

1 Ancient plant DNA in lake sediments by Parducci et al.

2

3 New Phytologist (2017) 214: 924–942

4 doi: 10.1111/nph.14470

5

6 Author final copy

7

8 This document is provided solely in order to be compliant with HEFCE requirements for REF. As

9 it may contain errors subsequently corrected in proof, NOTHING in it should be trusted.

10

11 Please use the final version on the publisher's website, or contact the authors for a final

12 published PDF that can be considered reliable.

13

14 **Ancient plant DNA in lake sediments**

15

16 Laura Parducci<sup>1</sup>, Keith D. Bennett<sup>2,3</sup>, Gentile Francesco Ficetola<sup>4,5</sup>, Inger Greve Alsos<sup>6</sup>,  
17 Yoshihisa Suyama<sup>7</sup>, Jamie R. Wood<sup>8</sup>, Mikkel Winther Pedersen<sup>9</sup>

18

19 <sup>1</sup>Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University,  
20 Norbyvägen 18D, 75236 Uppsala, Sweden; <sup>2</sup>Department of Geography & Sustainable  
21 Development, School of Geography & Geosciences, University of St Andrews, KY16 9AL,  
22 Scotland; <sup>3</sup>Queen's University Belfast, Marine Laboratory, Portaferry, BT22 1LS, Northern  
23 Ireland, UK; <sup>4</sup>Université Grenoble-Alpes, CNRS, Laboratoire d'Ecologie Alpine (LECA), F-38000  
24 Grenoble, France; <sup>5</sup>Department of Biosciences, Università degli Studi di Milano, 20133 Milano  
25 Italy. <sup>6</sup>Tromsø Museum, UiT – The Arctic University of Norway, NO-9037 Tromsø, Norway;  
26 <sup>7</sup>Field Science Center, Graduate School of Agricultural Science, Tohoku University, 232-3  
27 Yomogida, Naruko-onsen, Osaki, Miyagi 989-6711 Japan; <sup>8</sup>Long-term Ecology Lab, Landcare  
28 Research, PO Box 69040, Lincoln, Canterbury 7640, New Zealand; <sup>9</sup>Centre for GeoGenetics,  
29 Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen,  
30 Denmark.

31

32 Authors for correspondence:

33 Laura Parducci, +46 18 471 64 14, [laura.parducci@ebc.uu.se](mailto:laura.parducci@ebc.uu.se); [@lauparducci](https://twitter.com/lauparducci)

34 Mikkel Winther Pedersen, +45 29 27 53 42, [mwpedersen@snm.ku.dk](mailto:mwpedersen@snm.ku.dk); [@miwipe81](https://twitter.com/miwipe81)

35

36 Total word count for the main body of the text: 8995

37 Number of figures: 7 (all should be published in colour)

38 Number of tables: 1

39 Supporting information: Note S1

40

41 **Summary**

42 Recent advances in sequencing technologies now permit analyses of plant DNA from fossil  
43 samples (ancient plant DNA, plant aDNA), and thus enable molecular reconstruction of  
44 palaeofloras. Hitherto, ancient frozen soils have proved excellent in preserving DNA  
45 molecules, and have thus been the most commonly used source of plant aDNA. However,  
46 DNA from soil mainly represents taxa growing few meters from the sampling point. Lakes  
47 have larger catchment areas and recent studies suggest that plant aDNA from lake sediments  
48 is a more powerful tool for palaeofloristic reconstruction. Furthermore, lakes can be found  
49 globally in nearly all environments and are therefore not limited to perennially frozen areas.  
50 Here we review the latest approaches and methods for studying plant aDNA from lake  
51 sediments and discuss the progress made up to present. We argue that aDNA analyses add  
52 new and additional perspectives for studying ancient plant populations and in time will  
53 provide higher taxonomic resolution and more precise estimation of abundance. Despite this,  
54 key questions and challenges remain for such plant aDNA studies. Finally, we provide  
55 guidelines on technical issues, including lake selection and we suggest directions for future  
56 research on plant aDNA studies in lake sediments.

57

58 Keywords: ancient plant DNA, bioinformatics, environmental DNA, high-throughput DNA  
59 sequencing, lake sediments, metabarcoding, shotgun sequencing, pollen, taphonomy

60

## 61 Introduction

62

63 Ancient DNA (aDNA) analysis is a young, but rapidly developing research field. Since the  
64 pioneering work in the 1980s (Higuchi *et al.*, 1984; Pääbo, 1984) there has been an  
65 exponential increase in aDNA studies investigating evolution and ecology of the last 800  
66 thousand years before present (kyr BP).

67 A Web of Science search on aDNA studies (January 1984-August 2016, keyword ‘ancient  
68 DNA’), detected 2104 papers. Most focus on vertebrates, especially humans (50%), while only  
69 5% focus on plants and ecology (Fig. 1). Ancient DNA from skeletal remains has been so much  
70 easier to study than plant macrofossils, and the difficulty in getting aDNA from charred  
71 specimens that represent 95% of the plant archaeological record, may partially explain the  
72 lack of plant aDNA papers. Plant aDNA studies have been also delayed by the difficulty in  
73 finding standard barcode regions useful for all species (universal) and sufficiently variable to  
74 discriminate among taxa. Three important trends are evident from Fig. 1: (i) the total number  
75 of studies has increased steadily through the last three decades; (ii) the number of opinion  
76 and methodological papers accounts for most of the increase since ca 2004; (iii) after three  
77 decades of aDNA research, plants continue to receive limited attention in relation to their  
78 importance in the landscape.

79 Despite this, plant aDNA research has made dramatic progress with: (i) the recent adoption of  
80 high-throughput DNA sequencing (HTS) technology; (ii) the discovery that DNA can be  
81 isolated from fossil pollen (Suyama *et al.*, 1996; Parducci *et al.*, 2005); and (iii) the finding that  
82 plant aDNA can be extracted from ice-cores, permafrost soil, lake sediments, coprolites and  
83 peat cores (see Rawlence *et al.*, 2014). Since the first research (Willerslev *et al.*, 2003), a  
84 number of studies on past biodiversity have been published using aDNA from a variety of  
85 palaeoenvironments (ref. 3-18 in Pedersen *et al.* 2015). These studies (see also Table 1) show  
86 how DNA can often identify more species and at higher taxonomical resolution than those  
87 identified by pollen and macrofossil analyses, thus providing important ecological and climatic  
88 information on the investigated sites otherwise difficult or impossible to infer (e.g. minimum  
89 July temperature or nutrient conditions for taxa identified at species levels; Sønstebo *et al.*  
90 2010, Parducci *et al.* 2015). DNA studies generally identify more herbs (Willerslev *et al.* 2014,  
91 Alsos *et al.* 2016) and have greater taxonomic resolution for grasses, thus providing better  
92 information on local biodiversity and for reconstruction of palaeoenvironments. The latter is  
93 an advantage compared to pollen especially at high latitudes/altitudes where local pollen  
94 productivity is low and long-distance pollen dispersal is more common.

95 The DNA field however, is not without technical challenges, for which specialised techniques  
96 and protocols have been developed (Hofreiter *et al.*, 2001; Chan *et al.*, 2005; Shapiro &  
97 Hofreiter, 2012). It is therefore timely to review recent progress on plant aDNA studies and  
98 to suggest new directions for the future. We focus our review on lake sediments, because  
99 their geological context provides a robust archive for retrieval of plant aDNA through time,  
100 and because lakes can be found in all environments around the world. We synthesize and  
101 discuss recent key findings on DNA from ancient pollen and sediment samples (see Table 1  
102 form a complete list of studies) and in particular the progress achieved using HTS  
103 technologies. Finally we discuss issues relating to false positives and the need for  
104 authentication (i.e. replication, use of controls), taphonomic processes, and bioinformatic  
105 challenges relating to the newest taxonomic identification and authentication of aDNA  
106 sequences.

107

108

#### 109 **Environmental and temporal limits for survival of aDNA**

110 The temporal limit of DNA preservation is dictated by the rate of degradation, which varies  
111 depending on the source material and micro-environmental conditions. The most favourable  
112 conditions (anoxic and frozen/cold) occur in permafrost and ice; here DNA can persist in  
113 biotic remains and environmental samples (e.g. soils) for hundreds of thousands of years  
114 (Lindhal, 1993; Hofreiter *et al.*, 2001; Allentoft *et al.*, 2012; Dabney *et al.*, 2013; Orlando *et al.*,  
115 2014). Currently, the oldest authenticated plant aDNA sequences are from frozen  
116 environments dated between 450 and 800 kyr (Willerslev *et al.*, 2007). Such favourable  
117 conditions however, are restricted to polar regions and high alpine environments. Plant aDNA  
118 has also been successfully extracted and analysed from arid and hot environments (Hofreiter  
119 *et al.*, 2003; da Fonseca *et al.* 2015; Mascher *et al.* 2016), temperate middens and coprolites  
120 (see Rawlence *et al.*, 2014), suggesting that warm temperatures are not necessarily a barrier  
121 for preservation of DNA molecules.

122 Several studies have investigated the post mortem processes affecting DNA molecules in  
123 different tissues. We know that after the death of an organism, several intra and extra-  
124 cellular processes (e.g. enzymatic, hydrolytic and oxidative processes) cause DNA damage  
125 (seen as misincorporation of C to T and G to A transitions primarily toward the ends of the  
126 DNA molecules) (Briggs *et al.*, 2007; Jónsson *et al.*, 2013), eventually leading to fragmentation  
127 of the DNA molecules. The highest success rate for aDNA isolation is normally obtained from

128 frozen, anoxic or arid areas, environments with limited bacterial abundance and therefore  
129 presence of nucleases, which reduces longer nucleic acids to short molecules (Hofreiter *et al.*,  
130 2001). Nevertheless, exogenous processes will inevitably lead eventually to the  
131 destabilization, fragmentation and damage of DNA, even in good preservation conditions.

132 To determine the rate of DNA decay, Allentoft *et al.* (2012) quantified the temporal survival of  
133 DNA molecules in bones. They found a half-life of 521 years for short [242 base pairs (bp)]  
134 mitochondrial DNA (mtDNA) fragments preserved at ca. 13 °C. The fragmentation rate  
135 however was strongly reduced at lower temperatures, and it was assumed that under very  
136 cold and dry conditions short fragments should be retrievable from bone more than 1 million  
137 years old. No similar studies exist for ancient plant tissues, and there are currently no  
138 indications of an age limit for plant aDNA. Nevertheless, damage for plant DNA also  
139 accumulates with age (Pedersen *et al.*, 2016). So far, plant DNA has been recovered and  
140 analysed from sediments >55 ka BP (Willerslev *et al.*, 2014), from ice cores dated between  
141 450 and 800 kya BP (Willerslev *et al.*, 2007) and from pollen ca. ~150 ka BP (Suyama *et al.*,  
142 1996), which suggests that DNA studies from lake sediments can potentially span the Late  
143 Pleistocene and Holocene.

144

145

#### 146 **Lake sediments**

147 During the last decade, molecular ecologists have increasingly begun investigating Late  
148 Quaternary floristic history using DNA isolated from pollen and DNA extracted from lake  
149 sediments (pollen DNA and *sed*aDNA, respectively) for a number of reasons. We review these  
150 below.

151

#### 152 Lake sediment characteristics

153 Lakes are excellent sources of sediments containing aquatic and terrestrial environmental  
154 components accumulated over time and preserved in robust stratigraphic contexts and  
155 anoxic conditions. They offer the best settings for preservation of plant aDNA, as the  
156 sediments accumulate continuously; allowing establishment of high-resolution molecular  
157 records using available geo-chronological methods (Fig. 2a, b, c) and thus providing a robust  
158 record for interpretation of the molecular history. In particular, sediment records from small  
159 lakes (Fig. 3) are excellent archives for molecular studies, as the effects of disturbances are

160 low and seem to represent the surrounding terrestrial environment well. Lake sediments  
161 consist of variable proportions of autochthonous/allochthonous organic material (OM) and in-  
162 washed inorganic material. Microbial degradation of OM frequently creates anoxic conditions  
163 in the bottom water and below 1-2 cm sediment depth (Sobek *et al.*, 2009). Such conditions  
164 preclude the presence of burrowing animals and thus minimise bioturbation, water  
165 percolation and sediment reworking (Pansu *et al.*, 2015), all of which are major concerns for  
166 palaeoecological studies. Water has maximum density at 4°C, so water less dense than this,  
167 both warmer and colder, floats. The net result is that bottom waters become insulated from  
168 the atmosphere, thereby favouring the development of anoxia and temperature stability.  
169 However, the temperature of lake water also depends upon geography and depth  
170 (Hutchinson, 1957; Wetzel, 2001). If lakes are deep enough, the water column becomes  
171 thermally stratified. During the summer, surface water warms and establishes a gradient  
172 down to cooler lower water. During the winter, surface water cools until it reaches 4°C  
173 (maximum density), at which point no more cooling occurs until the whole lake is mixed at  
174 4°C. Tropical lakes with little seasonal variation of temperature may have more or less  
175 constant temperatures year-round, at a level similar to ambient temperatures. In temperate  
176 and cold regions, the bottom waters will normally be colder than surface water in the  
177 summer, and at a similar temperature in the winter, when the whole lake is cold (and may be  
178 ice covered). Sediments are thus in contact with the coldest water, and become insulated  
179 from the atmosphere, favouring the development of anoxia, and have greater temperature  
180 stability, increasing the probability of DNA survival.

181 The sediments of lakes where anoxic conditions dominate can also be 'laminated' (i.e. layers  
182 of different composition reflecting seasonal environmental differences). In some cases,  
183 laminated sediments are sufficiently continuous to provide a temporal record of layers with  
184 annual resolution (Larsen & Macdonald, 1993), similar to tree rings. While vertical migration  
185 (leaching) of DNA has been observed in cave sediments (Haile *et al.*, 2007) and non-frozen  
186 soils (Andersen *et al.*, 2011), leaching has not been observed in lake sediments (Anderson-  
187 Carpenter *et al.*, 2011; Pansu *et al.*, 2015; Sjögren *et al.*, 2016). Once imbedded in the  
188 sediments, plant macrofossils and pollen grains, as well as silica, clay or organic matter, to  
189 which extracellular DNA attach (Pietramellara *et al.*, 2009; Poté *et al.*, 2009; Taberlet *et al.*,  
190 2012a), are unlikely to move vertically. On the other hand, re-deposition of sediments can  
191 occur in lakes, contaminating the micro and macrofossil record with older material. While  
192 Pedersen *et al.* (2016) recently found pre-Quaternary re-deposited microfossils in lake  
193 sediments from the Peace River drainage basin in North America, they also found that the

194 DNA record remained un-affected, probably due to the significant older age and smaller  
195 quantities of the re-deposited material. Nevertheless, re-deposition of material should not be  
196 neglected as a possible source of DNA.

197

198 Coring techniques for aDNA work

199 The collection of sediment cores for plant aDNA can be done with the same coring devices  
200 used for conventional palaeoecological work, but some special precautions and procedures  
201 may be necessary to avoid contamination in the field as far as possible. Corers that enclose  
202 the sediment, such as piston or percussion corers (eg. Nesje, 1992; Wright *et al.*, 1984) are to  
203 be preferred. Other frequently used coring systems for palaeolimnology, such as the Kajak  
204 corer or HTH gravity corer (Renberg & Hansson, 2008) can also be used, and be essential for  
205 collection of surface sediments. If the coring is done in winter, there is minimal risk of  
206 contamination by airborne DNA (e.g. in pollen), but even this can be eliminated by sealing the  
207 core tubes (both piston and gravity systems), for transport to the laboratory, directly in the  
208 field. Additional procedures, including equipment sterilisation, are also available (Feek *et al.*,  
209 2006; 2011). On extruding and opening the cores, however, it must be assumed that the  
210 surface is contaminated, so subsamples must be taken from inside the undisturbed centre.  
211 During sub-sampling, it is therefore important to remove or avoid the first 4-10 mm of outer  
212 sediment in a clean laboratory setting, using sterile tools, full bodysuit and gloves to obtain an  
213 uncontaminated sample from within the centre of the core (Fig. 4a,b,c). Hence the core  
214 should have sufficient diameter to allow this. Extra precautions can also be taken by applying  
215 a DNA tracer to the coring equipment, which allows testing for infiltration by DNA molecules  
216 from the outer layers into the inner sampled sediments (Pedersen *et al.*, 2016), or by having a  
217 DNA-free water sample exposed to air in the laboratory during subsampling as a negative  
218 control.

219

220

221 **Perspective for plant aDNA research**

222 Traditionally, lake sediment records have been routinely analysed for pollen and other plant  
223 fossils, which have been the basis for our understanding of landscape-scale distribution of  
224 plants on millennial timescales. Plant aDNA now offers the potential of helping elucidation of  
225 long-standing ecological questions that classical palaeoecological techniques may not be able



226 to retrieve. It has been through collaborative work between molecular ecologists and classical  
227 palaeoecologists that robust and reliable results have been produced recently (Willerslev *et al.*,  
228 2014; Pedersen *et al.*, 2016; Alsos *et al.*, 2016), and this type of collaboration is strongly  
229 advocated (Hu *et al.*, 2009; Anderson-Carpenter *et al.*, 2011; Brown *et al.*, 2014). Examples of  
230 problems being elucidated by combining the two disciplines include Reid's paradox of rapid  
231 plant migration (Clark *et al.*, 1998) and the question of whether or not trees survived the Last  
232 Glacial Maximum (LGM) at high-latitudes in Europe (Parducci *et al.*, 2012). Here, combined  
233 information from fossils and molecular studies (modern and ancient) has changed our  
234 traditional view of post-glacial migration of trees from southern European refugia (Hewitt,  
235 2000; Stewart *et al.*, 2010). Another example is the limited taxonomic resolution generally  
236 achieved by pollen analysis. Even if such analyses have recently reached better taxonomic  
237 resolution due to improved identification keys, the work is still time demanding, and  
238 identifications are often at genus or family level, rarely species (Faegri *et al.*, 1989). We thus  
239 have little understanding about specific diversity, and even less information about ancient  
240 plant populations. The new possibilities offered by the latest HTS technologies will likely  
241 increase our ability to resolve plant taxa at species level and additionally elucidate ancestry  
242 and genetic composition of ancient plant populations. Likewise, metabarcoding and  
243 metagenomic analyses of *seda*DNA will provide more detailed insights on ancient ecosystems  
244 and link changes more tightly to past climate shifts (see section on HTS).

245

246

#### 247 **Sources of pollen, macrofossil and DNA**

248 The majority of lake sediment deposits contain both plant remains and non-biological  
249 material originating from the lake catchment. In small boreal lakes it is estimated that ca 70%  
250 of the deposited pollen is from vegetation growing within few km of the lake (Jacobson &  
251 Bradshaw, 1981). While the majority of the pollen often derives from high-pollen producing  
252 wind-pollinating plants, which are distributed regionally through the air, the proportions of  
253 pollen from insect-pollinated plants may vary between sites. Furthermore, in areas of local  
254 low pollen production the pollen record may be highly affected by long distance exotic pollen  
255 that is not part of the regional environment (Hyvärinen, 1970). Despite this, pollen records  
256 generally represent the regional flora. Plant macrofossils, however, are of local origin, as they  
257 are large and have low dispersal and transport capacity (Allen & Huntley, 1999). The  
258 macrofossil record is often dominated by the aquatic and wetland plants growing in and

259 around the lake, and terrestrial taxa often under-represented (Birks, 2003). *SedaDNA* seems  
260 to have a similar source of origin as macrofossils (Jørgensen *et al.*, 2012; Pedersen *et al.*,  
261 2013; Alsos *et al.*, 2015; Parducci *et al.*, 2015). To date, only a few studies have focussed on  
262 the release and deposition of DNA in the environment (Poté *et al.*, 2007; Pietramellara *et al.*,  
263 2009; Poté *et al.*, 2009; Barnes & Turner, 2016) and much therefore remains unknown about  
264 the ecology of the DNA – e.g. all processes occurring from source to deposition (taphonomy).  
265 Current *sedaDNA* results suggest that DNA extracted from sediments does not derive from  
266 actual pollen grains (Pedersen *et al.*, 2016; Sjögren *et al.*, 2016), but from other components  
267 imbedded in the sediment matrix, thus pollen DNA needs to be extracted directly from single  
268 or multiple isolated grains (see below).

269

## 270 Pollen DNA

271 Fossil pollen is often abundant and can be well preserved in lake sediments. Once deposited,  
272 pollen remains in situ in the sediments and there is therefore a high degree of certainty to its  
273 stratigraphic context. Further, aerial transport and rapid burial rates in sediment results in  
274 minimal physical damage to pollen grains and minimal exposure of the grains to biotic  
275 degradation. In addition the outer part of pollen grains (exine) is composed of cellulose and  
276 sporopollenin, an acid-resistant polymer that contains saturated and unsaturated  
277 hydrocarbons and phenolics (Southworth, 1974), and which protects the grains from physical  
278 and chemical attack, aiding pollen preservation in the sediments (Bennett & Willis, 2001).  
279 Inside the pollen, after maturation, there are 2-3 cells (a large vegetative cell and 1-2  
280 generative cells) that comprise the male gametophyte (Fig. 5). The vegetative cell comprises  
281 the cytoplasm, and numerous plastids and mitochondria, which are responsible for the  
282 development of the pollen tube and delivery of the generative cells to the embryo sac  
283 together with the nuclear haploid DNA. Generative cells of pollen from the majority of plant  
284 species contain multiple organelles, including several copies of organelle DNA [chloroplast  
285 (cpDNA) and mtDNA], regardless of the type of inheritance of these genomes (maternal or  
286 paternal). Some nuclear repetitive regions, like ITS ribosomal repeats, are also present in  
287 multiple copies in the nucleus of both cell types. During pollen maturation however, there is a  
288 selective increase or decrease in the amount of organelle DNA in the generative cells (not in  
289 the vegetative) depending on the inheritance pattern (Nagata *et al.*, 1999; Zhang & Liu, 2003).  
290 For example, mature pollen from species with paternal cpDNA inheritance (the majority of  
291 conifers), contains a regular amount of mtDNA and cpDNA in the vegetative cell and an  
292 increased amount of cpDNA in the generative cell/s. On the other hand, pollen from species

293 with maternal inheritance of cpDNA and mtDNA (most of the angiosperms) contains at  
294 maturation a regular amount of both genomes in the vegetative cell and a decreased amount  
295 in the generative cell/s (Fig. 5). All three plant genomes are therefore present in both pollen  
296 types, but the cpDNA and the mtDNA may be present in different amount (in the generative  
297 cell/s) in different taxa.

298

#### 299 Single-pollen genotyping

300 Petersen *et al.* (1996) were the first to amplify cpDNA from single pollen grains (*Hordeum* and  
301 *Secale*), while Suyama *et al.* (1996) were the first to amplify DNA from peat sediment pollen  
302 (fir, *Abies*) ca. 150 kyr old. Successively, Parducci *et al.* (2005; 2012) succeeded in retrieving  
303 short cpDNA and mtDNA fragments from Holocene *Pinus* and *Picea* pollen. A description of  
304 the techniques used for isolating and direct amplifying from single grains is presented in Box 1  
305 and in Parducci *et al.* (2005). Later, the same technique was used to sequence cpDNA from  
306 angiosperm fossil pollen from the Venice Lagoon (Paffetti *et al.*, 2007) and conifer pollen from  
307 glaciers (Nakazawa *et al.*, 2013). Using multiplex PCR and single-pollen genotyping methods  
308 on fresh pollen (Isagi & Suyama, 2010) it is also possible to perform paternity analysis and  
309 infer pattern and distance of pollen dispersal in modern plant populations (Matsuki *et al.*,  
310 2007; 2008; Hasegawa *et al.*, 2009; Hirota *et al.*, 2013; Hasegawa *et al.*, 2015). The potential  
311 of single-pollen analysis on fossil pollen however has not been explored further using  
312 traditional PCR-based Sanger sequencing technologies, because the PCR success-rate is low  
313 and the time required to handle and prepare the grains is high. With the advent of HTS  
314 technology, and in particular the recent availability of methodologies to directly construct  
315 HTS libraries from single cells (single cell sequencing technologies, SCS), it will now be  
316 possible to investigate more efficiently individual fossil pollen grains and hence conduct plant  
317 aDNA studies more effectively even at the population level (see section on HTS).

318

319

#### 320 Sedimentary ancient DNA

321 Different sediment types show distinctive physical and chemical characteristics that will  
322 differentially affect DNA preservation, thus DNA survival will vary between different locations.  
323 Likewise, it seems that extraction of DNA and removal of inhibiting substrates requires  
324 strategies adapted to differences in the sediment content (Taberlet *et al.*, 2012a; Pedersen *et*

325 *al.*, 2016). Furthermore, extraction of aDNA molecules requires optimized protocols and  
326 special laboratory precautions.

327

328 Proxy overlapping or complementary

329 Before a robust inference of past vegetation (presence and abundance of taxa) can be based  
330 on *sedaDNA*, it is critical to consider the origin and taphonomy of plant fossil assemblages in  
331 the sediments and the influences of environmental, biological and physical factors affecting  
332 the presence and relative abundance of their DNA molecules (Jackson, 2012; Barnes &  
333 Turner, 2016).

334 Typically a low overlap has been found in lakes between pollen flora and DNA identifications,  
335 which has led to the inference that DNA in lake sediments originates locally (Jørgensen *et al.*,  
336 2012; Pedersen *et al.*, 2013; Parducci *et al.*, 2013; 2015; Sjögren *et al.*, 2016) (Fig. 6).

337 Most plant macrofossils found in lakes are from the local vegetation (Birks, 2013). Despite  
338 this, most studies show a low to medium overlap (12-56%) between taxa recorded by  
339 *sedaDNA* and macrofossils (Jørgensen *et al.*, 2012; Parducci *et al.*, 2012; Boessenkool *et al.*,  
340 2013; Porter *et al.*, 2013; Pedersen *et al.*, 2013), while only one study has showed an overlap  
341 close to 100% (Alsos *et al.*, 2016). These differences and the limited overlap found between  
342 proxies may be due to: (i) differences in taphonomic processes between sites; (ii)  
343 incompleteness of reference genomic databases; (iii) robustness of the experimental design  
344 (Ficetola *et al.*, 2016); and (iv) number of pollen/macrofossils counted vs. sequencing depth  
345 (see sections below for further discussion of all points).

346 The most stringent test for identifying the origin of *sedaDNA* is by direct comparison with  
347 modern vegetation survey or with detailed historical vegetation maps. Yoccoz *et al.* (2012)  
348 were the first to demonstrate that plant diversity detected from environmental DNA  
349 extracted from boreal soil was consistent with plant taxonomic diversity estimated from  
350 conventional aboveground surveys. To date, we still do not know if such a good relationship  
351 holds with DNA in lake sediments. In a study from a high altitude crater lakes in Africa,  
352 Boessenkool *et al.* (2013) showed that *sedaDNA* largely reflected local flora. Similarly, Sjögren  
353 *et al.* (2016) compared DNA, pollen and historical vegetation maps in two Scottish lakes and  
354 found that *sedaDNA* was of local origin in contrast to regionally dispersed deciduous tree  
355 pollen. Finally, in a comparison of *sedaDNA* with vegetation surveys of 11 lakes in Northern  
356 Norway, the majority of taxa recorded in the *sedaDNA* were growing within 2 m of the lake (I.

357 G. Alsos, unpublished; Alsos *et al.*, 2015). Thus, we conclude that DNA deposition in lakes is  
358 more similar to that of macrofossils than pollen, and represents flora from within the  
359 catchment area. However with future improved DNA reference databases, methods and  
360 understanding, we expect increased information gained will lead to an almost complete  
361 overlap between DNA and macrofossils, but likely not between DNA and pollen (Fig. 6).

362

363 Taphonomic processes in lake sediments

364 Taphonomic processes (i.e. dispersal, transport, incorporation, and preservation of fossils and  
365 molecules in sediments) can affect assemblages recovered from sediments (Barnes & Turner,  
366 2016). How organisms or parts of organisms preserve in sediments, and the fact that some  
367 preserve better than others can influence the range of taxa identified and therefore result in  
368 molecular and fossil indicators that are 'silent'. These factors influence pollen, macrofossil  
369 and DNA records in lake sediments (Fig. 2d, e). Moreover, taphonomic processes clearly vary  
370 in their impact for different indicators and the relative intensity of the suite of processes  
371 influencing pollen preservation may be different from those affecting plant DNA and  
372 macrofossils.

373 Lake sediments contain DNA from many different organisms, which include cellular DNA from  
374 tissues and intact cells, as well as extracellular DNA. When a plant tissue is degraded and a  
375 cell is lysed, it releases its content in the surrounding environment and extracellular DNA can  
376 bind to charged mineralogical and organic particles or remain unbound. Pietramellara *et al.*  
377 (2009) showed that in modern soils, once the DNA binds to a particle, it is immediately  
378 protected against nuclease degradation. This implies that that clay-rich soil can be highly  
379 suitable for protection of DNA against degradation, as clay particles have relatively large and  
380 charged surface area (Huang, 2014). Extracellular DNA molecules in soils can also be taken up  
381 by competent prokaryotic cells in a process called natural transformation. Although it is  
382 unlikely that this will be a large source of 'plant' DNA, it remains however a possibility  
383 (Pedersen *et al.*, 2015). Whether the plant DNA in lake sediments primarily are present as  
384 extracellular DNA or as small plant parts like leaves, root caps cells, stem or fruits, remains  
385 however still unclear.

386

387 Inferring taxa abundance from PCR-based *seda*DNA analyses

388 As with macrofossils, aquatic plants are well represented in *seda*DNA from lake sediments.  
389 Detection of a species in modern or semi-modern sediments depends on both distance to the  
390 lake shore and its abundance in the vegetation (Alsos *et al.*, 2015; Sjögren *et al.*, 2016). For  
391 example, the dominant species growing around lakes are found with a high proportion of  
392 DNA reads and present in more PCR repeats for most cases. This also seems to be the case for  
393 ancient samples, as all species represented from at least one macrofossil are also detected in  
394 more than one PCR replicate (Alsos *et al.*, 2016). Nevertheless, for the time being, we should  
395 be very cautious about interpreting quantities of DNA beyond rough estimates when using  
396 PCR-based methods, as several metabarcoding processes may cause bias (e.g. primer binding  
397 site, amplicon length, taxonomy and diversity of extract) (Pornon *et al.*, 2016).

398

399

#### 400 **Recent key findings and future methods using HTS techniques**

401 HTS expands upon traditional PCR-based Sanger sequencing techniques and has facilitated a  
402 rapid development of aDNA research during the last decade. Ancient DNA molecules  
403 converted to sequencing libraries can now be parallel sequenced massively on HTS platforms  
404 like the Illumina HiSeq or the Complete Genomics platforms. HTS approaches have become  
405 increasingly affordable and are now routinely used by most aDNA laboratories, allowing  
406 screening of *seda*DNA from a wide range of complex ancient substrates. With the newest  
407 generation of desktop HTS platforms, e.g. NextSeq, even small laboratories can now sequence  
408 their own datasets. There are, in principle, three HTS methodological strategies for analysing  
409 plant aDNA in lake sediments: metabarcoding or shotgun sequencing of *seda*DNA and HTS of  
410 pollen DNA. Of the two latter methods, ‘metabarcoding’ is a relative established method,  
411 which relies on the information on one single locus, while the second is newer and relies on  
412 shotgun sequencing, i.e. sequencing a non-discriminated pool of aDNA. For clarity, it is  
413 important to define this new method. Here we suggest the term ‘shotgun metabarcoding’  
414 when shotgun sequencing of environmental DNA is used to identify taxa, and ‘metagenomics’  
415 when shotgun is used for functional analyses of the environments. This will allow researchers  
416 to discriminate between studies focusing on taxon identification and studies focusing on  
417 functional and attribute analysis. In the following paragraphs we review and discuss the  
418 progress made and future of these three methodologies.

419

420 Metabarcoding on *sedaDNA*

421 DNA metabarcoding has received enormous attention in the last decade in many ecological  
422 fields (Soininen *et al.*, 2009; Valentini *et al.*, 2009; Taberlet *et al.*, 2012b). The ability of  
423 different barcodes to target groups of organisms has been investigated in different  
424 environments and under different conditions (in silico and in vitro PCR) leading to the  
425 conclusion that metabarcoding is highly applicable for biodiversity screening of modern  
426 samples, which implies a good potential in palaeoecological studies too (Thomsen &  
427 Willerslev, 2015).

428 Metabarcoding has already been used in a variety of palaeoenvironmental studies (Willerslev  
429 *et al.*, 2003; Valentini *et al.*, 2009; Taberlet *et al.*, 2012b), including permafrost (Lydolph *et al.*,  
430 2005; Jørgensen *et al.*, 2011; 2012; Willerslev *et al.*, 2014), mid to high latitude/altitude lake  
431 sediments (Coolen & Gibson, 2009; Parducci *et al.*, 2012; Alsos *et al.*, 2015; Pansu *et al.*, 2015;  
432 Epp *et al.*, 2015; Paus *et al.*, 2015; Alsos *et al.*, 2016); tropical lake sediments (Boessenkool *et al.*  
433 *et al.*, 2013) (F. M. Ficetola, unpublished), and deposits preserved under ice-sheets, even in the  
434 absence of visible macrofossil remains (Willerslev *et al.*, 1999; 2007).

435 A main advantage of metabarcoding is the possibility of simultaneously amplifying and  
436 identifying a large number of taxa for limited cost. Metabarcoding data on plant assemblages  
437 from *sedaDNA* in an Alpine catchment over the last 6.4 kyr years is comparable to those  
438 obtained using relevées performed on modern vegetation so it was possible to identify  
439 vegetation changes through time (e.g. shifts from shade-tolerant to heliophilous vegetation)  
440 (Pansu *et al.*, 2015). Plant metabarcoding data can also be combined with other proxies of  
441 past environments, to identify potential drivers of such changes using approaches derived  
442 from community ecology (Giguet-Covex *et al.*, 2014).

443 The performance of the metabarcoding approach is often limited if markers are not able to  
444 amplify all the present taxa (universality), and by their capacity to differentiate and identify  
445 also closely related species (resolution) (Ficetola *et al.*, 2010; Sønstebo *et al.*, 2010; Furlan *et al.*  
446 *et al.*, 2016). Therefore, the use of primers with appropriate features (i.e. "universal" primers  
447 amplifying with high specificity all plants, and achieving a high taxonomic resolution) (Ficetola  
448 *et al.*, 2010; Furlan *et al.*, 2016) is essential for the success of metabarcoding studies. In aDNA  
449 studies the choice of the marker is particularly difficult, as prerequisites like minimal bias in  
450 amplification of different taxa and short sequence length, drastically reduce the ability to  
451 resolve taxa (Taberlet *et al.*, 2007). Plant metabarcoding studies can use either a single  
452 marker or a combination of multiple primers to resolve taxa. For instance, the *trnL* g/h

453 primers (Taberlet *et al.*, 2007) can first be used to obtain an overall analysis of plant diversity  
454 with a good resolution for most plant families, and additional primers (e.g. ITS1-F/ITS1Poa-R  
455 for Poaceae and ITS1-F/TS1Ast-R for Asteraceae) may be later added to increase resolution  
456 within families (Baamrane *et al.*, 2012). We should however remember that different primers  
457 might favour amplification of different taxa, which may lead to biases in the final results  
458 (Yoccoz, 2012).

459 The power of metabarcoding for vegetation reconstruction is also dependant on the  
460 availability of comprehensive taxonomic reference libraries needed to identify *sedaDNA*  
461 sequences (Taberlet *et al.*, 2012b). Unfortunately, for many plant species there are at the  
462 moment no sequences deposited in publicly available databases such as GenBank. The  
463 number of such sequences however is now rapidly increasing, boosting the utility of this  
464 technique for investigating past plant history and population response to environmental  
465 change.

466 Metabarcoding thus provides a complementary tool to classical palaeoecological analysis and  
467 the three main proxies (DNA, pollen and macrofossils) can be used in combination or  
468 singularly depending on the aim of each study. If the aim is local vegetation reconstruction,  
469 we can either choose DNA or macrofossils, as the two overlap to a large degree. If the aim is  
470 to get a more regional signal from taxa, pollen analysis is probably appropriate if the  
471 taxonomic resolution is sufficient for the aim of the study. If we want to obtain a general view  
472 of plant communities through time, metabarcoding alone can also be sufficient because the  
473 ecological signal of metabarcoding is often good, and data generation fast and cheaper.

474

475 Shotgun metabarcoding on *sedaDNA*

476 Shotgun metabarcode analysis is a newer alternative to the traditional metabarcoding  
477 approach. It relies on shotgun sequencing DNA (Orlando *et al.*, 2015) from a non-  
478 discriminated genetic pool obtained from environmental samples and then computational  
479 investigation, e.g. bioinformatic analysis, is used to decipher the taxonomic composition.  
480 While the shotgun sequencing approach traditionally was used for studying the fraction of un-  
481 cultivable microbes in modern environmental samples (Vos *et al.*, 2013; Sharon & Banfield,  
482 2013) by genome assembly and functional analysis of the microbial fraction (Mackelprang *et al.*,  
483 2011). More recently shotgun sequencing has been recognized as an important tool for  
484 bypassing DNA barcode amplification biases (Ziesemer *et al.*, 2015; Pedersen *et al.*, 2016),  
485 and also appear to give a more comprehensive insight into the community composition from



486 all trophic layers represented by the environmental sample (Pedersen *et al.*, 2016). While this  
487 approach has played an expanding role in studying ancient prokaryotic communities in soil,  
488 sediments (>90% of DNA molecules in a sample are likely to be prokaryote), but also teeth,  
489 coprolites, gut and ice (e.g. Warinner *et al.*, 2015; Wood & Wilmshurst, 2016), only one study  
490 so far has published results that are accepted as authentic using shotgun sequencing data for  
491 palaeoenvironmental reconstruction (Pedersen *et al.*, 2016). However, there exists a  
492 considerable potential for applying low-coverage shotgun sequencing of genomic DNA  
493 (genome skimming) for studying plant aDNA (Coissac *et al.*, 2016), but the absence of a  
494 consensus data processing and lack of experience interpreting such datasets makes it prone  
495 to misinterpretations and false positives (Bennett, 2015; Weiß *et al.*, 2015) (see section on  
496 Bioinformatics). Species identification using genomic plant data outside chloroplast barcode  
497 regions can be particularly problematic. For instance, of ca. 391,000 vascular plant species  
498 existing on earth (number constantly changing through new discoveries and taxonomic  
499 revisions) only 1,092 are represented by complete chloroplast genomes (NCBI RefSeq  
500 database, <ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plastid/> - accessed 31 August 2016). In  
501 addition, taxonomic identifications outside the chloroplast genome are skewed towards  
502 overrepresented taxa with fully sequenced genomes, especially of commercially important  
503 species such as *Zea mays*, *Triticum aestivum*, *Solanum lycopersicum*, *Hordeum vulgare*, *Oryza*  
504 *sativa*, *Nicotiana tabacum* and the model organism *Arabidopsis thaliana*. The latter are all  
505 amongst the 20th most sequenced organisms (no. of bases, NCBI,  
506 <ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt> - accessed 29 March 2016). While we can limit false  
507 positives and confirm aDNA authenticity using bioinformatic techniques (see section on  
508 Metagenomic bioinformatics), the lack of genomic references and the limited experience in  
509 interpreting such datasets makes proper contextualized biological interpretation a necessity.  
510 However, reference databases are quickly improving and on-going projects (e.g. PhyloAlps,  
511 <https://www.france-genomique.org/spip/spip.php?article112&lang=fr> and NorBol,  
512 <http://norbol.org/>) are currently assembling the entire chloroplast and nuclear ribosomal  
513 genomes of the whole floras of respectively the Alps, Norway, and parts of the Arctic (Coissac  
514 *et al.*, 2016). Shotgun metabarcoding will therefore become an important tool in the future  
515 years, as it will allow for detecting organismal diversity and potentially differentiate  
516 population structures, and will further help bridge the gap between different scientific  
517 disciplines in palaeoecology.

518 An important advantage of shotgun sequencing is the possibility of quantifying the degree of  
519 DNA degradation exerted by the sample, e.g. fragmentation and hydrolysis of cytosines

520 (C), which results in accumulating misincorporation frequencies of thymine (T) instead of C at  
521 the ends of the DNA molecules during sequencing (Jónsson *et al.*, 2013). Lastly, it was shown  
522 recently that DNA damage correlates with age of the lake sediments (Pedersen *et al.*, 2016)  
523 and should therefore be used as a independent tool for aDNA authentication.

524

525 HTS on pollen

526 Recently Suyama & Matsuki (2015) developed a method for constructing HTS libraries and  
527 genotyping genome-wide single-nucleotide polymorphism (SNP) from low-quantity DNA  
528 templates termed ‘multiplexed ISSR (inter-simple sequence repeat) genotyping by sequencing  
529 or ‘MIG-seq’. Unlike standard methods based on restriction enzyme steps that require large  
530 amounts of good quality DNA templates, the MIG-seq procedure is based on an initial PCR  
531 step and can therefore discover and genotype *de novo* SNPs starting from reduced amounts  
532 of DNA. The technique has been recently applied on modern pollen of *Heimerocallis* and  
533 tested also on Pinus pollen a few years old collected from subsurface snow layers on a glacier  
534 (Y. Suyama, unpublished). The MIG-seq technique on the glacier pollen was first combined  
535 with a whole-genome amplification (WGA) step and successfully detected good-quality SNPs  
536 in an ancient pine population.

537 A more efficient way of analysing ancient pollen is to construct HTS libraries directly from  
538 single pollen grains employing the SCS technology. SCS is a cost- and time-effective method to  
539 sequence in parallel hundreds of single-cells isolated and assayed in tiny (nl) reaction  
540 chambers for PCR and sequencing (Wang & Navin, 2015). Limiting reactions to few nl offers  
541 the advantages of increased throughputs, improved reaction sensitivity (higher percentages  
542 of grains successfully genotyped) and increased degree of PCR specificity (reduced DNA  
543 contamination). SCS methodology on fossil pollen seems at the moment the best available  
544 approach to investigate efficiently a large number of pollen grains and a good alternative to  
545 the more time-consuming single-pollen genotyping technique (Parducci *et al.*, 2005). The  
546 method offers the unprecedented opportunity of analysing in a time-effective way the  
547 genetic structures of large number of single plant individuals on millennial time scales.

548 SCS protocols for pollen can be broken down into four main steps: (i) pollen isolation from  
549 sediments and cleaning; (ii) pollen walls disruption; (iii) PCR amplification and HTS library  
550 construction; (iv) DNA sequencing. For a description of the technique used for pollen isolation  
551 and cleaning see Box 1 and Parducci & Suyama (2011). Before step 2, pollen can be screened  
552 for DNA content by staining pollen suspensions with NST-DAPI buffer so that grains are gated

553 by total DNA content using fluorescence-based flow cytometry (L. Parducci pers. comm.).  
554 Prior to amplification, each grain is encapsulated in SCS micro reaction droplets and different  
555 microfluidic liquid handling techniques can be used to automatically trap the grains in such  
556 droplets and to dispense and analyse them efficiently (Kong *et al.*, 2012). The droplet-based  
557 microfluidic approach (Brouzes *et al.*, 2009) uses a 2-phase system, in which each assay is  
558 compartmentalized in an aqueous micro droplet of 1 pl to 10 nl surrounded by an immiscible  
559 oil. For disrupting pollen walls, two main physical lysis methods can be used: bead beating  
560 with glass or ceramic beads (Roberts, 2007) and ultrasonic homogenization (Dong *et al.*,  
561 2015). These methods avoid the use of chemicals or enzymes that may later interfere during  
562 amplification, however they appear to be challenging with SCS methods, because it is easy to  
563 perturb monodispersed water-in-oil emulsions. Alternative methods to disrupt pollen walls  
564 and release genomic DNA involve the use of detergent-based or enzymatic lysis agents. The  
565 use of an extraction buffer containing proteinase K, sodium dodecyl sulfate (SDS), tris-HCl and  
566 EDTA has been previously used successfully to extract DNA from manually-crushed fossil  
567 pollen grains (Parducci & Suyama, 2011), and should also work well in micro reaction  
568 droplets. Alternatively, the use of chemicals like ethanolamine, which specifically disrupt  
569 pollen wall components, can also be tested (Southworth, 1974). After breaking the pollen  
570 wall, the SCS workflow involves library construction through direct PCR amplification of short  
571 (50-200 bp) regions of interest [e. g. the *trnL g/h* cpDNA region (Taberlet *et al.*, 2007)] or  
572 other mini-barcodes for plants (Little, 2014)] and amplicon tagging of each grain using unique  
573 ID tags with sequencing adaptors using (for example, the Drop-seq strategy). Sequence  
574 capture technique can also be used to specifically enrich for target cpDNA or mtDNA regions  
575 of interest prior to sequencing (Stull *et al.*, 2013). In this case custom self-made 'baits' are  
576 used to first to capture DNA regions of interests from ancient samples and successively to  
577 NGS sequence the hybridized fragments.

578

### 579 **Challenges when studying aDNA from lake sediments**

580 Contamination, laboratory analyses and experimental setup

581 Contamination of low-concentration aDNA samples with high-concentration modern DNA  
582 poses a challenge that should not be taken lightly and special precautions should always be  
583 taken in all steps of analyses (Fulton, 2012). There is not a single strategy valid for avoiding all  
584 possible contamination sources occurring from: (i) laboratory facilities; (ii) cross  
585 contamination; and (iii) reagents used during extraction and downstream preparation prior to

586 sequencing. In general, multi-strategy procedures should be employed (Champlot *et al.*,  
587 2010) to avoid contamination. However, while contamination from laboratory facilities and  
588 cross contamination are strictly related to facilities and experience/training of the workers  
589 and can be more easily taken under control, contamination from reagents, which are known  
590 to contain DNA especially of common food plants, is more difficult to control. Firstly, not all  
591 reagents can be filtered, UV-lighted, bleached or DNase-treated (the preferred  
592 decontaminating method). Secondly, even when treated, short DNA molecules can still persist  
593 and be a source of genetic material during extraction and downstream handling. We  
594 therefore stress the importance for having always experimental controls covering all reagents  
595 and all steps during handling. Equally important is that the experimental setup is properly  
596 designed thus enabling understanding of pre-analysis workflows and results for better and  
597 future-proof data production. This involves steps from sample collection, laboratory work to  
598 eventually multiplexing the DNA libraries for sequencing, and in which vital discoveries have  
599 been made (Murray *et al.*, 2015, Schnell *et al.*, 2015, Ficetola *et al.*, 2016).

600 Several publications have addressed the issues of contamination and how to authenticate  
601 aDNA (Hebsgaard *et al.*, 2005; Gilbert *et al.*, 2005; Sawyer *et al.*, 2012; Jónsson *et al.*, 2013)  
602 and two important authentication methods are now used to make aDNA inferences highly  
603 robust: (i) replication and use of controls; and (ii) DNA damage estimates.

604

605 Replication and use of controls

606 Extensive use of controls and replicated analyses is an important strategy ensuring the quality  
607 of aDNA results. First, multiple extraction and PCR controls must be performed and  
608 sequenced to detect sporadic contaminants. Taxa that are detected at a significant rate  
609 within controls (usually from known food or exotic plants) are usually easy to recognise and  
610 must be removed from analyses (Cooper & Poinar, 2000; Ficetola *et al.*, 2016). Alternatively,  
611 if a taxon is sporadically present within controls, but is abundant in test samples, statistical  
612 tests can assess whether the detection within HTS samples is significantly higher than the  
613 contamination rate (Champlot *et al.*, 2010). The use of positive controls is a further strategy  
614 to limit false positives and false negatives. De Barba *et al.* (2014) added to their analyses  
615 positive controls made by mixing DNA of known concentration from four known plant  
616 species. After HTS, they detected in positive controls several sequences at very low  
617 frequencies, which did not belong to any of the species actually present. This result was then  
618 used to identify a frequency threshold, which allowed detection and removal of sequences

619 representing low-frequency noise, without removing species that were actually present in the  
620 sample. The use of positive controls is however risky in aDNA analyses and should be avoided  
621 when possible, as it is itself a potential source of contamination (Cooper & Poinar, 2000).  
622 Species that are exotic to the study area (e.g. tropical species in studies focusing on the  
623 boreal flora) can be suitable positive controls, as their eventual contamination of the sample  
624 can be easily spotted.

625 Reproducibility is another key criterion to ensure the quality of results (Cooper & Poinar,  
626 2000). Alsos *et al.* (2016) compared aDNA with macrofossils and concluded that all common  
627 species could be detected using one DNA extraction and one PCR per sample independently  
628 of sample age. However, increasing the number of extractions or PCR repeats increased the  
629 chances for detecting rare species. When the probability of detecting the species of interest is  
630 low due to either low biomass in the local environment or high degradation due to age or  
631 temperature, multiple PCR replicates are needed for a more complete description of  
632 communities. Simulation studies show that performing 6-12 PCR reaction from the same  
633 extract may produce robust results (Ficetola *et al.*, 2015). Unfortunately, increasing repeats  
634 can have the drawback of increasing the probability of having false positives (Ficetola *et al.*,  
635 2015), but the benefit of detecting the species generally outweighs the problems of false  
636 positives. When multiple samples are analysed multiple times (e.g. in studies analysing time  
637 series), site-occupancy detection models allow estimation of the true frequency of the  
638 species, its detection probability, and the false positive rate (Ficetola *et al.*, 2015; Lahoz  
639 Monfort *et al.*, 2016). Bayesian models can therefore allow integration of prior information  
640 and the contamination rate of a specific taxon obtained for example from the analysis of  
641 controls (Lahoz Monfort *et al.*, 2016).

642

643 Authentication of ancient origin (DNA damage)

644 In all aDNA sequences, an excess of C to T transitions is observed at the 5' and 3' end of  
645 molecules. This pattern of post-mortem damage increases over time (Sawyer *et al.*, 2012;  
646 Pedersen *et al.*, 2016), and the increased frequency of C to T transitions can be used as a tool  
647 for distinguishing aDNA sequences from modern contaminants (Briggs *et al.*, 2007). The  
648 bioinformatic tool package MapDamage2.0 (Jónsson *et al.*, 2013) provides a way to quantify  
649 the rate of DNA damage. This led Weiß *et al.* (2015) to develop a specific computational  
650 approach for comparing DNA damage patterns of putative aDNA with modern DNA and thus

651 confirming whether a sequences was of ancient origin or not. The method may be particularly  
652 useful for validation of future metagenomic studies.

653

654

## 655 **Bioinformatic processing**

656

### 657 Metabarcoding bioinformatics

658 HTS generates large DNA data sets that require dedicated programs for analyses. There are a  
659 number of bioinformatic steps to follow after DNA metabarcoding sequencing on Illumina  
660 platforms. The first three steps are assembling paired-end reads, assigning HTS reads to  
661 samples or demultiplexing (metabarcoding studies generally analyse multiple samples within  
662 each HTS run) (Coissac, 2012 and subsequent papers), and filtering erroneous sequences  
663 originated from non-specific amplification and PCR/sequencing errors. The latter task is often  
664 critical for ensuring the quality of metabarcoding data, and can be achieved by removing  
665 sequences that are too short/too long relative to the known features of the used barcodes  
666 and chimera sequences or sequences that likely represent punctual errors (e.g. nucleotide  
667 substitutions, small insertions/deletions) originating during PCR. Furthermore, sequences  
668 with just one read (singletons) or two reads in one sample can be artefacts (De Barba *et al.*,  
669 2014; Elbrecht & Leese, 2015), and should generally be filtered. If sequencing depth is high, it  
670 is even possible to find thousands of reads which do not correspond to real barcodes.  
671 Therefore treatment of rare sequences remains a major challenge of metabarcoding data  
672 analysis (Ficetola *et al.*, 2016). The fourth bioinformatic step is clustering, which merges  
673 sequences belonging to the same molecular operational taxonomic units (MOTU). The fifth  
674 and final step is taxonomic assignment of sequences to currently known taxa. This is generally  
675 done by comparing the retrieved sequences with reference databases, which can be either  
676 broad databases like GenBank (Benson *et al.*, 2013) or ENA-EMBL (Leinonen *et al.*, 2011) or  
677 high-quality reference databases containing the verified and curated sequences of potentially  
678 present species, such as the arctic plant database (Sønstebo *et al.*, 2010), or the on-going  
679 PhyloAlps and NorBol projects (Coissac *et al.*, 2016). Multiple bioinformatics tools are  
680 currently available and a more detailed description can be found in Note S1 in the supporting  
681 information.

682

683 Shotgun metabarcoding bioinformatics

684 The idea of processing shotgun metabarcoding data is to take all DNA sequences present in a  
685 sample dataset, align them against a reference database, parse the alignment information,  
686 assign a taxonomic label to each read and hereby generate the taxonomic profile of the  
687 metagenome/sample. For short-read alignment standard programmes such as *BWA* (Li &  
688 Durbin, 2009) and *Bowtie2* (Langmead & Salzberg, 2012) are often used to align reads to  
689 reference sequences, however they offer no tools to decipher between the alignments or  
690 downstream handling, and additional tools and further analyses are therefore required. Over  
691 recent years several tools for taxonomic profiling of shotgun sequence data have therefore  
692 been developed. Overall these can be categorised as tools using all available sequences  
693 (GenBank) such as *MEGAN* (Huson *et al.*, 2007), which can also be used for metabarcoding,  
694 the metagenomic *MG-RAST* server (Wilke *et al.*, 2015), *EBI-metagenomics* (Mitchell, *et al.*,  
695 2016 ) *CLARK* (Ounit *et al.*, 2015), *Holi* (Pedersen *et al.* 2016), *Kraken* (Wood & Salzberg,  
696 2014), *Kaiju* (Menzel *et al.*, 2016), and then tools using selected marker genes such as  
697 *MetaPhlan* (Segata *et al.*, 2012), *mOTU* (Sunagawa *et al.*, 2013) and *metaBit* (Louvel *et al.*,  
698 2016). Pipelines such as *MG-RAST* and *EBI-metagenomics* accept raw and untrimmed datasets  
699 and are able to parse these through piped-programmes aligning against in-house databases  
700 and eventually DNA doing sequence classification and graphical presentation. The majority of  
701 the tools however, are designed and optimized for fast and accurate alignments (e.g. *BWA*  
702 and *Bowtie2*) or for alignment and eventual DNA sequence classification (e.g. *Kraken*, *Holi*,  
703 *CLARK* and *Kaiju*) against custom-build databases. Alternatively, unique marker-gene  
704 databases are used to lower computational time and for robust abundance estimates and  
705 taxa classification (e.g. *MetaPhlan* and *mOTU*) and graphical presentation (*metaBit*).

706 While the accuracy and speed of these tools make shotgun metabarcoding dataset analysis  
707 faster and easy to perform [for more details see Lindgreen, *et al.* (2015)], a common issue is  
708 that they are designed for modern sequencing datasets and for analysis of the microbial  
709 fraction and only two of the tools (i.e. *MG-RAST* and *EBI-metagenomics*) contain plant  
710 reference sequences within their databases unless the database is custom build. However,  
711 none of these tools are designed for aDNA analysis and therefore lack the stringent and  
712 robust criteria required for taxonomic consideration of short and damaged reads. Schubert *et al.*  
713 (2012) compared different computational methods for improving accuracy and sensitivity  
714 of aDNA sequence identification and showed that using reads  $\geq 30$  bp increase the quality of  
715 alignments to modern reference genomes and lower the number of false positives. Currently,  
716 from the limited number of studies of ancient plant shotgun metabarcoding, much indicates

717 that the sequence-to-reference similarity for taxonomic consideration should be as high as  
718 100%. An in silico test, modelling shotgun metabarcode libraries inferred with sequencing  
719 errors, found that errors or nucleotide substitutions can lead to false positives in ancient  
720 metagenomic datasets (Pedersen *et al.*, 2016). However, these false-positives appear as low  
721 abundance random 'back-ground' noise, and can thus be avoided by setting a minimum  
722 number of reads as a threshold. Although the size of this threshold seems to depend on the  
723 number of reads sequenced, the exact size and nature for setting this threshold remains  
724 unknown and likely varies with the genomic composition in the sample. No clear-cut choice  
725 exists for metagenomic analysis tool but with above suggested criteria taxonomic inference  
726 will become more robust. We furthermore urge that biological interpretation is always  
727 accompanied by aDNA authentication (Jónsson *et al.*, 2013).

728 One major bioinformatic challenge lies in the choice of reference database, which affects the  
729 taxonomic profiling of a metagenome and therefore should be selected with care. Ideally, if  
730 all organismal DNA was sequenced, we could rely on alignments against all reference  
731 sequences and eventually use a lowest common ancestor algorithm to resolve reads with  
732 alignment against multiple species. Such a method, with no 'a priori' assumption about  
733 environment or species composition in a sample, is objective and would be the preferred  
734 methodology. Pedersen *et al.* (2016) employed this methodology to data from ancient lake  
735 sediments using NCBI's nucleotide database and showed that, by using stringent assignment  
736 criteria even to a non-complete database, the plant metagenomic profile at genus level was in  
737 line with taxa found by pollen, macrofossils and faunal record. However, the full effect of the  
738 choice of database still remains unexplored and future studies are needed to investigate the  
739 potential consequences of this.

740 Finally, it is important to emphasize that due to the nature of an environmental DNA sample,  
741 the majority of the sequenced DNA obtained using a shotgun metabarcoding approach  
742 cannot be usually identified with the databases currently available (Pedersen *et al.*, 2016).  
743 Often, more than 90 % of the reads produced cannot be aligned to a reference, and in most  
744 cases less than 2% of the reads are unique to any taxa. However, as more genomes become  
745 available these proportions will improve.

746

747

748 **Conclusions and directions for future research**

749



750 Despite early challenges, the field of aDNA has lately experienced a massive improvement  
751 methodologically, technologically, and in the understanding of the underlying processes by  
752 which environmental DNA information is transferred and transformed in lake sediments. This  
753 has resulted in new and better use of the technique in lakes and a refined understanding of  
754 important long-standing palaeoecological issues. We expect that the applicability of both  
755 shotgun and conventional metabarcode analysis of *sed*aDNA and SCS on pollen will continue  
756 to improve in the coming years, as HTS methods become more refined, less expensive, and  
757 the genomic reference databases improves. In Fig. 2 we show our interpretation of how the  
758 biotic palaeoenvironmental proxies in lake sediments originate, accumulate and develop  
759 through time. In Fig.7 we show our understanding of the chain of processes involved in the  
760 transformation of pollen, macrofossils and aDNA. Below we present a number of conclusions  
761 drawn from this review and on papers from Table 1 that we hope will be useful for plant  
762 aDNA researchers.

763 1. With the methodological, technological and experience improvements achieved over the  
764 past decade, plant DNA from lake sediments has now become an established tool for  
765 analysing past vegetation in combination with classical palaeoecological analyses. At the same  
766 time, as it provides a local proxy, it will play a key role for identifying 'fossil silent diversity'  
767 useful for understanding past vegetation change and for modelling vegetation response to  
768 future climate changes.

769 2. Lake sediments will continue to provide continuous archives with a fine temporal and  
770 spatial resolution, allowing establishment of good molecular records for past vegetation  
771 history and the possibility for distinguishing origin, dispersal and ancestry of plant species and  
772 populations through time.

773 3. Ancient plant DNA from lake sediments will in time be more precise at determining local  
774 vegetation relative to macrofossil and pollen analysis.

775 4. Improved understanding of DNA taphonomy from lake sediments now allow a better  
776 understanding of the origin and fate of plant aDNA molecules during and after deposition in  
777 lakes. Further understanding of these processes is crucial, particularly those involved in DNA  
778 preservation (temperature, pH, adsorption onto mineral surfaces, and oxygen availability) for  
779 improve determination of the power and limitations of the new tools presented in this  
780 review.

781 5. SCS profiling of pollen from lake sediments will likely grow in the coming years and become  
782 an important tool for investigating histories and dynamics of plants at the population level.

783

784

785 **Acknowledgements**

786 The authors thank Sebastian Sobek for constructive comments on an earlier version of this  
787 manuscript and Pierre Taberlet for a good discussion about the ancient DNA terminology. The  
788 work was supported by the Swedish Research Council (grant no. 2013-D0568401), SciLifeLab  
789 Stockholm and the Carl Triggers' Foundation (grant no. 14:371) to LP and the Research  
790 Council of Norway to IGA (grant no. 213692/F20).

791

792

793 **References**

794

795 **Allen JRM, Huntley B. 1999.** Estimating past floristic diversity in montane regions from  
796 macrofossil assemblages. *Journal of Biogeography* **26**: 55–73.

797 **Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA,**  
798 **Gilbert MTP, Willerslev E, et al. 2012.** The half-life of DNA in bone: measuring decay kinetics in  
799 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* **279**: 4724–4733.

800 **Alsos IG, Sjögren P, Edwards ME, Landvik JY, Gielly L, Forwick M, Coissac E, Brown AG, Jakobsen**  
801 **LV, Føreid MK, et al. 2016.** Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing  
802 the resilience of arctic flora to Holocene climate change. *The Holocene* **26**: 1–16.

803 **Alsos IG, Coissac E, Edwards ME, Merkel M, Gielly L. 2015.** Plant DNA in sediments: to which  
804 degree do they represent the flora? *Genome* **58**: 163–303.

805 **Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjær KH, Orlando L, Gilbert**  
806 **MTP, Willerslev E. 2011.** Meta-barcoding of ‘dirt’ DNA from soil reflects vertebrate  
807 biodiversity. *Molecular Ecology* **21**: 1966–1979.

808 **Anderson-Carpenter L, Mclachlan J, Jackson S, Kuch M, Lumibao C, Poinar H. 2011.** Ancient  
809 DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC*  
810 *Evolutionary Biology* **11**: 1–15.

811 **Baamrane MAA, Shehzad W, Ouhammou A, Abbad A, Naimi M, Coissac E, Taberlet P, Znari M.**  
812 **2012.** Assessment of the Food Habits of the Moroccan Dorcas Gazelle in M’Sabih Talaa, West  
813 Central Morocco, Using the *trnL* Approach (LAN Amaral, Ed.). *PLOS ONE* **7**: e35643.

814 **Barnes MA, Turner CR. 2016.** The ecology of environmental DNA and implications for  
815 conservation genetics. *Conservation Genetics* **17**: 1–17.

816 **Belle S, Parent C, Frossard V, Verneaux V, Millet L, Chronopoulou P-M, Sabatier P, Magny M.**  
817 **2014.** Temporal changes in the contribution of methane-oxidizing bacteria to the biomass of  
818 chironomid larvae determined using stable carbon isotopes and ancient DNA. *Journal of*  
819 *Paleolimnology* **52**: 215–228.

820 **Bennett KD. 2015.** Comment on ‘Sedimentary DNA from a submerged site reveals wheat in  
821 the British Isles 8000 years ago’. *Science* **349**: 247–247.

822 **Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013.**  
823 GenBank. *Nucleic Acids Research* **41**: D36–42.

824 **Birks HH. 2003.** The importance of plant macrofossils in the reconstruction of Lateglacial  
825 vegetation and climate: examples from Scotland, western Norway, and Minnesota, USA.  
826 *Quaternary Science Reviews* **22**: 453–473.

827 **Birks HH. 2013.** Plant macrofossil introduction. Mock HP, Elias SA eds. *Encyclopedia of*  
828 *Quaternary Science*. Amsterdam: Elsevier, 593–612.

829 **Birks HJB. 2014.** Challenges in the presentation and analysis of plant-macrofossil  
830 stratigraphical data. *Vegetation History and Archaeobotany* **23**: 309–330.

- 831 Birks HJB, Birks HH. 2016. How have studies of ancient DNA from sediments contributed to  
832 the reconstruction of Quaternary floras? *The New Phytologist* 209: 499–506.
- 833 Bissett A, Gibson J, Jarman SN. 2005. Isolation, amplification, and identification of ancient  
834 copepod DNA from lake sediments. *Limnology and Oceanography: Methods* 3: 533–542.
- 835 Boessenkool S, MCGlynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, Memomissa S, Brochmann  
836 C, Popp M. 2013. Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-  
837 Altitude Tropical Biodiversity. *Conservation Biology* 28: 446–455.
- 838 Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT,  
839 Lachmann M, *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neandertal.  
840 *Proceedings of the National Academy of Sciences USA* 104: 14616–14621.
- 841 Brouzes E, Medkova M, Savenelli N, Marran D, Twardowski M, Hutchison JB, Rothberg JM, Link  
842 DR, Perrimon N, Samuels ML. 2009. Droplet microfluidic technology for single-cell high-  
843 throughput screening. *Proceedings of the National Academy of Sciences USA* 106: 14195–  
844 14200.
- 845 Brown TA, Cappellini E, Kistler L, Lister DL, Oliveira HR, Wales N, Schlumbaum A. 2014. Recent  
846 advances in ancient DNA research and their implications for archaeobotany. *Vegetation*  
847 *History and Archaeobotany* 24: 207–214.
- 848 Capo E, Debroas D, Arnaud F, Domaizon I. 2015. Is Planktonic Diversity Well Recorded in  
849 Sedimentary DNA? Toward the Reconstruction of Past Protistan Diversity. *Microbial Ecology*  
850 70: 865–875.
- 851 Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl E-M. 2010. An efficient  
852 multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR  
853 applications. *PLOS ONE* 5: e13042.
- 854 Chan YL, Lacey EA, Pearson OP, HADLY EA. 2005. Ancient DNA reveals Holocene loss of genetic  
855 diversity in a South American rodent. *Biology Letters* 272: 423–426.
- 856 Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, Lynch J, Pacala S, Prentice  
857 C, *et al.* 1998. Reid's Paradox of Rapid Plant Migration Dispersal theory and interpretation of  
858 paleoecological records. *BioScience* 48: 13–24.
- 859 Coissac E. 2012. OligoTag: a program for designing sets of tags for next-generation sequencing  
860 of multiplexed samples. Pompanon F, Bonin A eds. *Methods in Molecular Biology. Data*  
861 *Production and Analysis in Population Genomics*. New York, Dordrecht, Heidelberg, London,  
862 13–31.
- 863 Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. 2016. From barcodes to genomes:  
864 extending the concept of DNA barcoding. *Molecular Ecology* 25: 1423–1428.
- 865 Coolen M, Overmann J. 1998. Analysis of subfossil molecular remains of purple sulfur bacteria  
866 in a lake sediment. *Applied and Environmental Microbiology* 64: 4513–4521.
- 867 Coolen M, Muyzer G, Rijpstra W, Schouten S, Volkman JK, Sinninghe Damsté JS. 2004.  
868 Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and  
869 diatom populations in an Antarctic lake. *Earth and Planetary Science Letters* 223: 225–239.

- 870 Coolen M, Talbot HM, Abbas BA, Ward C, Schouten S, Volkman JK, Damsté JSS. 2008. Sources  
871 for sedimentary bacteriohopanepolyols as revealed by 16S rDNA stratigraphy. *Environmental*  
872 *Microbiology* 10: 1783–1803.
- 873 Coolen M, Gibson J. 2009. Ancient DNA in lake sediment records. *PAGES news* 17, 104-106.
- 874 Cooper A, Poinar HN. 2000. Ancient DNA: Do it Right or Not at All. *Science* 289: 1139–1141.
- 875 Dabney J, Meyer M, Pääbo S. 2013. Ancient DNA Damage. Cold Spring Harbor Perspectives in  
876 Biology: doi: 10.1101–cshperspect.a012567.
- 877 da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe  
878 C, Ávila-Arcos MAC, Hufnagel DE, et al. 2015. The origin and evolution of maize in the  
879 Southwestern United States. *Nature Plants* 1: 14003–14008.
- 880 D'Andrea WJ, Lage M, Martiny J. 2006. Alkenone producers inferred from well-preserved 18S  
881 rDNA in Greenland lake sediments. *Journal of Geographic Research* 111: G03013.
- 882 De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P. 2014. DNA  
883 metabarcoding multiplexing and validation of data accuracy for diet assessment: application  
884 to omnivorous diet. *Molecular Ecology Resources* 14: 306–323.
- 885 Domaizon I, Savichtcheva O, Debroas D, Arnaud F, Villar C, Pignol C, Alric B, Perga ME. 2013.  
886 DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus*  
887 assemblages. *Biogeosciences* 10: 3817–3838.
- 888 Dong J, Gao K, Wang K, Xu X, Zhang H. 2015. Cell Wall Disruption of Rape Bee Pollen Treated  
889 with Combination of Protamex Hydrolysis and Ultrasonication. *Food Research International*  
890 75: 123–130.
- 891 Elbrecht V, Leese F. 2015. Can DNA-Based Ecosystem Assessments Quantify Species  
892 Abundance? Testing Primer Bias and Biomass - Sequence Relationships with an Innovative  
893 Metabarcoding Protocol. *PLOS ONE* 10: e0130324.
- 894 Epp LS, Stoof-Leichsenring KR, Trauth MH, Tiedemann R. 2010. Historical genetics on a  
895 sediment core from a Kenyan lake: intraspecific genotype turnover in a tropical rotifer is  
896 related to past environmental changes. *Journal of Paleolimnology* 43: 939–954.
- 897 Epp LS, Stoof-Leichsenring KR, Trauth MH, Tiedelman R. 2011. Molecular profiling of diatom  
898 assemblages in tropical lake sediments using taxon-specific PCR and Denaturing High-  
899 Performance Liquid Chromatography (PCR-DHPLC). *Molecular Ecology Resources* 11: 842–  
900 853.
- 901 Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, Schrøder-Nielsen A, Ludikova A, Hassel K,  
902 Stenøien HK, Funder S, et al. 2015. Lake sediment multi-taxon DNA from North Greenland  
903 records early post-glacial appearance of vascular plants and accurately tracks environmental  
904 changes. *Quaternary Science Reviews* 117: 152–163.
- 905 Faegri K, Iversen J, Kaland PE, Krzywinski K. 1989. Textbook of Pollen Analysis. 4th edition.  
906 Wiley, Chichester.
- 907 Etienne D, Destas M, Lyautey E, Marti R, Ruffaldi P, Georges-Leroy M, Dambrine E, Topp E.  
908 2015. Two thousand-year reconstruction of livestock production intensity in France using

- 909 sediment-archived fecal Bacteroidales and source-specific mitochondrial markers. *The*  
910 *Holocene* **25**: 1384–1393.
- 911 Feek DT, Flenley JR, Chester PI, Welikala N, Matisoo-Smith EA, Tanno GW. 2006. A modified  
912 sampler for uncontaminated DNA cores from soft sediments. *Journal of Archaeological*  
913 *Science* **33**: 573–574.
- 914 Feek DT, Horrocks M, Baisden WT, Flenley J. 2011. The Mk II sampler: a device to collect  
915 sediment cores for analysis of uncontaminated DNA. *Journal of Paleolimnology* **45**: 115–119.
- 916 Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P, Pompanon F. 2010.  
917 An in silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**: 434–1572.
- 918 Ficetola GF, Pansu J, Bonin A, Coissac E, Giguët-Covex C, De Barba M, Gielly L, Lopes CM, Boyer  
919 F, Pompanon F, *et al.* 2015. Replication levels, false presences and the estimation of the  
920 presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* **15**: 543–  
921 556.
- 922 Ficetola GF, Taberlet P, Coissac E. 2016. How to limit false positives in environmental DNA and  
923 metabarcoding? *Molecular Ecology Resources* **16**: 604–607.
- 924 Fulton TL. 2012. Setting Up an Ancient DNA Laboratory. Shapiro B, Hofreiter M eds. Methods  
925 in Molecular Biology. Ancient DNA: methods and protocols. New York, Dordrecht, Heidelberg,  
926 London: Springer, 1–12.
- 927 Furlan EM, Gleeson D, Hardy CM, Duncan RP. 2016. A framework for estimating the sensitivity  
928 of eDNA surveys. *Molecular Ecology Resources* **16**: 641–654.
- 929 Giguët-Covex C, Pansu J, Arnaud F, Rey P-J, Griggo C, Gielly L, Domaizon I, Coissac E, David F,  
930 Choler P, *et al.* 2014. Long livestock farming history and human landscape shaping revealed by  
931 lake sediment DNA. *Nature Communications* **5**: 3211.
- 932 Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. 2005. Assessing ancient DNA studies.  
933 *Theoretical and Applied Genetics* **20**: 541–544.
- 934 Haile J, Holdaway R, Oliver K, Bunce M, Gilbert MTP, Nielsen R, Munch K, Ho SYW, Shapiro B,  
935 Willerslev E. 2007. Ancient DNA Chronology within Sediment Deposits: Are Paleobiological  
936 Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution* **24**:  
937 982–989.
- 938 Hasegawa Y, Suyama Y, Seiwa K. 2009. Pollen donor composition during the early phases of  
939 reproduction revealed by DNA genotyping of pollen grains and seeds of *Castanea crenata*.  
940 *New Phytologist* **182**: 994–1002.
- 941 Hasegawa Y, Suyama Y, Seiwa K. 2015. Variation in pollen-donor composition among  
942 pollinators in an entomophilous tree species, *Castanea crenata*, revealed by single-pollen  
943 genotyping. *PLOS ONE* **10**: e0120393.
- 944 Hebsgaard MB, Phillips MJ, Willerslev E. 2005. Geologically ancient DNA: fact or artefact?  
945 *Trends in Microbiology* **13**: 212–220.
- 946 Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

- 947 Higuchi R, Bowman B, Freiberger M, Ryder O, Wilson A. 1984. DNA sequences from the  
948 quagga, an extinct member of the horse family. *Nature* **312**: 282–284.
- 949 Hirota SK, Nitta K, Suyama Y, Kawakubo N, Yasumoto AA, Yahara T. 2013. Pollinator-mediated  
950 selection on flower color, flower scent and flower morphology of *Hemerocallis*: evidence  
951 from genotyping individual pollen grains on the stigma. *PLOS ONE* **8**: e85601.
- 952 Hofreiter M, Mead JI, Martin P, Poinar H. 2003. Molecular caving. *Current Biology* **13**: R693–  
953 R695.
- 954 Hofreiter M, Serre D, Poinar H, Kuch M, Pääbo S. 2001. Ancient DNA. *Nature Reviews Genetics*  
955 **2**: 353–359.
- 956 Hou W, Dong H, Li G, Yang J, Coolen MJL, Liu X, Wang S, Jiang H, Wu X, Xiao H, et al. 2014.  
957 Identification of Photosynthetic Plankton Communities Using Sedimentary Ancient DNA and  
958 Their Response to late-Holocene Climate Change on the Tibetan Plateau. *Scientific Reports* **4**:  
959 6648–9.
- 960 Hu FS, Hampe A, Petit RJ. 2009. Paleoecology meets genetics: deciphering past vegetational  
961 dynamics. *Frontiers in Ecology and the Environment* **7**: 371–379.
- 962 Huang YT. 2014. Studies on carbon and DNA preservation in allophanic soils and paleosols on  
963 Holocene tephras in New Zealand. PhD thesis, University of Waikato, NZ.
- 964 Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome*  
965 *Research* **17**: 377–386.
- 966 Hutchinson, G. E. (1957). A Treatise on Limnology, Vol. 1. Geography, Physics and Chemistry.  
967 Wiley, New York.
- 968 Isagi Y, Suyama Y. 2010. Single-Pollen Genotyping (Y Isagi and Y Suyama, Eds.). Tokyo: Springer  
969 Science & Business Media.
- 970 Jackson ST. 2012. Representation of flora and vegetation in Quaternary fossil assemblages:  
971 known and unknown knowns and unknowns. *Quaternary Science Reviews* **49**: 1–15.
- 972 Jacobson GL, Bradshaw RHW. 1981. The selection of sites for paleovegetational studies.  
973 *Quaternary Research* **16**: 80–96.
- 974 Jiang H, Dong H, Yu B, Liu X, Li Y, Ji S. 2007. Microbial response to salinity change in Lake  
975 Chaka, a hypersaline lake on Tibetan plateau. *Environmental Microbiology* **9**: 2603–2621
- 976 Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast  
977 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**:  
978 1682–1684.
- 979 Jørgensen T, Haile J, Møller P, Andreev V A, Boessenkool S, Rasmussen M, Kienast F, Coissac E,  
980 Taberlet P, Brochmann C, et al. 2012. A comparative study of ancient sedimentary DNA, pollen  
981 and macrofossils from permafrost sediments of northern Siberia reveals long-term  
982 vegetational stability. *Molecular Ecology* **21**: 1989–2003.
- 983 Jørgensen T, Kjær KH, Haile J, Rasmussen M, Boessenkool S, Andersen K, Coissac E, Taberlet P,  
984 Brochmann C, Orlando L, et al. 2011. Islands in the ice: detecting past vegetation on

- 985 Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding.  
986 *Molecular Ecology* **21**: 1980–1988.
- 987 **Kyle M, Haande S, Sønstebø JH, Rohrlack T. 2015.** Amplification of DNA in sediment cores to  
988 detect historic *Planktothrix* occurrence in three Norwegian lakes. *Journal of Paleolimnology*  
989 **53**: 61–72.
- 990 Kong F, Yuan L, Zheng YF, Chen W. 2012. Automatic liquid handling for life science: a critical  
991 review of the current state of the art. *Journal of Laboratory Automation* **17**: 169–185.
- 992 **Kojima H, Iwata T, Fikui M. 2009.** DNA-based analysis of planktonic methanotrophs in a  
993 stratified lake. *Freshwater Biology* **54**: 1501–1509.
- 994 **Lahoz Monfort JJ, Guillera Arroita G, Tingley R. 2016.** Statistical approaches to account for  
995 false-positive errors in environmental DNA samples. *Molecular Ecology Resources* **16**: 673–  
996 685.
- 997 **Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:  
998 357–359.
- 999 **Larsen CPS, Macdonald GM. 1993.** Lake morphometry, sediment mixing and the selection of  
1000 sites for fine resolution palaeoecological studies. *Quaternary Science Reviews* **12**: 781–792.
- 1001 **Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N,**  
1002 **Goodgame N, Gibson R, et al. 2011.** The European Nucleotide Archive. *Nucleic Acids Research*  
1003 **39**: D28–31.
- 1004 **Li H, Durbin R. 2009.** Fast and accurate short read alignment with Burrows-Wheeler  
1005 transform. *Bioinformatics* **25**: 1754–1760.
- 1006 **Li G, Dong H, Hou W, Wang S, Jiang H, Yang J, Wu G. 2016.** Temporal Succession of Ancient  
1007 Phytoplankton Community in Qinghai Lake and Implication for Paleo-environmental Change.  
1008 *Scientific Reports* **6**: 19769.
- 1009 **Lindgreen S, Adair KL, Gardner PP. 2016.** An evaluation of the accuracy and speed of  
1010 metagenome analysis tools. *Scientific Reports* **6**: 19233.
- 1011 Lindhal T. 1993. Instability and decay of the primary structure of DNA. **362**: 709–715.
- 1012 **Limburg PA, Weider LJ. 2002.** ‘Ancient’ DNA in the resting egg bank of a microcrustacean can  
1013 serve as a palaeolimnological database. *Proceedings of the Royal Society B: Biological Sciences*  
1014 **269**: 281–287.
- 1015 Little DP. 2014. A DNA mini-barcode for land plants. *Molecular Ecology Resources* **14**: 437–  
1016 446.
- 1017 **Louvel G, Sarkissian Der C, Hanghøj K, Orlando L. 2016.** metaBIT, an integrative and automated  
1018 metagenomic pipeline for analysing microbial profiles from high-throughput sequencing  
1019 shotgun data. *Molecular Ecology Resources* **16**: 1415–1427
- 1020 **Lydolph MC, Jacobsen J, Arctander P, Gilbert MTP, Gilichinsky DA, Hansen AJ, Willerslev E,**  
1021 **Lange L. 2005.** Beringian Paleoecology Inferred from Permafrost-Preserved Fungal DNA.  
1022 *Applied and Environmental Microbiology* **71**: 1012–1017.



- 1023 Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, Rubin EM,  
1024 Jansson JK. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid  
1025 response to thaw. *Nature* 480: 368–371.
- 1026 **Madeja J, Wacnik A, Zyga A, Stankiewicz E, Wypasek E, Guminski W, Harmata K. 2009. Bacterial**  
1027 **ancient DNA as an indicator of human presence in the past: its correlation with palynological**  
1028 **and archaeological data. *Journal of Quaternary Science* 24: 317–321.**
- 1029 **Madeja J, Wacnik A, Wypasek E, Chandran A, Stankiewicz E. 2010. Integrated palynological and**  
1030 **molecular analyses of late Holocene deposits from Lake Miłkowskie (NE Poland): Verification**  
1031 **of local human impact on environment. *Quaternary International* 220: 147–152.**
- 1032 **Madeja J. 2015. A new tool to trace past human presence from lake sediments: the human-**  
1033 **specific molecular marker *Bacteroides* strain HF 183. *Journal of Quaternary Science* 30: 349–**  
1034 **354.**
- 1035 **Magyari EK, Major A, Balint M, Nedli J, Braun M, Racz I, Parducci L. 2011. Population dynamics**  
1036 **and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient**  
1037 **DNA analyses. *BMC Evolutionary Biology* 11: 66.**
- 1038 **Marková S, Černý M, Rees D, Stuchlík E. 2006. Are they still viable? Physical conditions and**  
1039 **abundance of *Daphnia pulicaria* resting eggs in sediment cores from lakes in the Tatra**  
1040 **Mountains. *Biologia* 61: 135–146.**
- 1041 Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A,  
1042 **David M, Reiter E, Riehl S, et al. 2016. Genomic analysis of 6,000-year-old cultivated grain**  
1043 **illuminates the domestication history of barley. *Nature Genetics* 48: 1089–1093.**
- 1044 **Matisoo-Smith E, Roberts K, Welikala N, Tannock G, Chester P, Feek D, Flenley J. 2008.**  
1045 **Recovery of DNA and pollen from New Zealand lake sediments. *Quaternary International* 184:**  
1046 **139–149.**
- 1047 **Matsuki Y, Isagi Y, Suyama Y. 2007. The determination of multiple microsatellite genotypes**  
1048 **and DNA sequences from a single pollen grain. *Molecular Ecology Notes* 7: 194–198.**
- 1049 **Matsuki Y, Tateno R, Shibata M, Isagi Y. 2008. Pollination efficiencies of flower-visiting insects**  
1050 **as determined by direct genetic analysis of pollen origin. *American Journal of Botany* 95: 925–**  
1051 **930.**
- 1052 **Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics**  
1053 **with Kaiju. *Nature Communications* 7: 11257.**
- 1054 **Mergeay J, Vanoverbeke J, Verschuren D, De Meester L. 2007. Extinction, recolonization, and**  
1055 **dispersal through time in a planktonic crustacean. *Ecology* 88: 3032–3043.**
- 1056 Mitchell A, Bucchini F, Cochrane G, Denise H, Hoopen Ten P, Fraser M, Pesseat S, Potter S,  
1057 **Scheremetjew M, Sterk P, et al. 2016. EBI metagenomics in 2016--an expanding and evolving**  
1058 **resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research* 44:**  
1059 **D595–603.**
- 1060 **Murray DC, Coghlan ML & Bunce M, 2015. From Benchtop to Desktop: Important**  
1061 **Considerations when Designing Amplicon Sequencing Workflows. *PLoS ONE*, 10, p.e0124671.**

- 1062 Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T. 1999. The selective increase or decrease of  
 1063 organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic  
 1064 inheritance. *Planta* **209**: 53–65.
- 1065 Nakazawa F, Uetake J, Suyama Y, Kaneko R, Takeuchi N, Fujita K, Motoyama H, Imura S, Kanda H.  
 1066 2013. DNA analysis for section identification of individual *Pinus* pollen grains from Belukha glacier,  
 1067 Altai Mountains, Russia. *Environmental Research Letters* **8**: 014032.
- 1068 Nesje A. 1992. A Piston Corer for Lacustrine and Marine Sediments. *Arctic and Alpine Research*  
 1069 **24**: 257.
- 1070 Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes.  
 1071 *Nature Reviews Genetics* **16**: 395–408.
- 1072 Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E,  
 1073 Petersen B, Moltke I, *et al.* 2014. Recalibrating *Equus* evolution using the genomes sequence of  
 1074 an early Middle Pleistocene horse. *Nature* **498**: 74–78.
- 1075 Ounit R, Wanamaker S, Close TJ, Lonardi S. 2015. CLARK: fast and accurate classification of  
 1076 metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* **16**: 236.
- 1077 Paffetti D, Vettori C, Caramelli D, Vernesi C, Lari M, Paganelli A, Paule L, Giannini R. 2007.  
 1078 Unexpected presence of *Fagus orientalis* complex in Italy as inferred from 45,000-year-old  
 1079 DNA pollen samples from Venice lagoon. *BMC Evolutionary Biology* **7**: S6.
- 1080 Pal S, Gregory-Eaves I, Pick FR. 2015. Temporal trends in cyanobacteria revealed through DNA  
 1081 and pigment analyses of temperate lake sediment cores. *Journal of Paleolimnology* **54**: 87–  
 1082 101.
- 1083 Pansu J, Giguët-Covex C, Ficetola GF, Gielly L, Boyer F, Zinger L, Arnaud F, Poulenard J, Taberlet  
 1084 P, Choler P. 2015. Reconstructing long-term human impacts on plant communities: an  
 1085 ecological approach based on lake sediment DNA. *Molecular Ecology* **24**: 1485–1498.
- 1086 Parducci L, Suyama Y. 2011. Single-Pollen Genotyping of Holocene Lake Sediments. Isagi Y,  
 1087 Suyama Y eds. Ecological Research Monographs. Single-Pollen Genotyping. Tokyo: Springer  
 1088 Japan, 101–109.
- 1089 Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J,  
 1090 Matetovici I, Suyama Y, *et al.* 2012. Glacial survival of boreal trees in northern Scandinavia.  
 1091 *Science* **335**: 1083–1086.
- 1092 Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, Kjær KH, Larsen NK, Drouzas  
 1093 AD, Willerslev E. 2013. Molecular- and pollen-based vegetation analysis in lake sediments  
 1094 from central Scandinavia. *Molecular Ecology* **22**: 3511–3524.
- 1095 Parducci L, Suyama Y, Lascoux M, Bennett KD. 2005. Ancient DNA from pollen: a genetic record  
 1096 of population history in Scots pine. *Molecular Ecology* **14**: 2873–2882.
- 1097 Parducci L, Väliiranta M, Salonen JS, Ronkainen T, Matetovici I, Fontana SL, Eskola T, Sarala P,  
 1098 Suyama Y. 2015. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant  
 1099 DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20130382.
- 1100 Poulain AJ, Aris-Brosou SEP, Blais JM, Brazeau M, Keller WB, Paterson AM. 2015. Microbial DNA  
 1101 records historical delivery of anthropogenic mercury. *The ISME Journal* **9**: 2541–2550.

- 1102 Paus A, Boessenkool S, Brochmann C, Epp LS, Fabel D, Hafliðason H, Linge H. 2015. Lake Store  
1103 Finnsjøen – a key for understanding Lateglacial/early Holocene vegetation and ice sheet  
1104 dynamics in the central Scandes Mountains. *Quaternary Science Reviews* **121**: 36–51.
- 1105 Pääbo S. 1984. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* **314**: 644–645.
- 1106 Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, Holm J, Funder S, Willerslev E, Kjær  
1107 KH. 2013. A comparative study of ancient environmental DNA to pollen and microfossils from  
1108 lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science  
1109 Review* **75**: 161–168.
- 1110 Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, Spens J,  
1111 Thomsen PF, Bohmann K, Cappellini E, *et al.* 2015. Ancient and modern environmental DNA.  
1112 *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20130383.
- 1113 Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, Mendoza MLZ, Beaudoin  
1114 AB, Zutter C, Larsen NK, *et al.* 2016. Postglacial viability and colonization in North America's  
1115 ice-free corridor. *Nature* **537**: 45–49.
- 1116 Petersen G, Johansen B, Sedeberg O. 1996. PCR and sequencing form a sing pollen grain.  
1117 *Plant Molecular Biology* **31**: 189–191.
- 1118 Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P. 2009. Extracellular  
1119 DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils* **45**: 219–  
1120 235.
- 1121 Pollard HG, Colbourne JK, Keller W. 2003. Reconstruction of centuries-old Daphnia  
1122 communities in a lake recovering from acidification and metal contamination. *AMBIO: A  
1123 Journal of the Human Environment* **32**: 214–218.
- 1124 Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, Pellizzari C, Iribar A,  
1125 Etienne R, Taberlet P, *et al.* 2016. Using metabarcoding to reveal and quantify plant-pollinator  
1126 interactions. *Scientific Reports* **6**: 27282.
- 1127 Porter TM, Golding GB, King C, Froese D, Zazula G, Poinar HN. 2013. Amplicon pyrosequencing  
1128 late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale  
1129 reproducibility. *Molecular Ecology Resources* **13**: 798–810.
- 1130 Poté J, Ackermann R, Wildi W. 2009. Plant leaf mass loss and DNA release in freshwater  
1131 sediments. *Ecotoxicology and Environmental Safety* **72**: 1378–1383.
- 1132 Poté J, Rosselli W, Wigger A, Wildi W. 2007. Release and leaching of plant DNA in unsaturated  
1133 soil column. *Ecotoxicology and Environmental Safety* **68**: 293–298.
- 1134 Randlett M-È, Coolen M, Stockhecke M, Pickarski N, Litt T, Balkema C, Kwiecien O, Tomonaga Y,  
1135 Wehrli B, Schubert CJ. 2014. Alkenone distribution in Lake Van sediment over the last 270 ka:  
1136 influence of temperature and haptophyte species composition. *Quaternary Science Reviews*  
1137 **104**: 53–62.
- 1138 Ravasi DF, Peduzzi S, Guidi V, Peduzzi R, Wirth SB, Gilli A, Tonolla M. 2012. Development of a  
1139 real-time PCR method for the detection of fossil 16S rDNA fragments of phototrophic sulfur  
1140 bacteria in the sediments of Lake Cadagno. *Geobiology* **10**: 196–204.

- 1141 Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, Huang Y-T, Cooper A. 2014. Using  
1142 palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal*  
1143 *of Quaternary Science* **29**: 610–626.
- 1144 Renberg I, Hansson H. 2008. The HTH sediment corer. *Journal of Paleolimnology* **40**: 655–659.
- 1145 Reid VA, Carvalho GR, George DG, Griffiths HI. 2002. A technique for the molecular genetic  
1146 analysis of *Daphnia* resting eggs from sub-recent lake sediments. *Journal of Paleolimnology*  
1147 **27**: 481–486.
- 1148 Roberts AV. 2007. The use of bead beating to prepare suspensions of nuclei for flow  
1149 cytometry from fresh leaves, herbarium leaves, petals and pollen. *Cytometry Part A* **71**: 1039–  
1150 1044.
- 1151 Savichtcheva O, Debroas D, Kurmayer R, Villar C, Jenny JP, Arnaud F, Perga ME, Domaizon I.  
1152 2011. Quantitative PCR Enumeration of Total/Toxic *Planktothrix rubescens* and Total  
1153 Cyanobacteria in Preserved DNA Isolated from Lake Sediments. *Applied and Environmental*  
1154 *Microbiology* **77**: 8744–8753.
- 1155 Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal Patterns of  
1156 Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS ONE* **7**: e34131.
- 1157 Schnell I B, Bohmann K, Gilbert M T P. 2015. Tag jumps illuminated - reducing sequence-to-  
1158 sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* **15**: 1289–  
1159 1303.
- 1160  
1161 Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, Willerslev E, Krogh A,  
1162 Orlando L. 2012. Improving ancient DNA read mapping against modern reference genomes.  
1163 *BMC Genomics* **13**: 178.
- 1164 Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012.  
1165 Metagenomic microbial community profiling using unique clade-specific marker genes.  
1166 *Nature Methods* **9**: 811–814.
- 1167 Shapiro B, Hofreiter M. 2012. Ancient DNA: methods and protocols (B Shapiro and M  
1168 Hofreiter, Eds.). New York Dordrecht Heidelberg London: Springer.
- 1169 Sharon I, Banfield JF. 2013. Microbiology. Genomes from metagenomics. *Science* **342**: 1057–  
1170 1058.
- 1171 Sjögren P, Edwards ME, Gielly L, Langdom CT, Croudace IW, Merkel MKF, Fonville T, Alsos IG.  
1172 2016. Lake sedimentary DNA accurately records 20th century introductions of exotic conifers  
1173 in Scotland. *New Phytologist*: doi: 10.1111/nph.14199
- 1174 Sobek S, Durisch-Kaiser E, Zurbrügg R. 2009. Organic carbon burial efficiency in lake sediments  
1175 controlled by oxygen exposure time and sediment source. *Limnology and Oceanography* **54**:  
1176 2243–2254.
- 1177 Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH,  
1178 Ims RA, Yoccoz NG, *et al.* 2009. Analysing diet of small herbivores: the efficiency of DNA  
1179 barcoding coupled with high-throughput pyrosequencing for deciphering the composition of  
1180 complex plant mixtures. *Frontiers in Zoology* **6**: 16.

- 1181 Southworth D. 1974. Solubility of pollen exines. *American Journal of Botany* 61: 36–44.
- 1182 **Stager JC, Sporn LA, Johnson M, Regalado S. 2015.** Of Paleo-Genes and Perch: What if an  
1183 'Alien' Is Actually a Native? *PLOS ONE* 10 (3): e0119071
- 1184 **Stewart JR, Lister AM, Barnes I, Dalén L. 2010.** Refugia revisited: individualistic responses of  
1185 species in space and time. *Proceedings of the Royal Society B: Biological Sciences* 277: 661–  
1186 671.
- 1187 **Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedelman R. 2012.** Hidden diversity in diatoms of  
1188 Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular Ecology* 21:  
1189 1918–1930.
- 1190 **Stoof-Leichsenring KR, Herzschuh U, Pestryakova LA, Klemm J, Epp LS, Tiedelman R. 2015.**  
1191 Genetic data from algae sedimentary DNA reflect the influence of environment over  
1192 geography. *Scientific Reports* 5: 12924.
- 1193 **Stull GW, Moore MJ, Mandala VS, Douglas NA, Kates H-R, Qi X, Brockington SF, Soltis PS, Soltis  
1194 DE, Gitzendanner MA. 2013.** A targeted enrichment strategy for massively parallel sequencing  
1195 of angiosperm plastid genomes. *Applications in Plant Sciences* 1: 1200497.
- 1196 **Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, Coelho LP,  
1197 Arumugam M, Tap J, Nielsen HB, et al. 2013.** Metagenomic species profiling using universal  
1198 phylogenetic marker genes. *Nature Methods* 10: 1196–1199.
- 1199 **Suyama Y, Matsuki Y. 2015.** MIG-seq: an effective PCR-based method for genome-wide single-  
1200 nucleotide polymorphism genotyping using the next-generation sequencing platform.  
1201 *Scientific Reports* 5: 16963.
- 1202 **Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996.** DNA sequence  
1203 from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes & Genetic Systems* 71: 145–  
1204 149.
- 1205 **Sønstebo JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J, Willerslev E, Coissac E, Rioux D,  
1206 Sannier J, et al. 2010.** Using next-generation sequencing for molecular reconstruction of past  
1207 Arctic vegetation and climate. *Molecular Ecology Resources* 10: 1009–1018.
- 1208 **Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012a.** Environmental DNA. *Molecular  
1209 Ecology* 21: 1789–1793.
- 1210 **Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012b.** Towards next-  
1211 generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21: 2045–  
1212 2050.
- 1213 **Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G,  
1214 Brochmann C, Willerslev E. 2007.** Power and limitations of the chloroplast *trnL* (UAA) intron  
1215 for plant DNA barcoding. *Nucleic Acids Research* 35: e14.
- 1216 **Thomsen PF, Willerslev E. 2015.** Environmental DNA—an emerging tool in conservation for  
1217 monitoring past and present biodiversity. *Biological Conservation* 183: 4–18.
- 1218 **Valentini A, Pompanon F, Taberlet P. 2009.** DNA barcoding for ecologists. *Theoretical and  
1219 Applied Genetics* 24: 110–117.

- 1220 Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. 2013. Micro-scale determinants of bacterial  
1221 diversity in soil. *FEMS Microbiology Reviews* **37**: 936–954.
- 1222 Vuillemin A, Ariztegui D, Leavitt PR, Bunting L, the PASADO Science Team. 2016. Recording of  
1223 climate and diagenesis through sedimentary DNA and fossil pigments at Laguna Potrok Aike,  
1224 Argentina. *Biogeosciences* **13**: 2475–2492.
- 1225 Wang Y, Navin NE. 2015. Advances and applications of single-cell sequencing technologies.  
1226 *Molecular Cell* **58**: 598–609.
- 1227 Warinner C, Speller C, Collins MJ. 2015. A new era in palaeomicrobiology: prospects for  
1228 ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical  
1229 Transactions of the Royal Society B: Biological Sciences* **370**: 20130376.
- 1230 Weiß CL, Dannemann M, Prufer K, Burbano HA, Pickrell JK. 2015. Contesting the presence of  
1231 wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-  
1232 coverage data. *eLife* **4**: e10005.
- 1233 Wilke A, Bischof J, Harrison T, Brettin T, D'Souza M, Gerlach W, Matthews H, Paczian T,  
1234 Wilkening J, Glass EM, *et al.* 2015. A RESTful API for accessing microbial community data for  
1235 MG-RAST. *PLoS Computational Biology* **11**: e1004008.
- 1236 Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB, Hofreiter M, Bunce  
1237 M, Poinar HN, Johnsen S, *et al.* 2007. Ancient biomolecules from deep ice cores reveal a  
1238 forested southern Greenland. *Science* **317**: 111–114.
- 1239 Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, Lorenzen ED, Vestergård M,  
1240 Gussarova G, Haile J, *et al.* 2014. Fifty thousand years of Arctic vegetation and megafaunal  
1241 diet. *Nature* **506**: 47–51.
- 1242 Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, Bunce M, Wiuf C,  
1243 Gilichinsky DA, Cooper A. 2003. Diverse plant and animal genetic records from Holocene and  
1244 Pleistocene sediments. *Science* **300**: 791–795.
- 1245 Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P. 1999. Diversity of Holocene  
1246 life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences USA* **96**: 8017–  
1247 8021.
- 1248 Wetzel, R. G. 2001. Limnology. Lake and River Ecosystems, 3rd ed., Academic press, San  
1249 Diego.
- 1250 Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using  
1251 exact alignments. *Genome Biology* **15**: R46.
- 1252 Wooller MJ, Gaglioti B, Fulton TL, Lopez A, Shapiro B. 2015. Post-glacial dispersal patterns of  
1253 Northern pike inferred from an 8800 year old pike (*Esox cf. lucius*) skull from interior Alaska.  
1254 *Quaternary Science Reviews* **120**: 118–125.
- 1255 Wood JR, Wilmshurst JM. 2016. A protocol for subsampling Late Quaternary coprolites for  
1256 multi-proxy analysis. *Quaternary Science Reviews* **138**: 1–5.
- 1257 Wright HE, Mann DH, Glaser PH. 1984. Piston Corers for Peat and Lake Sediments. *Ecology* **65**:  
1258 657–659.

- 1259 Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J, Yang Y, Yao T, Knight R, Chu H. 2012. Geographic  
1260 distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau.  
1261 *Environmental Microbiology* **14**: 2457–2466.
- 1262 Xu Z-H, Jiang X-D, Wang G-Z, He J-F, Cai M-H, Wu L-S, Jiang J-L, Chen X-L. 2011. DNA extraction,  
1263 amplification and analysis of the 28S rRNA portion in sediment-buried copepod DNA in the  
1264 Great Wall Bay and Xihu Lake, Antarctica. *Journal of Plankton Research* **33**: 917–925.
- 1265 Yang J, Jiang H, Dong H, Hou W, Li G, Wu G. 2015. Sedimentary archaeal amoA gene  
1266 abundance reflects historic nutrient level and salinity fluctuations in Qinghai Lake, Tibetan  
1267 Plateau. *Scientific Reports* **5**: 1–9.
- 1268 Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Stedingk Von H, Brysting AK,  
1269 Coissac E, Pompanon F, *et al.* 2012. DNA from soil mirrors plant taxonomic and growth form  
1270 diversity. *Molecular Ecology* **21**: 3647–3655.
- 1271 Yoccoz N. 2012. The future of environmental DNA in ecology. *Molecular Ecology* **21**:  
1272 2031–2038.
- 1273 Zhang Q, Liu Y. 2003. Examination of the cytoplasmic DNA in male reproductive cells to  
1274 determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant and Cell*  
1275 *Physiology* **44**: 1–11.
- 1276 Ziesemer KA, Mann AE, Sankaranarayanan K, Schroeder H, Ozga AT, Brandt BW, Zaura E,  
1277 Waters-Rist A, Hoogland M, Salazar-García DC, *et al.* 2015. Intrinsic challenges in ancient  
1278 microbiome reconstruction using 16S rRNA gene amplification.
- 1279
- 1280

1281 **Supporting Information**

1282 Additional Information may be found in the online version of this article

1283

1284 **Note S1**

1285 Additional information on bioinformatic tools for metabarcoding datasets relevant to this  
1286 article.

1287

1288



1289 **Figure legends**

1290

1291 **Figure 1.**

1292 Number of publications retrieved from Web of Science data bank in August 2016, using  
1293 'ancient DNA' as search term, categorized according to the study organisms.

1294

1295 **Figure 2.**

1296 Biotic palaeoenvironmental proxies in lake sediments. **a**, sequential environmental  
1297 development for a temperate region, in which the lake sediments start accumulating as  
1298 glacial ice retreats, incorporating glacially eroded debris and the sparse pioneering biota (1),  
1299 which later is replaced by a tundra-steppe community (2), then the boreal forest establishes  
1300 (3) before eventually being replaced by a temperate forest (4). **b**, by identifying organisms  
1301 detectable by DNA, macro- and microfossils and accumulated and preserved in the lake  
1302 sediments, **c**, it is possible to reconstruct the environments through time. It is important to  
1303 notice that preservation or rate of degradation is strongly correlated with the age of the  
1304 sediments and that the input concentration **d**, varies in different climatic environments from  
1305 these three proxies. **e**, In addition the resulting DNA profile, as well as macro and microfossils  
1306 is influenced by taphonomic processes such as differences in biomass production and the  
1307 distance from source to deposit. This is why a combination of all these three proxies makes a  
1308 more robust palaeoenvironmental reconstruction. kyr BP: thousand years before present.

1309

1310 **Figure 3.**

1311 Spatial distribution of larger lakes (blue squares) in the world (source: Natural Earth). The  
1312 map shows that lakes are widely distributed geographically and present in many different  
1313 environments. Photos exemplify different types of lake environments, (1) Lake Comarum,  
1314 South Greenland. (2) Fiskvatn, Troms, Norway. (3) Olive-backed lake, Kamchatka. (4) Lake  
1315 Milluni, Bolivia. (5) Buyan, Bali. (6) Lake Waikaremoana, North Island, New Zealand.

1316

1317 **Figure 4.**

1318 A full bodysuit, shoe cover, hairnet, facemask, hood, gloves, and sleeve guards are necessary  
1319 during DNA extraction from a sediment core for aDNA analyses in an aDNA lab (a). During

1320 subsampling in a clean lab, surface contamination from the sediment core is removed with  
1321 sterilized razors (b), and non-contaminated material from within the intact cores is extracted  
1322 for DNA extraction (c).

1323

1324 **Figure 5.**

1325 Schematic illustration of the changes in the organellar DNA from the four types of generative  
1326 cells present in pollen. When chloroplast (p) or mitochondrial (m) DNA is present in the  
1327 mature generative pollen cells (m+ or p+), the DNA content per organelle increases after  
1328 pollen mitosis one (PMI) and during developmental stages from uni-nucleate to bi- (or tri-)  
1329 cellular pollen (stages 1-5) and it decreases when organellar DNA is absent (m- or p-). All the  
1330 changes in organellar DNA in the generative cells, whether an increase or a decrease, occur  
1331 just after PMI, and the mitochondrial DNA and plastid DNA contents are regulated  
1332 independently in the vegetative cells. An increase in the organellar DNA occurs with  
1333 biparental/paternal inheritance (B/P) of the corresponding organelle, while a decrease occurs  
1334 with maternal inheritance (M) of the corresponding organelle. Modified from Nagata *et al.*  
1335 (1999).

1336

1337 **Figure 6.**

1338 Venn diagrams showing the proportion of plant taxa commonly detected by different proxies:  
1339 metabarcoding and shotgun metabarcoding of *sedDNA* (blue), vegetation survey (light  
1340 green), pollen (yellow) and macrofossils (dark green). These proportions have so far varied  
1341 among different studies and the lower panel shows what we can theoretically expect with  
1342 shotgun metabarcoding when a full genome reference library is developed.

1343

1344 **Figure 7.**

1345 Representation of the chain of processes involved in the transformation of plant information  
1346 present in the three lake sediment assemblages types: pollen (P), macrofossils (M), ancient  
1347 DNA (aDNA). Current understanding of the processes is indicated as good (+++), reasonable  
1348 (++) or poor (+). The publications upon which the levels of understanding are based are  
1349 shown in Table 1. Modified from Birks & Birks (2016) and based on Jackson's (2012) general  
1350 conceptual model for the representation of floristic material in palaeoecological assemblages.

1351

1352 **Box 1.**

1353 **Isolating single pollen grains for direct PCR**

1354 1) Sieving sediment. Place a small amount (ca. 0.2-0.5 ml) of sediment onto a sterile filter with  
1355 mesh size slightly larger than the target pollen grains. Add water and stir sediment with a  
1356 small spatula. Wash sediment through the filter into a petri dish. Top up the petri dish with  
1357 water.

1358 2) Isolating pollen grains. Dilute sample so grains are adequately spaced in the petri dish, and  
1359 scan at 100-200x magnification (depending on size of pollen). Once a target grain has been  
1360 found, switch to a lower magnification, ensuring the grain can still be seen in the field of view.  
1361 Place the tip of a glass pipette into the field of view, and slowly lower it down into the water  
1362 beside the pollen grain. Capillary action will ensue. Capture the pollen grain as quickly as  
1363 possible and remove pipette from the water. Transfer water from the pipette onto a  
1364 microscope slide. Check the droplet at 200-400x magnification to ensure the correct pollen  
1365 grain is present. For larger pollen grains and plant fragments, a 1-10  $\mu$ l pipette can be used to  
1366 isolate the specimen.

1367 3) Dilution. Use a 1-10  $\mu$ l pipette to add several droplets of water to the microscope slide.  
1368 Capture the pollen in a glass pipette, avoiding as much debris as possible. Transfer the water  
1369 from the pipette into a clean water droplet. Check to ensure the pollen grain is still present.  
1370 Repeat this process until the pollen grain is isolated (i.e. no debris is transferred with the  
1371 grain). For the final step, capture the grain in the glass pipette, and transfer to a PCR tube.  
1372 The remainder of the final droplet can be transferred to another PCR tube as a PCR control  
1373 for that pollen grain.

1374



1375

1376

1377 Left to right: Petri dishes containing filtered sediment samples; using a glass pasteur pipette  
1378 to isolate pollen grain from petri dish; microscope slide with water droplets for sequential  
1379 dilution of pollen grain.  
1380

1381 **Table 1.**

1382 Ancient DNA related publications from lake sediments sorted by year. Studies have been  
 1383 organized according to sample type, method used for DNA extraction [environmental DNA  
 1384 (eDNA) or DNA isolated from tissues] and research environment investigated.

1385

Author	Sample type	Method	Region	Environment
Coolen & Overmann (1998)	bacteria	eDNA	Mahoney Lake, British Columbia, Canada	temperate, low altitude
Limburg & Weider (2002)	daphnia	isolation	Belauer See Lake, Germany	temperate, low altitude
Reid <i>et al.</i> (2002)	daphnia	isolation	English Lake District, UK	temperate, low altitude
Pollard <i>et al.</i> (2003)	daphnia	isolation	Hannah Lake, Sudbury, Ontario, Canada	temperate, low altitude
Coolen <i>et al.</i> (2004)	diatoms	eDNA	Ace Lake, eastern Antarctica	antarctic, low altitude
Bissett <i>et al.</i> (2005)	copepods	isolation	Lake Terrasovoje, Antarctica	antarctic, low altitude
Parducci <i>et al.</i> (2005)	pollen	isolation	Holtjärnen, central Sweden	temperate, low altitude
D'Andrea <i>et al.</i> (2006)	algae	filtration> eDNA	Søndre Strømfjord, western Greenland	arctic, low altitude
Marková <i>et al.</i> (2006)	daphnia	isolation	Lakes on Tatra Mts, Carpathians, Hungary	temperate, high altitude
Jiang <i>et al.</i> (2007)	bacteria	eDNA	Lake Chaka, Tibetan plateau	temperate, high altitude
Mergeay <i>et al.</i> (2007)	daphnia	isolation	Lake Naivasha, Kenya	temperate, low altitude

Matisoo-Smith <i>et al.</i> (2008)	plants and metazoan	eDNA	Round Lake, New Zealand	temperate, low altitude
Coolen <i>et al.</i> (2008)	bacteria	eDNA	Ace Lake, Antarctica	antarctic low altitude
Madeja <i>et al.</i> (2009)	bacteria > human	eDNA	Great Mazurian Lake District, NE Poland	temperate, low altitude
Kojima <i>et al.</i> (2009)	bacteria	filtration > eDNA	Lake Mizugaki, Central Japan	temperate, low altitude
Epp <i>et al.</i> (2010)	rotifers	eDNA	Lake Sonachi, Rift Valley, Kenya	temperate, high altitude
Madeja <i>et al.</i> (2010)	bacteria > human	eDNA	Lake Miłkowskie, north-eastern Poland	temperate, low altitude
Anderson-Carpenter <i>et al.</i> (2011)	plants	isolation	Great Lakes, North America	temperate, low altitude
Savichtcheva <i>et al.</i> (2011)	bacteria	eDNA + isolation	Lake Geneva, Bourget, Annecy, French Alps	temperate, low altitude
Xu <i>et al.</i> (2011)	bacteria	eDNA	Great Wall Bay, Xihu Lake, Antarctica	antarctic low altitude
Epp <i>et al.</i> (2011)	diatom	eDNA	Lake Naivasha, Kenya	tropical, high altitude
Magyari <i>et al.</i> (2011)	plants	isolation	Retezat Mountains, South Carpathians, Hungary	temperate, high altitude
Parducci <i>et al.</i> (2012)	plants	eDNA	Central Norway and Sweden	temperate, low altitude
Stoof-Leichsenring <i>et al.</i> (2012)	diatoms	eDNA	Kenya, Lake Naivasha	tropical, high altitude
Xiong <i>et al.</i> (2012)	bacteria	eDNA	Tibetan Plateau	temperate, high altitude
Ravasi <i>et al.</i> (2012)	bacteria	eDNA?	Lake Cadagno, Swiss Alps	temperate, high altitude

Pedersen <i>et al.</i> (2013)	plants	eDNA	Southern Greenland	arctic, low altitude
Parducci <i>et al.</i> (2013)	plants	eDNA	Central Norway and Sweden	temperate, low altitude
Domaizon <i>et al.</i> (2013)	bacteria	eDNA	Lake Bourget, French Alps	temperate, low altitude
Giguet-Covex <i>et al.</i> (2014)	plants and metazoan	eDNA	Lake Anterne, French Alps	temperate, high altitude
Boessenkool <i>et al.</i> (2013)	plants	eDNA	Mt Gahinga, eastern Africa	temperate, high altitude
Belle <i>et al.</i> (2014)	bacteria	eDNA	Lake Narlay, Jura Mountains, France	temperate, low altitude
Hou <i>et al.</i> (2014)	plankton	eDNA	Tibetan Plateau	temperate, high altitude
Randlett <i>et al.</i> (2014)	phytoplankton	eDNA	Lake Van, Turkey	temperate, high altitude
Madeja <i>et al.</i> (2015)	bacteria > human	eDNA	Great Mazurian Lake District, Poland	temperate, low altitude
Kyle <i>et al.</i> (2015)	bacteria	eDNA	Bjørkelangen, Gjersjøen, Hemnessjø lakes, Norway	temperate, low altitude
Pal <i>et al.</i> (2015)	bacteria	eDNA	Gatineau Park lakes, western Quebec	temperate, low altitude
Pansu <i>et al.</i> (2015)	plants	eDNA	Lake Anterne, French Alps	temperate, high altitude
Paus <i>et al.</i> (2015)	plants	eDNA	Dovre region, central Norway	temperate, low altitude
Stoof-Leichsenring <i>et al.</i> (2015)	algae	eDNA	North Siberia	arctic, low altitude
Epp <i>et al.</i> (2015)	algae vascular plants	eDNA	North Greenland	arctic, low altitude

Capo <i>et al.</i> (2015)	bacteria	eDNA	Lake Bourget, French Alps	temperate, low altitude
Yang <i>et al.</i> (2015)	bacteria	eDNA	Qinghai Lake, Tibetan Plateau	temperate, high altitude
Poulain <i>et al.</i> (2015)	bacteria	eDNA	Hudson Bay, Ontario, Canada,	arctic, low altitude
Wooller <i>et al.</i> (2015)	fish	isolation	Quartz Lake, Alaska	arctic, low altitude
Etienne <i>et al.</i> (2015)	bacteria> human	eDNA	Lorraine Plateau, north-eastern France	temperate, low altitude
Stager <i>et al.</i> (2015)	fish	eDNA	Lower Saint Regis Lake, Franklin County, NY, US	temperate, low altitude
Li <i>et al.</i> (2016)	phytoplankton	eDNA	Qinghai Lake, Tibetan Plateau	temperate, high altitude
Vuillemin <i>et al.</i> (2016)	bacteria >taphonomy	eDNA	Laguna Potrok Aike, Argentina	temperate, low altitude
Alsos Greve <i>et al.</i> (2016)	plants	eDNA	Lake Skartjørna, Svalbard	arctic, low altitude
Pedersen <i>et al.</i> (2016)	plants	eDNA	Peace River, Alberta and British Columbia, Canada	temperate, low altitude
Sjögren <i>et al.</i> (2016)	plants	eDNA	Spectacle Loch, Loch of the Lowes, Scotland, UK	temperate, low altitude

1386