Manuscript bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.
Click here to download Manuscript Alsos et al main text PlosOne.docx

1 # Metabarcoding lake sediments: taphonomy and representation of
2 # contemporary vegetation in environmental DNA (eDNA) records

3

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

6

7 [1]Tromsø Museum, University of Tromsø – The Arctic University of Norway, NO-9037 Tromsø,
8 Norway

9

10 [2]Department of Arctic and Marine Biology, University of Tromsø – The Arctic University of Norway,
11 NO-9037 Tromsø, Norway

12

13 [3]University Grenoble Alpes, LECA, F-38000 Grenoble, France

14

15 [4]CNRS, LECA, F-38000 Grenoble, France

16

17 [5]Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

18

19 *Corresponding author
20 Email: inger.g.alsos@uit.no

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21 # Abstract

22 Metabarcoding of lake sediments may reveal current and past biodiversity, but little is known
23 about the degree to which taxa growing in the vegetation are represented in environmental DNA
24 (eDNA) records. We analysed composition of lake and catchment vegetation and vascular plant eDNA
25 at 11 lakes in northern Norway. Out of 489 records of taxa growing within 2 m from the lake shore,
26 17-49% (mean 31%) of the identifiable taxa recorded were detected with eDNA. Of the 217 eDNA
27 records, 73% and 12% matched taxa recorded in vegetation surveys within 2 m and up to about 50 m
28 away from the lakeshore, respectively, whereas 16% were not recorded in the vegetation surveys of
29 the same lake. The latter include taxa likely overlooked in the vegetation surveys or growing outside
30 the survey area. The percentages detected were 61, 47, 25, and 15 for dominant, common, scattered,
31 and rare taxa, respectively. Similar numbers for aquatic plants were 88, 88, 33 and 62%, respectively.
32 Detection rate and taxonomic resolution varied among plant families and functional groups with good
33 detection of e.g. Ericaceae, Roseaceae, deciduous trees, ferns, club mosses and aquatics. The
34 representation of terrestrial taxa in eDNA depends on both their distance from the sampling site and
35 their abundance and is sufficient for recording vegetation types. For aquatic vegetation, eDNA may be
36 comparable with, or even superior to, in-lake vegetation surveys and therefore be used as an tool for
37 biomonitoring. For reconstruction of terrestrial vegetation, technical improvements and more intensive
38 sampling is needed to detect a higher proportion of rare taxa although DNA of some taxa may never
39 reach the lake sediments due to taphonomical constrains. Nevertheless, eDNA performs similar to
40 conventional methods of pollen and macrofossil analyses and may therefore be an important tool for
41 reconstruction of past vegetation.

42

43

44

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

45 # Introduction

46

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

58

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

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

71 and the number of DNA reads may even reflect the biomass of species [6, 7, 21], making this a
72 promising method for biodiversity monitoring.
73
74 When applied to late-Quaternary sediments, eDNA analysis may help disclose hitherto
75 inaccessible information, thus providing promising new avenues of palaeoenvironmental
76 reconstruction [22, 23]. Lake sediments are a major source of palaeoenvironmental information [24]
77 and, given good preservation, DNA in lake sediments can provide information on biodiversity change
78 over time [4, 22, 25]. However, sedimentary ancient DNA is still beset by authentication issues [2,
79 10]. For example, the authenticity and source of DNA reported in several recent studies have been
80 questioned [26-30]. As with pollen and macrofossils [31, 32], we need to understand the source of the
81 DNA retrieved from lake sediments and know which portion of the flora is represented in DNA
82 records.
83
84 The P6 loop of the plastid DNA *trn*L (UAA) intron [33] is the most widely applied marker for
85 identification of vascular plants in environmental samples such as Pleistocene permafrost samples [34-
86 36], late-Quaternary lake sediments [15, 22, 27, 37-41], sub-modern or modern lake sediments [42],
87 animal faeces [43, 44], and sub-modern or modern soil samples [6, 45]. While some studies include
88 comparator proxies to assess the ability of DNA to represent species diversity (e.g., [35, 41, 46, 47],
89 only one study has explicitly tested how well the floristic composition of eDNA assemblages reflect
90 the composition of extant plant communities [6], and similar tests are urgently needed for lake
91 sediments. Yoccoz *et al.* found most common species and some rare species in the vegetation were
92 represented in the soil eDNA at a subarctic site in northern Norway. The present study attempts a
93 similar vegetation-DNA calibration in relation to lake sediments.
94
95 We retrieved sedimentary eDNA and recorded the vegetation at 11 lakes that represent a
96 gradient from boreal to alpine vegetation types in northern Norway. We chose this area because DNA
97 is best preserved in cold environments and because an almost complete reference library is available
98 for the relevant DNA sequences for arctic and boreal taxa [34, 36]. Our aims were to 1) increase our

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99 understanding of eDNA taphonomy by determining how abundance in vegetation and distance from
100 lake shore affect the detection of taxa, and 2) examine variation in detection of DNA among lakes and
101 taxa. Based on this, we discuss the potential of eDNA from lake sediments as a proxy for modern and
102 past floristic richness.
103
104

105 Materials and Methods

106 Study sites

107 Eleven lakes were selected using the following criteria: 1) lakes size within the range of lakes
108 studied for pollen in the region and with limited inflow and outflow streams; 2) a range of vegetation
109 types from boreal forest to alpine heath was represented; and 3) lakes sediments are assumed to be
110 undisturbed by human construction activity. Six of the lakes were selected also for the availability of
111 pollen, macro and/or ancient DNA analyses [27, 48-52]. Data on catchment size, altitude, yearly mean
112 temperature, mean summer temperature and yearly precipitation were gathered using NEVINA
113 (http://nevina.nve.no/) from the Norwegian Water Resources and Energy Directorate (NVE,
114 https://www.nve.no). Lake size was calculated using http://www.norgeibilder.no/. Number and size of
115 inlets and outlets were noted during fieldwork.
116
117 **Fig 1. Study lakes in northern Norway.** a) A-tjern, b) Brennskogtjørna, c) Einletvatnet, d)
118 Finnvatnet, e) Gauptjern, f) Jula Jävrí, g) Lakselvhøgda, h) Lauvås, i) Øvre Æråsvatnet, j) Paulan
119 Jávri, k) Rottjern, l) Tina Jørgensen sampling surface sediments with Kajak corer. Photo: I.G. Alsos.
120
121

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

122 ## Vegetation surveys

123 We attempted to record all species growing within 2 m from the lakeshore. This was a
124 practically achievable survey, and data are comparable among sites. Aquatics were surveyed from the
125 boat using a "water binocular" and a long-handled rake, while rowing all around smaller lakes and at
126 least half way around the three largest lakes. We also surveyed a larger part of the catchment
127 vegetation. For this, we used aerial photos (http://www.norgeibilder.no) to identify polygons of
128 relatively homogeneous vegetation (including the area within 2 m). In the field we surveyed each
129 polygon and classified observed species as rare (only a few ramets), scattered (ramets occur
130 throughout but at low abundance), common, or dominant. The area covered and intensity of these
131 broad-scale vegetation surveys varied among lakes due to heterogeneity of the vegetation, catchment
132 size and time constraints. They mainly represent the vegetation within 50 m of the lakeshore. Sites
133 were revisited several times during the growing season to increase the detection rate. For each lake our
134 dataset consisted of a taxon list for 1) the <2-m survey, 2) the extended survey consisting of
135 observations from <2 m and the polygons, 3) an abundance score based on the highest abundance
136 score from any polygon at that lake. Taxonomy follows [53, 54].
137

138 ## Sampling lake sediments

139 Surface sediments were collected from the centres of the lakes between September 21st and
140 October 1st, 2012, using a Kajak corer (mini gravity corer) modified to hold three core tubes spaced 15
141 cm apart, each with a diameter of 3 cm and a length of 63 cm (Fig 1, Table 1). The core tubes were
142 washed in Deconex®22 LIQ-x and bleached prior to each sampling. The top 8 cm sediments were
143 extruded in field. Samples of ca. 25 mL were taken in 2-cm increments and placed in 50-ml falcon
144 tubes using a sterilized spoon. All samples were frozen until extraction.
145

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

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

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

[a,b]Named A-tjern and B-tjern in Jensen& Vorren 2008. B-tjern later official named Rottjern. A-tjern named “Vesltjønna” on NEVINA but this name is not official.

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

147
148

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

149 # DNA extraction and amplification

150 For each lake, we analysed the top 0-2 cm of sediment separately from two of the three core
151 tubes (n=22). Twenty extra samples from lower in the cores were also analysed. The main down-core
152 results will be presented in a separate paper in which we compare eDNA records with the pollen
153 analyses by [49]. Taxa that were only identified from lower levels in the cores are noted in S1 Table.
154 Samples were thawed in a refrigerator over 24-48 hours, and 4-10 g were subsampled for DNA. The
155 42 samples and 6 negative controls underwent extraction at the Department for Medical Biology,
156 University of Tromsø, in a room where no previous plant DNA work had been done. A PowerMax
157 Soil DNA Isolation kit (MO BIO Laboratories, Carlsbad, CA, USA) was used following the
158 manufacturer´s instructions, with water bath at 60°C and vortexing for 40 min.
159
160 All PCRs were performed at LECA (Laboratoire d'ECologie Alpine, University Grenoble
161 Alpes), using the *g* and *h* universal plant primers for the short and variable P6 loop region of the
162 chloroplast *trn*L (UAA) intron [33]. Primers include a unique flanking sequence of 8 bp at the 5' end
163 (tag, each primer pair having the same tag) to allow parallel sequencing of multiple samples [55, 56].
164 PCR and sequencing on an Illumina 2500 HiSeq sequencing platform follows [41], using six PCR
165 negative controls and two positive controls, and six different PCR replicates for each of the 56
166 samples, giving a total of 336 PCR samples, of which 216 represent the upper 0-2 cm.
167

168 # DNA sequences analysis and filtering

169 Initial filtering steps were done using OBITools [57] following the same criteria as in [41, 42]
170 (S2 Table). We then used *ecotag* program [57] to assign the sequences to taxa by comparing them
171 against a local taxonomic reference library containing 2445 sequences of 815 arctic [34] and 835
172 boreal [36] vascular plant taxa; the library also contained 455 bryophytes [44]. We also made
173 comparisons with a second reference library generated after running *ecopcr* on the global EMBL
174 database (release r117 from October 2013). Only sequences with 100% match to a reference sequence

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

175 were kept. We excluded sequences matching bryophytes as we did not include them in the vegetation
176 surveys. We used BLAST (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/blast/)
177 to check for potential wrong assignments of sequences.
178
179 When filtering next-generation sequencing data, there is a trade-off between losing true
180 positives (TP, sequences present in the samples and correctly identified) and retaining false positives
181 (FP, sequences that originate from contamination, PCR or sequencing artefacts, or wrong match to
182 database) [17, 20, 58]. We therefore assessed the number of TP and FP when applying different last
183 step filtering criteria. We initially used two spatial levels of comparison with the DNA results: i) data
184 from our vegetation surveys and ii) the regional flora (i.e., species in the county of Nordland and
185 Troms as listed by the Norwegian Bioinformation Centre (http://www.biodiversity.no/). For any lake,
186 both datasets are likely incomplete, as inconspicuous species may be lacking in the regional records
187 [59] and our vegetation surveys did not include the entire catchment area. Nevertheless, the exercise is
188 useful for evaluating how many FPs and TPs are lost by applying different filtering criteria. We
189 defined true positives as sequences that matched a species recorded in the vegetation surveys at the
190 same lake, being aware that this is an under-representation, as the vegetation surveys likely missed
191 species. We defined false positives as species recorded neither in the vegetation surveys nor the
192 regional flora. We tested the effect of different rules of sequence removal: 1) found as ≤1,≤5 or ≤10
193 reads in a PCR repeat, 2) found as ≤1,≤2 or ≤3 PCR repeats for a lake sample, 3) occurring in more
194 than one of 72 negative control PCR replicates, 4) on average, higher number of PCR repeats in
195 negative controls than in sample, and 5) on average a higher number of reads in negative controls than
196 in samples (S2 Table). The filtering criteria resulting in overall highest number of true positives kept
197 compared to false positives lost were applied to all lakes. These were removing sequences with less
198 than 10 reads, less than 2 PCR repeats in lake samples, and on average a lower number of reads in lake
199 samples than in negative controls.
200

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

201 **Data analyses and statistics**

202 After data filtering, we compared taxon assemblages from DNA amplifications with the taxa
203 recorded in the vegetation surveys. To make this comparison, taxa in the vegetation surveys were
204 lumped according to the taxonomic resolution of the P6 loop (S1 Table), and the comparison was done
205 at the lowest resultant taxonomic level. The majority of results explore only present/absence (taxa
206 richness); quantitative data are given in tables (including Supporting Information).
207
208 Multivariate ordinations (Correspondence Analysis and Non-symmetric Correspondence
209 Analysis, the latter giving more weight to abundant species; [60, 61], were run independently on the
210 vegetation data (present/absent using only taxa recorded within 2m) and eDNA data (present/absent).
211 The similarity between ordinations of vegetation and eDNA data was assessed using Procrustes
212 analysis [62], as implemented in the functions procrustes() and protest() in R library vegan [63].
213
214 To estimate the percentages of false negatives and positives in the DNA data and in the
215 vegetation survey, we used the approach described in [64]. If we define the probability of a DNA false
216 positive as p_{DNA_0}, the detectability by DNA as p_{DNA_1}, the detectability in the vegetation survey as
217 p_{VEG_1}, and the probability that a species is present as p_{OCC}, we can state that the four probabilities of
218 observing Presence(1)/Absence(0) in the DNA and Vegetation are as follows:
219
220 1. $Prob(DNA{=}0, Vegetation{=}0) = (1- p_{OCC})(1- p_{DNA_0}) + p_{OCC}(1- p_{DNA_1})(1- p_{VEG_1})$
221 In this case, if the species is absent in both the DNA and vegetation, it is either absent with probability
222 $(1- p_{OCC})$ and no false positive has occurred with probability $(1- p_{DNA_0})$, or it is present with
223 probability p_{OCC}, but was not detected both in the DNA with probability $(1- p_{DNA_1})$ and in the
224 vegetation with probability $(1- p_{VEG_1})$.
225
226 2. $Prob(DNA{=}0, Vegetation{=}1) = p_{OCC}(1- p_{DNA_1})\, p_{VEG_1}$
227 In this case, the species is present, not detected in DNA but detected in the vegetation survey.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

228

229 *3. Prob(DNA=1,Vegetation=0)* $= (1- p_{OCC})\, p_{DNA_0} + p_{OCC}\, p_{DNA_1}\, (1- p_{VEG_1})$

230 In this case, the species is either absent and is a false DNA positive, or is present, detected by DNA

231 but not in the vegetation survey.

232

233 *4. Prob(DNA=1, Vegetation=1)* $= p_{OCC}\, p_{DNA_1}\, p_{VEG_1}$

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

235

236 We assumed the four probabilities varied only among lakes, not among species. We also

237 restricted the analyses to species that were detected at least once using DNA, because for species that

238 were never detected using eDNA, different processes might be important. For p_{DNA_1} , we also

239 considered a model assuming a logistic relationship between p_{DNA_1} and lake characteristics, such as

240 lake depth or catchment area, that is: logit(p_{DNA_1}) = $b_0 + b_1$ Lake Covariate. We fitted these models

241 using Bayesian methods, using uninformative priors (uniform distributions on the [0,1] interval) for

242 the false positive/negative rates for DNA, and an informative prior for the detectability in the

243 vegetation survey (uniform prior on the [0.8,1] interval, as detectability was high in the vegetation

244 survey, but we had no repeated surveys or time to detection available to estimate it). We used the R

245 package rjags to run the MCMC simulations [64]. Model convergence was assessed using the Gelman-

246 Rubin statistics [65], values of which were all ~1.0.

247

248 Results

249 Vegetation records

250 The vegetation surveys provided 2316 observations of 268 taxa, including hybrids, subspecies,

251 and uncertain identifications. Of these, 97 taxa share sequences with one or more other taxa (e.g., 20

252 taxa of *Carex* and 15 of *Salix*). Another nine taxa were not in the reference library (e.g. *Cicerbita*

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

253 *alpina*), and eight taxa could not be matched due to incomplete identification in the vegetation survey.
254 Eight taxa of *Equisetum* were filtered out due to short sequence length. This left 171 taxa that could
255 potentially be recognized by the technique we used (S1 Table). For the 11 sites, between 31 and 58
256 taxa were potentially identifiable (Table 2), and this value was positively correlated with vegetation
257 species richness (y=0.67x+10.3, r^2=0.93, p<0.0001, n=11). Taxonomic resolution at species level was
258 77-93% (mean 88) and 65-79 (mean 74%) for the <2 m and extended (i.e., combined) vegetation
259 surveys, respectively.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

265 Of 489 records <2 m from the lakeshore, the majority were rare (148) or scattered (146) in the
266 vegetation; fewer were common (131) or dominant (64). An additional 245 observations of 46 taxa
267 came from >2 m from the lakeshore (156 rare, 68 scattered, 19 common and 2 dominant).
268

269 Molecular data

270 The numbers of sequences matching entries in the regional arctic-boreal and EMBL-r117
271 databases were 227 and 573 at 98% identity, respectively. For sequences matching both databases, we
272 retained the arctic-boreal identification; this resulted in 11,236,288 reads of 301 sequences matched at
273 100% similarity with at least 10 reads in total (S2 Table). There were 244 and 181 records of taxa that
274 with certainty could be defined as true or false positive, respectively (see methods). We found no
275 combination of filtering criteria that only filtered out the false positives without any loss of true
276 positives (S3 Table). The best ratio was obtained when retaining sequences that were on average more
277 common in samples than in negative controls, plus with at least two replicates in one sample and at
278 least 10 reads per replicate. Applying these criteria filtered out 163 false positives leaving only 18
279 records of three false positive taxa (Annonaceae, Meliaceae and Solanaceae), which were then
280 removed as obvious contamination. However, it also removed 61 (25%) true positives, e.g., *Pinus*,
281 which had high read numbers at lakes in pine forest and low ones at lakes where it is probably brought
282 in as firewood, but which also occurred with high read numbers in two of the negative controls (S4
283 Table). After matching against the local vegetation, 2,500,138 reads of 56 sequences remained.
284 Sequences representing the same taxa were merged, resulting in 47 final taxa (Table 3). Taking into
285 account some taxa shared sequences, e.g. *Carex* and *Salix*, these may potentially represent 81 taxa (S1
286 Table).
287

288 **Table 3. Read numbers per taxa and per lake.**

289

Family	Taxa	A-tj	Bren	Einl	Finn	Gaup	Jula	Laks	Lauv	Ovre	Paul	Rott	Sum
Asteraceae	*Crepis paludosa*									455			455
Betulaceae	*Alnus incana*	48 183	117 855			40 802					131	15 710	222 681
Betulaceae	*Betula* spp.	126 727	120 369	40 991	5 630	101 688		144	32	31 639	3 263	16 283	446 766
Caryophyllaceae	*Sagina* sp.		46	10	37		18			10	24		145
Cornaceae	*Chamaepericlymenum suecicum*			338									338
Cupressaceae	*Juniperus communis*	261	752						45		27		1 085
Cyperaceae	*Carex lasiocarpa*	47				76						84	207
Cyperaceae	*Carex* spp.	34		48	33	72							187
Dryopteridaceae	*Dryopteris* spp.	10 088	16 947	6 406	6 781	5 882	87	1 886	1 141	6 252	216	5 239	60 925
Ericaceae	*Andromeda polifolia*	191		235		244			23			310	1 003
Ericaceae	*Calluna vulgaris*			1 384	357								1 741
Ericaceae	*Cassiope tetragona*	181	86								163		430
Ericaceae	*Chamaedaphne calyculata*	31		29		46						41	147
Ericaceae	*Empetrum nigrum*	3 466	12 736	2 266	4 714	2 807	6 813		14	3 149	13 507	1 758	51 230
Ericaceae	*Oxycoccus microcarpus*					538							538
Ericaceae	*Phyllodoce caeruela*	1 386	305				165						1 856
Ericaceae	*Vaccinium vitis-idaea/myrtillus*	2 005	2 042	916	308	1 286				189	815	394	7 955
Ericaceae	*Vaccinium uliginosum*	1 073	2 325	1 045		2 726	431	30		1 233	1 014	873	10 750
Geraniaceae	*Geranium sylvaticum*									68	145		213
Haloragaceae	*Myriophyllum alterniflorum*	11 389		273 929						226 753			512 071
Isoetaceae	*Isoetes* spp.			27 136						14 411			41 547
Lentibulariaceae	*Utricularia minor*											893	893
Lycopodiaceae	*Huperzia selago*	783	710						10		27	195	1 725
Lycopodiaceae	Lycopodiaceae	9 226	32 590	1 016	2 360	4 285	299	270	217	1 196	5 082	3 381	59 922
Menyanthaceae	*Menyanthes trifoliata*	26 842	467	17 384	1 173	18 978		98	871	378		42 408	108 599
Numphaeaceae	*Numphar pumila*											63 844	63 844
Plantaginaceae	*Callitriche hermaphroditica*			951			5 598						6 549
Plantaginaceae	*Hippuris vulgare*			238						107			345

Poaceae	*Festuca* spp.	30			2 724								2 754
Polygonaceae	*Oxyria digyna*										429		429
Polypodiaceae	*Athyrium* sp.	6 266	33 588	10 557	2 098	1 258		743	539	10 851	1 239	466	67 605
Potamogetonaceae	*Potamogeton praelongus*	1 754								254		9 268	11 276
Potamogetonaceae	*Potamogeton* sp.	28		19 281						12 817		1 335	33 461
Potamogetonaceae	*Stuckenia filiformis*	4 964	183			7 023						246	12 416
Ranunculaceae	*Caltha palustris*			1 131						5 080			6 211
Rosaceae	*Comarum palustre*	258				1 058				222			1 538
Rosaceae	*Dryas octopetala*		750				37					394	1 181
Rosaceae	*Filipendula ulmaria*	850		957		2 293				2 520		6 019	12 639
Rosaceae	*Rubus chamaemorus*	1 453		75		197						317	2 042
Rosaceae	*Sorbus aucuparia*	1 198	894	1 915		1 953				1 468			7 428
Salicaceae	*Populus tremula*	2 009	1 671		1 225		27			1 152		48 201	54 285
Salicaceae	*Salicaceae*	4 488	182 354	1 212	246	68 186	148 060	141		15 658	149 450	2 542	572 337
Saxifragaceae	*Saxifraga aizoides*		585	30									615
Saxifragaceae	*Saxifraga oppositifolia*		922										922
Sparganiaceae	*Sparganium* spp.				958			258	74				1 290
Thelypteridaceae	*Phegopteris connectilis*	4 776	13 594	1 104	1 357	100		546	132	2 085	1 014	366	25 074
Woodsiaceae	*Gymnocarpium dryopteris*	10 290	42 766	1 339	1 287	18 254	336	764	355	3 029	1 986	2 082	82 488
Sum DNA reads		280 277	584 537	411 923	31 288	279 752	161 871	4 880	3 453	340 976	178 532	222 649	2 500 138

DNA and vegetation < 2m

DNA and vegetation > 2m

DNA only

Vegetation <2m and potentially >2m

Vegetation only > 2m

No DNA, no vegetation

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

290 The read numbers are sum of two DNA extractions with 6 PCR replicates for each. All read
291 numbers are after the filtering steps in S2 Table. Note that the records of *Chamaeodaphne calyculata*
292 are likely to represent false positives. For taxa only recorded in vegetation and/or filtered out of the
293 eDNA records, see S1 Table. The lakes names are A-tjern (A-tj), Brennskogtjørna (Bren), Einletvatnet
294 (Einl), Finnvatnet (Finn), Gauptjern (Gaup), Jula Jávri (Jula), Lakselvhøgda (Laks), Lauvås (Lauv),
295 Øvre Æråsvatnet (Ovre), Paulan Jávri (Paul), and Rottjern (Rott).
296
297
298 In our positive control, 7 out of 8 species were detected in all replicates. Only *Aira praecox*,
299 which was added with the lowest DNA concentration, could not be detected. This indicates that the
300 PCR and sequencing was successful for taxa with an extracted DNA concentration of ≥0.03 ng/µL (S5
301 Table).
302
303 The gain in number of taxa when analysing two cores instead of one was 2.5±1.2 per lake. All
304 data presented below are based on the upper 0-2 cm of sediment of two cores combined (but not from
305 deeper levels as these were not sampled at all sites). This gave an average of 19.7±6.9 taxa (10-30) per
306 lake (Table 2). Samples from below 2-cm depth provide an additional 14 records of 42 taxa, some not
307 recorded in 0-2 cm samples (S1 Table).
308

309 Detection of taxa in eDNA

310 Of the 217 eDNA records, the majority matched taxa recorded within 2 m of the lake shore
311 (Figure 2a). Higher proportions of dominant or common taxa were detected in DNA compared with
312 scattered or rare ones (Fig 2b). Most dominant taxa, such as *Betula, Empetrum nigrum, Vaccinium*
313 *uliginosum,* and *Salix*, were correctly detected at most or all lakes (Table 3), whereas some were
314 filtered out (*Equisetum* spp., *Pinus sylvestris,* many *Poa,* S1 Table). Of dominants, only two *Juncus*
315 and two *Eriophorum* species were not recorded. Many taxa that were rare or scattered were filtered out
316 (S1 and S4 Table).

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

317

318 **Fig 2. Match between records of taxa in the sedimentary eDNA in relation to vegetation surveys**.
319 a) Number of records in the sedimentary eDNA in relation to vegetation survey distance. b)
320 Percentage records in eDNA in relation to abundance in vegetation surveys. c) Variation in percentage
321 data among families with >11 eDNA records. d) Variation in percentage of taxa detected among lakes.
322 Percentages in b), c) and d) refers to percentage of taxa recorded in the vegetation that potentially
323 could be identified with the DNA barcode used. Note that DNA of more taxa were likely recorded but
324 filtered out (S1-S4 Tables) – these numbers are only shown in figure b).

325

326

327 Detection success and taxonomic resolution in the eDNA varied among families (Table 3, Fig
328 2c). High success and resolution characterise Ericaceae and Rosaceae as they were identified to
329 species level and successfully detected at most sites. Ferns (Dryopteridaceae, Thelypteridaceae,
330 Woodsiaceae) and club mosses (Lycopodiaceae) were almost always detected, even when only
331 growing >2 m from the lake shore. Aquatics (Haloragaceae, Lentibulariaceae, Menyanthaceae,
332 Numphaeaceae, Plantaginaceae, Potamogetonaceae, Sparganiaceae) were also well detected, often also
333 when not recorded in the vegetation surveys. Deciduous trees and shrubs (Betulaceae, Salicaceae)
334 were also correctly identified at most lakes although often at genus level. In contrast, Poaceae and
335 Cyperaceae, which were common to dominant around most lakes, were underrepresented in the DNA
336 records. Juncaceae and Asteraceae, which were present at all lakes, although mainly scattered or rare,
337 were mainly filtered out due to presence in only one PCR repeat or only in samples from 2-8 cm depth
338 (S1-S4 Tables).

339

340 The numbers of taxa recorded in vegetation, in eDNA, and as match between them varied two-
341 to six-fold among lakes (Table 2, Fig 2d). Jula Jávri had the lowest match between eDNA and
342 vegetation with only four taxa in common. Lakselvhøgda and Lauvås had extremely low read numbers
343 after filtering. For Lauvås, Finnvatnet and Lakselvhøgda, 84%, 30% and 20%, respectively, of raw
344 reads were allocated to algae. If we assume that a big unidentified sequence cluster also represents

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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).

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 3a-b). 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, □=0.50, 20.4%, Axis 2, □=0.37, 15.1%; eDNA: Axis 1, □=0.24, 18.9%, Axis 2, □=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).

Fig 3. **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

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

373 forbs birch/pine mixed forest (A-tjern, Rottjern, Gauptjern are clustered together in both plots; so are
374 also Einletvatnet and Øvre Æråsvatnet (both mire/birch forest at the island Andøya), whereas some
375 lake with poorer DNA records show some differences in clustering.
376

377 Probability of detecting taxa in vegetation and DNA records

378 The posterior probability that all local taxa were recorded during the vegetation survey varied
379 from 0.85-0.95 (S6 Table). Thus, on average, about three species may have been overlooked at each
380 lake. The posterior probability that taxa recorded in the vegetation surveys and detected at least once
381 by eDNA were also recorded in the DNA in a given lake (true positives) was 0.33-0.90, whereas the
382 posterior probability of any DNA records representing a false positive varied from 0.06-0.33 per lake
383 (S6 Table). There was evidence that the probability of detecting a species using eDNA (p_{DNA_1}) was
384 higher for deeper lakes (slope b_1 = 0.58, 95% CI = [0.20; 0.98], Fig 4). Not surprisingly a similar
385 effect was found for lake size (slope b_1=0.25 [0.10, 0.41]) as lake size and depth were highly
386 correlated (r=0.81). Catchment area (b_1=0.06 [-0.15, 0.27]) and mean annual temperature (b_1=-0.03 [-
387 0.14, 0.08]) did not appear to influence probability of detection by eDNA.
388

389 **Fig 4. Lake depth versus detection probability.** Relationship between lake depth and probability
390 that a species present in the vegetation and detected at least once by eDNA is detected by eDNA in a
391 given lake. The relationship is modelled as a logit function and back-transformed to the probability
392 scale.
393

394 Discussion

395 Taking into account the limitation of taxonomic resolution due to sequence sharing or taxa
396 missing in the reference library, we were able to detect about one third of the taxa growing in the
397 immediate vicinity of the lake using only two small sediment samples from the lake centre. The large

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

398 number of true positives lost (S1 Appendix) suggests that this proportion may be further improved.
399 Nevertheless, the current approach was sufficient to distinguish the main vegetation types.
400

401 Taphonomy of environmental plant DNA

402 The high proportion of taxa in the <2 m survey detected with eDNA than in the extended
403 surveys indicates that eDNA is mainly locally deposited. The observation of taxa not recorded in the
404 vegetation surveys but common in the region (Fig 3, S1 Table) indicates that some DNA does
405 originate from some hundreds of meters or even a few km distant. Indeed, a higher correlation
406 between catchment relief and total eDNA (R^2=0.42) than eDNA matching records in the vegetation
407 (R^2=0.34), may suggest that runoff water from snow melt or material blown in also contributes. Thus,
408 the taphonomy of eDNA may be similar to that of macrofossils [66, 67], except that eDNA may also
409 be transported via non-biological particles (e.g. fine mineral grains). From other studies, pollen does
410 not appear to contribute much to local eDNA records [15, 35, 37, 42, 47]. This is probably due to its
411 generally low biomass compared with stems, roots and leaves, and to the resilience of the
412 sporopollenin coat, which requires a separate lysis step in extraction of DNA [68].
413
414 The higher proportion of eDNA taxa that matched common or dominant taxa in the
415 vegetation, compared with taxa that were rare or scattered, was as expected, as higher biomass should
416 be related to a greater chance for deposition and preservation in the lake sediments [9]. Yoccoz *et al.*
417 [6] found the same in their comparison of soil eDNA with standing vegetation. While some dominant
418 taxa were filtered out in our study, their DNA was mainly present (S1 Appendix, S1-4 Tables), and
419 most dominant taxa were recorded in all PCR replicates (not shown). Thus, for studies where the focus
420 is on detecting dominant taxa, running costs may be reduced by performing fewer PCR replicates.
421

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

422 **Variation among lakes**

423 The variation among lakes seen in DNA-based detection of taxa shows that even when
424 identical laboratory procedures are followed, the ability to detect taxa can vary. Our sample size of 11
425 lakes does not allow a full evaluation of the reasons for this variation. Factors such as low pH or
426 higher temperature may increase DNA degradation [16], but the two lakes with lowest numbers of
427 reads after filtering in our study, Lakselvhøgda and Lauvås, had pH values close to optimal for DNA
428 preservation (7.2 and 6.8, respectively, I.G. Alsos and A.G. Brown, pers. obs. 2016), and variation in
429 temperature was low among our sites. The lack of an inflowing stream at Lakselvhøgda may reduce
430 the supply of eDNA, but Lauvås has two inflows. For these two lakes, and to a lesser extent
431 Finnvatnet, we suspect high algal abundance might have caused PCR competition [69]. PCR
432 competition may also occurred in samples from Jula Jávri, but in this case we were not able to identify
433 the most dominant cluster of sequences. These lakes are also small and shallow. Variation among
434 eDNA qualities has also been observed in a study of 31 lakes on Taymyr Peninsula in Siberia [70]. We
435 suspect that high algae production may be a limiting factor as we also have seen poor aDNA results in
436 samples with high Loss on Ignition values, but this should be studied further. A potential solution to
437 avoid solution to avoid PCR competition may be to design a primer to block amplification of algae as
438 has been done for human DNA in studies of mammals eDNA [71].
439

440 **Variation among taxa**

441 The variation we observed among plant families, both in taxonomic resolution and likelihood
442 of detection, is a general problem when using generic primers [45, 72, 73]. For example, the poor
443 detection of the Cyperaceae may be due to the long sequence length of *Carex* and *Eriophorum* (>80
444 bp), and most studies only detect it at genus or family level [38, 42, 74]. The low representation of
445 Asteraceae may be due to its rare or scattered representation in the vegetation and/or its poor
446 amplification. While some studies successfully amplify Asteraceae [15, 37, 38, 42, 75], others do not,
447 even when other proxies indicate its presence in the environment [46]. This may be due to the high

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

448 percentage of Asteraceae taxa that have a one base-pair mismatch in the reverse primer [34]. Poaceae,
449 which has no primer mismatch, is regularly detected in ancient DNA studies [15, 36-38, 41], and was
450 present in nine lakes, although most records were filtered out due to occurrence in negative controls.
451 To avoid any bias due to primer match and potentially increase the overall detection of taxa, one
452 solution would be to use family-specific primers, such as ITS primers developed for Cyperaceae,
453 Poaceae, and Asteraceae [36]. Alternatively, shotgun sequencing could be tested as this minimizes
454 PCR biases [76, 77].
455
456 The common woody deciduous taxa *Betula* and *Salix,* as well as most common dwarf shrubs
457 such as *Andromeda polifolia, Empetrum nigrum,* and *Vaccinium uliginosum,* were correctly detected
458 in most cases. They are also regularly recorded in late-Quaternary lake-sediment samples [15, 25, 37,
459 41, 70, 74]. These are ecologically important taxa in many northern ecosystems, and their reliable
460 detection in eDNA could be expected to extend to other types of samples, e.g., samples relating to
461 herbivore diet [44].
462
463 The general over-representation of spore plants in eDNA among taxa only found >2 m from
464 the lake and those not recorded in the catchment vegetation raises the question as to whether eDNA
465 can originate from spores. Spore-plant DNA is well represented in some studies [42, 78], is lacking in
466 other studies [15, 37] and has been found as an exotic in one study [41]. As with pollen, the protective
467 coat and low biomass of spores suggest that they are an unlikely source of the eDNA. This inference is
468 supported by clear stratigraphic patterns shown by fern DNA in two lake records from Scotland.
469 Records are ecologically consistent with other changes in vegetation, whereas spores at the same sites
470 show no clear stratigraphy [42]. Preferential amplification could be an alternative explanation, but this
471 is not likely as the amplification of fern DNA from herbarium specimens is poor [34]. It is possible
472 that in some cases, including this study, we are detecting the minute but numerous gametophytes
473 present in soil, which would not be visible in vegetation surveys.
474

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

475 Aquatic taxa were detected in all lakes, and they have been regularly identified in eDNA
476 analyses of recent [42] and late-Quaternary lake sediments [15, 37, 38]. eDNA may be superior to
477 vegetation surveys in some cases, e.g., *Potamogeton praelongus*, which is characteristic of deeper
478 water https://www.brc.ac.uk/plantatlas/) and was likely overlooked in surveys due to poor visibility.
479 *Callitriche hermaphroditica* was observed in two lakes (Einleten and Jula javri), whereas *C. palustris*
480 was observed at Einleten. We cross-checked the herbarium voucher and the DNA sequence and both
481 seems correct, so potentially both were present but detected only in either eDNA or vegetation
482 surveys. Overall, eDNA appears to detect aquatic plants more efficiently than terrestrial plants, which
483 is not unexpected as the path from plant to sediment is short.
484

485 The use of eDNA for reconstruction of present and past plant 486 richness

487 In contrast to water samples, from which eDNA has been shown to represent up to 100 % of
488 fish and amphibian taxa living in a lake [7, 79], one or two small, surficial sediment samples do not
489 yield enough DNA to capture the full richness of vascular plants growing around a lake; the same
490 limitation may apply in attempts to capture Holocene mammalian richness [22]. This is likely largely
491 due to taphonomic limitations affecting preservation and transport on land, as aquatics were generally
492 well detected. Also, surface samples are typically flocculent and represent a short time span, e.g. a few
493 centimetre may represent 10-25 years ([49]; pers. obs.). Increasing the amount of material analysed,
494 the amount of time sampled (by combining the top several cm of sediment), and/or the number of
495 surface samples may improve detection rates for species that are rare, have low biomass and/or grow
496 at some distance from the lake. In this study we identified more taxa when we used two surface
497 samples and/or material from deeper in the sediment cores. Nevertheless, taphonomic constraints may
498 mean that DNA of some species rarely reaches the lake sediment. On the technical side, both
499 improvements in laboratory techniques and in bioinformatics could increase detection of rare species.
500 In this study, DNA of many of the rarer taxa was recorded but was filtered out. As the rarest species

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

501 are also difficult to detect in vegetation surveys [59], combining conventional and DNA-based surveys
502 may produce optimal estimates of biodiversity.
503
504 The potential taxonomic resolution (i.e., for eDNA taxa to be identified to species level) was
505 similar or higher than that for macrofossils [80] or pollen [81, 82]. The potential taxonomic resolution
506 of any of these methods depends on how well the local flora is represented in the available reference
507 collection/library, site-specific characteristics, such as the complexity and type of the vegetation [34,
508 82], and the morphological or genetic variation displayed by different taxonomic groups. In our case,
509 only 3% of the taxa found in the vegetation surveys were missing in the reference database which
510 likely improved the resolution. To reach 100% resolution, a different genetic marker is needed to
511 avoid the problem of sequence sharing. Using longer barcodes may improve resolution [45, 83] and
512 may work for modern samples, bur for taxa with cpDNA sharing as e.g. *Salix*, nuclear regions should
513 be explored. For ancient samples with highly degraded DNA, taxonomic resolution may potentially be
514 increased by using a combining several markers, hybridization capture RAD probe techniques, or full-
515 genome approach [77, 84-86].
516
517 The actual proportion of taxa in the vegetation detected in the eDNA records (average 28%
518 and 18% for <2 m and extended surveys, respectively, not adjusting for taxonomic resolution) is
519 similar to the results of various macrofossil [80, 87-89] and pollen studies [81, 82]. This contrasts with
520 five previous studies of late-Quaternary sediments that compared aDNA with macrofossils and seven
521 that did so with pollen; these showed rather poor richness in aDNA compared to other approaches
522 (reviewed in [10]). We think a major explanation may be the quality and size of available reference
523 collections/libraries, as the richness found in studies done prior to the publication of the boreal
524 reference library (e.g. [15, 27, 35, 37]) was lower than in more recent studies, including this one [42,
525 70, 90, 91]. The variation in laboratory procedures, the number and size of samples processed and the
526 number of replicates also affect the results [4, 82, 86]. Nevertheless, the correlation between eDNA
527 and vegetation found in the Procrust analyses show that the current standard of the method is sufficient
528 to detect major vegetation types.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

529

530 # Conclusion

531 Our study supports previous conclusions that eDNA mainly detects vegetation from within a
532 lake catchment area. Local biomass is important, as dominant and common taxa showed the highest
533 probability of detection. For aquatic vegetation, eDNA may be comparable with, or even superior to,
534 in-lake vegetation surveys. Lake-based eDNA detection is currently not good enough to monitor
535 modern terrestrial plant biodiversity because too many rare species are overlooked. The method can,
536 however, detect a similar percentage of the local flora as is possible with macrofossil or pollen
537 analyses. As many true positives are lost in the filtering process, and as even higher taxonomic
538 resolution could be obtained by adding genetic markers or doing full genome analysis, there is the
539 potential to increase detection rates. Similarly, results will improve as we learn more about how
540 physical conditions influence detection success among lakes, and how sampling strategies can be
541 optimized.
542

543 # Acknowledgements

544 We thank Marie Kristine Føreid Merkel, Veronica Rystad, Premasany Kanapathippillai, Chris
545 Ware, Sarah Lovibond, Martin Årseth-Hansen, Torbjørn Alm and Antony G. Brown for field
546 assistance, Department for Medical Biology at UiT for use of laboratory for extraction, Frederic Boyer
547 for raw data handling, Torbjørn Alm for help with identifying plant specimens, and H. John B. Birks
548 for valuable comments on an earlier draft of this manuscript.
549

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

550 # Conflict of Interest

551 The authors would like to mention that LG is one of the co-inventors of patents related to g-h
552 primers and the subsequent use of the P6 loop of the chloroplast *trn*L (UAA) intron for plant
553 identification using degraded template DNA. These patents only restrict commercial applications and
554 have no impact on the use of this locus by academic researchers.

555

556

557 # References

558

559 1. Orlando L, Cooper A. Using ancient DNA to understand evolutionary and ecological
560 processes. Annu Rev Ecol Evol Syst. 2014;45(1):573-98. doi: 10.1146/annurev-ecolsys-
561 120213-091712.
562 2. Brown TA, Barnes IM. The current and future applications of ancient DNA in Quaternary
563 science. J Quat Sci. 2015;30(2):144-53. doi: 10.1002/jqs.2770.
564 3. Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, et
565 al. Ancient and modern environmental DNA. Philos Trans R Soc London Ser B.
566 2015;370(1660):20130383. Epub 2014/12/10. doi: 10.1098/rstb.2013.0383. PubMed
567 PMID: 25487334; PubMed Central PMCID: PMCPMC4275890.
568 4. Thomsen PF, Willerslev E. Environmental DNA - An emerging tool in conservation for
569 monitoring past and present biodiversity. Biol Conserv. 2015;183:4-18. doi:
570 10.1016/j.biocon.2014.11.019.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

571 5. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-
572 generation biodiversity assessment using DNA metabarcoding. Mol Ecol.
573 2012;21(8):2045-50. doi: 10.1111/j.1365-294X.2012.05470.x.
574 6. Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, et al. DNA from soil
575 mirrors plant taxonomic and growth form diversity. Mol Ecol. 2012;21(15):3647-55. doi:
576 10.1111/j.1365-294X.2012.05545.x.
577 7. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, et al. Next-
578 generation monitoring of aquatic biodiversity using environmental DNA metabarcoding.
579 Mol Ecol. 2016;25(4):929-42. doi: 10.1111/mec.13428.
580 8. Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, et al.
581 Understanding environmental DNA detection probabilities: A case study using a stream-
582 dwelling char Salvelinus fontinalis. Biol Conserv. 2016;194:209-16. doi:
583 http://dx.doi.org/10.1016/j.biocon.2015.12.023.
584 9. Barnes MA, Turner CR. The ecology of environmental DNA and implications for
585 conservation genetics. Conserv Genet. 2016;17(1):1-17. doi: 10.1007/s10592-015-0775-
586 4.
587 10. Birks HJB, Birks HH. How have studies of ancient DNA from sediments contributed to
588 the reconstruction of Quaternary floras? New Phytol. 2016;209:499-506. doi:
589 10.1111/nph.13657.
590 11. Torti A, Lever MA, Jørgensen BB. Origin, dynamics, and implications of extracellular
591 DNA pools in marine sediments. Marine Genomics. 2015;24, Part 3:185-96. doi:
592 http://dx.doi.org/10.1016/j.margen.2015.08.007.
593 12. Deiner K, Altermatt F. Transport distance of invertebrate environmental DNA in a
594 natural river. PLoS ONE. 2014;9(2):e88786. doi: 10.1371/journal.pone.0088786.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

595 13. Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM.
596 Environmental conditions influence eDNA persistence in aquatic systems. Environmental
597 Science & Technology. 2014;48(3):1819-27. doi: 10.1021/es404734p.
598 14. Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjær KH, et al.
599 Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Mol Ecol.
600 2012;21:1966-79. doi: 10.1111/j.1365-294X.2011.05261.x.
601 15. Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, et al. Molecular-
602 and pollen-based vegetation analysis in lake sediments from central Scandinavia. Mol
603 Ecol. 2013;22:3511-24. doi: 10.1111/mec.12298.
604 16. Strickler KM, Fremier AK, Goldberg CS. Quantifying effects of UV-B, temperature, and
605 pH on eDNA degradation in aquatic microcosms. Biol Conserv. 2015;183(0):85-92. doi:
606 http://dx.doi.org/10.1016/j.biocon.2014.11.038.
607 17. Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, et al.
608 Replication levels, false presences and the estimation of the presence/absence from
609 eDNA metabarcoding data. Mol Ecol Res. 2015;15(3):543-56. doi: 10.1111/1755-
610 0998.12338.
611 18. Deiner K, Walser J-C, Mächler E, Altermatt F. Choice of capture and extraction methods
612 affect detection of freshwater biodiversity from environmental DNA. Biol Conserv.
613 2015;183:53-63. doi: http://dx.doi.org/10.1016/j.biocon.2014.11.018.
614 19. Lahoz-Monfort JJ, Guillera-Arroita G, Tingley R. Statistical approaches to account for
615 false-positive errors in environmental DNA samples. Mol Ecol Res. 2016;16:673-85. doi:
616 10.1111/1755-0998.12486.
617 20. Nguyen NH, Smith D, Peay K, Kennedy P. Parsing ecological signal from noise in next
618 generation amplicon sequencing. New Phytol. 2015;205(4):1389-93. doi:
619 10.1111/nph.12923.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

620 21. Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, et al. Quantification of
621 mesocosm fish and amphibian species diversity via environmental DNA metabarcoding.
622 Mol Ecol Res. 2016;16(1):29-41. doi: 10.1111/1755-0998.12433.
623 22. Giguet-Covex C, Pansu J, Arnaud F, Rey P-J, Griggo C, Gielly L, et al. Long livestock
624 farming history and human landscape shaping revealed by lake sediment DNA. Nat
625 Commun. 2014;5. doi: 10.1038/ncomms4211.
626 23. Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, Huang Y-T, et al. Using
627 palaeoenvironmental DNA to reconstruct past environments: progress and prospects. J
628 Quat Sci. 2014;29(7):610-26. doi: 10.1002/jqs.2740.
629 24. Smol JP, Birks HJ, Last WM, editors. Terrestrial, algal, and siliceous indicators.
630 Heidelberg: Springer; 2001.
631 25. Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, et al. Postglacial
632 viability and colonization in North America's ice-free corridor. Nature. 2016;537:45-9.
633 doi: 10.1038/nature19085
634 http://www.nature.com/nature/journal/vaop/ncurrent/abs/nature19085.html#supplementary-
635 information.
636 26. Birks HH, Giesecke T, Hewitt GM, Tzedakis PC, Bakke J, Birks HJB. Comment on
637 "Glacial survival of boreal trees in northern Scandinavia". Science. 2012;338(6108):742.
638 doi: 10.1126/science.1225345.
639 27. Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, et al. Glacial
640 survival of boreal trees in northern Scandinavia. Science. 2012;335(6072):1083-6. doi:
641 10.1126/science.1216043.
642 28. Parducci L, Edwards ME, Bennett KD, Alm T, Elverland E, Tollefsrud MM, et al.
643 Response to Comment on "Glacial Survival of Boreal Trees in Northern Scandinavia".
644 Science. 2012;338(6108):742. doi: 10.1126/science.1225476.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

645 29. Smith O, Momber G, Bates R, Garwood P, Fitch S, Pallen M, et al. Sedimentary DNA
646 from a submerged site reveals wheat in the British Isles 8000 years ago. Science.
647 2015;347(6225):998-1001. doi: 10.1126/science.1261278.
648 30. Weiß CL, Dannemann M, Prüfer K, Burbano HA. Contesting the presence of wheat in
649 the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-
650 coverage data. eLife. 2015;4. doi: 10.7554/eLife.10005.
651 31. Birks HH, Bjune AE. Can we detect a west Norwegian tree line from modern samples of
652 plant remains and pollen? Results from the DOORMAT project. Veg Hist Archaeobot.
653 2010;19(4):325-40. doi: 10.1007/s00334-010-0256-0. PubMed PMID:
654 WOS:000280825200007.
655 32. Jackson ST. Representation of flora and vegetation in Quaternary fossil assemblages:
656 known and unknown knowns and unknowns. Quat Sci Rev. 2012;49:1-15. doi:
657 http://dx.doi.org/10.1016/j.quascirev.2012.05.020.
658 33. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and
659 limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids
660 Res. 2007;35(3):e14. doi: 10.1093/nar/gkl938. PubMed PMID: ISI:000244429800001.
661 34. Sønstebø JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J, et al. Using next-
662 generation sequencing for molecular reconstruction of past Arctic vegetation and climate.
663 Mol Ecol Res. 2010;10(6):1009-18. doi: 10.1111/j.1755-0998.2010.02855.x.
664 35. Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, Rasmussen M, et al. A
665 comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost
666 sediments of northern Siberia reveals long-term vegetational stability. Mol Ecol.
667 2012;21(8):1989-2003. doi: 10.1111/j.1365-294X.2011.05287.x.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

668 36. Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, et al. Fifty
669 thousand years of Arctic vegetation and megafaunal diet. Nature. 2014;506(7486):47-51.
670 doi: 10.1038/nature12921
671 http://www.nature.com/nature/journal/v506/n7486/abs/nature12921.html#supplementary-
672 information.
673 37. Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, et al. A
674 comparative study of ancient environmental DNA to pollen and macrofossils from lake
675 sediments reveals taxonomic overlap and additional plant taxa. Quat Sci Rev.
676 2013;75(0):161-8. doi: http://dx.doi.org/10.1016/j.quascirev.2013.06.006.
677 38. Boessenkool S, McGlynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, et al. Use of
678 ancient sedimentary DNA as a novel conservation tool for high-altitude tropical
679 biodiversity. Conserv Biol. 2014;28(2):446-55. doi: 10.1111/cobi.12195.
680 39. Pansu J, Giguet-Covex C, Ficetola GF, Gielly L, Boyer F, Zinger L, et al. Reconstructing
681 long-term human impacts on plant communities: an ecological approach based on lake
682 sediment DNA. Mol Ecol. 2015;24:1485-98. doi: 10.1111/mec.13136.
683 40. Paus A, Boessenkool S, Brochmann C, Epp LS, Fabel D, Haflidason H, et al. Lake Store
684 Finnsjøen – a key for understanding Lateglacial/early Holocene vegetation and ice sheet
685 dynamics in the central Scandes Mountains. Quat Sci Rev. 2015;121:36-51. doi:
686 http://dx.doi.org/10.1016/j.quascirev.2015.05.004.
687 41. Alsos IG, Sjögren P, Edwards ME, Landvik JY, Gielly L, Forwick M, et al. Sedimentary
688 ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to
689 Holocene climate change. The Holocene. 2016;26(4):627-42. doi:
690 10.1177/0959683615612563.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

691 42. Sjögren P, Edwards ME, Gielly L, Langdon CT, Croudace IW, Merkel MKF, et al. Lake
692 sedimentary DNA accurately records 20th Century introductions of exotic conifers in
693 Scotland. New Phytol. 2017;213(2):929-41. doi: 10.1111/nph.14199.
694 43. De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, et al. DNA
695 metabarcoding multiplexing and validation of data accuracy for diet assessment:
696 application to omnivorous diet. Mol Ecol Res. 2014;14(2):306-23. doi: 10.1111/1755-
697 0998.12188.
698 44. Soininen EM, Gauthier G, Bilodeau F, Berteaux D, Gielly L, Taberlet P, et al. Highly
699 overlapping diet in two sympatric lemming species during winter revealed by DNA
700 metabarcoding. Plos One. 2015;10:e0115335.
701 45. Fahner NA, Shokralla S, Baird DJ, Hajibabaei M. Large-scale monitoring of plants
702 through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation
703 of four DNA markers. PLoS ONE. 2016;11(6):e0157505. doi:
704 10.1371/journal.pone.0157505.
705 46. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, et al. Analysing
706 diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput
707 pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in
708 Zoology. 2009;6:16. doi: doi:10.1186/1742-9994-6-16.
709 47. Parducci L, Väliranta M, Salonen JS, Ronkainen T, Matetovici I, Fontana SL, et al. Proxy
710 comparison in ancient peat sediments: pollen, macrofossil and plant DNA. Philos Trans R
711 Soc London Ser B. 2015;370(1660):20130382. doi: 10.1098/rstb.2013.0382.
712 48. Alm T. Øvre Æråsvatn – palynostratigraphy of a 22,000 to 10,000 B.P. lacustrine record
713 on Andøya, Northern Norway. Boreas. 1993;22:171-88.
714 49. Jensen C, Kunzendorf H, Vorren K-D. Pollen deposition rates in peat and lake sediments
715 from the *Pinus sylvestris* L. forest-line ecotone of northern Norway. Review of

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

716 Palaeobotany and Palynology. 2002;121(2):113-32. doi: 10.1016/s0034-6667(02)00077-
717 5.
718 50. Jensen C, Kuiper JGJ, Vorren K-D. First post-glacial establishment of forest trees: early
719 Holocene vegetation, mollusc settlement and climate dynamics in central Troms, North
720 Norway. Boreas. 2002;31(3):285-301. doi: 10.1111/j.1502-3885.2002.tb01074.x.
721 51. Jensen C, Vorren K-D. Holocene vegetation and climate dynamics of the boreal alpine
722 ecotone of northwestern Fennoscandia. J Quat Sci. 2008;23(8):719-43. doi:
723 10.1002/jqs.1155.
724 52. Vorren TO, Vorren K-D, Aasheim O, Dahlgren KIT, Forwick M, Hassel K.
725 Palaeoenvironment in northern Norway between 22.2 and 14.5 cal. ka BP. Boreas. 2013:
726 876–95 doi: 10.1111/bor.12013.
727 53. Elven R. J. Lid & D.T. Lid. Norsk flora. 7th edition. Oslo: Det Norske Samlaget; 2005.
728 54. Elven R, Murray DF, Razzhivin VY, Yurtsev BA. Annotated checklist of the Panarctic
729 Flora (PAF). Vascular plants. Natural History Museum, University of Oslo:
730 CAFF/University of Oslo; 2011 [cited 2013]. Available from: http://nhm2.uio.no/paf/.
731 55. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, et al. The use of
732 coded PCR primers enables high-throughput sequencing of multiple homolog
733 amplification products by 454 parallel sequencing. PLoS ONE. 2007;2(2):e197. doi:
734 10.1371/journal.pone.0000197.
735 56. Valentini A, Miquel C, Nawaz M, Bellemain E, Coissac E, Pompanon F, et al. New
736 perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the
737 trnL approach. Mol Ecol Res. 2009;24(2):110 - 7.
738 57. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. OBITOOLS: a unix-
739 inspired software package for DNA metabarcoding. Mol Ecol Res. 2016;16(1):176-82.
740 doi: 10.1111/1755-0998.12428.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

741 58. Yoccoz NG. The future of environmental DNA in ecology. Mol Ecol. 2012;21(8):2031-8.
742 doi: 10.1111/j.1365-294X.2012.05505.x.
743 59. Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, et al. Using
744 niche-based models to improve the sampling of rare species. Conserv Biol.
745 2006;20(2):501-11. doi: 10.1111/j.1523-1739.2006.00354.x.
746 60. Pélissier R, Couteron P, Dray S, Sabatier D. Consistency between ordination techniques
747 and diversity measurements: Two strategies for species occurrence data. Ecology.
748 2003;84(1):242-51. doi: 10.1890/0012-9658(2003)084[0242:CBOTAD]2.0.CO;2.
749 61. Greenacre M. Correspondence analysis of raw data. Ecology. 2010;91(4):958-63. doi:
750 10.1890/09-0239.1.
751 62. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages
752 of a Procrustean superimposition approach over the Mantel test. Oecologia.
753 2001;129(2):169-78. doi: 10.1007/s004420100720.
754 63. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan:
755 Community Ecology Package. R package version 2.4-3. 2017.
756 64. Chambert T, Miller DAW, Nichols JD. Modeling false positive detections in species
757 occurrence data under different study designs. Ecology. 2015;96(2):332-9. doi:
758 10.1890/14-1507.1.
759 65. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative
760 simulations. Journal of Computational and Graphical Statistics. 1998;7:434-55.
761 66. Jackson ST, Booth RK. Validation of pollen studies. In: Elias SA, editor. Encyclopedia of
762 Quaternary science. London: Elsevier; 2007. p. 2413-22.
763 67. Dieffenbacher-Krall AC. Plant macrofossil methods and studies: Surface samples,
764 taphonomy, representation. In: Elias SA, editor. Encyclopedia of Quaternary Science.
765 Oxford: Elsevier; 2013. p. 684-9.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

766 68. Kraaijeveld K, de Weger LA, Ventayol García M, Buermans H, Frank J, Hiemstra PS, et
767 al. Efficient and sensitive identification and quantification of airborne pollen using next-
768 generation DNA sequencing. Mol Ecol Res. 2015;15(1):8-16. doi: 10.1111/1755-
769 0998.12288.
770 69. Piñol J, Mir G, Gomez-Polo P, Agustí N. Universal and blocking primer mismatches
771 limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of
772 arthropods. Mol Ecol Res. 2015;15(4):819-30. doi: 10.1111/1755-0998.12355.
773 70. Niemeyer B, Epp LS, Stoof-Leichsenring KR, Pestryakova LA, Herzschuh U. A
774 comparison of sedimentary DNA and pollen from lake sediments in recording vegetation
775 composition at the Siberian treeline. Mol Ecol Res. 2017. doi: 10.1111/1755-0998.12689.
776 71. Boessenkool S, Epp LS, Haile J, Bellemain EVA, Edwards M, Coissac E, et al. Blocking
777 human contaminant DNA during PCR allows amplification of rare mammal species from
778 sedimentary ancient DNA. Mol Ecol. 2011:no-no. doi: 10.1111/j.1365-
779 294X.2011.05306.x.
780 72. CBOL PWG, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S,
781 et al. A DNA barcode for land plants. Proceedings of the National Academy of Sciences.
782 2009;106(31):12794-7. doi: 10.1073/pnas.0905845106.
783 73. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode.
784 PLoS ONE. 2011;6(5):e19254. doi: 10.1371/journal.pone.0019254.
785 74. Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, Schrøder-Nielsen A, et al. Lake
786 sediment multi-taxon DNA from North Greenland records early post-glacial appearance
787 of vascular plants and accurately tracks environmental changes. Quat Sci Rev.
788 2015;117(0):152-63. doi: http://dx.doi.org/10.1016/j.quascirev.2015.03.027.

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

789 75. Hiiesalu I, ÖPik M, Metsis M, Lilje L, Davison J, Vasar M, et al. Plant species richness
790 belowground: higher richness and new patterns revealed by next-generation sequencing.
791 Mol Ecol. 2011;21:2004-16. doi: 10.1111/j.1365-294X.2011.05390.x.
792 76. Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, et al. Genome skimming
793 by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol Ecol
794 Res. 2014;14(5):966-75. doi: 10.1111/1755-0998.12246.
795 77. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes:
796 extending the concept of DNA barcoding. Mol Ecol. 2016:1423-8. doi:
797 10.1111/mec.13549.
798 78. Wilmshurst JM, Moar NT, Wood JR, Bellingham PJ, Findlater AM, Robinson JJ, et al.
799 Use of pollen and ancient DNA as conservation baselines for offshore islands in New
800 Zealand. Conserv Biol. 2014;28(1):202-12. doi: 10.1111/cobi.12150.
801 79. Lopes CM, Sasso T, Valentini A, Dejean T, Martins M, Zamudio KR, et al. eDNA
802 metabarcoding: a promising method for anuran surveys in highly diverse tropical forests.
803 Mol Ecol Res. 2016:n/a-n/a. doi: 10.1111/1755-0998.12643.
804 80. Allen JRM, Huntley B. Estimating past floristic diversity in montane regions from
805 macrofossil assemblages. J Biogeogr. 1999;26(1):55-73. doi: 10.1046/j.1365-
806 2699.1999.00284.x.
807 81. Meltsov V, Poska A, Odgaard BV, Sammul M, Kull T. Palynological richness and pollen
808 sample evenness in relation to local floristic diversity in southern Estonia. Review of
809 Palaeobotany and Palynology. 2011;166(3–4):344-51. doi:
810 http://dx.doi.org/10.1016/j.revpalbo.2011.06.008.
811 82. Birks HJB, Felde VA, Bjune AE, Grytnes J-A, Seppä H, Giesecke T. Does pollen-
812 assemblage richness reflect floristic richness? A review of recent developments and

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

813 future challenges. Review of Palaeobotany and Palynology. 2016;228:1-25. doi:
814 http://dx.doi.org/10.1016/j.revpalbo.2015.12.011.
815 83. Lamb EG, Winsley T, Piper CL, Freidrich SA, Siciliano SD. A high-throughput
816 belowground plant diversity assay using next-generation sequencing of the trnL intron.
817 Plant and Soil. 2016;404(1):361-72. doi: 10.1007/s11104-016-2852-y.
818 84. Schmid S, Genevest R, Gobet E, Suchan T, Sperisen C, Tinner W, et al. HyRAD-X, a
819 versatile method combining exome capture and RAD sequencing to extract genomic
820 information from ancient DNA. Methods in Ecology and Evolution. 2017;8(10):1374-88.
821 doi: 10.1111/2041-210X.12785.
822 85. Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, et al.
823 Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic
824 analyses on collection specimens. PLoS ONE. 2016;11(3):e0151651. doi:
825 10.1371/journal.pone.0151651.
826 86. Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, et al. Transley
827 Reviews: Ancient plant DNA from lake sediments. New Phytol. 2017;214(3):924-42. doi:
828 10.1111/nph.14470.
829 87. Dunwiddie PW. Macrofossil and pollen representation of coniferous trees in modern
830 sediments from Washington. Ecology. 1987;68(1):1-11. doi: 10.2307/1938800.
831 88. McQueen DR. Macroscopic plant remains in recent lake sediments. Tuatara.
832 1969;17(1):13-9.
833 89. Drake H, Burrows CJ. The influx of potential macrofossils into Lady Lake, north
834 Westland, New Zealand. New Zealand Journal of Botany. 1980;18(2):257-74. doi:
835 10.1080/0028825X.1980.10426924.
836 90. Zimmermann H, Raschke E, Epp L, Stoof-Leichsenring K, Schirrmeister L, Schwamborn
837 G, et al. The history of tree and shrub taxa on Bol'shoy Lyakhovsky Island (New Siberian

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

838 Archipelago) since the Last Interglacial uncovered by sedimentary ancient DNA and
839 pollen data. Genes. 2017;8(10):273. PubMed PMID: doi:10.3390/genes8100273.
840 91. Zimmermann HH, Raschke E, Epp LS, Stoof-Leichsenring KR, Schwamborn G,
841 Schirrmeister L, et al. Sedimentary ancient DNA and pollen reveal the composition of
842 plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya
843 Peninsula (north-eastern Siberia). Biogeosciences. 2017;14(3):575-96. doi: 10.5194/bg-
844 14-575-2017.

845 # Supporting Information

846 **S1 Appendix. Comments on true and false positives.**

847

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

860

861 **S2 Table. Number of sequence reads remaining after each filtering step** for 42 samples from
862 modern lake sediment collected in northern Norway, 6 extraction negative controls, 6 PCR negative

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

863 controls and 2 PCR positive controls. Six individually tagged PCR repeats were run for each sample,
864 giving a total of 336 PCR samples. Numbers of sequences and unique sequences are given for
865 applying the criteria to all sequences.
866
867 **S3 Table. Effect of different filtering criteria on the number of True Positives.** True positive (TP,
868 defined as species also detected in vegetation surveys thus lower than the numbers given in Table 2)
869 and False Positives (FP, defined as species not found in the regional flora; including 15 potential food
870 plants) per lake and in total. The criteria used in this study, which gives the highest ratio between TP
871 kept and FP lost, is shown in bold. 1) Minimum number of reads in lake samples, 2) minimum number
872 of PCR repeats in lake samples, 3) "1" if occurring in more than 1 PCR repeat of negative control
873 samples, 4) "1" if number of PCR repeats in lake sample > PCR repeats in negative control samples,
874 5) "1" if mean number of reads in lake samples > mean number of reads in negative control samples.
875 The lakes names are A-tjern (A-tj), Brennskogtjørna (Bren), Einletvatnet (Einl), Finnvatnet (Finn),
876 Gauptjern (Gaup), Jula Jávri (Jula), Lakselvhøgda (Laks), Lauvås (Lauv), Øvre Æråsvatnet (Ovre),
877 Paulan Jávri (Paul), and Rottjern (Rott).
878
879 **S4 Table. Taxa removed during filtering**. All DNA reads that have 100% match to the reference
880 libraries and have been removed during the second last step of filtering (see Table S2).
881
882 **S5 Table. Retrieval of positive controls from raw Orbitool output file.** The file consisted of
883 12706536 reads of 581 sequences, 98% match). Note that not all taxa used in the positive controls
884 were present in the reference library but they match to closely related taxa.
885
886 **S6 Table. The probability of detection in eDNA and vegetation.** The probability that all taxa in the
887 vegetation were recorded (Vegetation), and that the DNA records represents true and false positives.
888 Mean probability, standard deviation (SD) are given for each lake.
889
890

bioRxiv preprint doi: https://doi.org/10.1101/264903; this version posted February 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

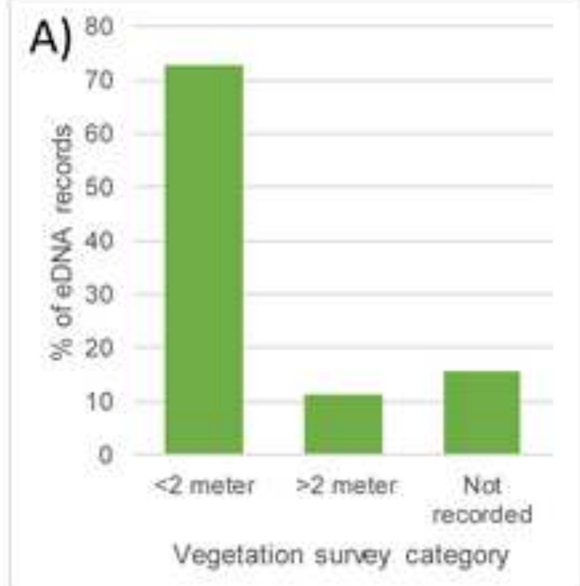
891

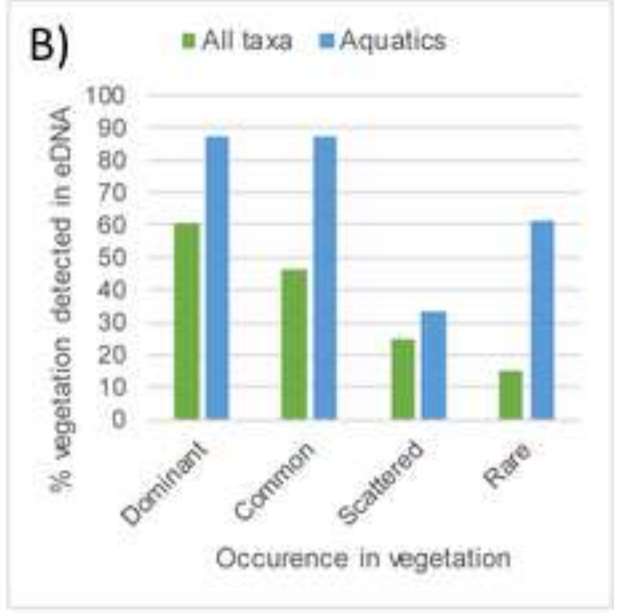
Fig 1

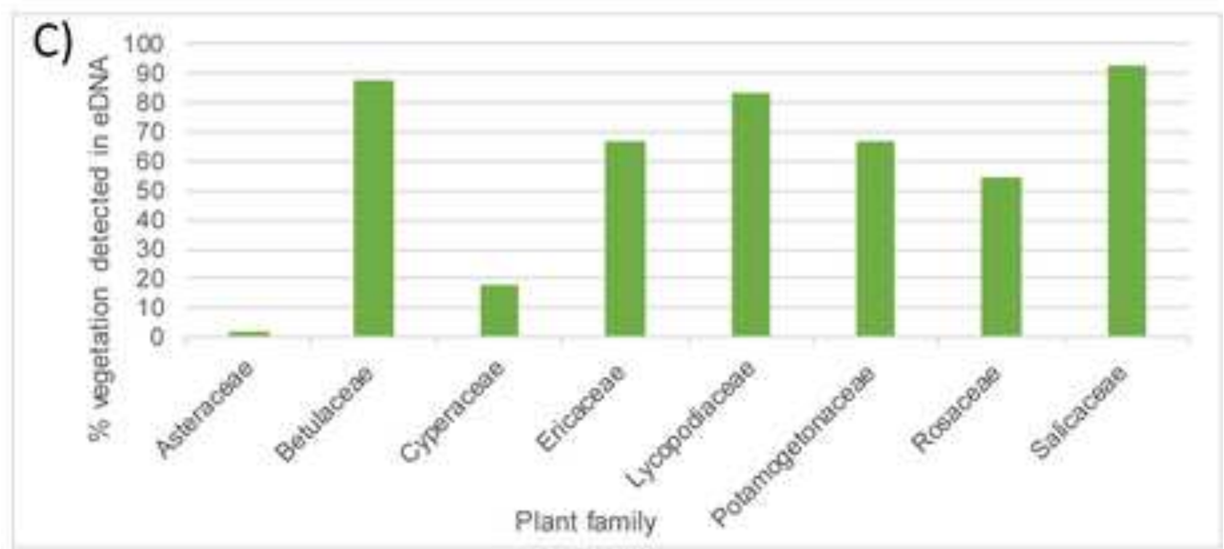
[Click here to download Figure Fig1.tif](#)

Fig 2

Click here to download Figure Fig2.tif







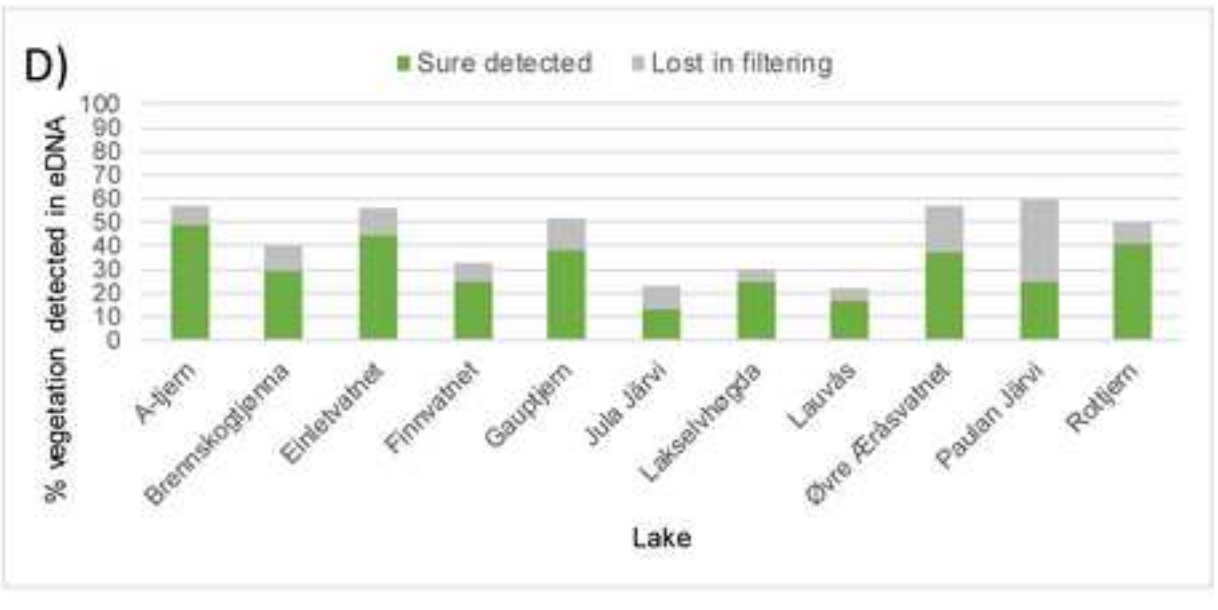


Fig 3

Click here to download Figure Fig3.tif

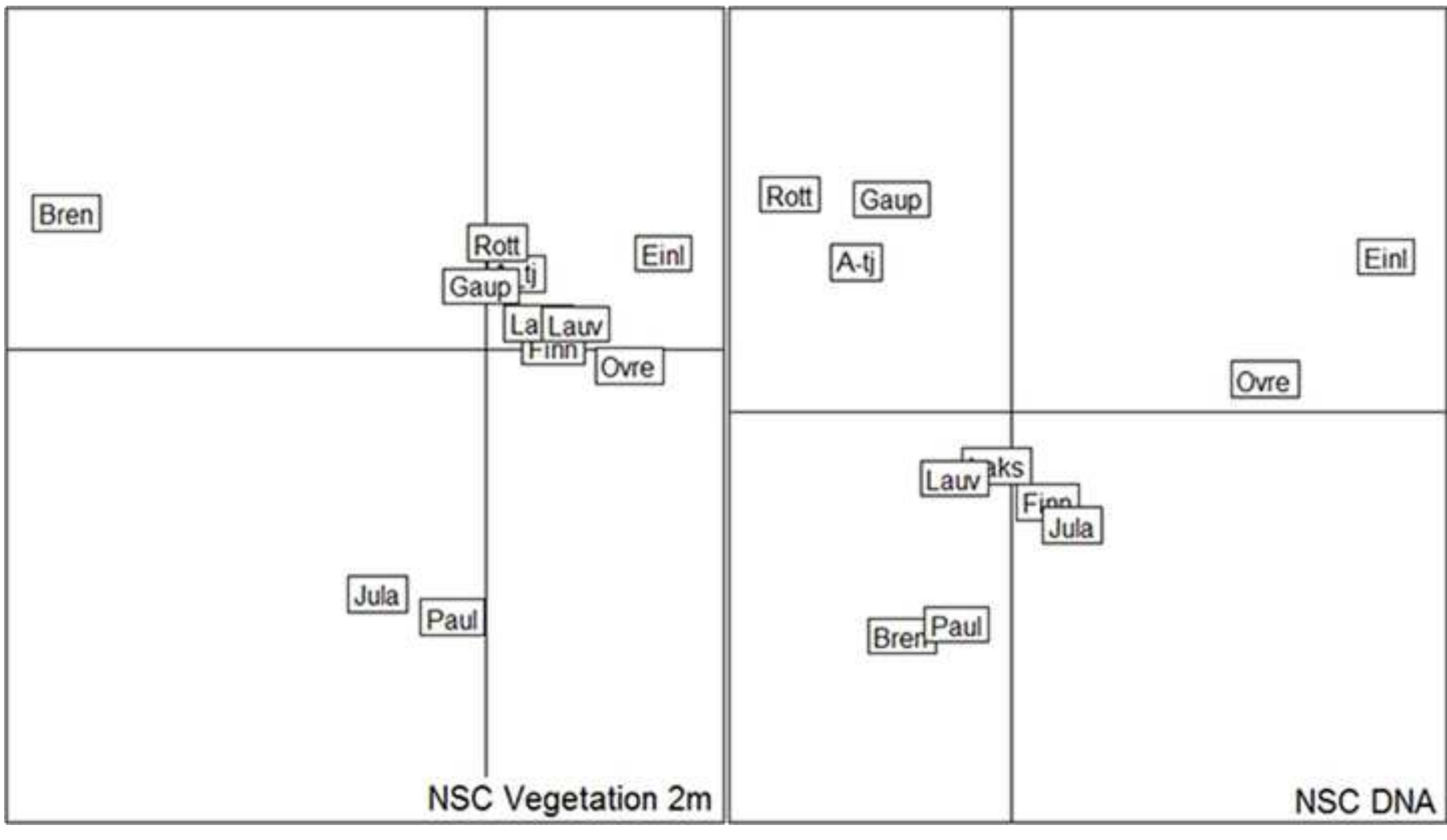
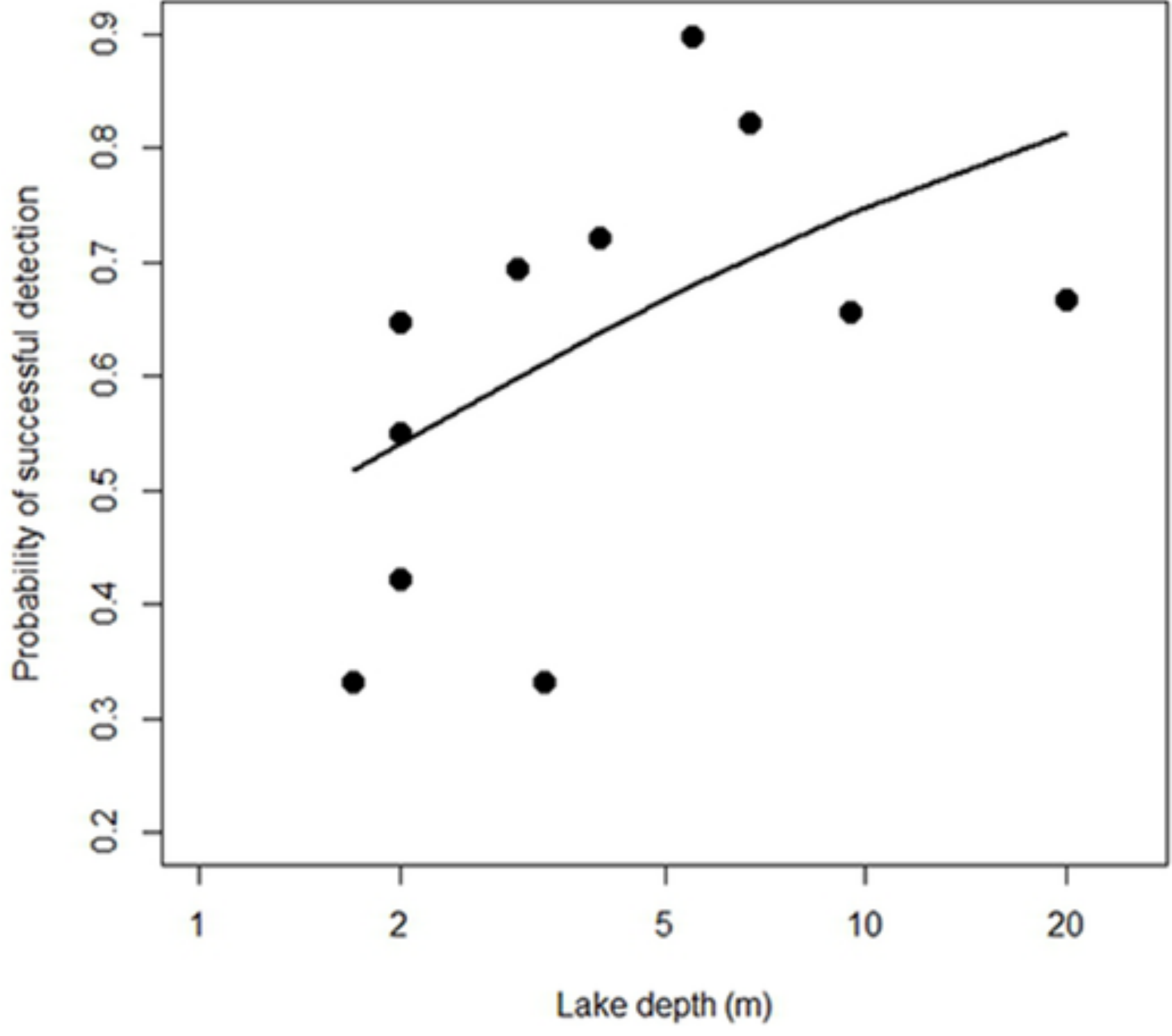


Fig 4

Click here to download Figure Fig4.tif



Click here to access/download
Supporting Information
S1_Appendix.docx

Click here to access/download
Supporting Information
S1_Table.docx

Click here to access/download
Supporting Information
S2_Table.docx

Click here to access/download
Supporting Information
S3_Table.docx

Click here to access/download
Supporting Information
S4_Table.xlsx

Click here to access/download
Supporting Information
S5_Table.docx

Click here to access/download
Supporting Information
S6_Table.docx