

RESEARCH ARTICLE

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

Inger Greve Alsos<sup>1\*</sup>, Youri Lammers<sup>1</sup>, Nigel Giles Yoccoz<sup>2</sup>, Tina Jørgensen<sup>1</sup>, Per Sjögren<sup>1</sup>, Ludovic Gielly<sup>3,4</sup>, Mary E. Edwards<sup>1,5</sup>

**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](mailto: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

**Copyright:** © 2018 Alsos et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**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](https://doi.org/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 a 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 constraints. 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 *trnL* (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.norgebilder.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) Gaupstjern, f) Jula Jávri, g) Lakselvhogda, h) Lauvås, i) Øvre Ærásvatnet, 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.norgebilder.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 lake-shore. 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 <sup>2</sup> )	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 <sup>a</sup>	Dividalen	Mixed forest/mire, tall herbs	0.17	125	1.70	5.5	-0.8	6.9	636	3	68.996	19.486
Brennskogtjønn	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 <sup>c</sup>	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 <sup>b</sup>	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.

<sup>a</sup>Named A-tjern in Jensen & Vorren 2008. Named “Vesltjønn” on NEVINA but this name is not official.

<sup>b</sup>Named B-tjern in Jensen & Vorren 2008, but later official named Rottjern.

<sup>c</sup>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 21<sup>st</sup> and October 1<sup>st</sup>, 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 *trnL* (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<sub>2</sub>, 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:

$$\begin{aligned} Prob(DNA = 0, Vegetation = 0) \\ = (1 - P_{OCC})(1 - P_{DNA_0}) + P_{OCC}(1 - P_{DNA_1})(1 - P_{VEG_1}) \end{aligned} \quad 1$$

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  $(1 - p_{VEG_1})$ .

$$Prob(DNA = 0, Vegetation = 1) = P_{OCC}(1 - P_{DNA_1})P_{VEG_1} \quad 2$$

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

$$Prob(DNA = 1, Vegetation = 0) = (1 - P_{OCC})P_{DNA_0} + P_{OCC}P_{DNA_1}(1 - P_{VEG_1}) \quad 3$$

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.

$$Prob(DNA = 1, Vegetation = 1) = P_{OCC}P_{DNA_1}P_{VEG_1} \quad 4$$

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  $p_{DNA_I}$ , we also considered a model assuming a logistic relationship between  $p_{DNA_I}$  and lake characteristics, such as lake depth or catchment area, that is:  $\text{logit}(p_{DNA_I}) = b_0 + b_1 \text{Lake Covariate}$ . 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 Æräsvatnet	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 an 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), Gaupstjern (Gaup), Jula Jávri (Jula), Lakselvhogda (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  $\geq 0.03$  ng/ $\mu$ L (S5 Table).

The gain in number of taxa when analysing two cores instead of one was  $2.5 \pm 1.2$  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  $19.7 \pm 6.9$  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	<i>Crepis paludosa</i>									455			455
Betulaceae	<i>Alnus incana</i>	48 183	117 855			40 802					131	15 710	222 681
Betulaceae	<i>Betula</i> spp.	126 727	120 369	40 991	5 630	101 688		144	32	31 639	3 263	16 283	446 766
Caryophyllaceae	<i>Sagina</i> sp.		46	10	37		18			10	24		145
Cornaceae	<i>Chamaepericlymenum suecicum</i>			338									338
Cupressaceae	<i>Juniperus communis</i>	261	752						45		27		1 085
Cyperaceae	<i>Carex lasiocarpa</i>	47				76						84	207
Cyperaceae	<i>Carex</i> spp.	34		48	33	72							187
Dryopteridaceae	<i>Dryopteris</i> spp.	10 088	16 947	6 406	6 781	5 882	87	1 886	1 141	6 252	216	5 239	60 925
Ericaceae	<i>Andromeda polifolia</i>	191		235		244			23			310	1 003
Ericaceae	<i>Calluna vulgaris</i>			1 384	357								1 741
Ericaceae	<i>Cassiope tetragona</i>	181	86								163		430
Ericaceae	<i>Chamaedaphne calyculata</i>	31		29		46						41	147
Ericaceae	<i>Empetrum nigrum</i>	3 466	12 736	2 266	4 714	2 807	6 813		14	3 149	13 507	1 758	51 230
Ericaceae	<i>Oxycoccus microcarpus</i>					538							538
Ericaceae	<i>Phyllodoce caerulea</i>	1 386	305				165						1 856
Ericaceae	<i>Vaccinium vitis-idaea/myrtillus</i>	2 005	2 042	916	308	1 286				189	815	394	7 955
Ericaceae	<i>Vaccinium uliginosum</i>	1 073	2 325	1 045		2 726	431	30		1 233	1 014	873	10 750
Geraniaceae	<i>Geranium sylvaticum</i>									68	145		213
Haloragaceae	<i>Myriophyllum alterniflorum</i>	11 389		273 929						226 753			512 071
Isoetaceae	<i>Isoetes</i> spp.			27 136						14 411			41 547
Lentibulariaceae	<i>Utricularia minor</i>											893	893
Lycopodiaceae	<i>Huperzia selago</i>	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	<i>Menyanthes trifoliata</i>	26 842	467	17 384	1 173	18 978		98	871	378		42 408	108 599
Nymphaeaceae	<i>Nuphar pumila</i>											63 844	63 844
Plantaginaceae	<i>Callitriche hermaphroditica</i>			951			5 598						6 549
Plantaginaceae	<i>Hippuris vulgare</i>			238						107			345
Poaceae	<i>Festuca</i> spp.	30			2 724								2 754
Polygonaceae	<i>Oxyria digyna</i>										429		429
Polypodiaceae	<i>Athyrium</i> sp.	6 266	33 588	10 557	2 098	1 258		743	539	10 851	1 239	466	67 605
Potamogetonaceae	<i>Potamogeton praelongus</i>	1 754								254		9 268	11 276
Potamogetonaceae	<i>Potamogeton</i> sp.	28		19 281						12 817		1 335	33 461
Potamogetonaceae	<i>Stuckenia filiformis</i>	4 964	183			7 023						246	12 416
Ranunculaceae	<i>Caltha palustris</i>			1 131						5 080			6 211
Rosaceae	<i>Comarum palustre</i>	258				1 058				222			1 538
Rosaceae	<i>Dryas octopetala</i>		750				37					394	1 181
Rosaceae	<i>Filipendula ulmaria</i>	850		957		2 293				2 520		6 019	12 639
Rosaceae	<i>Rubus chamaemorus</i>	1 453		75		197						317	2 042
Rosaceae	<i>Sorbus aucuparia</i>	1 198	894	1 915		1 953				1 468			7 428
Salicaceae	<i>Populus tremula</i>	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	<i>Saxifraga aizoides</i>		585	30									615

(Continued)

Table 3. (Continued)

Family	Taxa	A-tj	Bren	Einl	Finn	Gaup	Jula	Laks	Lauv	Ovre	Paul	Rott	Sum
Saxifragaceae	<i>Saxifraga oppositifolia</i>		922										922
Sparganiaceae	<i>Sparganium</i> spp.				958			258	74				1 290
Thelypteridaceae	<i>Phegopteris connectilis</i>	4 776	13 594	1 104	1 357	100		546	132	2 085	1 014	366	25 074
Woodsiaceae	<i>Gymnocarpium dryopteris</i>	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												

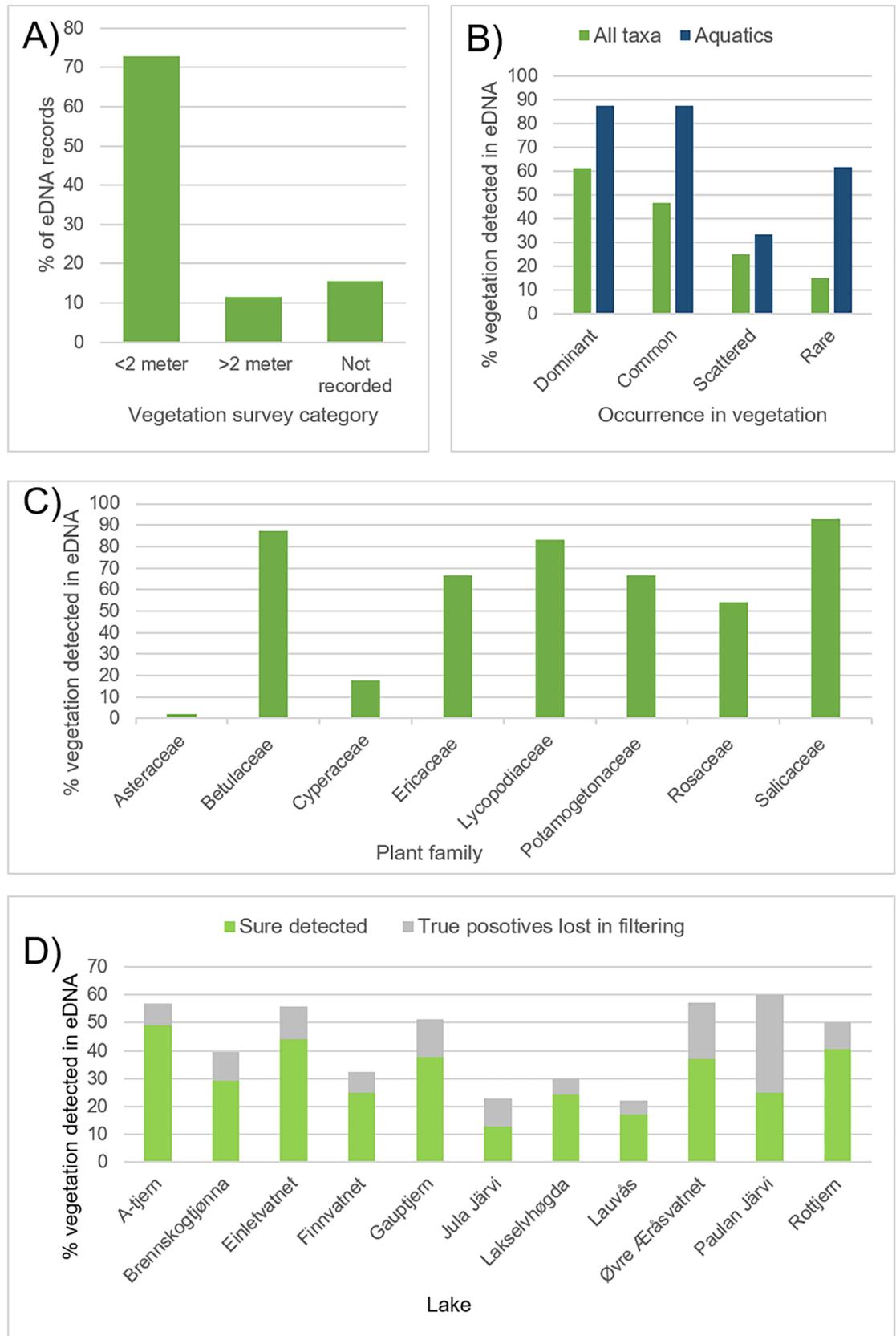
<https://doi.org/10.1371/journal.pone.0195403.t003>

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, Nymphaeaceae, 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).



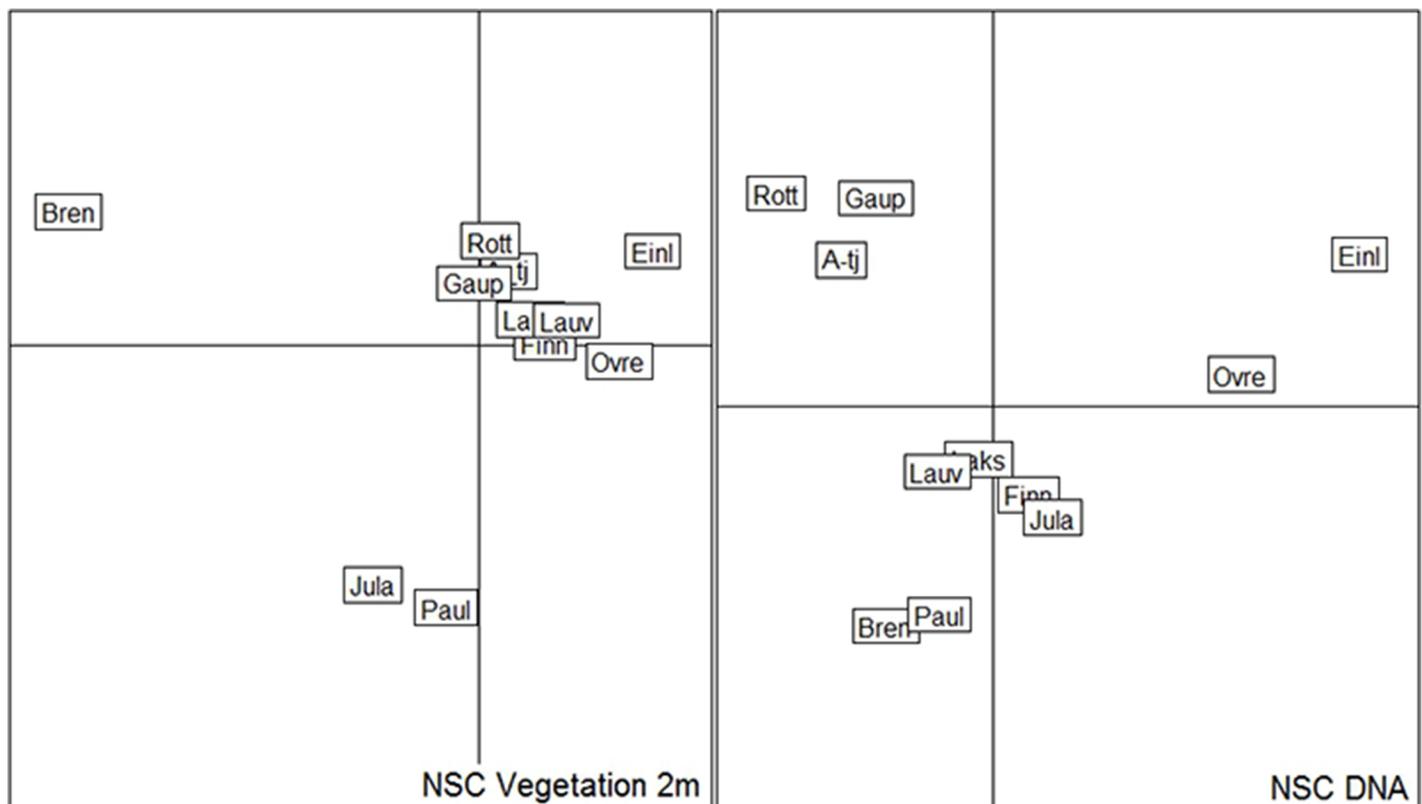
**Fig 3. Match between records of taxa in the sedimentary eDNA in relation to vegetation surveys.** a) Number of records in the sedimentary eDNA in relation to vegetation survey distance. b) Percentage records in eDNA in relation to abundance in vegetation surveys. c) Variation in percentage data among 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,  $\lambda = 0.50$ , 20.4%, Axis 2,  $\lambda = 0.37$ , 15.1%; eDNA: Axis 1,  $\lambda = 0.24$ , 18.9%, Axis 2,  $\lambda = 0.24$ , 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, Gaupjern are clustered together in both plots; so are also Einletvatnet and Øvre Ærâsvatnet (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>

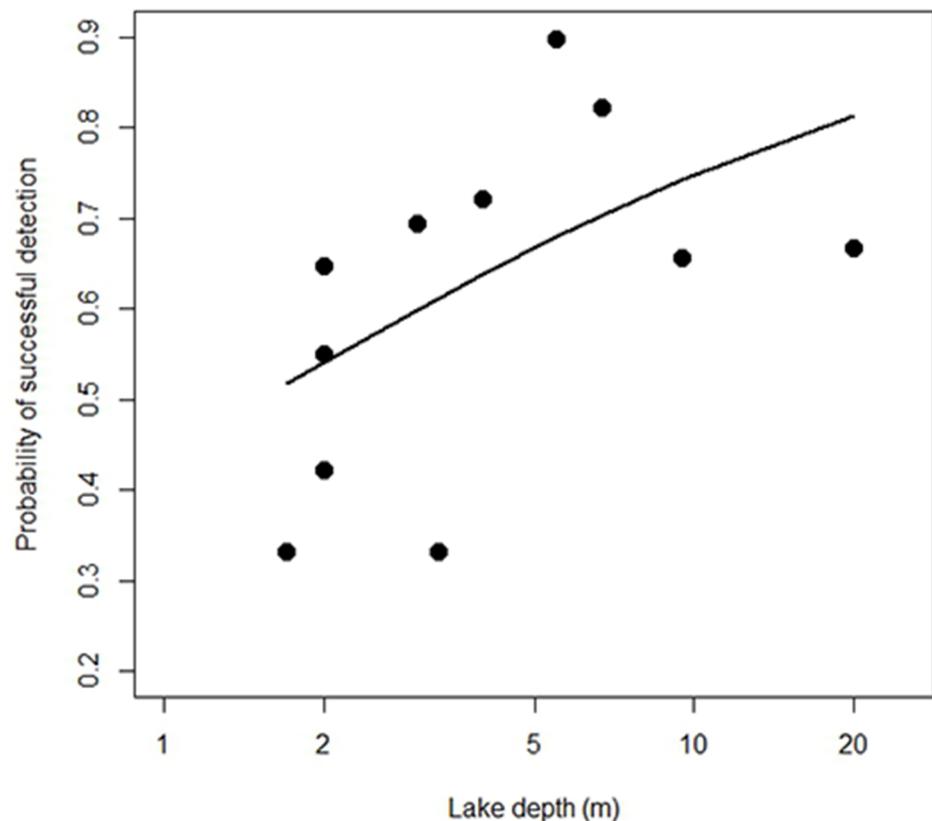
at least once by eDNA were also recorded in the DNA in a given lake (true positives) was 0.33–0.90, whereas the posterior probability of any DNA records representing a false positive varied from 0.06–0.33 per lake (S6 Table). There was evidence that the probability of detecting a species using eDNA ( $p_{DNA-I}$ ) 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 incom = 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), Gaupjern (Gaup), Jula Jávri (Jula), Lakselvhogda (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), Gaupjern (Gaup), Jula Jávri (Jula), Lakselvhogda (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 Sjøgren, 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 Sjøgren, 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, Giguët-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-Aroita 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. Giguët-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, Haffidason 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, Väiliranta 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](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 north-western 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](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, McGlenn 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, Pestyakova 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, Schröder-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, Stooß-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, Stooß-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>