in response to climatic change and have indicated that plant taxa have expanded and contracted their ranges many times during the last glacial and interglacial cycles. Pollen assemblages, however, include pollen grains from both local and distant sources, and their ability to detect the presence of low-density populations or taxa producing low amounts of pollen is in most of the cases limited. In such cases, plant macrofossils offer a better taxonomical and spatial resolution and have proven to be an excellent complementary tool to pollen analysis (Birks & Birks 2000; Binney *et al.* 2009; Paus *et al.* 2011; Väliranta *et al.* 2011); however, their sample size is often limited.

Plant macrofossils, when preserved under optimal conditions, may also offer a good source of material for molecular analysis. In a recent study, Anderson-Carpenter *et al.* (2011) successfully amplified DNA from macrofossil samples up to 4600 years old from Holocene lake sediments in the western Great Lakes region of North America, matching the taxonomic identity of the macrofossils from which they were extracted, 79% of the time. Plant ancient DNA (aDNA) has been extracted from a different number of other fossil tissues (Hofreiter *et al.* 2000; Gugerli *et al.* 2005; Palmer *et al.* 2012), including pollen (Parducci *et al.* 2005) and even from mixed sources like Pleistocene coprolites (fossilized faeces) (Poinar *et al.* 1998; Gilbert *et al.* 2008). More recently, DNA has also been extracted from soil and lake sediments [(environmental DNA (eDNA)], providing important information on species presence in cold and temperate environments during the late Pleistocene and Holocene and revealing, in some cases, the presence of taxa that were not detectable by fossil analysis (Willerslev *et al.* 2003, 2007; Lydolph *et al.* 2005; Taberlet *et al.* 2007; Munch *et al.* 2008a; Sønstebø *et al.* 2010; Jørgensen *et al.* 2012; Parducci *et al.* 2012). Plant DNA from sediments is believed to originate principally from somatic tissues like rootlets and leafs, and less so from pollen (Jørgensen *et al.* 2012), as DNA in the pollen is only present in small amounts and the success rate of direct amplification of pollen is low (Parducci *et al.* 2005), particularly when extracted from mixed environmental samples. Ancient eDNA records seem therefore to directly reflect the local plant composition (Jørgensen *et al.* 2012; see also Yoccoz *et al.* 2012).

These considerations open new possibilities for aDNA-based plant assessments using eDNA. The metabarcoding analysis, or the biological identification of taxa from environmental samples against a database/library of reference sequences (Hebert *et al.* 2003), is a promising method in such contexts as, in principle, local presence of organisms can be tested, even when these are present at very low density, or when they do not produce pollen (e.g. during harsh environmental conditions). Recently, the method has been further improved by the introduction of next-generation sequencing of amplified pieces of DNA (amplicons) (Binladen *et al.* 2006; Haile *et al.* 2009), allowing the sequencing of thousands of reads from a single DNA extract. The choice of the fragment to amplify for barcode identification, however, has been more problematic in plants than in animals, due to extensive occurrences of phenomena like hybridization, introgression and incomplete sorting of ancestral polymorphisms among plant species. Hollingsworth *et al.* (2009) recommended the two-locus combination of *rbc*L and *mat*K plastid markers as a generic barcode for plants. This approach has proven to be the best so far, but it relies on the amplification of a long DNA region that is unrealistic to amplify from degraded aDNA molecules. For this reason, Taberlet *et al.* (2007) proposed the short P6 loop region of the *trn*L intron plastid DNA fragment as the marker of choice for ancient biodiversity assessment in mixed sediments. Together with the recent establishment of an arctic reference database containing 842 plant species representing all widespread and/or ecologically important taxa of the arctic flora [GenBank accession nos GQ244527 to GQ245667 (Sønstebø *et al.* 2010)], and new statistical tools for sequence identifications (Munch *et al.* 2008a,b), the metabarcoding technique reportedly allows identification at lower taxonomic level than a traditional pollen-based analysis. For example, using the *trn*L primers on old permafrost soil samples from Russia dated 27 900 and 19 000 calibrated years before present (cal. yr BP), Sønstebø *et al.* (2010) were able to assess a set of species present in the respective local (regional) flora and identify as many as 65 different molecular taxonomic units [i.e. clusters of specimens based on sequence identity (MTUs)] at different taxonomic levels. The recent application of the metabarcoding technique to 18 ancient permafrost samples spanning the Late Pleistocene (46 000–12 500 cal. yr BP) from the Taymyr Peninsula in northern Siberia showed that the pollen, macrofossils and eDNA approaches are complementary rather than overlapping (Jørgensen *et al.* 2012) and, together, reveal more detailed information on the composition of past plant communities than can be achieved by each individual approach. Also, in layers dated *c.* 22 000 cal. yr BP from Andøya (northern Norway), the metabarcoding approach recently allowed the identification of tree taxa (spruce and pine) not detected by pollen or macrofossils, suggesting glacial survival of these taxa in the region (Parducci *et al.* 2012).

The short *trn*L marker is therefore suitable for plant identification, particularly where a limited number of taxa are present within families, like in arctic and boreal environments or in old samples (late glacial or early

Holocene) from temperate regions, as here fewer taxa may be present. The strength of this marker resides in the universality of its primers (designed to be generalist and amplify DNA from as many species as possible) and in their ability to amplify short fragments (the length of the *trn*L fragment varies from 10 to 143 base pairs (bp) in the arctic database) (Taberlet *et al.* 2007), thus increasing the chance of amplifying fragmented DNA molecules from ancient sediments. In addition, plastid DNA is present with multiple copies per cell and is therefore very common in many plant tissues (Preuten *et al.* 2010; Rauwolf *et al.* 2010), improving the chances of DNA survival and detection in old sediments. However, the robustness of the system in environments with a larger number of taxa (e.g. younger sediments from Arctic and boreal regions or sediments from temperate and tropical regions) remains to be tested. Jørgensen *et al.* (2012) concluded in their study that the *trn*L system suffers from sequence selectiveness owing to primer binding biases and to differences in the abundance of target sequences among plant species and tissues within species. Indeed, it has been demonstrated that generic primers may fail to detect rare species, particularly when their DNA is mixed with DNA from other more targeted taxa as they will preferably bind to common DNA templates (Murray *et al.* 2011).

In this study, we evaluated the power of the metabarcoding technique to assess the presence of plant taxa through time on lake sediments from a boreal environment, and the ability of the *trn*L primers to reveal the presence of taxa not detected by pollen analysis. We tested these generic markers on bulk DNA sediments from two lakes and compared the results with those previously obtained in an earlier study on the same sediments at one of the two sites, using species-specific primers (Parducci *et al.* 2012). The aim was to explore the extent to which the metabarcoding technique can retrieve DNA signal from old sediments and to investigate the selectiveness of the systems in identifying different plant taxa. Our attempt focused on the Scandes Mountains in central Scandinavia, as here Holocene vegetation change has already been established by numerous palynological studies (Birks & Birks 2000; Seppä & Bennett 2003; Giesecke & Bennett 2004; Giesecke 2005; Eide *et al.* 2006). This region is also an area of major palaeoecological interest due to megafossil finds of boreal trees dated to more than 11 500 cal. yr BP, when most of central Scandinavia was believed to be still covered by ice (Kullman 2002). These finds include remains of birch (*Betula pubescens*), pine (*Pinus sylvestris*) and spruce (*Picea abies*), which gave rise to speculations about the survival of some of these species off the Norwegian coast during the last glacial maximum. Scattered *Picea* pollen occurs in late-glacial and early-Holocene sediments in northern Scandinavia; however, these low pollen values have typically been considered a result of long-distance transport from the same regions (Huntley & Birks 1983). In an earlier study, Parducci *et al.* (2012) showed the presence of *P. abies* on the Scandes Mountains using species-specific primers on bulk lake sediments dating back to 10 400 cal. yr BP, where no pollen was observed. These results combined with results from modern and aDNA from other sites showed that *P. abies* was present in Norway during the early Holocene, where it probably occurred in few numbers in the landscape and likely reproduced by vegetative propagation, so leaving no trace in the palynological record.

Materials and methods

Sampling sites

Our samples came from two lakes, Rundtjørna and Klocka, situated *c.* 35 km apart, on different sides of the Scandes Mountains at the Norwegian–Swedish border in central Scandinavia (Fig. 1). Klocka (325 m a.s.l., 63°17′59″ N 12°29′58″ E) is the more southerly and the larger (*c.* 1000 m × 150 m), located in the Jämtland region in central Sweden, and has a maximum water depth of 10 m. Rundtjørna is smaller (*c.* 200 m × 100 m; 526 m a.s.l., 63° 22′ 32″ N, 11°49′44″ E) and located in north Trøndelag (central Norway), with a maximum water depth of 3 m. The lakes are situated within the northern boreal zone nowadays dominated by *Picea abies* and *Pinus sylvestris*.

Sampling and radiocarbon dating

The sediments were obtained in 2008 using a Livingstone corer (Wright *et al.* 1984) at maximum water depth by coring from winter ice. The corer used was similar to the device described by Feek *et al.* (2011) to collect sediment cores for analysis of uncontaminated DNA, except that it was water-filled (rather than air-filled). Initial surface sterilization was not carried out because subsequently filling the tube with nonsterile media (whether air or water) and movement of the sediment through the core tube (in both designs) cause surface contamination, which must be removed in the laboratory in any case. In addition, dissolved DNA from lake water is not a contaminant. The degradation of DNA in water has been investigated experimentally, and it has been shown that even short 100-bp eDNA sequences degrade to subdetectable levels within days (Thomsen *et al.* 2012a,b).

Fieldwork and all laboratory analyses were performed during winter, to avoid the period of pollen

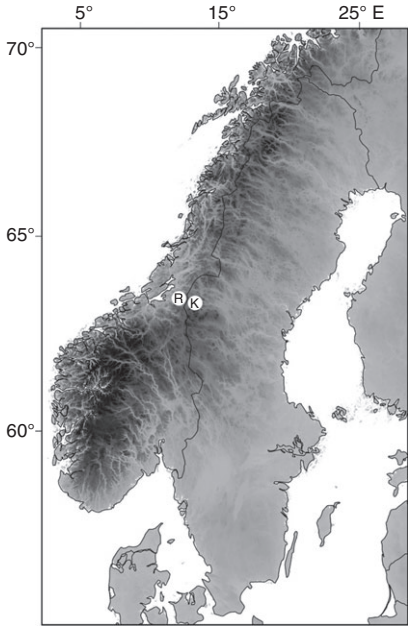

Fig. 1 Map of Norway and Sweden showing the location of the study sites. R: Lake Rundtjørna (Trøndelag, Norway); K: Lake Klocka (Jämntland, Sweden).

release of extant vegetation. After the cores were extruded, they were immediately wrapped in plastic film and aluminium foil, kept cold in the field, transported to the palaeoecological facility at the Department of Earth Sciences of Uppsala University, a building free from DNA research, and stored at 5 °C. Here, sampling and manipulation of the cores were performed in a room equipped for pollen analysis (GEO room). Disposable and sterilized tools were used and changed between samples to avoid cross-contamination.

For molecular analyses, we collected 18 samples (nine from each core; *c.* 10 g wet weight; Table 1). The external 2-cm part from the surface of the sediment was discarded during core processing, following normal palaeoecological practice (e.g. Moore *et al.* 1991) to avoid all contamination, including from exogenous plant DNA (Feek *et al.* 2011), and the rest was stored in a sterile plastic tube at −20 °C until DNA extraction.

For comparison, we collected *c.* 0.5-g samples at the same levels sampled for molecular analysis, and a minimum of 500 pollen grains and spores of terrestrial vascular plants were counted and identified, if possible, in each. Pollen processing followed the methods described in Bennett & Willis (2001). Pollen of aquatic plants and mosses (*Sphagnum*) spores were excluded from the total pollen sum. Pollen taxonomy follows Bennett (2004), modified for Sweden using the checklist by Karlsson (1998).

The lake sediment core from Rundtjørna was dated at the Tandem Laboratory, Uppsala University, Sweden. The age model was based on radiocarbon dates of seven macrofossils and two bulk samples that have been converted into calendar years using OXCAL 4.1 with IntCal09 (Ramsey 2008; Reimer *et al.* 2009) and are quoted as cal. yr BP. The model is constructed using the depositional model in OXCAL 4.1 with a k-value of 100 yielding A_{model} = 81%. All molecular, pollen and dated samples derived from highly organic sediment retrieved at various depths (Table 2).

DNA extraction, PCR setup and sequencing

We extracted DNA in dedicated aDNA facilities at the Centre for GeoGenetics of the University of Copenhagen, following established aDNA precautions (Willerslev *et al.* 2004). The 18 samples were extracted all in one batch that included two negative controls to monitor for possible contamination during extraction (extraction blank). We used a combined Sergey Bulat protocol and Cambio PowerMax™ Soil DNA isolation kit protocol (Mo Bio Laboratories, Cambridge, UK), which employs a silica clean-up method (Haile *et al.* 2009). Approximately 10 g wet weight of sediment was subsampled, placed in PowerMax Soil PowerBead tubes (Cambio) and dissolved in 24 mL lysis buffer (Bulat *et al.* 2000). The tubes were then agitated vigorously for 1 min and left to incubate overnight at 65 °C under gentle agitation. Following extraction, eDNA was purified using the PowerMax™ Soil DNA Isolation Kit protocol following manufacturer's instructions.

PCR amplifications of the *trn*L fragment with the g and h primers (Taberlet *et al.* 2007) were performed independently in two experiments at Uppsala and Copenhagen University following established aDNA methodologies (Willerslev *et al.* 2004). At Uppsala, we cloned and Sanger-sequenced the PCR products, and in Copenhagen, we used the next-generation sequencing technique.

At Uppsala University, we set up two replicates of PCRs. The first was performed in the GEO room and the second in a laboratory dedicated to aDNA analysis of pollen at the Evolutionary Biology Centre (EBC room). Both GEO and EBC rooms were physically separated from the laboratory where we later performed the cloning experiments. In total, we ran 102 PCRs in multiple batches that included a maximum of five PCRs and a negative control (PCR blank). We used the Qiagen Multiplex PCR kit, following the protocol from the manufacturer and with 40 amplification cycles. The volume of the reaction mixtures was 20 μL, comprising 3 μL of DNA template, 1× Multiplex PCR Master Mix (Qiagen), 0.2 μM of each primer and water for adjusting the final volume. After amplification, 5 μL of PCR

Table 1 Sediment samples used for palynological and barcoding analyses at Rundtjørna (RD) and Klocka (KL), with summary of the results obtained after multiple PCR amplifications of the *trn*L (g/h) fragment using the cloning and Sanger sequencing technique. Depths are given in centimetres from the top of the sediments. Only amplifications showing fragments of expected size, with no amplifications in the corresponding PCR control, were cloned

Sample	Depth (cm)	PCR runs	Fragments	Clones	Plant sequences assigned	Unknown plant sequences	Bad-quality sequences
RD1	1	8	5	113	97	5	11
RD2	25.5	19	5	25	23	0	2
RD3	50	—	—	—	—	—	—
RD4	69.5	6	2	19	12	2	5
RD5	99.5	10	5	23	16	2	5
RD6	146	—	—	—	—	—	—
RD7	184	13	5	26	18	5	3
RD8	220	—	—	—	—	—	—
RD9	258	4	1	41	25	15	1
Total RD		60	23	247	191	29	27
KL1	1	—	—	—	—	—	—
KL2	20	—	—	—	—	—	—
KL3	40	—	—	—	—	—	—
KL4	65	8	1	1	1	0	0
KL5	110	12	4	3	1	2	0
KL6	140	—	—	—	—	—	—
KL7	180	14	5	6	4	2	0
KL8	215	6	5	30	23	7	0
KL9	270	2	2	16	7	2	7
Total KL		42	17	56	36	13	7
Total		102	40	303	227	42	34

Table 2 Radiocarbon dates from Rundtjørna calibrated for the age model. Depths of the samples are given in cm from the top of the sediments

Laboratory no.	Depth (cm)	Material	14C age BP	Model age (cal. yr BP)
Ua-39895	13–14	Macro	1369 ± 32	1290
Ua-39896	46–47	Macro	2197 ± 31	2230
Ua-39897	76–77	Bulk	3151 ± 30	3380
Ua-39898	105–106	Macro	3624 ± 34	3940
Ua-39899	134–135	Macro	4412 ± 31	4990
Ua-39900	165–166	Macro	5226 ± 38	5990
Ua-39901	193–194	Macro	5930 ± 41	6760
Ua-39902	222–223	Macro	6930 ± 52	7770
Ua-39903	256–257.5	Bulk	9233 ± 52	10390

products were screened on 2% agarose gels. From the positive amplifications, we purified 2 μL using ExoSAP-IT (Affymetrix, Inc.) and used this for cloning with the CloneJet™ PCR Cloning Kit (Fermentas), following manufacturer's instructions. All clones with inserts of expected sizes were sequenced using Macrogen DNA Sequencing service (Korea).

At Copenhagen University, PCR amplifications were performed with the *trn*L primers containing a 46-bp flanking sequence for the GS FLX sequencer and an 8-bp 'barcoding' tag to enable differentiation of the samples after sequencing. Amplifications were performed in a 25-μL final volume, using 1 μL of DNA as template in an amplification mixture containing 1.0 U Platinum® Taq High-Fidelity DNA Polymerase (Invitrogen, Carlsbad, CA), with 1× HiFi buffer, 5 mM $MgSO_4$, 0.2 mM dNTP, 1 mg/mL bovine serum albumin (BSA) and 0.4 mM of each primer. The DNA was subjected to 55–60 cycles of PCR (4-min initial denaturation at 97 °C, 45 s at 94 °C, 45 s at 55 °C, 45 s at 68 °C and a final cycle at 72 °C for 10 min). PCR products were checked on 2% agarose gels, and for each sample, 2–4 PCR replicates were purified using the E.Z.N.A.® Gel Extraction kit (Omega). Amplification products were sequenced on the Roche GS FLX DNA sequencing platform following the manufacturer's guidelines for amplicon sequencing.

Taxonomical assignment

Sanger sequences were aligned using BIOEDIT v. 7.1.3.0 (Hall 1999) to identify errors due to base call misidentifi-

cations, postdamage lesions and possible polymorphism. After alignment, they were compared to GenBank and to the arctic database [GenBank accession nos GQ244527 to GQ245667 (Sønstebø *et al.* 2010)]. As the arctic database includes few boreal species, we also created a database that included part of the arctic database and boreal plant taxa occurring at our study site (45 families, 126 genera and 190 boreal plant species in total) (Table S1, Supporting information). For taxonomic assignment, we adopted a conservative approach to avoid incorrect conclusions due to sequencing or PCR error. We assigned only those sequences that differed by a maximum of two nucleotides from those matching from the databases and retained also sequences with type 1 and type 2 transitions substitutions (adenine ↔ guanine and thymine ↔ cytosine) that are typically present in fragments amplified from damaged aDNA templates (Hansen *et al.* 2001; Binladen *et al.* 2006). Sequences with more than two nucleotides difference were not assigned and were considered of unknown plant origin, even if they could be identified to family or higher taxonomic levels.

The next-generation sequences were aligned and sorted based on their barcoding tag (zero mismatches permitted) using GENEIOUS v. 5.4.1 (Biomatters, New Zealand). Alignments were performed using BIOEDIT v. 7.1.3.0 (Hall 1999), and only sequences containing PCR primers (two mismatches allowed) were retained (Margulies *et al.* 2005; Balzer *et al.* 2010). As the Roche GS FLX platform can be inaccurate reading homopolymers longer than 5–8 bp (Balzer *et al.* 2010), all marker sequences shorter than 20 bp were filtered out. Assignments were performed manually following the same procedure used for the Sanger sequences but we retained only sequences with a 100% match. To account for sequencing or PCR errors, all sequences represented by less than three reads (copies) were removed from the data set.

Results

Chronological analysis

Chronological control of the Rundtjørna sediment record is presented in Table 2. The core spanned the 11 000 years of the Holocene. The radiocarbon-dated macrofossils and bulk samples showed an increasing age trend with increasing depth, and given the good agreement between bulk and plant macrofossil dates, the bulk samples were also included in the constructed age model. Age determinations were internally consistent, and the constructed age-depth model suggested a relatively constant sediment accumulation rate during the last 11 000 cal. yr BP (Fig. 2).

The pollen record from Klocka and Rundtjørna followed similar trends (see 'Pollen analysis' below), reflecting vegetation changes synchronous with other sites from the region (Segerström & von Stedingk 2003; Giesecke 2005); therefore, dates for the Klocka samples can be approximately estimated on the basis of the Rundtjørna data, by comparison of pollen spectra.

Pollen analysis

Pollen analysis revealed *P. sylvestris*, *Betula* sp. and alder (*Alnus* sp.) as the taxa that dominated the forest around the two lakes throughout the Holocene (Fig. 3). The lowermost two samples from Rundtjørna captured the spread of *Alnus* sp. at around 9500 cal. yr. BP. The lowest sample at Klocka already contained a high proportion of *Alnus* pollen indicating that the sample dated to after 9500 cal. yr. BP. The decline in the abundance of *Alnus* sp. during the mid-Holocene was compensated by an increase in *Pinus* sp. mainly at Rundtjørna, while Klocka showed a larger abundance of *Betula* sp.

Picea abies pollen was only important in the upper samples from both sites, while its amount only indicated abundance of the tree for the uppermost sample from Rundtjørna.

In total, we identified 33 families and 46 taxa by pollen and spore analysis: eight families included deciduous trees and shrubs, two families of conifers (Pinaceae and Cupressaceae), three families of Pteridophytes

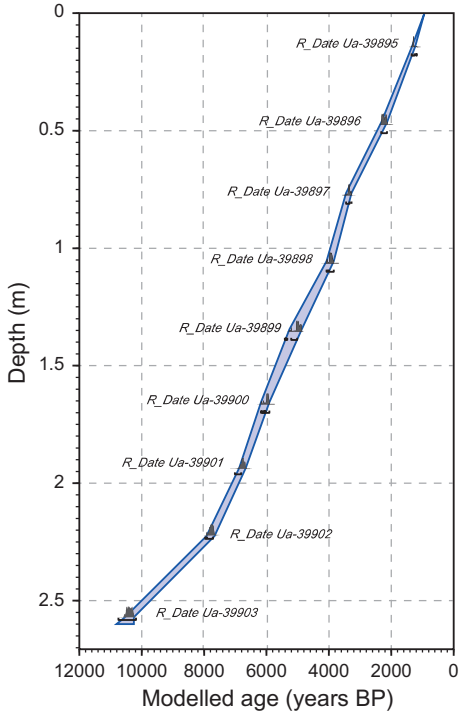

Fig. 2 Age model for Rundtjørna.[14]C ages are calibrated with OXCAL v 4.1.7 (Ramsey 2008).

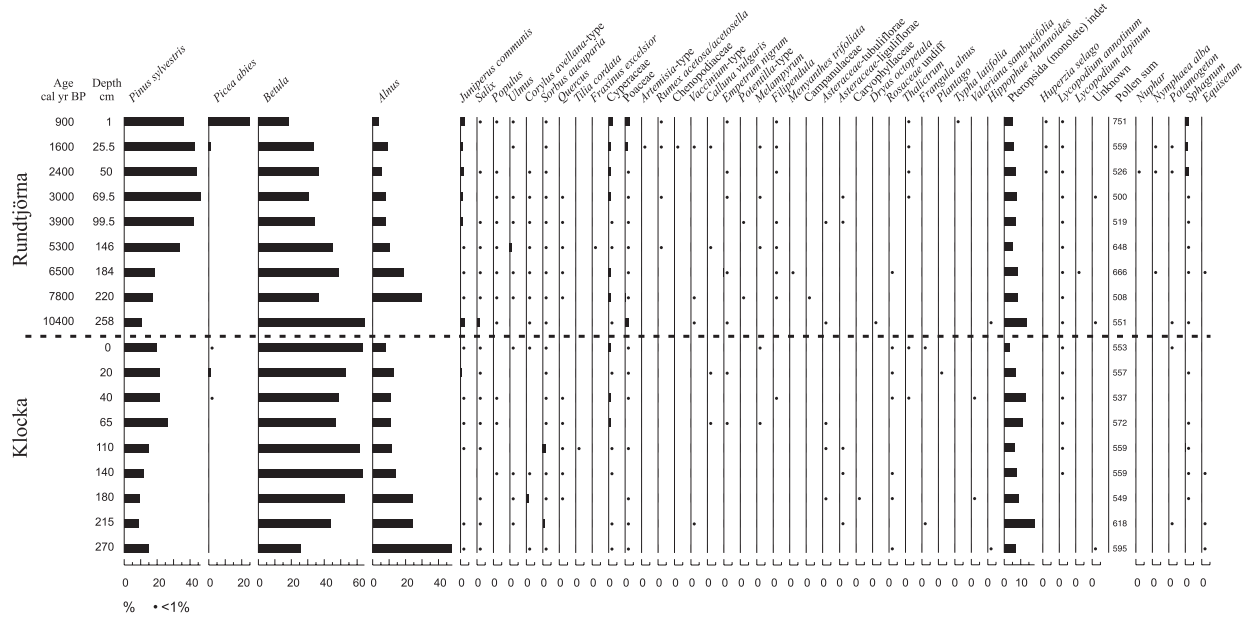

Fig. 3 Pollen percentage values found at Rundtjørna and Klocka. Dots indicate taxa occurring with <1%.

(Lycopodiaceae, Equisetaceae and Huperziaceae) and one bryophyte family (Sphagnaceae). The other nineteen families were represented mainly by herbaceous plants (Fig. 3).

Molecular analyses

At Uppsala University, we obtained DNA amplifications from a total of 11 samples from the two lakes, six at Rundtjørna and five at Klocka (Table 1). In total, we performed 102 PCR amplifications that yielded 40 PCR fragments (23 from Rundtjørna and 17 from Klocka), with no visible bands on agarose gels in the corresponding PCR blanks. After cloning, we obtained 384 colonies in total, of which 303 gave correct inserts after screening on agarose. After sequencing analysis and filtering for sequencing and PCR errors, the 303 colonies yielded 227 plant sequences (191 at Rundtjørna and 36 at Klocka). Thirty-four were rejected because of bad quality, showing no cloning insertion or chimera products (i.e. spurious sequences derived from more than one DNA template). The 227 plant sequences could be assigned to 10 MTUs from eight different families (Table 3), matching known plant taxa from our database, all currently present at our study region (Table S1, Supporting information). Four MTUs were identified at the species level: rowan (*Sorbus aucuparia*), grey alder (*Alnus incana*), common aspen (*Populus tremula*) and common heather (*Calluna vulgaris*), while the rest were identified to species group or genus level.

At Copenhagen University, the PCR was successful for a total of 12 samples from the two lakes, with extraction and PCR controls remaining clean. Next-generation sequencing however, worked for eight samples, one from Rundtjørna and seven from Klocka (Table 4), that generated sequences with a mean of 4886 reads per sample. After filtering for products shorter than 20 bp, we obtained between 141 and 704 sequences per barcode tag. Rejection of sequences represented by less than three reads resulted in a total of 526 unique sequences with a number of copies varying from three to 7089. The selected sequences were assigned to eight different families and 10 MTUs, all matching known plant taxa from our database (Table 4). Four MTUs were identified at the species level [*S. aucuparia*, *A. incana*, *P. tremula* and common mare's-tail (*Hippuris vulgaris*)], one at the family level (Fabaceae) and the rest at the genus level.

In total, from the two lakes, we identified 14 MTUs and 11 families with the barcoding analyses (Table 5). Betulaceae sequences occurred with high frequency (the large majority belonging to *A. incana* and *Betula* sp.), while the rest of the sequences belonged to the temperate and arctic families Salicaceae (with three sequences identified as *P. tremula*), Brassicaceae and Poaceae. At Klocka, we also detected DNA signal from other tree taxa, also identified by pollen, like *Betula* sp., willow (*Salix* sp.), *Populus* sp. and *S. aucuparia*. At Rundtjørna, most of the sequences matched with *Nymphaea/Nuphar* spp. (Nymphaeaceae) (in 79% of the cases with the Sanger technique and in 71% of the cases with the next-generation sequencing technique), while the rest were assigned to species or a group of species belonging to the four major temperate and arctic families Rosaceae, Brassicaceae, Asteraceae and

Table 3 Number of sequences obtained after cloning and Sanger sequencing the plastid *trn*L fragment from Rundtjørna (RD) and Klocka (KL) with respective taxonomic assignment

Sample age (cal. yr BP)														
900	1600	3000	3900	6500	10400							Taxonomic assignment		
RD1	RD2	RD4	RD5	RD7	RD9	KL4	KL5	KL7	KL8	KL9	Sum	Family	Genus	Species
97	20	7	8	18	25				9		184	Nymphaeaceae	*Nymphaea* *Nuphar*	*N. tetragona*/ *N. pumila*/*N. lutea*
	3										3	Rosaceae	*Sorbus*	*S. aucuparia*
								2	5	5	12	Betulaceae	*Alnus*	*A. incana*
						1			4	2	7		*Betula*	*B. pubescens*/*B. nana*
		1	4					2			7	Brassicaceae	*Arabidopsis*	*A. arenosa*/*A. lyrata* subsp. *petraea*
									3		3	Salicaceae	*Populus*	*P. tremula*
			4									Asteraceae	*Saussurea*	*S. alpina*/
											4		*Lactuca* *Taraxacum*	*L. sibirica*/ *T. croceum*
		1									1	Ericaceae	*Calluna*	*Calluna vulgaris*
		1									1	Poaceae	*Poa*	*P. alpina*/*P. pratensis*/ *P. trivialis*
		2					1		2		5		*Elymus* *Leymus*	*E. repens*/*E. mutabilis*/ *L. interior*/*L. arenarius*
5		2	2	5	15		2	2	7	2	42	Plant sequences not assigned		
102	23	14	18	23	40	1	3	6	30	9				
220						49					269			

Poaceae. In general, and with the exception of *S. aucuparia*, the taxa at Rundtjørna belonged to herbaceous plants like grasses (*Elymus*/*Leymus* sp. and *Poa* sp.), rock-cress (*Arabidopsis* sp.), dandelion (*Taraxacum* sp.), lettuce (*Lactuca* sp.) and alpine saw-wort (*Saussurea alpina*), regardless of the sequencing technique used (Tables 3 and 4). The large amount of Nymphaeaceae sequences obtained with Sanger/cloning at Rundtjørna should not be taken as indication of the amount of DNA present in the sediments. With the Sanger technique, one should refer to the number of fragments obtained from each sample rather than to the number of sequences obtained after cloning. This is because a fragment from younger templates may give a larger number of colonies compared to a fragment from older sediments, as the number of recombinant cells with correct insertion tends to decrease with poorer/older amplifications. Nevertheless, we find the molecular results to agree sufficiently well with the pollen data at Rundtjørna where water lily (*Nymphaea alba*) had three temporal occurrences and *Nuphar* sp. one. Both species were absent from the pollen records at Klocka, and here barcoding with Sanger/cloning revealed only one occurrence. Because Nymphaeaceae are rooted aquatics with a large amount of biomass in the lake and its sediment, but the flowers and pollen are emergent, it is reasonable that DNA may be well represented in sediment, while pollen is less.

Comparison of the barcoding and the pollen results

Overall, the combined analyses (pollen and metabarcoding) recognized 52 different taxa, identified at different taxonomic levels, from species to family. The level of taxonomic classification varied within proxy, and each identified a varying number of taxa within five different plant groups: trees and shrubs, herbs, graminoids, ferns and aquatic plants (Table 5). The two proxies were only partially overlapping as only 25% of the taxa were detected by both techniques. Of the taxa identified by pollen analyses (46), the majority (78%) could not be detected by metabarcoding, but the latter technique identified taxa to a lower taxonomic level. Mosses, ferns and some families within the group of trees and shrubs (e.g. Pinaceae) were identified by pollen or spores only. Within herbs and graminoids, three families were detected by both techniques (Asteraceae, Ericaceae and Poaceae), and the taxonomic level differed extensively between them. For aquatic plants, the DNA and the pollen assignments were similar.

Table 4 Number of unique identical sequences (minimum 3) obtained after amplification and next-generation sequencing analysis of the plastid *trn*L fragment from Rundtjørna (RD) and Klocka (KL) with respective taxonomic assignment. The range of identical sequences varied between three and 7089

Sample age (cal. yr BP)											
7800									Taxonomic assignment		
RD8	KL1	KL2	KL4	KL5	KL6	KL 8	KL9	Sum	Family	Genus	Species
75								75	Nymphaeaceae	*Nymphaea Nuphar*	*N. tetragona/ N. pumila/N. lutea*
					49			49	Rosaceae	*Sorbus*	*S. aucuparia*
		61	30		15	46	22	174	Betulaceae	*Alnus*	*A. incana*
30								30	Fabaceae		
							2	2	Salicaceae	*Populus*	*P. tremula*
						1		1		*Salix*	*S. reticulata/S. glauca*
							4	4	Hippuridaceae	*Hippuris*	*H. vulgaris*
			3					3	Poaceae	*Poa*	*P. alpina/P. pratensis/P. trivialis*
	100		52					152		*Elymus Leymus*	*E. repens/E. mutabilis/ L. interior/L. arenarius*
				36				36	Apiaceae	*Angelica*	*A. sylvestris/*
										Heracleum	*H. sphondylium*
105	100	61	85	36	64	47	28				
	421							526			

Discussion

Contamination from contemporary vegetation

Recently, several DNA analyses on fossil remains have shown that ancient sequences can be distinguished from contaminants on the basis of nucleotide misincorporation rates (Lindahl 1993; Höss *et al.* 1996; Olivieri *et al.* 2010) and that the majority of aDNA damage is reflected by type 2 (cytosine→thymine/guanine→adenine) transitions (Sawyer *et al.* 2012), while type 1 transitions are essentially PCR artefacts. In our study, the alignment of all cloned ancient sequences identified several characteristic damage-induced errors of type 1 and 2 (see Data accessibility), and the positions of these substitutions differed in bacterially cloned PCR amplicons. In addition, negative results obtained from an array of multiple negative controls used during DNA extraction and PCRs indicate that our sequences come from endogenous DNA. More importantly, at both lakes, we found no substantial difference between the molecular data obtained in two independent experiments, performed in separate laboratories (Uppsala and Copenhagen), and using different techniques (Sanger and next-generation sequencing technique, respectively). Finally, were contamination present, taxa currently abundant in the catchment as blueberry and lingonberry (*Vaccinium* spp.) should have been found also abundantly through the core, but they were not. In addition, we followed all conventional protocols for avoiding contamination: careful core storage, cleaning of cores before sampling and sampling from the undisturbed interior of a core in clean laboratories and with clean instruments. For all these reasons, we exclude that contemporary vegetation was a source for eDNA in our samples.

Comparison of the barcoding and the pollen results

The comparison between pollen and barcoding results provided important information that can be summarized as follows: (i) the pollen and barcoding analyses resulted in largely nonoverlapping species/taxon lists, (ii) the two analyses each detected likely a subset of the total flora present in each period; hence, none alone was totally comprehensive; (iii) pollen-limited plants were more easily found in the DNA record, and barcoding was therefore a significant complement to pollen analysis; (iv) the barcoding analysis allowed in general identification at lower taxonomic level than pollen-based analysis; (v) unlike the pollen results, the barcoding analysis showed different results at the two lakes, with a better correspondence between methods found at Klocka; (vi) there was no substantial difference between barcoding data obtained with the Sanger vs. next-generation sequencing technique; (vii) the molecu-

Table 5 Taxa detected from Rundtjørna (RD) and Klocka (KL) after barcoding and pollen analysis. Taxa indicated by an asterisk were found in both proxies

			Barcoding		
	Family	Taxon	Sanger	Next-gen.	Pollen
Trees and shrubs	Pinaceae	*Pinus sylvestris*			KL RD
		Picea abies			KL RD
	Betulaceae	*Betula* sp.*	KL		KL RD
		Alnus sp.*	KL	KL	KL RD
	Corylaceae	*Corylus avellana*-type			KL RD
	Cupressaceae	*Juniperus communis*			KL RD
	Salicaceae	*Salix* sp.*		KL	KL RD
		Populus sp.*	KL	KL	KL RD
	Ulmaceae	*Ulmus* sp.			KL RD
	Oleaceae	*Fraxinus excelsior*			RD
	Tiliaceae	*Tilia cordata*			KL
	Elaeagnaceae	*Hippophae rhamnoides*			KL RD
	Rhamnaceae	*Frangula alnus*			KL RD
	Fagaceae	*Quercus* sp.			KL RD
	Rosaceae	*Sorbus aucuparia**	RD KL	KL	KL RD
		Filipendula sp.			KL RD
		Dryas octopetala			RD
		Potentilla-type			RD
		Rosaceae indiff.			KL RD
Herbs	Fabaceae			RD	
	Plantaginaceae	*Plantago* sp.			KL
	Brassicaceae	*Arabidopsis* sp.	RD KL		
	Apiaceae	*Heracleum sphondylium/*		KL	
		Angelica sylvestris		KL	
	Asteraceae	*Taraxacum croceum/**	RD		KL RD
		*Lactuca sibirica/**	RD		Asteraceae-liguliflorae
		Saussurea alpina	RD		
		Artemisia-type			RD
		Asteraceae-tubuliflorae			KL RD
	Polygonaceae	*Rumex acetosa*			RD
	Chenopodiaceae				RD
	Ericaceae	*Vaccinium*-type			KL RD
		*Calluna vulgaris**	RD		KL RD
	Empetraceae	*Empetrum nigrum*			KL RD
	Scrophulariaceae	*Melampyrum* sp.			KL RD
	Menyanthaceae	*Menyanthes trifoliata*			RD
	Campanulaceae				RD
	Caryophyllaceae				KL
	Ranunculaceae	*Thalictrum* sp.			KL RD
	Valerianaceae	*Valeriana sambucifolia*			KL
Graminoids	Cyperaceae				KL RD
	Typhaceae	*Typha latifolia*			RD
	Poaceae	*Poa alpina/pratensis/trivialis**	RD KL	KL	KL RD Poaceae undiff.
		*Elymus repens/mutabilis/**	RD KL	KL	
		*Leymus interior/arenarius**	RD KL	KL	
Mosses	Lycopodiaceae	*Huperzia selago*			KL RD
	Lycopodiaceae	*Lycopodium annotinum*			KL RD
		Lycopodium alpinum			RD
	Sphagnaceae	*Sphagnum* sp.			RD KL
Ferns	Indet.				KL RD
Horstails	Equisetaceae	*Equisetum* sp.			RD KL
Aquatic plants	Hippuridaceae	*Hippuris vulgaris*		KL	
	Nymphaeaceae	*Nuphar* sp.*	RD KL	RD	KL
		Nymphaea sp.*	RD KL	RD	RD
	Potamogetonaceae	*Potamogeton* sp.			RD KL
Unknown/not assigned			RD KL		KL RD
Total number of taxa or MTU			14		46
Total number of families			11		33
Total number of taxa; both proxies			52		

lar analyses failed to detect some major tree taxa detected in the pollen record (e.g. *P. sylvestris* and *P. abies*) previously identified in the same sediments using species-specific primers (Parducci *et al.* 2012).

The main result was that the pollen and molecular analyses resulted in largely nonoverlapping species/taxon lists, and this was in good agreement with recent results from Jørgensen *et al.* (2012). Our comparison showed complementary data rather than overlapping, with each proxy able to detect likely a subset of the flora around the lakes, and confirming that neither of the two methods is totally comprehensive. An example of good complementarity between proxies was the ability of barcoding to detect pollen-limited plants like *H. vulgaris*, *Taraxacum* sp., *S. alpina*, wild angelica (*Angelica sylvestris*) and common hogweed (*Heracleum sphondylium*) that are usually rare, or difficult to identify, in the pollen record. In this case, the molecular approach was a significant complement to pollen analysis, enhancing the information on changes in plant composition and leading to a better representation of the fossil record. The number of taxa that might be expected from a DNA approach is a very new area, so we suggest that there is insufficient comparative data for saying whether our results are reasonable. For the pollen data, our finding of 46 taxa seems reasonable. The number of pollen types identified increases as one counts more and more pollen, and this study counted relatively little pollen in total (>500 grains at 18 levels = *c.* 9000 grains), but the gain in taxa levels off as the total count increases. In our study, we suggest that it would be reasonable to find *c.* 20–30 taxa in one pollen count of 500 grains, and subsequent counts would have all the same common taxa and a slightly different subset of less common and rare taxa.

Our pollen results were consistent with previous palaeoecological studies conducted on the central part of the Scandes Mountains (Giesecke 2005; von Stedingk *et al.* 2008), showing a similar vegetation at both lakes dominated by forest tree species, *P. sylvestris*, *Betula* sp. and *Alnus* sp., and a limited presence of thermophilous species (*Corylus avellana*, *Alnus glutinosa* and *Ulmus glabra*), unusual at these latitudes but previously also detected (Giesecke 2005). In contrast, the barcoding analysis suggested an almost treeless environment at Rundtjørna (with the exception of *Sorbus aucuparia*) and a predominance of Nymphaeaceae and other herbaceous plants (Brassicaceae, Asteraceae, Ericaceae, Fabaceae and Poaceae). At Klocka instead, barcoding results suggested an environment dominated by trees (*Alnus* sp., *Betula* sp. *Salix* sp., *Populus* sp. and *S. aucuparia*) and as such, the results were more in accordance with the pollen analysis. However, at both lakes, we found no DNA signal from major tree taxa detected by pollen like *P. sylvestris*, *P. abies*, elm (*Ulmus* sp.) and oak (*Quercus* sp.). The little overlap found between proxies (25%) was also evident at the family level, as the molecular approaches could detect less than one-third of the families identified by pollen. Yet, in the majority of the cases, the barcoding provided identification at a low taxonomic level (genus or species), while many of the taxa found by pollen could be identified at family level only. As such, our findings confirm previous conclusions that barcoding and pollen should be used in combination to achieve a detailed palaeovegetational reconstruction in ancient environments (Jørgensen *et al.* 2012).

There are several reasons why the barcoding vs. traditional approaches could differ: (i) the pollen and the molecular record could contain different plant tissues and truly represent different plant communities, (ii) taxonomic resolution could differ between the two approaches, (iii) amplification and/or sequencing of *trn*L DNA could be biased for/against certain species, (iv) the pollen record could be biased for/against certain species. Taphonomy (i.e. the study of fossilization processes) and DNA preservation conditions are likely the two major factors affecting the molecular record in the sediments and, hence, responsible for the difference in results across techniques. The applications of the barcoding technique to eDNA from lake sediments need therefore particular attention to taphonomic issues, including the sources of the aDNA. We unfortunately know too little about decay processes that occur in lacustrine environments. Current estimates suggest that optimal conditions for aDNA survival are cold environments, such as permafrost or ice; however, nonfrozen and anaerobic environments such as lakes also seem to preserve DNA well for several thousand years (Anderson-Carpenter *et al.* 2011). However, it is not clear what proportion of DNA damage occurs within the aquatic and the sediment fraction of a lake and whether the lake depth plays any important role for DNA preservation. Our barcoding results seem to suggest a better preservation of DNA at Klocka (10 m depth) than at Rundtjørna (3 m depth); hence depth may favour preservation, but our study lacks respective replication to test this hypothesis. The DNA results might be influenced by physical processes like vertical migration or leaching of DNA molecules through sediments, as periodic downward percolation of water can move DNA in porous, granular sediments of cave profiles (Haile *et al.* 2007). Lake sediments, however, are permanently saturated so that vertical percolation of liquids does not occur and organic compounds are immobilized in the sediment matrix (Anderson-Carpenter *et al.* 2011).

The source of plant DNA in lake sediments is also uncertain. It is believed that bulk sediments contain plant aDNA from a variety of sources (seed, roots,

leaves, fruits, pollen, etc.), but it is not clear which of these tissues contribute most to the total DNA yield and how this is later preserved. Seeds, needles and bud scales can be blown or washed over long distances like pollen, especially in open ice-covered landscapes. In Arctic regions, the long-distance proportion of plant material deposited in lakes can be high compared with that of any local component, but it is low for boreal and temperate lakes because of the abundant wind-dispersed pollen produced by trees (Gajewski 1995). The failure to detect *P. sylvestris* in samples where pollen was particularly abundant suggests that eDNA is from sources other than pollen. Jørgensen *et al.* (2012) also failed to retrieve DNA signals from major tree taxa (*Picea* sp., *Pinus* sp., *Larix* sp., *Corylus* sp., *Betula* sp. and *Alnus* sp.) all present in the pollen record in permafrost samples from northern Siberia. Indeed, DNA from pollen is difficult to amplify, particularly from mixed sediments, as the grains contain small amounts of DNA and are difficult to open for amplification (Parducci *et al.* 2005). This is also in agreement with previous assumptions of eDNA originating from the local environment around the lake rather than being transported over long distances (Willerslev *et al.* 2007; Haile *et al.* 2009; Jørgensen *et al.* 2012; Parducci *et al.* 2012). Several other studies on different soil types have also demonstrated that eDNA reflects contemporaneous plant biomass derived from above ground and provides a good proxy for species identification (e.g. Andersen *et al.* 2012; Yoccoz *et al.* 2012).

If we therefore assume that DNA from lake sediments is local in origin, is well enough preserved for DNA studies within a time frame of *c.* 20 000 years and does not move through the sediments, our inability to barcode major tree taxa like *P. sylvestris* and *P. abies* in Holocene sediments is intriguing. In a previous study, in the absence of fossil evidence, Parducci *et al.* (2012) succeeded in obtaining *P. abies* DNA signal using the same barcoding primers (*trn*L) in late-glacial sediments (17 700 cal. yr BP) from Andøya, in northern Norway. However, in the same study, the same DNA signal could be retrieved from younger sediments (10 400 cal. yr BP) at Rundtjørna only when using species-specific primers. This result demonstrates the power of the molecular technique in tracing taxa with no pollen or macrofossil evidence, but it also highlights the problems associated with the generic barcoding primers when used in different lacustrine environments.

Likely, one crucial issue while barcoding ancient plant taxa is the different amplification preference among DNA templates due to selectiveness of the generic markers, as the *trn*L primers are not completely conserved between species (*quasi* universal). In addition, because eDNA of small lakes is local in origin and represents trees growing within a few metres of the lake margin, only a small number of individuals are probably represented in molecular assemblages from different lakes. These two facts, in combination with the varying levels of decay processes associated with each lake (due to water depth, rate of deposition, etc.), will inevitably cause differences between taxa recovered at different sites and, consequently, also differences in results between proxies. Finally, because ancient eDNA cannot be dated directly, it should be recognized that the temporal accuracy of the barcoding technique relies on the sediment chronology for dating the DNA remains, which in turn relies on it being contemporaneous with the associated dated material. This appeared to be the case in our study, as shown by the age-depth model suggesting an approximately constant sediment accumulation rate at Rundtjørna, which indicates that the sediments remained undisturbed after deposition.

Conclusions

On the whole, the pollen and barcoding analyses resulted in largely nonoverlapping species/taxon lists (25% overlapping); hence, the plant composition in the two Holocene lake records could be satisfactorily detected by combining results from the two techniques. The two proxies provided different levels of taxonomic resolution and, in combination, revealed more detailed information on plant composition than could be achieved by each proxy individually. Some of the gain in taxa from the DNA over pollen was mainly through finer taxonomic resolution (e.g. Poaceae, Asteraceae). Our conclusion is, however, that metabarcoding analysis, although very powerful in tracing taxa with no pollen or macrofossil evidence, if run alone, is more risky and results may be biased across lakes. As it is likely that ancient eDNA originates from remains other than pollen and is local in origin, the combination of the two techniques is particularly useful for past vegetation reconstruction in the case of low pollen productivity of plants (e.g. insect-pollinated plants that often tend to result in poor representation of many species in the palynological data), or in the case of low-density plant populations, or plants living in extreme environments where they may not produce pollen, but instead reproduce vegetatively (e.g. *Picea abies*) (Parducci *et al.* 2012).

These findings, coupled with previous results from Jørgensen *et al.* (2012) and Parducci *et al.* (2012), also suggest that in the metabarcoding of eDNA samples, the use of generic primers, like the *trn*L, should be taken as a complementary approach to pollen analysis for flora reconstruction. In particular, they should be used in combination with species-specific primers whenever particular taxa are desired to be targeted or if one wants to test

for species not present in either records (pollen and barcoding), which could thus complement the incomplete species list of the flora present in a given area and period. This is because generic primers may have a limited capability of recovering all taxa, especially when some are underrepresented in an environment where others dominate. This can be due to either original species richness in the catchment of the lake or taphonomic issues associated with the lake environment.

Acknowledgements

This study was supported by the Swedish Research Council (grant 2007-4490) and the Carl Trygger's Foundation (grant 08:303) to LP; a Royal Society–Wolfson Merit Award to KDB; the Environment Research and Technology Development Fund (S9) of the Ministry of the Environment, Japan, and KAKENHI (Nos. 22658046 and 24248025) to YS; and the Danish National Research Foundation to EW. The authors declare no competing financial interests.

References

Andersen K, Bird KL, Rasmussen M *et al.* (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21**, 1966–1979.

Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY, Poinar HN (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology*, **11**, 30.

Balzer S, Malde K, Lanzén A, Sharma A, Jonassen I (2010) Characteristics of 454 pyrosequencing data-enabling realistic simulation with flowsim. *Bioinformatics*, **26**, i420–i425.

Bennett KD (2004) Pollen catalogue of the British Isles. Available from http://chrono.qub.ac.uk/pollen/pc-intro.html.

Bennett KD, Willis KJ (2001) Pollen. In: *Tracking Environmental Change Using Lake Sediments. Vol. 3. Terrestrial, Algal, and Siliceous Indicators* (eds Smol JP, Birks HJB, Last WM), pp. 5–32. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Binladen J, Wiuf C, Gilbert MTP *et al.* (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics*, **172**, 733–741.

Binney HA, Willis KJ, Edwards ME *et al.* (2009) The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, **28**, 2445–2464.

Birks HH, Birks HJB (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31–35.

Bulat SA, Lübeck M, Alekhina IA, Jensen DF, Knudsen IMB, Lübeck PS (2000) Identification of a universally primed-PCR-derived sequence-characterized amplified region marker for an antagonistic strain of *Clonostachys rosea* and development of a strain-specific PCR detection assay. *Applied Environmental Microbiology*, **66**, 4758–4763.

Eide W, Birks HH, Bigelow NH, Peglar SM, Birks HJB (2006) Holocene forest development along the Setesdal valley, southern Norway, reconstructed from macrofossil and pollen evidence. *Vegetation History and Archaeobotany*, **15**, 65–85.

Feek DT, Horrocks M, Baisden WT, Flenley J (2011) The Mk II sampler: a device to collect sediment cores for analysis of uncontaminated DNA. *Journal of Paleolimnology*, **45**, 115–119.

Gajewski K (1995) Modern and Holocene pollen assemblages from some small arctic lakes on Somerset Island, NWT, Canada. *Quaternary Research*, **44**, 228–236.

Giesecke T (2005) Holocene forest development in the central Scandes Mountains, Sweden. *Vegetation History and Archaeobotany*, **14**, 133–147.

Giesecke T, Bennett KD (2004) The Holocene spread of *Picea abies* (L.) Karst. in Fennoscandia and adjacent areas. *Journal of Biogeography*, **31**, 1523–1548.

Gilbert MT, Kivisild T, Grønnow B *et al.* (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science*, **320**, 1787–1789.

Gugerli F, Parducci L, Petit R (2005) Ancient plant DNA: review and prospects. *New Phytologist*, **166**, 409–418.

Haile J, Holdaway R, Oliver K, Bunce M *et al.* (2007) Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution*, **24**, 982–989.

Haile J, Froese DG, MacPhee RDE *et al.* (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences, USA*, **106**, 22363–22368.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution*, **18**, 262–265.

Hebert PDN, Ratnasingham S, de Waard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.

Hofreiter M, Poinar HN, Spaulding WG *et al.* (2000) A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology*, **9**, 1975–1984.

Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA*, **106**, 12794–12797.

Höss M, Jaruga P, Zastawny T, Dizdaroglu M, Pääbo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research*, **24**, 1304–1307.

Huntley B, Birks HJB (1983) *An Atlas of Past and Present Pollen Maps of Europe: 0–13,000 years ago*. Cambridge University Press, Cambridge.

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.

Karlsson T (1998) Förteckning över svenska kärlväxter [The vascular plants of Sweden – a checklist]. *Svensk Botanisk Tidskrift*, **91**, 241–560. (in Swedish).

Kullman L (2002) Boreal tree taxa in the central Scandes during the Late-Glacial: implications for Late-Quaternary forest history. *Journal of Biogeography*, **29**, 1117–1124.

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

Lydolph MC, Jacobsen J, Arctander P *et al.* (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology*, **71**, 1012–1017.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Moore PD, Webb JA, Collinson ME (1991) *Pollen Analysis*. Blackwell Scientific Publications, Oxford.

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008a) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology*, **57**, 750–757.

Munch K, Boomsma W, Willerslev E, Nielsen R (2008b) Fast phylogenetic DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences*, **363**, 3997–4002.

Murray DC, Bunce M, Cannell BL *et al.* (2011) DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS ONE*, **6**, e25776.

Olivieri C, Ermini L, Rizzi E *et al.* (2010) Characterization of nucleotide misincorporation patterns in the iceman's mitochondrial DNA. *PLoS ONE*, **5**, e8629.

Palmer SA, Smith O, Allaby RG (2012) The blossoming of plant archaeogenetics. *Annals of Anatomy*, **194**, 146–156.

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

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

Paus A, Velle G, Berge J (2011) The Lateglacial and early Holocene vegetation and environment in the Dovre mountains, central Norway, as signalled in two Lateglacial nunatak lakes. *Quaternary Science Reviews*, **30**, 1780–1796.

Poinar HN, Hofreiter M, Spaulding WG *et al.* (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **281**, 402–406.

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T (2010) Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal*, **64**, 948–959.

Ramsey CB (2008) Deposition models for chronological records. *Quaternary Science Reviews*, **27**, 42–60.

Rauwolf U, Golczyk H, Greiner S, Herrmann R (2010) Variable amounts of DNA related to the size of chloroplasts III. Biochemical determinations of DNA amounts per organelle. *Molecular Genetics and Genomics*, **283**, 35–47.

Reimer PJ, Baillie MGL, Bard E *et al.* (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon*, **51**, 1111–1150.

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

Segerström U, von Stedingk H (2003) Early-Holocene spruce, *Picea abies* (L.) Karst., in west central Sweden as revealed by pollen analysis. *The Holocene*, **13**, 897–906.

Seppä H, Bennett KD (2003) Quaternary pollen analysis: recent progress in palaeoecology and palaeoclimatology. *Progress in Physical Geography*, **27**, 548–579.

Sønstebø JH, Gielly L, Brysting AK *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, **10**, 1009–1018.

von Stedingk H, Fyfe RM, Allard A (2008) Pollen productivity estimates from the forest—tundra ecotone in west-central Sweden: implications for vegetation reconstruction at the limits of the boreal forest. *The Holocene*, **18**, 323–332.

Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, e14.

Thomsen P, Kielgast J, Iversen L *et al.* (2012a) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565–2573.

Thomsen P, Kielgast J, Iversen L, Møller P *et al.* (2012b) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, **7**, e41732.

Väliranta M, Kaakinen A, Kuhry P, Kultti S, Sakari Salonen J, Seppä H (2011) Scattered late-glacial and early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia. *Journal of Biogeography*, **38**, 922–932.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

Willerslev E, Hansen AJ, Poinar HN (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution*, **19**, 141–147.

Willerslev E, Cappellini E, Boomsma W *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, **317**, 111–114.

Wright HE, Mann DH, Glaser PH (1984) Piston corers for peat and lake sediments. *Ecology*, **65**, 657–659.

Yoccoz NG, Bråthen KA, Gielly L *et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, **21**, 3647–3655.

L.P. conceived and designed the work with contributions from other authors; L.P., K.D.B. and Y.S. participated in the fieldwork; L.P., I.M., Y.S., J.H., A.D.D. and E.W. contributed to the molecular data; S.L.F. and K.D.B. contributed to the palaeoecological data; L.P., K.H.K. and N.K.L. contributed to analyses and interpretation of sediments. All authors contributed to the interpretation of the results; L.P. and I.M. wrote the manuscript with contributions from other authors.

Data accessibility

The core pollen and age data, the *trn*L sequences and taxonomic assignments for Copenhagen (NGS), and the *trn*L sequences and taxonomic assignments for Uppsala (Sanger/cloning) Dryad doi: 10.5061/dryad.9n815.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Sub-set of the Arctic database [17] and other boreal plant species currently occurring at the sites of this study.