1 **Ancient plant DNA in lake sediments**

2

3 Laura Parducci[1], K. D. Bennett[2], Gentile Francesco Ficetola[3,4], Inger Greve Alsos[5], Yoshihisa
4 Suyama[6], Jamie R. Wood[7], Mikkel Winther Pedersen[8]

5

6 [1]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University,
7 Norbyvägen 18D, 75236 Uppsala, Sweden; [2]Department of Geography & Sustainable
8 Development, School of Geography & Geosciences, University of St Andrews, Scotland;
9 [3]Université Grenoble-Alpes, CNRS, Laboratoire d'Ecologie Alpine (LECA), F-38000 Grenoble,
10 France; [4]Department of Biosciences, Università degli Studi di Milano, 20133 Milano Italy.
11 [5]Tromsø Museum, UiT – The Arctic University of Norway, NO-9037 Tromsø, Norway; [6]Field
12 Science Center, Graduate School of Agricultural Science, Tohoku University, 232-3 Yomogida,
13 Naruko-onsen, Osaki, Miyagi 989-6711 Japan; [7]Long-term Ecology Lab, Landcare Research,
14 PO Box 69040, Lincoln, Canterbury 7640, New Zealand; [8]Centre for GeoGenetics, Natural
15 History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark.

16

17 Authors for correspondence:

18 Laura Parducci, +46 18 471 64 14, laura.parducci@ebc.uu.se;

19 Mikkel Winther Pedersen, +45 29 27 53 42, mwpedersen@snm.ku.dk

20

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

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

23 Number of tables: 1

24 Supporting information: Note S1

25

For Peer Review

Summary

Recent advances in sequencing technologies now permit analyses of plant DNA from fossil samples (ancient plant DNA, plant aDNA), and thus enable molecular reconstruction of palaeofloras. Hitherto, ancient frozen soils have proved excellent in preserving DNA molecules, and have thus been the most commonly used source of plant aDNA. However, DNA from soil mainly represents taxa growing few meters from the sampling point. Lakes have larger catchment areas and recent studies suggest that plant aDNA from lake sediments is a more powerful tool for palaeofloristic reconstruction. Furthermore, lakes can be found globally in nearly all environments and are therefore not limited to perennially frozen areas.

Here we review the latest approaches and methods for studying plant aDNA from lake sediments and discuss the progress made up to present. We argue that aDNA analyses add new and additional perspectives for studying ancient plant populations and in time will provide higher taxonomic resolution and more precise estimation of abundance. Despite this, key questions and challenges remain for such plant aDNA studies. Finally, we provide guidelines on technical issues, including lake selection and we suggest directions for future research on plant aDNA studies in lake sediments.

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

For Peer Review

Introduction

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

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

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

For Peer Review

latitudes/altitudes where local pollen productivity is low and long-distance pollen dispersal is more common.

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

Environmental and temporal limits for survival of aDNA

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

Several studies have investigated the post mortem processes affecting DNA molecules in different tissues. We know that after the death of an organism, several intra and extra-cellular processes (e.g. enzymatic, hydrolytic and oxidative processes) cause DNA damage (seen as misincorporation of C to T and G to A transitions primarily toward the ends of the

For Peer Review

DNA molecules) (Briggs *et al.*, 2007; Jónsson *et al.*, 2013), eventually leading to fragmentation of the DNA molecules. The highest success rate for aDNA isolation is normally obtained from frozen, anoxic or arid areas, environments with limited bacterial abundance and therefore presence of nucleases, which reduces longer nucleic acids to short molecules (Hofreiter *et al.*, 2001). Nevertheless, exogenous processes will inevitably lead eventually to the destabilization, fragmentation and damage of DNA, even in good preservation conditions.

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

Lake sediments

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

Lake sediment characteristics

Lakes are excellent sources of sediments containing aquatic and terrestrial environmental components accumulated over time and preserved in robust stratigraphic contexts and anoxic conditions. They offer the best settings for preservation of plant aDNA, as the sediments accumulate continuously; allowing establishment of high-resolution molecular

For Peer Review

records using available geo-chronological methods (Fig. 2a, b, c) and thus providing a robust record for interpretation of the molecular history. In particular, sediment records from small lakes (Fig. 3) are excellent archives for molecular studies, as the effects of disturbances are low and seem to represent the surrounding terrestrial environment well. Lake sediments consist of variable proportions of autochthonous/allochthonous organic material (OM) and in-washed inorganic material. Microbial degradation of OM frequently creates anoxic conditions in the bottom water and below 1-2 cm sediment depth (Sobek *et al.*, 2009). Such conditions preclude the presence of burrowing animals and thus minimise bioturbation, water percolation and sediment reworking (Pansu *et al.*, 2015), all of which are major concerns for palaeoecological studies. Water has maximum density at 4°C, so water less dense than this, both warmer and colder, floats. The net result is that bottom waters become insulated from the atmosphere, thereby favouring the development of anoxia and temperature stability. However, the temperature of lake water also depends upon geography and depth (Hutchinson, 1957; Wetzel, 2001). If lakes are deep enough, the water column becomes thermally stratified. During the summer, surface water warms and establishes a gradient down to cooler lower water. During the winter, surface water cools until it reaches 4°C (maximum density), at which point no more cooling occurs until the whole lake is mixed at 4°C. Tropical lakes with little seasonal variation of temperature may have more or less constant temperatures year-round, at a level similar to ambient temperatures. In temperate and cold regions, the bottom waters will normally be colder than surface water in the summer, and at a similar temperature in the winter, when the whole lake is cold (and may be ice covered). Sediments are thus in contact with the coldest water, and become insulated from the atmosphere, favouring the development of anoxia, and have greater temperature stability, increasing the probability of DNA survival.

The sediments of lakes where anoxic conditions dominate can also be 'laminated' (i.e. layers of different composition reflecting seasonal environmental differences). In some cases, laminated sediments are sufficiently continuous to provide a temporal record of layers with annual resolution (Larsen & Macdonald, 1993), similar to tree rings. While vertical migration (leaching) of DNA has been observed in cave sediments (Haile *et al.*, 2007) and non-frozen soils (Andersen *et al.*, 2011), leaching has not been observed in lake sediments (Anderson-Carpenter *et al.*, 2011; Pansu *et al.*, 2015; Sjögren *et al.*, 2016). Once imbedded in the sediments, plant macrofossils and pollen grains, as well as silica, clay or organic matter, to which extracellular DNA attach (Pietramellara *et al.*, 2009; Poté *et al.*, 2009; Taberlet *et al.*, 2012a), are unlikely to move vertically. On the other hand, re-deposition of sediments can

For Peer Review

occur in lakes, contaminating the micro and macrofossil record with older material. While Pedersen *et al*. (2016) recently found pre-Quaternary re-deposited microfossils in lake sediments from the Peace River drainage basin in North America, they also found that the DNA record remained un-affected, probably due to the significant older age and smaller quantities of the re-deposited material. Nevertheless, re-deposition of material should not be neglected as a possible source of DNA.

Coring techniques for aDNA work

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

For Peer Review

Perspective for plant aDNA research

Traditionally, lake sediment records have been routinely analysed for pollen and other plant fossils, which have been the basis for our understanding of landscape-scale distribution of plants on millennial timescales. Plant aDNA now offers the potential of helping elucidation of long-standing ecological questions that classical palaeoecological techniques may not be able to retrieve. It has been through collaborative work between molecular ecologists and classical palaeoecologists that robust and reliable results have been produced recently (Willerslev *et al.*, 2014; Pedersen *et al.*, 2016; Alsos *et al.*, 2016), and this type of collaboration is strongly advocated (Hu *et al.*, 2009; Anderson-Carpenter *et al.*, 2011; Brown *et al.*, 2014). Examples of problems being elucidated by combining the two disciplines include Reid's paradox of rapid plant migration (Clark *et al.*, 1998) and the question of whether or not trees survived the Last Glacial Maximum (LGM) at high-latitudes in Europe (Parducci *et al.*, 2012). Here, combined information from fossils and molecular studies (modern and ancient) has changed our traditional view of post-glacial migration of trees from southern European refugia (Hewitt, 2000; Stewart *et al.*, 2010). Another example is the limited taxonomic resolution generally achieved by pollen analysis. Even if such analyses have recently reached better taxonomic resolution due to improved identification keys, the work is still time demanding, and identifications are often at genus or family level, rarely species (Faegri *et al.*, 1992). We thus have little understanding about specific diversity, and even less information about ancient plant populations. The new possibilities offered by the latest HTS technologies will likely increase our ability to resolve plant taxa at species level and additionally elucidate ancestry and genetic composition of ancient plant populations. Likewise, metabarcoding and metagenomic analyses of *sed*aDNA will provide more detailed insights on ancient ecosystems and link changes more tightly to past climate shifts (see section on HTS).

Sources of pollen, macrofossil and DNA

The majority of lake sediment deposits contain both plant remains and non-biological material originating from the lake catchment. In small boreal lakes it is estimated that ca 70% of the deposited pollen is from vegetation growing within few km of the lake (Jacobson & Bradshaw, 1981). While the majority of the pollen often derives from high-pollen producing wind-pollinating plants, which are distributed regionally through the air, the

For Peer Review

proportions of pollen from insect-pollinated plants may vary between sites. Furthermore, in areas of local low pollen production the pollen record may be highly affected by long distance exotic pollen that is not part of the regional environment (Hyvärinen, 1970). Despite this, pollen records generally represent the regional flora. Plant macrofossils, however, are of local origin, as they are large and have low dispersal and transport capacity (Allen & Huntley, 1999). The macrofossil record is often dominated by the aquatic and wetland plants growing in and around the lake, and terrestrial taxa often under-represented (Birks, 2003). *Sed*aDNA seems to have a similar source of origin as macrofossils (Jørgensen *et al*., 2012; Pedersen *et al*., 2013; Alsos *et al*., 2015; Parducci *et al*., 2015). To date, only a few studies have focussed on the release and deposition of DNA in the environment (Poté *et al*., 2007; Pietramellara *et al*., 2009; Poté *et al*., 2009; Barnes & Turner, 2016) and much therefore remains unknown about the ecology of the DNA – e.g. all processes occurring from source to deposition (taphonomy). Current *sed*aDNA results suggest that DNA extracted from sediments does not derive from actual pollen grains (Pedersen *et al*., 2016; Sjögren *et al*., 2016), but from other components imbedded in the sediment matrix, thus pollen DNA needs to be extracted directly from single or multiple isolated grains (see below).

Pollen DNA

Fossil pollen is often abundant and can be well preserved in lake sediments. Once deposited, pollen remains in situ in the sediments and there is therefore a high degree of certainty to its stratigraphic context. Further, aerial transport and rapid burial rates in sediment results in minimal physical damage to pollen grains and minimal exposure of the grains to biotic degradation. In addition the outer part of pollen grains (exine) is composed of cellulose and sporopollenin, an acid-resistant polymer that contains saturated and unsaturated hydrocarbons and phenolics (Southworth, 1974), and which protects the grains from physical and chemical attack, aiding pollen preservation in the sediments (Bennett & Willis, 2001). Inside the pollen, after maturation, there are 2-3 cells (a large vegetative cell and 1-2 generative cells) that comprise the male gametophyte (Fig. 5). The vegetative cell comprises the cytoplasm, and numerous plastids and mitochondria, which are responsible for the development of the pollen tube and delivery of the generative cells to the embryo sac together with the nuclear haploid DNA. Generative cells of pollen from the majority of plant species contain multiple organelles, including several copies of organelle DNA [chloroplast (cpDNA) and mtDNA], regardless of the type of inheritance of these genomes (maternal or paternal). Some nuclear repetitive regions, like ITS ribosomal repeats, are also present in

For Peer Review

multiple copies in the nucleus of both cell types. During pollen maturation however, there is a selective increase or decrease in the amount of organelle DNA in the generative cells (not in the vegetative) depending on the inheritance pattern (Nagata *et al*., 1999; Zhang & Liu, 2003). For example, mature pollen from species with paternal cpDNA inheritance (the majority of conifers), contains a regular amount of mtDNA and cpDNA in the vegetative cell and an increased amount of cpDNA in the generative cell/s. On the other hand, pollen from species with maternal inheritance of cpDNA and mtDNA (most of the angiosperms) contains at maturation a regular amount of both genomes in the vegetative cell and a decreased amount in the generative cell/s (Fig. 5). All three plant genomes are therefore present in both pollen types, but the cpDNA and the mtDNA may be present in different amount (in the generative cell/s) in different taxa.

Single-pollen genotyping

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

For Peer Review

Sedimentary ancient DNA

Different sediment types show distinctive physical and chemical characteristics that will differentially affect DNA preservation, thus DNA survival will vary between different locations. Likewise, it seems that extraction of DNA and removal of inhibiting substrates requires strategies adapted to differences in the sediment content (Taberlet *et al.*, 2012a; Pedersen *et al.*, 2016). Furthermore, extraction of aDNA molecules requires optimized protocols and special laboratory precautions.

Proxy overlapping or complementary

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

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

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

The most stringent test for identifying the origin of *sed*aDNA is by direct comparison with modern vegetation survey or with detailed historical vegetation maps. Yoccoz *et al.* (2012) were the first to demonstrate that plant diversity detected from environmental DNA extracted from boreal soil was consistent with plant taxonomic diversity estimated from

For Peer Review

conventional aboveground surveys. To date, we still do not know if such a good relationship holds with DNA in lake sediments. In a study from a high altitude crater lakes in Africa, Boessenkool *et al*. (2013) showed that *sed*aDNA largely reflected local flora. Similarly, Sjögren *et al*. (2016) compared DNA, pollen and historical vegetation maps in two Scottish lakes and found that *sed*aDNA was of local origin in contrast to regionally dispersed deciduous tree pollen. Finally, in a comparison of *sed*aDNA with vegetation surveys of 11 lakes in Northern Norway, the majority of taxa recorded in the *sed*aDNA were growing within 2 m of the lake (I. G. Alsos, unpublished; Alsos *et al*., 2015). Thus, we conclude that DNA deposition in lakes is more similar to that of macrofossils than pollen, and represents flora from within the catchment area. However with future improved DNA reference databases, methods and understanding, we expect increased information gained will lead to an almost complete overlap between DNA and macrofossils, but likely not between DNA and pollen (Fig. 6).

Taphonomic processes in lake sediments

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

Lake sediments contain DNA from many different organisms, which include cellular DNA from tissues and intact cells, as well as extracellular DNA. When a plant tissue is degraded and a cell is lysed, it releases its content in the surrounding environment and extracellular DNA can bind to charged minerogenic and organic particles or remain unbound. Pietramellara *et al*. (2009) showed that in modern soils, once the DNA binds to a particle, it is immediately protected against nuclease degradation. This implies that that clay-rich soil can be highly suitable for protection of DNA against degradation, as clay particles have relatively large and charged surface area (Huang, 2014). Extracellular DNA molecules in soils can also be taken up by competent prokaryotic cells in a process called natural

For Peer Review

374 transformation. Although it is unlikely that this will be a large source of 'plant' DNA, it
375 remains however a possibility (Pedersen *et al.*, 2015). Whether the plant DNA in lake
376 sediments primarily are present as extracellular DNA or as small plant parts like leaves, root
377 caps cells, stem or fruits, remains however still unclear.

378

379 Inferring taxa abundance from PCR-based *sed*aDNA analyses

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

391

392

393 **Recent key findings and future methods using HTS techniques**

394 HTS expands upon traditional PCR-based Sanger sequencing techniques and has facilitated a
395 rapid development of aDNA research during the last decade. Ancient DNA molecules
396 converted to sequencing libraries can now be parallel sequenced massively on HTS platforms
397 like the Illumina HiSeq or the Complete Genomics platforms. HTS approaches have become
398 increasingly affordable and are now routinely used by most aDNA laboratories, allowing
399 screening of *sed*aDNA from a wide range of complex ancient substrates. With the newest
400 generation of desktop HTS platforms, e.g. NextSeq, even small laboratories can now
401 sequence their own datasets. There are, in principle, three HTS methodological strategies for
402 analysing plant aDNA in lake sediments: metabarcoding or shotgun sequencing of *sed*aDNA
403 and HTS of pollen DNA. Of the two latter methods, 'metabarcoding' is a relative established
404 method, which relies on the information on one single locus, while the second is newer and
405 relies on shotgun sequencing, i.e. sequencing a non-discriminated pool of aDNA. For clarity,

For Peer Review

406 it is important to define this new method. Here we suggest the term 'shotgun
407 metabarcoding' when shotgun sequencing of environmental DNA is used to identify taxa,
408 and 'metagenomics' when shotgun is used for functional analyses of the environments. This
409 will allow researchers to discriminate between studies focusing on taxon identification and
410 studies focusing on functional and attribute analysis. In the following paragraphs we review
411 and discuss the progress made and future of these three methodologies.

412

413 Metabarcoding on *sed*aDNA

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

421 Metabarcoding has already been used in a variety of palaeoenvironmental studies
422 (Willerslev *et al.*, 2003; Valentini *et al.*, 2009; Taberlet *et al.*, 2012b), including permafrost
423 (Lydolph *et al.*, 2005; Jørgensen *et al.*, 2011; 2012; Willerslev *et al.*, 2014), mid to high
424 latitude/altitude lake sediments (Coolen & Gibson, 2009; Parducci *et al.*, 2012; Alsos *et al.*,
425 2015; Pansu *et al.*, 2015; Epp *et al.*, 2015; Paus *et al.*, 2015; Alsos *et al.*, 2016); tropical lake
426 sediments (Boessenkool *et al.*, 2013) (F. M. Ficetola, unpublished), and deposits preserved
427 under ice-sheets, even in the absence of visible macrofossil remains (Willerslev *et al.*, 1999;
428 2007).

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

437 The performance of the metabarcoding approach is often limited if markers are not able to
438 amplify all the present taxa (universality), and by their capacity to differentiate and identify

For Peer Review

439 also closely related species (resolution) (Ficetola *et al*., 2010; Sønstebø *et al*., 2010; Furlan *et*
440 *al*., 2016). Therefore, the use of primers with appropriate features (i.e. "universal" primers
441 amplifying with high specificity all plants, and achieving a high taxonomic resolution)
442 (Ficetola *et al*., 2010; Furlan *et al*., 2016) is essential for the success of metabarcoding
443 studies. In aDNA studies the choice of the marker is particularly difficult, as prerequisites like
444 minimal bias in amplification of different taxa and short sequence length, drastically reduce
445 the ability to resolve taxa (Taberlet *et al*., 2007). Plant metabarcoding studies can use either
446 a single marker or a combination of multiple primers to resolve taxa. For instance, the *trn*L
447 g/h primers (Taberlet *et al*., 2007) can first be used to obtain an overall analysis of plant
448 diversity with a good resolution for most plant families, and additional primers (e.g. ITS1-
449 F/ITS1Poa-R for Poaceae and ITS1-F/TS1Ast-R for Asteraceae) may be later added to increase
450 resolution within families (Baamrane *et al*., 2012). We should however remember that
451 different primers might favour amplification of different taxa, which may lead to biases in
452 the final results (Yoccoz, 2012).

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

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

468

469 Shotgun metabarcoding on *sed*aDNA

470 Shotgun metabarcode analysis is a newer alternative to the traditional metabarcoding
471 approach. It relies on shotgun sequencing DNA (Orlando *et al*., 2015) from a non-

For Peer Review

discriminated genetic pool obtained from environmental samples and then computational investigation, e.g. bioinformatic analysis, is used to decipher the taxonomic composition. While the shotgun sequencing approach traditionally was used for studying the fraction of un-cultivable microbes in modern environmental samples (Vos *et al*., 2013; Sharon & Banfield, 2013) by genome assembly and functional analysis of the microbial fraction (Mackelprang *et al*., 2011). More recently shotgun sequencing has been recognized as an important tool for bypassing DNA barcode amplification biases (Ziesemer *et al*., 2015; Pedersen *et al*., 2016), and also appear to give a more comprehensive insight into the community composition from all trophic layers represented by the environmental sample (Pedersen *et al*., 2016). While this approach has played an expanding role in studying ancient prokaryotic communities in soil, sediments (>90% of DNA molecules in a sample are likely to be prokaryote), but also teeth, coprolites, gut and ice (e.g. Warinner *et al*., 2015; Wood & Wilmshurst, 2016), only one study so far has published results that are accepted as authentic using shotgun sequencing data for palaeoenvironmental reconstruction (Pedersen *et al*., 2016). However, there exists a considerable potential for applying low-coverage shotgun sequencing of genomic DNA (genome skimming) for studying plant aDNA (Coissac *et al*., 2016), but the absence of a consensus data processing and lack of experience interpreting such datasets makes it prone to misinterpretations and false positives (Bennett, 2015; Weiß et al., 2015) (see section on Bioinformatics). Species identification using genomic plant data outside chloroplast barcode regions can be particularly problematic. For instance, of ca. 391,000 vascular plant species existing on earth (number constantly changing through new discoveries and taxonomic revisions) only 1,092 are represented by complete chloroplast genomes (NCBI RefSeq database, ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plastid/ - accessed 31 August 2016). In addition, taxonomic identifications outside the chloroplast genome are skewed towards overrepresented taxa with fully sequenced genomes, especially of commercially important species such as *Zea mays, Triticum aestivum, Solanum lycopersicum, Hordeum vulgare, Oryza sativa, Nicotiana tabacum* and the model organism *Arabidopsis thaliana*. The latter are all amongst the 20th most sequenced organisms (no. of bases, NCBI, ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt - accessed 29 March 2016). While we can limit false positives and confirm aDNA authenticity using bioinformatic techniques (see section on Metagenomic bioinformatics), the lack of genomic references and the limited experience in interpreting such datasets makes proper contextualized biological interpretation a necessity. However, reference databases are quickly improving and on-going projects (e.g. PhyloAlps,

For Peer Review

506 https://www.france-genomique.org/spip/spip.php?article112&lang=fr and NorBol,
507 http://norbol.org/) are currently assembling the entire chloroplast and nuclear ribosomal
508 genomes of the whole floras of respectively the Alps, Norway, and parts of the Arctic
509 (Coissac *et al.*, 2016). Shotgun metabarcoding will therefore become an important tool in
510 the future years, as it will allow for detecting organismal diversity and potentially
511 differentiate population structures, and will further help bridge the gap between different
512 scientific disciplines in palaeoecology.

513 An important advantage of shotgun sequencing is the possibility of quantifying the degree of
514 DNA degradation exerted by the sample, e.g. fragmentation and hydrolysation of cytosines
515 (C), which results in accumulating misincorporation frequencies of thymine (T) instead of C
516 at the ends of the DNA molecules during sequencing (Jónsson *et al.*, 2013). Lastly, it was
517 shown recently that DNA damage correlates with age of the lake sediments (Pedersen *et al.*,
518 2016) and should therefore be used as a independent tool for aDNA authentication.

519

520 HTS on pollen

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

532 A more efficient way of analysing ancient pollen is to construct HTS libraries directly from
533 single pollen grains employing the SCS technology. SCS is a cost- and time-effective method
534 to sequence in parallel hundreds of single-cells isolated and assayed in tiny (nl) reaction
535 chambers for PCR and sequencing (Wang & Navin, 2015). Limiting reactions to few nl offers
536 the advantages of increased throughputs, improved reaction sensitivity (higher percentages
537 of grains successfully genotyped) and increased degree of PCR specificity (reduced DNA
538 contamination). SCS methodology on fossil pollen seems at the moment the best available

For Peer Review

approach to investigate efficiently a large number of pollen grains and a good alternative to the more time-consuming single-pollen genotyping technique (Parducci *et al*., 2005). The method offers the unprecedented opportunity of analysing in a time-effective way the genetic structures of large number of single plant individuals on millennial time scales.

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

For Peer Review

Challenges when studying aDNA from lake sediments

Contamination, laboratory analyses and experimental setup

Contamination of low-concentration aDNA samples with high-concentration modern DNA poses a challenge that should not be taken lightly and special precautions should always be taken in all steps of analyses (Fulton, 2012). There is not a single strategy valid for avoiding all possible contamination sources occurring from: (i) laboratory facilities; (ii) cross contamination; and (iii) reagents used during extraction and downstream preparation prior to sequencing. In general, multi-strategy procedures should be employed (Champlot *et al*., 2010) to avoid contamination. However, while contamination from laboratory facilities and cross contamination are strictly related to facilities and experience/training of the workers and can be more easily taken under control, contamination from reagents, which are known to contain DNA especially of common food plants, is more difficult to control. Firstly, not all reagents can be filtered, UV-lighted, bleached or DNase-treated (the preferred decontaminating method). Secondly, even when treated, short DNA molecules can still persist and be a source of genetic material during extraction and downstream handling. We therefore stress the importance for having always experimental controls covering all reagents and all steps during handling. Equally important is that the experimental setup is properly designed thus enabling understanding of pre-analysis workflows and results for better and future-proof data production. This involves steps from sample collection, laboratory work to eventually multiplexing the DNA libraries for sequencing, and in which vital discoveries have been made (Murray *et al*., 2015, Schnell *et al*., 2015, Ficetola *et al*., 2016).

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

Replication and use of controls

Extensive use of controls and replicated analyses is an important strategy ensuring the quality of aDNA results. First, multiple extraction and PCR controls must be performed and sequenced to detect sporadic contaminants. Taxa that are detected at a significant rate

For Peer Review

within controls (usually from known food or exotic plants) are usually easy to recognise and must be removed from analyses (Cooper & Poinar, 2000; Ficetola *et al*., 2016). Alternatively, if a taxon is sporadically present within controls, but is abundant in test samples, statistical tests can assess whether the detection within HTS samples is significantly higher than the contamination rate (Champlot *et al*., 2010). The use of positive controls is a further strategy to limit false positives and false negatives. De Barba *et al*. (2014) added to their analyses positive controls made by mixing DNA of known concentration from four known plant species. After HTS, they detected in positive controls several sequences at very low frequencies, which did not belong to any of the species actually present. This result was then used to identify a frequency threshold, which allowed detection and removal of sequences representing low-frequency noise, without removing species that were actually present in the sample. The use of positive controls is however risky in aDNA analyses and should be avoided when possible, as it is itself a potential source of contamination (Cooper & Poinar, 2000). Species that are exotic to the study area (e.g. tropical species in studies focusing on the boreal flora) can be suitable positive controls, as their eventual contamination of the sample can be easily spotted.

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

For Peer Review

Authentication of ancient origin (DNA damage)

In all aDNA sequences, an excess of C to T transitions is observed at the 5' and 3' end of molecules. This pattern of post-mortem damage increases over time (Sawyer *et al*., 2012; Pedersen *et al*., 2016), and the increased frequency of C to T transitions can be used as a tool for distinguishing aDNA sequences from modern contaminants (Briggs *et al*., 2007). The bioinformatic tool package MapDamage2.0 (Jónsson *et al*., 2013) provides a way to quantify the rate of DNA damage. This led Weiß *et al*. (2015) to develop a specific computational approach for comparing DNA damage patterns of putative aDNA with modern DNA and thus confirming whether a sequences was of ancient origin or not. The method may be particularly useful for validation of future metagenomic studies.

Bioinformatic processing

Metabarcoding bioinformatics

HTS generates large DNA data sets that require dedicated programs for analyses. There are a number of bioinformatic steps to follow after DNA metabarcoding sequencing on Illumina platforms. The first three steps are assembling paired-end reads, assigning HTS reads to samples or demultiplexing (metabarcoding studies generally analyse multiple samples within each HTS run) (Coissac, 2012 and subsequent papers), and filtering erroneous sequences originated from non-specific amplification and PCR/sequencing errors. The latter task is often critical for ensuring the quality of metabarcoding data, and can be achieved by removing sequences that are too short/too long relative to the known features of the used barcodes and chimera sequences or sequences that likely represent punctual errors (e.g. nucleotide substitutions, small insertions/deletions) originating during PCR. Furthermore, sequences with just one read (singletons) or two reads in one sample can be artefacts (De Barba *et al*., 2014; Elbrecht & Leese, 2015), and should generally be filtered. If sequencing depth is high, it is even possible to find thousands of reads which do not correspond to real barcodes. Therefore treatment of rare sequences remains a major challenge of metabarcoding data analysis (Ficetola *et al*., 2016). The fourth bioinformatic step is clustering, which merges sequences belonging to the same molecular operational taxonomic units (MOTU). The fifth and final step is taxonomic assignment of sequences to currently known taxa. This is generally done by comparing the retrieved sequences with reference

For Peer Review

672 databases, which can be either broad databases like GenBank (Benson *et al*., 2013) or ENA-
673 EMBL (Leinonen *et al*., 2011) or high-quality reference databases containing the verified and
674 curated sequences of potentially present species, such as the arctic plant database
675 (Sønstebø *et al*., 2010), or the on-going PhyloAlps and NorBol projects (Coissac *et al*., 2016).
676 Multiple bioinformatics tools are currently available and a more detailed description can be
677 found in Note S1 in the supporting information.

678

679 Shotgun metabarcode bioinformatics

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

702 While the accuracy and speed of these tools make shotgun metabarcode dataset analysis
703 faster and easy to perform [for more details see Lindgreen, *et al*. (2015)], a common issue is
704 that they are designed for modern sequencing datasets and for analysis of the microbial

For Peer Review

fraction and only two of the tools (i.e. *MG-RAST* and *EBI-metagenomics*) contain plant reference sequences within their databases unless the database is custom build. However, none of these tools are designed for aDNA analysis and therefore lack the stringent and robust criteria required for taxonomic consideration of short and damaged reads. Schubert *et al*. (2012) compared different computational methods for improving accuracy and sensitivity of aDNA sequence identification and showed that using reads ≥ 30 bp increase the quality of alignments to modern reference genomes and lower the number of false positives. Currently, from the limited number of studies of ancient plant shotgun metabarcoding, much indicates that the sequence-to-reference similarity for taxonomic consideration should be as high as 100%. An in silico test, modelling shotgun metabarcode libraries inferred with sequencing errors, found that errors or nucleotide substitutions can lead to false positives in ancient metagenomic datasets (Pedersen *et al*., 2016). However, these false-positives appear as low abundance random 'back-ground' noise, and can thus be avoided by setting a minimum number of reads as a threshold. Although the size of this threshold seems to depend on the number of reads sequenced, the exact size and nature for setting this threshold remains unknown and likely varies with the genomic composition in the sample. No clear-cut choice exists for metagenomic analysis tool but with above suggested criteria taxonomic inference will become more robust. We furthermore urge that biological interpretation is always accompanied by aDNA authentication (Jónsson *et al*., 2013).

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

Finally, it is important to emphasize that due to the nature of an environmental DNA sample, the majority of the sequenced DNA obtained using a shotgun metabarcoding approach

For Peer Review

739 cannot be usually identified with the databases currently available (Pedersen *et al*., 2016).
740 Often, more than 90 % of the reads produced cannot be aligned to a reference, and in most
741 cases less than 2% of the reads are unique to any taxa. However, as more genomes become
742 available these proportions will improve.

743

744

745 **Conclusions and directions for future research**

746

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

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

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

For Peer Review

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

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

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

Acknowledgements

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

For Peer Review

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Birks HJB, Birks HH. 2016. How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? *The New Phytologist* **209**: 499–506.

For Peer Review

Boessenkool S, MCGlynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, Memomissa S, Brochmann C, Popp M. 2013. Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. *Conservation Biology* **28**: 446–455.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al*. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences USA* **104**: 14616–14621.

Brouzes E, Medkova M, Savenelli N, Marran D, Twardowski M, Hutchison JB, Rothberg JM, Link DR, Perrimon N, Samuels ML. 2009. Droplet microfluidic technology for single-cell high-throughput screening. *Proceedings of the National Academy of Sciences USA* **106**: 14195–14200.

Brown TA, Cappellini E, Kistler L, Lister DL, Oliveira HR, Wales N, Schlumbaum A. 2014. Recent advances in ancient DNA research and their implications for archaeobotany. *Vegetation History and Archaeobotany* **24**: 207–214.

Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, Geigl E-M. 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLOS ONE* **5**: e13042.

Chan YL, Lacey EA, Pearson OP, HADLY EA. 2005. Ancient DNA reveals Holocene loss of genetic diversity in a South American rodent. *Biology Letters* **272**: 423–426.

Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, Lynch J, Pacala S, Prentice C, *et al*. 1998. Reid's Paradox of Rapid Plant Migration Dispersal theory and interpretation of paleoecological records. *BioScience* **48**: 13–24.

Coissac E. 2012. OligoTag: a program for designing sets of tags for next-generation sequencing of multiplexed samples. Pompanon F, Bonin A eds. Methods in Molecular Biology. Data Production and Analysis in Population Genomics. New York, Dordrecht, Heidelberg, London, 13–31.

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. 2016. From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology* **25**: 1423–1428.

Coolen M, Gibson J. 2009. Ancient DNA in lake sediment records. *PAGES news*: 1–3.

Cooper A, Poinar HN. 2000. Ancient DNA: Do it Right or Not at All. *Science* **289**: 1139–1141.

Dabney J, Meyer M, Pääbo S. 2013. Ancient DNA Damage. Cold Spring Harbor Perspectives in Biology: doi: 10.1101–cshperspect.a012567.

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MAC, Hufnagel DE, *et al*. 2015. The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**: 14003–14008.

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P. 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources* **14**: 306–323.

Dong J, Gao K, Wang K, Xu X, Zhang H. 2015. Cell Wall Disruption of Rape Bee Pollen Treated with Combination of Protamex Hydrolysis and Ultrasonication. *Food Research*

For Peer Review

International **75**: 123–130.

Elbrecht V, Leese F. 2015. Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass - Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE* **10**: e0130324.

Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, Schrøder-Nielsen A, Ludikova A, Hassel K, Stenøien HK, Funder S, *et al*. 2015. Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews* **117**: 152–163.

Faegri K, Iversen J, Kaland PE, Krzywinski K. 1992. Textbook of Pollen Analysis (PE Kaland and K Krzywinski, Eds.). Wiley, 1992.

Feek DT, Flenley JR, Chester PI, Welikala N, Matisoo-Smith EA, Tannock GW. 2006. A modified sampler for uncontaminated DNA cores from soft sediments. *Journal of Archaeological Science* **33**: 573–574.

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

Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P, Pompanon F. 2010. An in silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**: 434–1572.

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

Ficetola GF, Taberlet P, Coissac E. 2016. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* **16**: 604–607.

Fulton TL. 2012. Setting Up an Ancient DNA Laboratory.Shapiro B, Hofreiter M eds. Methods in Molecular Biology. Ancient DNA: methods and protocols. New York, Dordrecht, Heidelberg, London: Springer, 1–12.

Furlan EM, Gleeson D, Hardy CM, Duncan RP. 2016. A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources* **16**: 641–654.

Giguet-Covex C, Pansu J, Arnaud F, Rey P-J, Griggo C, Gielly L, Domaizon I, Coissac E, David F, Choler P, *et al*. 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications* **5**: 3211.

Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. 2005. Assessing ancient DNA studies. *Theoretical and Applied Genetics* **20**: 541–544.

Haile J, Holdaway R, Oliver K, Bunce M, Gilbert MTP, Nielsen R, Munch K, Ho SYW, Shapiro B, Willerslev E. 2007. Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution* **24**: 982–989.

Hasegawa Y, Suyama Y, Seiwa K. 2009. Pollen donor composition during the early phases of

For Peer Review

reproduction revealed by DNA genotyping of pollen grains and seeds of Castanea crenata. New Phytologist **182**: 994–1002.

Hasegawa Y, Suyama Y, Seiwa K. 2015. Variation in pollen-donor composition among pollinators in an entomophilous tree species, Castanea crenata, revealed by single-pollen genotyping. *PLOS ONE* **10**: e0120393.

Hebsgaard MB, Phillips MJ, Willerslev E. 2005. Geologically ancient DNA: fact or artefact? *Trends in Microbiology* **13**: 212–220.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Higuchi R, Bowman B, Freiberger M, Ryder O, Wilson A. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**: 282–284.

Hirota SK, Nitta K, Suyama Y, Kawakubo N, Yasumoto AA, Yahara T. 2013. Pollinator-mediated selection on flower color, flower scent and flower morphology of Hemerocallis: evidence from genotyping individual pollen grains on the stigma. *PLOS ONE* **8**: e85601.

Hofreiter M, Mead JI, Martin P, Poinar H. 2003. Molecular caving. *Current Biology* **13**: R693–R695.

Hofreiter M, Serre D, Poinar H, Kuch M, Pääbo S. 2001. Ancient DNA. *Nature Reviews Genetics* **2**: 353–359.

Hu FS, Hampe A, Petit RJ. 2009. Paleoecology meets genetics: deciphering past vegetational dynamics. F*rontiers in Ecology and the Environment* **7**: 371–379.

Huang YT. 2014. Studies on carbon and DNA preservation in allophanic soils and paleosols on Holocene tephras in New Zealand. PhD thesis, University of Waikato, NZ.

Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377–386.

Hutchinson, G. E. (1957). A Treatise on Limnology, Vol. 1. Geography, Physics and Chemistry. Wiley, New York.

Isagi Y, Suyama Y. 2010. Single-Pollen Genotyping (Y Isagi and Y Suyama, Eds.). Tokyo: Springer Science & Business Media.

Jackson ST. 2012. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. *Quaternary Science Reviews* **49**: 1–15.

Jacobson GL, Bradshaw RHW. 1981. The selection of sites for paleovegetational studies. *Quaternary Research* **16**: 80–96.

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

Jørgensen T, Haile J, Moller P, AndreevV A, Boessenkool S, Rasmussen M, Kienast F, Coissac E, Taberlet P, Brochmann C, *et al.* 2012. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia

For Peer Review

reveals long-term vegetational stability. *Molecular Ecology* **21**: 1989–2003.

Jørgensen T, Kjær KH, Haile J, Rasmussen M, Boessenkool S, Andersen K, Coissac E, Taberlet P, Brochmann C, Orlando L, *et al*. 2011. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Molecular Ecology* **21**: 1980–1988.

Kong F, Yuan L, Zheng YF, Chen W. 2012. Automatic liquid handling for life science: a critical review of the current state of the art. *Journal of Laboratory Automation* **17**: 169–185.

Lahoz Monfort JJ, Guillera Arroita G, Tingley R. 2016. Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources* **16**: 673–685.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Larsen CPS, Macdonald GM. 1993. Lake morphometry, sediment mixing and the selection of sites for fine resolution palaeoecological studies. *Quaternary Science Reviews* **12**: 781–792.

Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, *et al*. 2011. The European Nucleotide Archive. *Nucleic Acids Research* **39**: D28–31.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lindgreen S, Adair KL, Gardner PP. 2016. An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports* **6**: 19233.

Lindhal T. 1993. Instability and decay of the primary structure of DNA. **362**: 709–715.

Little DP. 2014. A DNA mini-barcode for land plants. *Molecular Ecology Resources* **14**: 437–446.

Louvel G, Sarkissian Der C, Hanghøj K, Orlando L. 2016. metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. *Molecular Ecology Resources:* doi: 10.1111/1755-0998.12546

Lydolph MC, Jacobsen J, Arctander P, Gilbert MTP, Gilichinsky DA, Hansen AJ, Willerslev E, Lange L. 2005. Beringian Paleoecology Inferred from Permafrost-Preserved Fungal DNA. *Applied and Environmental Microbiology* **71**: 1012–1017.

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, Rubin EM, Jansson JK. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480**: 368–371.

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, *et al*. 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics* **48**: 1089–1093.

Matsuki Y, Isagi Y, Suyama Y. 2007. The determination of multiple microsatellite genotypes and DNA sequences from a single pollen grain. *Molecular Ecology Notes* **7**: 194–198.

For Peer Review

Matsuki Y, Tateno R, Shibata M, Isagi Y. 2008. Pollination efficiencies of flower-visiting insects as determined by direct genetic analysis of pollen origin. *American Journal of Botany* **95**: 925–930.

Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications 7: 11257.

Mitchell A, Bucchini F, Cochrane G, Denise H, Hoopen Ten P, Fraser M, Pesseat S, Potter S, Scheremetjew M, Sterk P, et al. 2016. EBI metagenomics in 2016--an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research* **44**: D595–603.

Murray DC, Coghlan ML & Bunce M, 2015. From Benchtop to Desktop: Important Considerations when Designing Amplicon Sequencing Workflows. *PLoS ONE*, **10**, p.e0124671.

Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T. 1999. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. *Planta* **209**: 53–65.

Nakazawa F, Uetake J, Suyama Y, Kaneko R, Takeuchi N, Fujita K, Motoyama H, Imura S, Kanda H. 2013. DNA analysis for section identification of individual Pinus pollen grains from Belukha glacier, Altai Mountains, Russia. *Environmental Research Letters* **8**: 014032.

Nesje A. 1992. A Piston Corer for Lacustrine and Marine Sediments. *Arctic and Alpine Research* **24**: 257.

Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics* **16**: 395–408.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, *et al.* 2014. Recalibrating Equus evolution using the genomesequence of an early Middle Pleistocene horse. *Nature* **498**: 74–78.

Ounit R, Wanamaker S, Close TJ, Lonardi S. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC *Genomics* **16**: 236.

Paffetti D, Vettori C, Caramelli D, Vernesi C, Lari M, Paganelli A, Paule L, Giannini R. 2007. Unexpected presence of Fagus orientalis complex in Italy as inferred from 45,000-year-old DNA pollen samples from Venice lagoon. *BMC Evolutionary Biology* **7**: S6.

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

Parducci L, Suyama Y. 2011. Single-Pollen Genotyping of Holocene Lake Sediments.Isagi Y, Suyama Y eds. Ecological Research Monographs. Single-Pollen Genotyping. Tokyo: Springer Japan, 101–109.

Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J, Matetovici I, Suyama Y, *et al.* 2012. Glacial survival of boreal trees in northern Scandinavia. *Science* **335**: 1083–1086.

Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, Haile J, Kjær KH, Larsen NK, Drouzas AD, Willerslev E. 2013. Molecular- and pollen-based vegetation analysis in lake

For Peer Review

sediments from central Scandinavia. *Molecular Ecology* **22**: 3511–3524.

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

Parducci L, Väliranta M, Salonen JS, Ronkainen T, Matetovici I, Fontana SL, Eskola T, Sarala P, Suyama Y. 2015. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20130382.

Paus A, Boessenkool S, Brochmann C, Epp LS, Fabel D, Haflidason H, Linge H. 2015. Lake Store Finnsjøen – a key for understanding Lateglacial/early Holocene vegetation and ice sheet dynamics in the central Scandes Mountains. *Quaternary Science Reviews* **121**: 36–51.

Pääbo S. 1984. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* **314**: 644–645.

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

Pe**dersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E, *et al*. 2015.** Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20130383.

Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, *et al*. 2016. Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**: 45–49.

Petersen G, Johansen B, Sedeberg O. 1996. PCR and sequencing form a sing pollen grain. *Plant Molecular Biology* **31**: 189–191.

Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P. 2009. Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils* **45**: 219–235.

Pornon A, Escaravage N, Burrus M, Holota H, Khimoun A, Mariette J, Pellizzari C, Iribar A, Etienne R, Taberlet P, *et al*. 2016. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports* **6**: 27282.

Porter TM, Golding GB, King C, Froese D, Zazula G, Poinar HN. 2013. Amplicon pyrosequencing late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale reproducibility. *Molecular Ecology Resources* **13**: 798–810.

Poté J, Ackermann R, Wildi W. 2009. Plant leaf mass loss and DNA release in freshwater sediments. *Ecotoxicology and Environmental Safety* **72**: 1378–1383.

Poté J, Rosselli W, Wigger A, Wildi W. 2007. Release and leaching of plant DNA in unsaturated soil column. *Ecotoxicology and Environmental Safety* **68**: 293–298.

Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, Huang Y-T, Cooper A. 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science* **29**: 610–626.

For Peer Review

Renberg I, Hansson H. 2008. The HTH sediment corer. *Journal of Paleolimnology* **40**: 655–659.

Roberts AV. 2007. The use of bead beating to prepare suspensions of nuclei for flow cytometry from fresh leaves, herbarium leaves, petals and pollen. *Cytometry* Part A **71**: 1039–1044.

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLOS ONE 7: e34131.

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

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, Willerslev E, Krogh A, Orlando L. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**: 178.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* **9**: 811–814.

Shapiro B, Hofreiter M. 2012. Ancient DNA: methods and protocols (B Shapiro and M Hofreiter, Eds.). New York Dordrecht Heidelberg London: Springer.

Sharon I, Banfield JF. 2013. Microbiology. Genomes from metagenomics. *Science* **342**: 1057–1058.

Sjögren P, Edwards ME, Gielly L, Langdom CT, Croudace IW, Merkel MKF, Fonville T, Alsos IG. 2016. Lake sedimentary DNA accurately records 20th century introductions of exotic conifers in Scotland. *New Phytologist*: doi: 10.1111/nph.14199

Sobek S, Durisch-Kaiser E, Zurbrügg R. 2009. Organic carbon burial efficiency in lake sediments controlled by oxygen exposure time and sediment source. *Limnology and Oceanography* **54**: 2243–2254.

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG, *et al*. 2009. Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology* **6**: 16.

Southworth D. 1974. Solubility of pollen exines. American Journal of Botany 61: 36–44.

Stewart JR, Lister AM, Barnes I, Dalén L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B*: *Biological Sciences* **277**: 661–671.

Stull GW, Moore MJ, Mandala VS, Douglas NA, Kates H-R, Qi X, Brockington SF, Soltis PS, Soltis DE, Gitzendanner MA. 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Applications in Plant Sciences **1**: 1200497.

Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger SA, Kultima JR, Coelho LP, Arumugam M, Tap J, Nielsen HB, *et al*. 2013. Metagenomic species profiling using universal

For Peer Review

phylogenetic marker genes. *Nature Methods* **10**: 1196–1199.

Suyama Y, Matsuki Y. 2015. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports* **5**: 16963.

Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996. DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. *Genes & Genetic Systems* **71**: 145–149.

Sønstebø JH, Gielly L, Brysting AK, Elven R, Edwards M, Haile J, Willerslev E, Coissac E, Rioux D, Sannier J, *et al*. 2010. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources* **10**: 1009–1018.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012a. Environmental DNA. *Molecular Ecology* **21**: 1789–1793.

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

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

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

Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. *Theoretical and Applied Genetics* **24**: 110–117.

Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. 2013. Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiology Reviews* **37**: 936–954.

Wang Y, Navin NE. 2015. Advances and applications of single-cell sequencing technologies. *Molecular Cell* **58**: 598–609.

Warinner C, Speller C, Collins MJ. 2015. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. P*hilosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20130376.

Weiß CL, Dannemann M, Prufer K, Burbano HA, Pickrell JK. 2015. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *eLife* **4**: e10005.

Wilke A, Bischof J, Harrison T, Brettin T, D'Souza M, Gerlach W, Matthews H, Paczian T, Wilkening J, Glass EM, *et al*. 2015. A RESTful API for accessing microbial community data for MG-RAST. *PLoS Computational Biology* **11**: e1004008.

Willerslev E, Cappellini E, Boomsma W, Nielsen R, Hebsgaard MB, Brand TB, Hofreiter M, Bunce M, Poinar HN, Johnsen S, *et al*. 2007. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**: 111–114.

For Peer Review

Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, Lorenzen ED, Vestergård M, Gussarova G, Haile J, *et al*. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**: 47–51.

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

Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P. 1999. Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences USA* 96: 8017–8021.

Wetzel, R. G. 2001. Limnology. Lake and River Ecosystems, 3rd ed., Academic press, San Diego.

Wood DE, **Salzberg SL. 2014.** Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15: R46.**

Wood JR, Wilmshurst JM. 2016. A protocol for subsampling Late Quaternary coprolites for multi-proxy analysis. *Quaternary Science Reviews* **138**: 1–5.

Wright HE, Mann DH, Glaser PH. 1984. Piston Corers for Peat and Lake Sediments. *Ecology* **65**: 657–659.

Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Stedingk Von H, Brysting AK, Coissac E, Pompanon F, *et al*. 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* **21**: 3647–3655.

Yoccoz N. 2012. The future of environmental DNA in ecology. *Molecular Ecology* **21**: 2031–2038.

Zhang Q, Liu Y. 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant and Cell Physiology* **44**: 1–11.

Ziesemer KA, Mann AE, Sankaranarayanan K, Schroeder H, Ozga AT, Brandt BW, Zaura E, Waters-Rist A, Hoogland M, Salazar-García DC, *et al*. 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* **5**: 16498.

For Peer Review

1168 **Figure legends**

1169

1170 **Figure 1.**

1171 Number of publications retrieved from Web of Science data bank in August 2016, using
1172 'ancient DNA' as search term, categorized according to the study organisms.

1173

1174 **Figure 2.**

1175 Biotic palaeoenvironmental proxies in lake sediments. **a**, sequential environmental
1176 development for a temperate region, in which the lake sediments start accumulating as
1177 glacial ice retreats, incorporating glacially eroded debris and the sparse pioneering biota (1),
1178 which later is replaced by a tundra-steppe community (2), then the boreal forest establishes
1179 (3) before eventually being replaced by a temperate forest (4). **b**, by identifying organisms
1180 detectable by DNA, macro- and microfossils and accumulated and preserved in the lake
1181 sediments, **c**, it is possible to reconstruct the environments through time. It is important to
1182 notice that preservation or rate of degradation is strongly correlated with the age of the
1183 sediments and that the input concentration **d**, varies in different climatic environments from
1184 these three proxies. **e**, In addition the resulting DNA profile, as well as macro and
1185 microfossils is influenced by taphonomic processes such as differences in biomass
1186 production and the distance from source to deposit. This is why a combination of all these
1187 three proxies makes a more robust palaeoenvironmental reconstruction.

1188

1189 **Figure 3.**

1190 Spatial distribution of larger lakes (blue squares) in the world (source: Natural Earth). The
1191 map shows that lakes are widely distributed geographically and present in many different
1192 environments. Photos exemplify different types of lake environments, (1) Lake Comarum,
1193 South Greenland. (2) Fiskvatn, Troms, Norway. (3) Olive-backed lake, Kamchatka. (4) Lake
1194 Milluni, Bolivia. (5) Buyan, Bali. (6) Lake Waikaremoana, North Island, New Zealand.

1195

1196 **Figure 4.**

1197 A full bodysuit, shoe cover, hairnet, facemask, hood, gloves, and sleeve guards are necessary
1198 during DNA extraction from a sediment core for aDNA analyses in an aDNA lab (a). During

For Peer Review

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

Figure 5.

Schematic illustration of the changes in the organellar DNA from the four types of generative cells present in pollen. When chloroplast (p) or mitochondrial (m) DNA is present in the mature generative pollen cells (m+ or p+), the DNA content per organelle increases after pollen mitosis one (PMI), and it decreases when organellar DNA is absent (m- or p-). All the changes in organellar DNA in the generative cells, whether an increase or a decrease, occur just after PMI, and the mitochondrial DNA and plastid DNA contents are regulated independently in the vegetative cells. An increase in the organellar DNA occurs with biparental/paternal inheritance (B/P) of the corresponding organelle, while a decrease occurs with maternal inheritance (M) of the corresponding organelle. Figure redrawn from Nagata *et al*. (1999).

Figure 6.

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

Figure 7.

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

For Peer Review

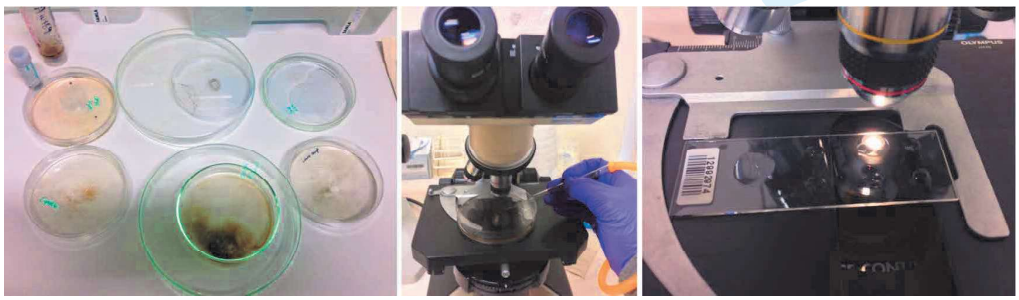
Box 1.

Isolating single pollen grains for direct PCR

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

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

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



For Peer Review

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

For Peer Review

1260 **Table 1.**

1261 Ancient DNA related publications from lake sediments sorted by year. Studies have been
1262 organized according to sample type, method used for DNA extraction [environmental DNA
1263 (eDNA) or DNA isolated from tissues] and research environment investigated.

1264

Author	Year	Sample type	Title	Method	Region	Environment
MJL Coolen	1998	bacteria	Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment	eDNA	Mahoney Lake, British Columbia, Canada	temperate, low altitude
PA Limburg	2002	daphnia	'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database	isolation	Northern German lake (Belauer See),	temperate, low altitude
VA Reid	2002	daphnia	A technique for the molecular genetic analysis of *Daphnia* resting eggs from sub-recent lake sediments	isolation	English Lake District, UK	temperate, low altitude
HG Pollard	2003	daphnia	Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination	isolation	Hannah Lake, Sudbury, Ontario, Canada	temperate, low altitude
MJL Coolen	2004	diatoms	Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake	eDNA	Ace Lake at the Vestfold Hills, eastern Antarctica	antarctic, low altitude
A Bissett	2005	copepods	Isolation, amplification, and identification of ancient copepod DNA from lake sediments	isolation	Lake Terrasovoje, Antarctica	antarctic, low altitude
L Parducci	2005	pollen	Ancient DNA from pollen: a genetic record of population history in Scots pine	isolation	Central northern Norway	temperate, low altitude
WJ D'Andrea	2006	algae	Alkenone producers inferred from well-preserved 18S rDNA in Greenland lake sediments	filtration > eDNA	Western Greenland	arctic, low altitude
S Marková	2006	daphnia	Are they still viable? Physical conditions and abundance of *Daphnia pulicaria* resting eggs in sediment cores from lakes in the Tatra Mountains	isolation	Lakes on Tatra Mts, Carpathians	temperate, high altitude
H Jiang	2007	bacteria	Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau	eDNA	Lake Chaka, Tibetan plateau	temperate, high altitude

J Mergeay	2007	daphnia	Extinction, recolonization, and dispersal through time in a planktonic crustacean	isolation	Lake Naivasha , Kenya	temperate, low altitude
E Matisoo-Smith	2008	plants and metazoan	Recovery of DNA and pollen from New Zealand lake sediments	eDNA	Round Lake, New Zealand	temperate, low altitude
MJL Coolen	2008	bacteria	Sources for sedimentary bacteriohopanepolyols as revealed by 16S rDNA stratigraphy	eDNA	Ace Lake, Antarctica	antarctic low altitude
J Madeja	2009	bacteria> human	Bacterial ancient DNA as an indicator of human presence in the past: its correlation with palynological and archaeological data	eDNA	Great Mazurian Lake District NE Poland	temperate, low altitude
H Kojima	2009	bacteria	DNA-based analysis of planktonic methanotrophs in a stratified lake	filtration > eDNA	Lake Mizugaki, central Japan	temperate, low altitude
LS Epp	2010	rotifers	Historical genetics on a sediment core from a Kenyan lake: intraspecific genotype turnover in a tropical rotifer is related to past environmental changes	eDNA	Lake Sonachi, Rift Valley in Kenya	temperate, high altitude
J Madeja	2010	bacteria> human	Integrated palynological and molecular analyses of late Holocene deposits from Lake Miłkowskie (NE Poland): verification of local human impact on environment	eDNA	Lake Miłkowskie, NE Poland	temperate, low altitude
LL Anderson-Carpenter	2011	plants	Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics	isolation	Great Lakes, N America	temperate, low altitude
O Savichtcheva	2011	bacteria	Quantitative PCR enumeration of total/toxic *Planktothrix rubescens* and total cyanobacteria in preserved DNA isolated from lake sediments	eDNA + isolataion	Lake Geneva, Bourget, Annecy, French Alps	temperate, low altitude
ZH Xu	2011	bacteria	DNA extraction, amplification and analysis of the 28S rRNA portion in sediment-buried copepod DNA in the Great Wall Bay and Xihu Lake, Antarctica	eDNA	Great Wall Bay, Xihu Lake, Antarctica	antarctic low altitude
LS Epp	2011	diatom	Molecular profiling of diatom assemblages in tropical lake sediments using taxon-specific PCR and Denaturing High-Performance Liquid Chromatography (PCR-DHPLC)	eDNA	Kenya, Lake Naivasha	tropical, high altitude
E Magyari	2011	plants	Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses	isolation	Retezat Mountains, South Carpathians	temperate, high altitude
L Parducci	2012	plants	Glacial survival of boreal trees in northern Scandinavia	eDNA	Central northern	temperate, low altitude

					Norway	
KR Stoof-Leichsenring	2012	diatoms	Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation	eDNA	Kenya, Lake Naivasha	tropical, high altitude
J Xiong	2012	bacteria	Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau	eDNA	Tibetan Plateau	temperate, high altitude
DF Ravasi	2012	bacteria	Development of a real-time PCR method for the detection of fossil 16S rDNA fragments of phototrophic sulfur bacteria in the sediments of Lake Cadagno.	eDNA?	Swiss Alps, Lake Cadagno	temperate, high altitude
M Winther Pedersen	2013	plants	A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa	eDNA	Greenland	arctic, low altitude
L Parducci	2013	plants	Molecular-and pollen-based vegetation analysis in lake sediments from central Scandinavia	eDNA	Central Norway	temperate, low altitude
I Domaizon	2013	bacteria	DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus* assemblages	eDNA	Lake Bourget, French Alps	temperate, low altitude
C Giguet-Covex	2014	plants and metazoan	Long livestock farming history and human landscape shaping revealed by lake sediment DNA	eDNA	Lake Anterne, Northern French Alps	temperate, high altitude
S Boessenkool	2014	plants	Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity	eDNA	Mt Gahinga, Albertine Rift, E Africa	temperate, high altitude
S Belle	2014	bacteria	Temporal changes in the contribution of methane-oxidizing bacteria to the biomass of chironomid larvae determined using stable carbon isotopes and ancient DNA	eDNA	Lake Narlay in the Jura Mountains	temperate, low altitude
W Hou	2014	plankton	Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan …	eDNA	Tibetan Plateau	temperate, high altitude
M Randlett	2014	phytoplankton	Alkenone distribution in Lake Van sediment over the last 270 ka: influence of temperature and haptophyte species composition	eDNA	Lake Van, Turkey	temperate, high altitude

J Madeja	2015	bacteria> human	A new tool to trace past human presence from lake sediments: the human-specific molecular marker Bacteroides strain HF 183	eDNA	Great Mazurian Lake District NE Poland	temperate, low altitude
M Kyle	2015	bacteria	Amplification of DNA in sediment cores to detect historic *Planktothrix* occurrence in three Norwegian lakes	eDNA	Bjørkelangen, Gjersjøen, Hemnessjø lakes, southern Norway	temperate, low altitude
S Pal	2015	bacteria	Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores	eDNA	Gatineau Park lakes,Western Quebec	temperate, low altitude
J Pansu	2015	plants	Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA	eDNA	Lake Anterne, Northern French Alps	temperate, high altitude
A Paus	2015	plants	Lake Store Finnsjøen – a key for understanding Lateglacial/early Holocene vegetation and ice sheet dynamics in the central Scandes Mountains	eDNA	Dovre, Central Norway.	temperate, low altitude
KR Stoof-Leichsenring	2015	algae	Genetic data from algae sedimentary DNA reflect the influence of environment over geography	eDNA	North Siberia	arctic, low altitude
L Epp	2015	algae vascular plants	Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes	eDNA	Greenland	arctic, low altitude
E Capo	2015	bacteria	Is Planktonic Diversity Well Recorded in Sedimentary DNA? Toward the Reconstruction of Past Protistan Diversity?	eDNA	Lake Bourget, Alps	temperate, low altitude
J Yang	2015	bacteria	Sedimentary archaeal amoA gene abundance reflects historic nutrient level and salinity fluctuations in Qinghai Lake, Tibetan Plateau	eDNA	Qinghai Lake, Tibetan Plateau	temperate, high altitude
AJ Poulain	2015	bacteria	Microbial DNA records historical delivery of anthropogenic mercury	eDNA	Aquatuk, Hawley, North Raft lakes, Ontario, Canada,	arctic, low altitude
MJ Wooller	2015	fish	Post-glacial dispersal patterns of Northern pike inferred from an 8800 year old pike (*Esox* cf. *lucius*) skull from interior Alaska	isolation	Quartz Lake, Alaska.	arctic, low altitude

D Etienne	2015	bacteria> human	Two thousand–year reconstruction of livestock production intensity in France using sediment-archived fecal Bacteroidales and source-specific mitochondrial markers	eDNA	Lorraine Plateau, NE France	temperate, low altitude
JC Stager	2015	fish	Of Paleo-Genes and Perch: What if an "Alien" Is Actually a Native?	eDNA	Lower Saint Regis Lake	temperate, low altitude
G Li	2016	phytoplankton	Temporal Succession of Ancient Phytoplankton Community in Qinghai Lake and Implication for Paleo-environmental Change	eDNA	Qinghai Lake, Tibetan Plateau	temperate, high altitude
GA Vuillemin	2016	bacteria >taphonomy	Recording of climate and diagenesis through sedimentary DNA and fossil pigments at Laguna Potrok Aike, Argentina.	eDNA	Laguna Potrok Aike, Argentina.	temperate, low altitude
I Alsos Greve	2016	plants	Sedimentary ancient DNA from Lake Skartjørna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change.	eDNA	Lake Skartjørna, Arctic archipelago of Svalbard	arctic, low altitude
M Winther Pedersen	2016	plants	Postglacial viability and colonization in North America's ice-free corridor	eDNA	Alberta and British Columbia, Canada	temperate, low altitude
P Sjögren	2016	plants	Lake sedimentary DNA accurately records 20th century introductions of exotic conifers in Scotland	eDNA	Spectacle Loch and Loch of the Lowes, Scotland, UK	temperate, low altitude

1265

Review

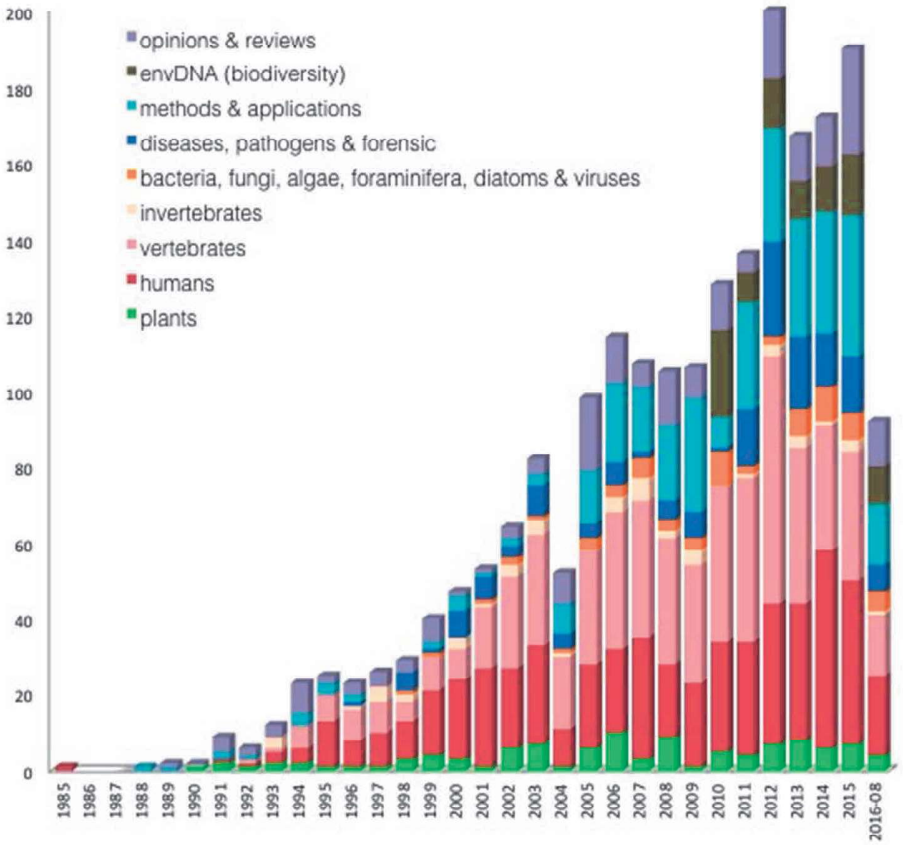


Figure 1.
Number of publications retrieved from Web of Science data bank in August 2016, using 'ancient DNA' as search term, categorized according to the study organisms.

Fig. 1
196x211mm (72 x 72 DPI)

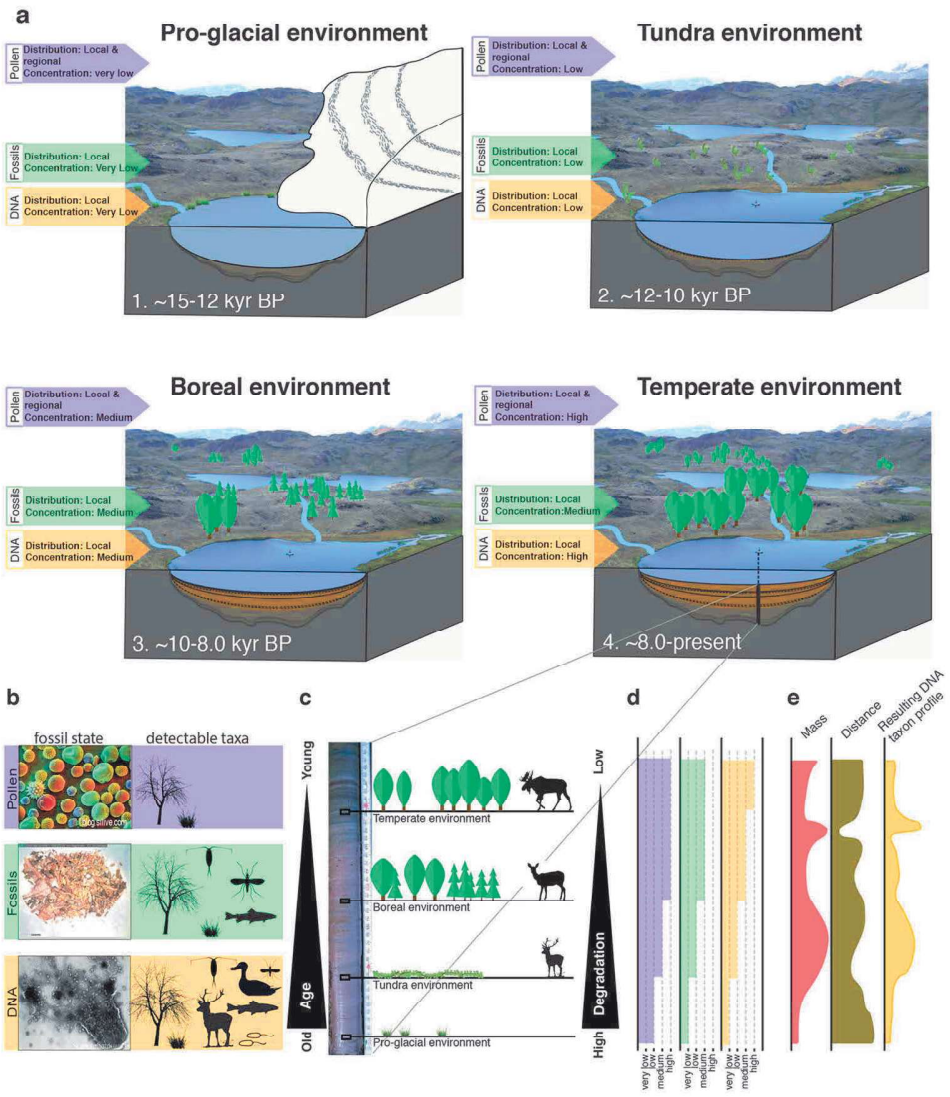


Figure 2.
Biotic palaeoenvironmental proxies in lake sediments. a, sequential environmental development for a temperate region, in which the lake sediments start accumulating as glacial ice retreats, incorporating glacially eroded debris and the sparse pioneering biota (1), which later is replaced by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

Fig. 2

127x144mm (300 x 300 DPI)

For Peer Review

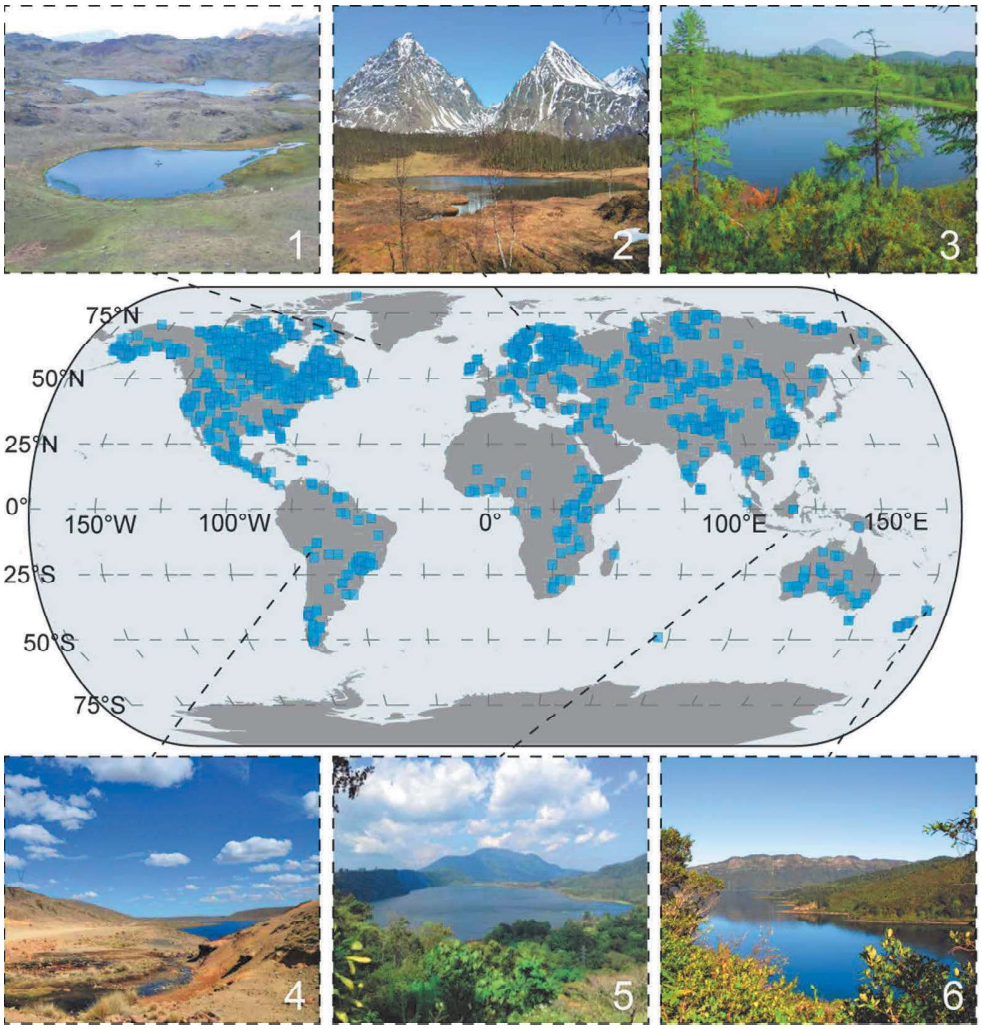


Figure 3.
Spatial distribution of larger lakes (blue squares) in the world (source: Natural Earth). The map shows that lakes are widely distributed geographically and present in many different environments. Photos exemplify different types of lake environments, (1) Lake Comarum, South Greenland. (2)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

Fig. 3
93x98mm (300 x 300 DPI)

Figure 4.
A full bodysuit, shoe cover, hairnet, facemask, hood, gloves, and sleeve guards are necessary during DNA extraction from a sediment core for aDNA analyses in an aDNA lab (a). During subsampling in a clean lab, surface contamination from the sediment core is removed with sterilized razors (b), and non-contaminated material from within the intact cores is extracted for DNA extraction (c).

Fig. 4
401x375mm (72 x 72 DPI)

Figure 4.
A full bodysuit, shoe cover, hairnet, facemask, hood, gloves, and sleeve guards are necessary during DNA extraction from a sediment core for aDNA analyses in an aDNA lab (a). During subsampling in a clean lab, surface contamination from the sediment core is removed with sterilized razors (b), and non-contaminated material from within the intact cores is extracted for DNA extraction (c).

Fig. 4
1151x863mm (72 x 72 DPI)

Figure 4.
A full bodysuit, shoe cover, hairnet, facemask, hood, gloves, and sleeve guards are necessary during DNA extraction from a sediment core for aDNA analyses in an aDNA lab (a). During subsampling in a clean lab, surface contamination from the sediment core is removed with sterilized razors (b), and non-contaminated material from within the intact cores is extracted for DNA extraction (c).

Fig. 4
1151x863mm (72 x 72 DPI)

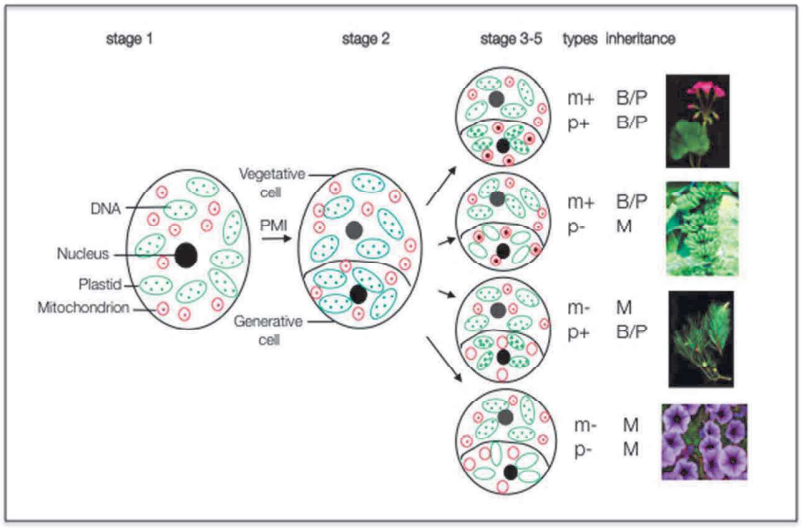


Figure 5.
Schematic illustration of the changes in the organellar DNA from the four types of generative cells present in pollen. When chloroplast (p) or mitochondrial (m) DNA is present in the mature generative pollen cells (m+ or p+), the DNA content per organelle increases after pollen mitosis one (PMI), and it decreases when organellar DNA is absent (m- or p-). All the changes in organellar DNA in the generative cells, whether an increase or a decrease, occur just after PMI, and the mitochondrial DNA and plastid DNA contents are regulated independently in the vegetative cells. An increase in the organellar DNA occurs with biparental/paternal inheritance (B/P) of the corresponding organelle, while a decrease occurs with maternal inheritance (M) of the corresponding organelle. Figure redrawn from Nagata et al. (1999).

Fig. 5
254x190mm (72 x 72 DPI)

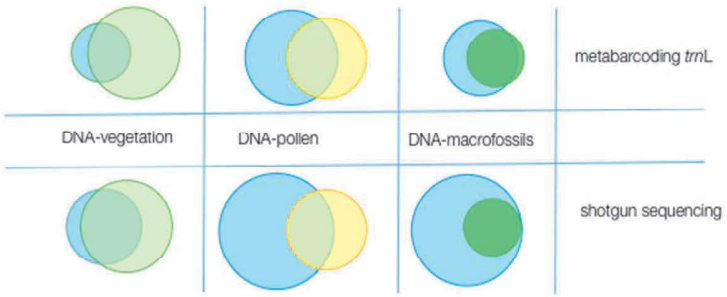


Figure 6.
Venn diagrams showing the proportion of plant taxa commonly detected by different proxies: metabarcoding and shotgun metabarcoding of sedaDNA (blue), vegetation survey (light green), pollen (yellow) and macrofossils (dark green). These proportions have so far varied among different studies and the lower panel shows what we can theoretically expect with shotgun metabarcoding when a full genome reference library is developed.

Fig. 6
254x142mm (72 x 72 DPI)

review

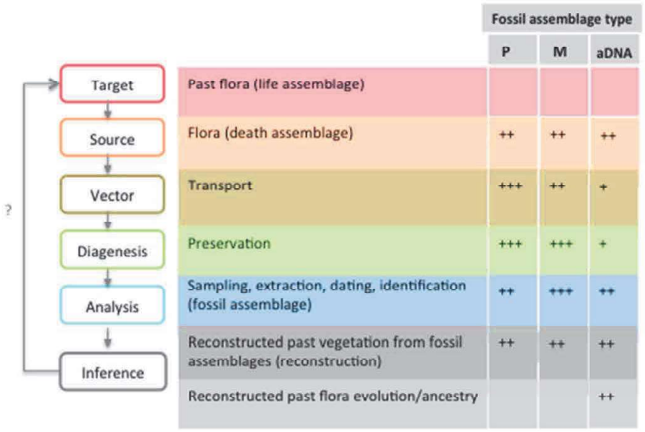


Figure 7.
Representation of the chain of processes involved in the transformation of plant information present in the three lake sediment assemblages types: pollen (P), macrofossils (M), ancient DNA (aDNA). Current understanding of the processes is indicated as good (+++), reasonable (++) or poor (+). The publications upon which the levels of understanding are based are shown in Table 1. Figure redrawn from & Birks (2016) and originally based on Jackson's (2012) general conceptual model for the representation of floristic material in palaeoecological assemblages.
Fig. 7
254x190mm (72 x 72 DPI)