



Tansley review

Ancient plant DNA in lake sediments

Authors for correspondence:

Laura Parducci

Tel: +46 18 471 64 14

Email: laura.parducci@ebc.uu.se

Mikkel Winther Pedersen

Tel: +45 29 27 53 42

Email: mwpedersen@snm.ku.dk

Received: 1 October 2016

Accepted: 7 December 2016

**Laura Parducci¹, Keith D. Bennett^{2,3}, Gentile Francesco Ficetola^{4,5},
Inger Greve Alsos⁶, Yoshihisa Suyama⁷, Jamie R. Wood⁸ and
Mikkel Winther Pedersen⁹**

¹Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, Uppsala 75236, Sweden;

²Department of Geography & Sustainable Development, School of Geography & Geosciences, University of St Andrews, St Andrews Fife, KY16 9AL, UK; ³Marine Laboratory, Queen's University Belfast, Portaferry, BT22 1LS, UK; ⁴CNRS, Université Grenoble-Alpes,

Laboratoire d'Ecologie Alpine (LECA), Grenoble F-38000, France; ⁵Department of Biosciences, Università degli Studi di Milano,

Milan 20133, Italy; ⁶Tromsø Museum, UiT – The Arctic University of Norway, Tromsø NO-9037, Norway; ⁷Field Science Center, Graduate School of Agricultural Science, Tohoku University, 232-3 Yomogida, Naruko-onsen, Osaki, Miyagi 989-6711, Japan; ⁸Long-

term Ecology Lab, Landcare Research, PO Box 69040, Lincoln Canterbury 7640, New Zealand; ⁹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen 1350, Denmark

Contents

Summary	1	VIII. Recent key findings and future methods using HTS techniques	10
I. Introduction	2	IX. Challenges when studying aDNA from lake sediments	12
II. Environmental and temporal limits for survival of aDNA	3	X. Bioinformatic processing	13
III. Lake sediments	4	XI. Conclusions and directions for future research	14
IV. Perspective for plant aDNA research	6	Acknowledgements	15
V. Sources of pollen, macrofossils and DNA	7	References	15
VI. Pollen DNA	7		
VII. Sedimentary aDNA	8		

Summary

Recent advances in sequencing technologies now permit the analyses of plant DNA from fossil samples (ancient plant DNA, plant aDNA), and thus enable the 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 a few metres from the sampling point. Lakes have larger catchment areas and recent studies have suggested 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 the study of plant aDNA from lake sediments and discuss the progress made up to the present. We argue that aDNA analyses add new and additional perspectives for the study of 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.

New Phytologist (2017)
doi: 10.1111/nph.14470

Key words: ancient plant DNA (aDNA), bioinformatics, environmental DNA (eDNA), high-throughput DNA sequencing, lake sediments, metabarcoding, pollen, shotgun sequencing, taphonomy.

I. 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 the 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 have focused on vertebrates, especially humans (50%), whereas only 5% have focused on plants and ecology (Fig. 1). aDNA from skeletal remains has been so much easier to study than plant macrofossils, and the difficulty in obtaining aDNA from charred specimens, which represent 95% of the plant archaeological record, may partially explain the lack of plant aDNA papers. Plant aDNA studies have also been delayed by the difficulty in finding standard barcode regions useful for all species (universal) and sufficiently variable to discriminate between taxa. Three important trends are evident from Fig. 1: (1) the total number of studies has increased steadily throughout the last three decades; (2) the number of opinion and methodological papers has accounted for most of the increase since *c.* 2004; (3) 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: (1) the recent adoption of high-throughput DNA sequencing (HTS) technology; (2) the discovery that DNA can be isolated from fossil pollen (Suyama *et al.*, 1996; Pardiucci *et al.*, 2005); and (3) 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 (references 3–18 in Pedersen *et al.*, 2015). These studies (see also Table 1) show how DNA can often identify more species and at higher taxonomic

resolution than those identified by pollen and macrofossil analyses, thus providing important ecological and climatic information on the investigated sites that is otherwise difficult or impossible to infer (Sønstebo *et al.*, 2010; Pardiucci *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 the reconstruction of palaeoenvironments. The latter is an advantage compared with pollen, especially at high 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 specialized 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 the 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 for 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 methodologies used for taxonomic identification and authentication of aDNA sequences.

II. 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

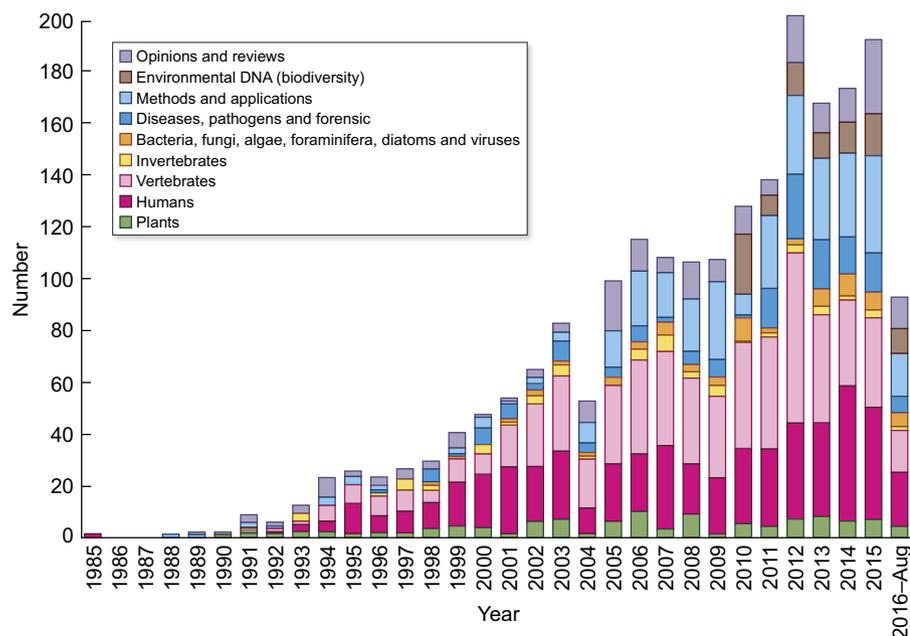


Fig. 1 Number of publications retrieved from the Web of Science data bank in August 2016, using ‘ancient DNA’ as search term, categorized according to the study organisms.

Table 1 Ancient DNA-related publications from lake sediments sorted by year

Author	Sample type	Method	Region	Environment
Coolen & Overmann (1998)	Bacteria	eDNA	Mahoney Lake, British Columbia, Canada	Temperate, low altitude
Limburg & Weider (2002)	Daphnia	Isolation	Belauer See Lake, Germany	Temperate, low altitude
Reid <i>et al.</i> (2002)	Daphnia	Isolation	English Lake District, UK	Temperate, low altitude
Pollard <i>et al.</i> (2003)	Daphnia	Isolation	Hannah Lake, Sudbury, Ontario, Canada	Temperate, low altitude
Coolen <i>et al.</i> (2004)	Diatoms	eDNA	Ace Lake, eastern Antarctica	Antarctic, low altitude
Bissett <i>et al.</i> (2005)	Copepods	Isolation	Lake Terrasovoje, Antarctica	Antarctic, low altitude
Parducci <i>et al.</i> (2005)	Pollen	Isolation	Holtjärnen, central Sweden	Temperate, low altitude
D'Andrea <i>et al.</i> (2006)	Algae	Filtration > eDNA	Søndre Strømfjord, western Greenland	Arctic, low altitude
Marková <i>et al.</i> (2006)	Daphnia	Isolation	Lakes on Tatra Mts, Carpathians, Hungary	Temperate, high altitude
Jiang <i>et al.</i> (2007)	Bacteria	eDNA	Lake Chaka, Tibetan plateau	Temperate, high altitude
Mergeay <i>et al.</i> (2007)	Daphnia	Isolation	Lake Naivasha, Kenya	Temperate, low altitude
Matisoo-Smith <i>et al.</i> (2008)	Plants and metazoa	eDNA	Round Lake, New Zealand	Temperate, low altitude
Coolen <i>et al.</i> (2008)	Bacteria	eDNA	Ace Lake, Antarctica	Antarctic, low altitude
Madeja <i>et al.</i> (2009)	Bacteria > human	eDNA	Great Mazurian Lake District, NE Poland	Temperate, low altitude
Kojima <i>et al.</i> (2009)	Bacteria	Filtration > eDNA	Lake Mizugaki, central Japan	Temperate, low altitude
Epp <i>et al.</i> (2010)	Rotifers	eDNA	Lake Sonachi, Rift Valley, Kenya	Temperate, high altitude
Madeja <i>et al.</i> (2010)	Bacteria > human	eDNA	Lake Miłkowskie, north-eastern Poland	Temperate, low altitude
Anderson-Carpenter <i>et al.</i> (2011)	Plants	Isolation	Great Lakes, North America	Temperate, low altitude
Savichtcheva <i>et al.</i> (2011)	Bacteria	eDNA + isolation	Lake Geneva, Bourget, Annecy, French Alps	Temperate, low altitude
Xu <i>et al.</i> (2011)	Bacteria	eDNA	Great Wall Bay, Xihu Lake, Antarctica	Antarctic, low altitude
Epp <i>et al.</i> (2011)	Diatom	eDNA	Lake Naivasha, Kenya	Tropical, high altitude
Magyari <i>et al.</i> (2011)	Plants	Isolation	Retezat Mountains, south Carpathians, Hungary	Temperate, high altitude
Parducci <i>et al.</i> (2012)	Plants	eDNA	Central Norway and Sweden	Temperate, low altitude
Stoof-Leichsenring <i>et al.</i> (2012)	Diatoms	eDNA	Kenya, Lake Naivasha	Tropical, high altitude
Xiong <i>et al.</i> (2012)	Bacteria	eDNA	Tibetan Plateau	Temperate, high altitude
Ravasi <i>et al.</i> (2012)	Bacteria	eDNA?	Lake Cadagno, Swiss Alps	Temperate, high altitude
Pedersen <i>et al.</i> (2013)	Plants	eDNA	Southern Greenland	Arctic, low altitude
Parducci <i>et al.</i> (2013)	Plants	eDNA	Central Norway and Sweden	Temperate, low altitude
Domaizon <i>et al.</i> (2013)	Bacteria	eDNA	Lake Bourget, French Alps	Temperate, low altitude
Boessenkool <i>et al.</i> (2013)	Plants	eDNA	Mt Gahinga, eastern Africa	Temperate, high altitude
Giguët-Covex <i>et al.</i> (2014)	Plants and metazoa	eDNA	Lake Anterne, French Alps	Temperate, high altitude
Belle <i>et al.</i> (2014)	Bacteria	eDNA	Lake Narlay, Jura Mountains, France	Temperate, low altitude
Hou <i>et al.</i> (2014)	Plankton	eDNA	Tibetan Plateau	Temperate, high altitude
Randlett <i>et al.</i> (2014)	Phytoplankton	eDNA	Lake Van, Turkey	Temperate, high altitude
Madeja (2015)	Bacteria > human	eDNA	Great Mazurian Lake District, Poland	Temperate, low altitude
Kyle <i>et al.</i> (2015)	Bacteria	eDNA	Bjørkelangen, Gjersjøen, Hemnessjø lakes, Norway	Temperate, low altitude
Pal <i>et al.</i> (2015)	Bacteria	eDNA	Gatineau Park lakes, western Quebec	Temperate, low altitude
Pansu <i>et al.</i> (2015)	Plants	eDNA	Lake Anterne, French Alps	Temperate, high altitude
Paus <i>et al.</i> (2015)	Plants	eDNA	Dovre region, central Norway	Temperate, low altitude
Stoof-Leichsenring <i>et al.</i> (2015)	Algae	eDNA	North Siberia	Arctic, low altitude
Epp <i>et al.</i> (2015)	Algae vascular plants	eDNA	North Greenland	Arctic, low altitude
Capo <i>et al.</i> (2015)	Bacteria	eDNA	Lake Bourget, French Alps	Temperate, low altitude
Yang <i>et al.</i> (2015)	Bacteria	eDNA	Qinghai Lake, Tibetan Plateau	Temperate, high altitude
Poulain <i>et al.</i> (2015)	Bacteria	eDNA	Hudson Bay, Ontario, Canada,	Arctic, low altitude
Wooller <i>et al.</i> (2015)	Fish	Isolation	Quartz Lake, Alaska	Arctic, low altitude
Etienne <i>et al.</i> (2015)	Bacteria > human	eDNA	Lorraine Plateau, north-eastern France	Temperate, low altitude
Stager <i>et al.</i> (2015)	Fish	eDNA	Lower Saint Regis Lake, Franklin County, NY, USA	Temperate, low altitude
Li <i>et al.</i> (2016)	Phytoplankton	eDNA	Qinghai Lake, Tibetan Plateau	Temperate, high altitude
Vuillemin <i>et al.</i> (2016)	Bacteria > taphonomy	eDNA	Laguna Potrok Aike, Argentina	Temperate, low altitude
Alsos <i>et al.</i> (2016)	Plants	eDNA	Lake Skartjørna, Svalbard	Arctic, low altitude
Pedersen <i>et al.</i> (2016)	Plants	eDNA	Peace River, Alberta and British Columbia, Canada	Temperate, low altitude
Sjøgren <i>et al.</i> (2017)	Plants	eDNA	Spectacle Loch, Loch of the Lowes, Scotland, UK	Temperate, low altitude

Studies have been organized according to sample type, method used for DNA extraction (environmental DNA (eDNA) or DNA isolated from tissues) and research environment investigated.

(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 BP

(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 the 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 the misincorporation of C to T and G to A transitions, primarily towards the ends of the 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 limited presence of nucleases, which reduce 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 yr for short (242 base pairs (bp)) mitochondrial DNA (mtDNA) fragments preserved at *c.* 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 > 1 million yr 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 kyr BP (Willerslev *et al.*, 2014), from ice cores dated between 450 and 800 kyr BP (Willerslev *et al.*, 2007) and from pollen *c.* 150 kyr BP (Suyama *et al.*, 1996), which suggests that DNA studies from lake sediments can potentially span the Late Pleistocene and Holocene.

III. Lake sediments

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

1. 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 the preservation of plant aDNA, as the sediments accumulate continuously, allowing the establishment of high-resolution molecular records using available geo-chronological methods (Fig. 2a–c), and thus

providing a robust record for the interpretation of the molecular history. In particular, sediment records from small lakes are excellent archives for molecular studies, as the effects of disturbances are low and seem to represent the surrounding terrestrial environment well (Fig. 3). 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 minimize bioturbation, water percolation and sediment reworking (Pansu *et al.*, 2015), all of which are major concerns for palaeoecological studies. Water has a maximum density at 4°C, and 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 on 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 in 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 in which 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. Although 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 embedded 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. However, re-deposition of sediments can occur in lakes, contaminating the micro- and macrofossil record with older material. Although 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 unaffected, probably as a result of 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.

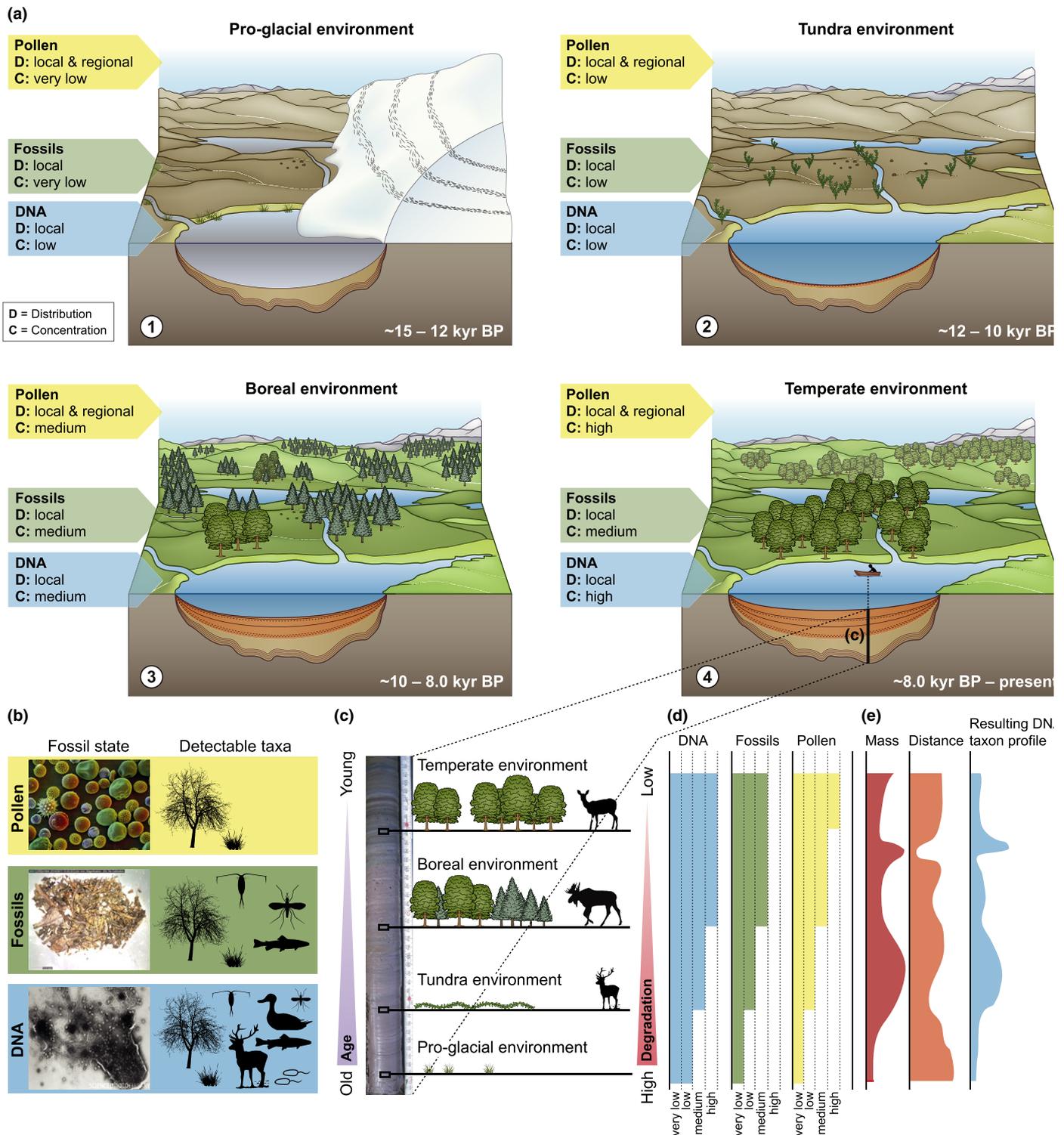


Fig. 2 Biotic palaeoenvironmental proxies in lake sediments. (a) Sequential environmental development for a temperate region, in which the lake sediments start to accumulate as the glacial ice retreats, incorporating glacially eroded debris and the sparse pioneering biota (1), which later is replaced by a tundra-steppe community (2); then the boreal forest establishes (3) before eventually being replaced by a temperate forest (4). By identifying organisms detectable by DNA, macro- and microfossils accumulated and preserved in the lake sediments (b), it is possible to reconstruct the environments through time (c). It is important to note that the 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. In addition, the resulting DNA profile (e), as well as macro- and microfossils, are influenced by taphonomic processes, such as differences in biomass production and the distance from source to deposit. This is why a combination of all three proxies makes a more robust palaeoenvironmental reconstruction. kyr BP, thousand years before present.

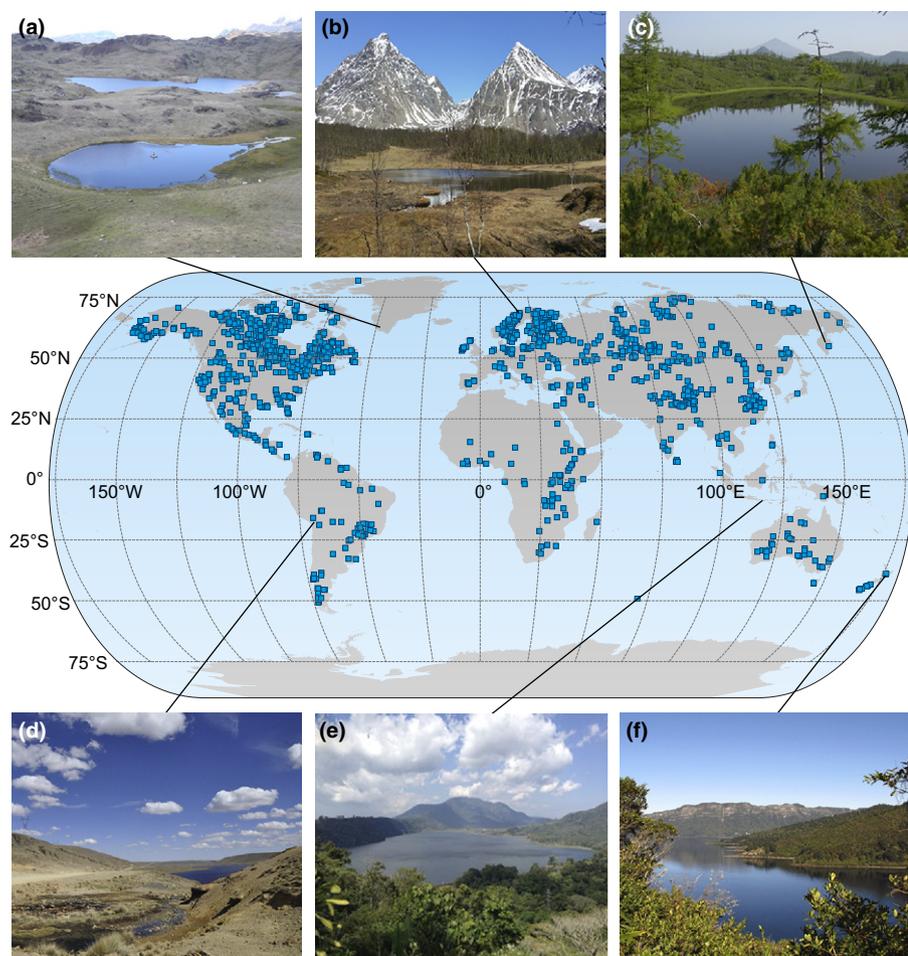


Fig. 3 Spatial distribution of larger lakes (blue squares) in the world (source: Natural Earth). The map shows that lakes are widely distributed geographically and present in many different environments. The photographs exemplify different types of lake environment: (a) Lake Comarum, south Greenland; (b) Fiskvatn, Troms, Norway; (c) Olive-Backed Lake, Kamchatka; (d) Lake Milluni, Bolivia; (e) Buyan, Bali; (f) Lake Waikaremoana, North Island, New Zealand.

2. Coring techniques for aDNA work

The collection of sediment cores for plant aDNA can be performed with the same coring devices as 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 (e.g. Wright *et al.*, 1984; Nesje, 1992), are 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 are essential for the collection of surface sediments. If the coring is performed 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 sterilization, are also available (Feek *et al.*, 2006, 2011). On extruding and opening the cores, however, it must be assumed that the surface is contaminated, and so subsamples must be taken from inside the undisturbed centre. During subsampling, 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–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 DNA-free water exposed to air in the laboratory during subsampling as a negative control.

IV. Perspective for plant aDNA research

Traditionally, lake sediment records have been routinely analysed for pollen and other plant fossils, which have been the basis of our understanding of the landscape-scale distribution of plants on millennial timescales. Plant aDNA now offers the potential of helping the elucidation of long-standing ecological questions which cannot be retrieved by classical palaeoecological techniques. 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; Alsos *et al.*, 2016; Pedersen *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



Fig. 4 A full bodysuit, shoe cover, hairnet, facemask, hood, gloves and sleeve guards are necessary during DNA extraction from a sediment core for ancient DNA (aDNA) analyses in an aDNA laboratory (a). During subsampling in a clean laboratory, 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).

latitudes in Europe (Parducci *et al.*, 2012). Here, combined information from fossils and molecular studies (modern and ancient) has changed our traditional view of the 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 though such analyses have recently reached better taxonomic resolution because of improved identification keys, the work is still time demanding, and identifications are often at the genus or family level, rarely species (Faegri *et al.*, 1989). 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 probably increase our ability to resolve plant taxa at the species level and, in addition, elucidate the ancestry and genetic composition of ancient plant populations. Likewise, metabarcoding and metagenomic analyses of *sedaDNA* will provide more detailed insights into ancient ecosystems and link changes more tightly to past climate shifts (see section VIII).

V. Sources of pollen, macrofossils 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 *c.* 70% of the deposited pollen is from vegetation growing within a few kilometres of the lake (Jacobson & Bradshaw, 1981). Although the majority of pollen often derives from high pollen-producing wind-pollinating plants, which is distributed regionally through the air, the 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 are often under-represented (Birks, 2003). *sedaDNA* seems to have a similar source of origin to macrofossils (Jørgensen *et al.*, 2012; Pedersen *et al.*, 2013; Alsos *et al.*, 2015; Parducci *et al.*,

2015). To date, only a few studies have focused on the release and deposition of DNA in the environment (Poté *et al.*, 2007, 2009; Pietramellara *et al.*, 2009; Barnes & Turner, 2016), and therefore much remains unknown about the ecology of the DNA – for example, all processes occurring from source to deposition (taphonomy). Current *sedaDNA* 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 embedded in the sediment matrix; thus, pollen DNA needs to be extracted directly from single or multiple isolated grains.

VI. 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 result 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 two to three cells (a large vegetative cell and one to two 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, such as internal transcribed spacer (ITS) ribosomal repeats, are also present in 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

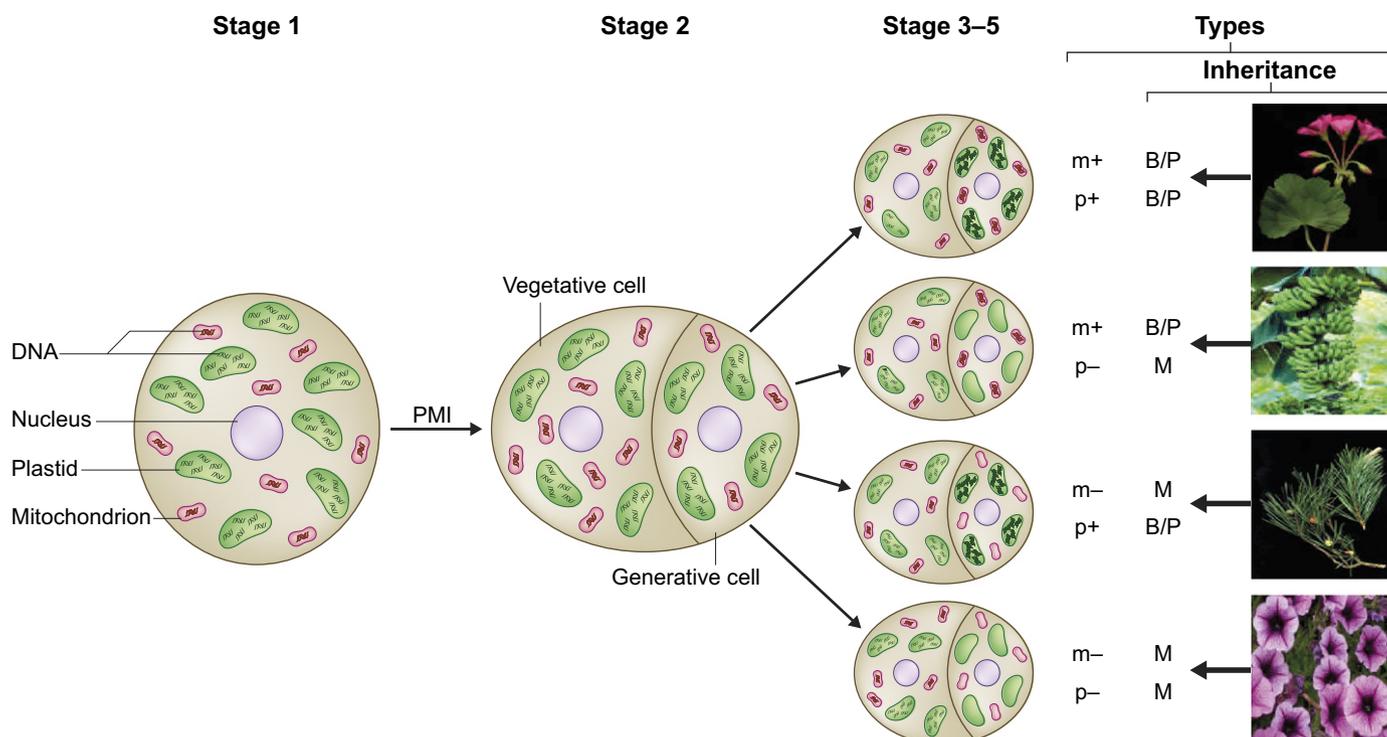


Fig. 5 Schematic illustration of the changes in the organellar DNA from the four types of generative cell present in pollen. When chloroplast (p) or mitochondrial (m) DNA is present in the mature generative pollen cells (m+ or p+), the DNA content per organelle increases after pollen mitosis one (PMI) and during developmental stages from uni-nucleate to bi- (or tri-) cellular pollen (stages 1–5), 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, whereas a decrease occurs with maternal inheritance (M) of the corresponding organelle. Figure modified from Nagata *et al.* (1999).

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. However, 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 mtDNA may be present in different amounts (in the generative cell/s) in different taxa.

1. Single-pollen genotyping

Petersen *et al.* (1996) were the first to amplify cpDNA from single pollen grains (*Hordeum* and *Secale*), whereas Suyama *et al.* (1996) were the first to amplify DNA from peat sediment pollen (fir, *Abies*) c. 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 the isolation and direct amplification 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-

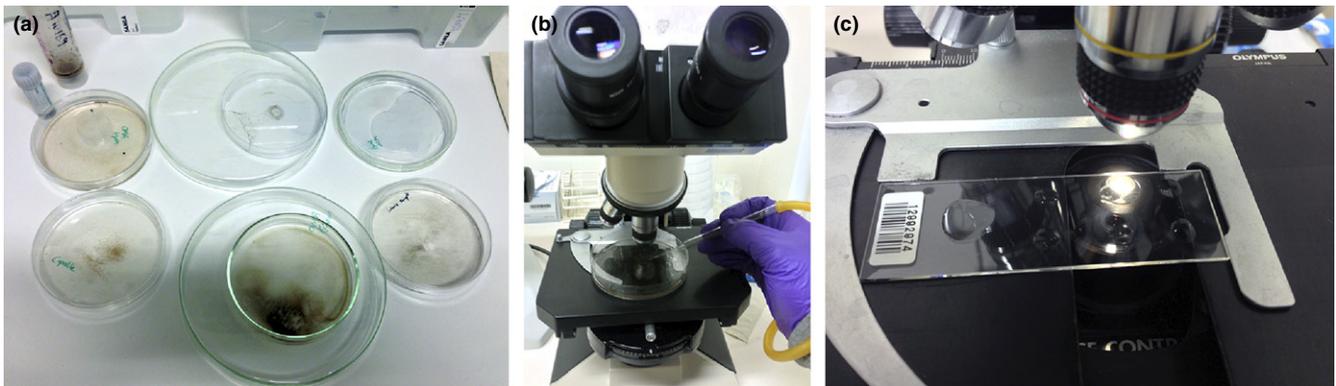
pollen genotyping methods on fresh pollen (Isagi & Suyama, 2010), it is also possible to perform paternity analysis and infer the pattern and distance of pollen dispersal in modern plant populations (Matsuki *et al.*, 2007, 2008; Hasegawa *et al.*, 2009, 2015; Hirota *et al.*, 2013). The potential of 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 should now be possible to investigate more efficiently individual fossil pollen grains and hence to conduct plant aDNA studies more effectively, even at the population level (see section VIII).

VII. Sedimentary aDNA

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 the extraction of DNA and the removal of inhibiting substrates require 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.

Box 1 Isolating single pollen grains for direct PCR

- (a) Sieving sediment. Place a small amount (*c.* 0.2–0.5 ml) of sediment onto a sterile filter with a 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.
- (b) Isolating pollen grains. Dilute sample so that grains are adequately spaced in the Petri dish, and scan at $\times 100$ – 200 magnification (depending on the size of the pollen). Once a target grain has been found, switch to a lower magnification, ensuring that 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 the pipette from the water. Transfer the water from the pipette onto a microscope slide. Check the droplet at $\times 200$ – 400 magnification to ensure that the correct pollen grain is present. For larger pollen grains and plant fragments, a 1–10 μ l pipette can be used to isolate the specimen.
- (c) Dilution. Use a 1–10 μ l 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 that 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.



(a–c) 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.

1. Proxy overlapping or complementary

Before a robust inference of past vegetation (presence and abundance of taxa) can be based on *sedaDNA*, 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; Parducci *et al.*, 2013, 2015; Pedersen *et al.*, 2013; 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 *sedaDNA* and macrofossils (Jørgensen *et al.*, 2012; Parducci *et al.*, 2012; Boessenkool *et al.*, 2013; Pedersen *et al.*, 2013; Porter *et al.*, 2013), whereas only one study has shown an overlap close to 100% (Alsos *et al.*, 2016). These differences and the limited overlap found between proxies may be caused by differences in taphonomic processes between sites, incompleteness of reference genomic databases, robustness of the experimental design (Ficetola *et al.*, 2016), and number of pollen/macrofossils counted vs sequencing depth (see sections VIII–X).

The most stringent test for the identification of the origin of *sedaDNA* 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 conventional above-ground surveys. To date, we still do not know whether such a good relationship holds with DNA in lake sediments. In a study from a high-altitude crater lake in Africa, Boessenkool *et al.* (2013) showed that *sedaDNA* largely reflected local flora. Similarly, Sjögren *et al.* (2016) compared DNA, pollen and historical vegetation maps in

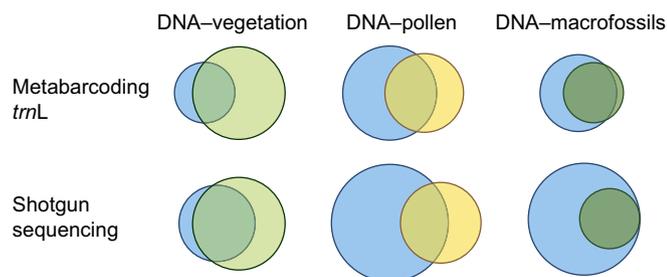


Fig. 6 Venn diagrams showing the proportion of plant taxa commonly detected by different proxies: metabarcoding and shotgun metabarcoding of *sedaDNA* (blue), vegetation survey (light green), pollen (yellow) and macrofossils (dark green). These proportions have so far varied among different studies and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

two Scottish lakes, and found that *seda*DNA was of local origin in contrast with regionally dispersed deciduous tree pollen. Finally, in a comparison of *seda*DNA with vegetation surveys of 11 lakes in northern Norway, the majority of taxa recorded in the *seda*DNA 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 that the increased information gained will lead to an almost complete overlap between DNA and macrofossils, but probably not between DNA and pollen (Fig. 6).

2. 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 are preserved in sediments, and the fact that some are preserved 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 intensities 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, including 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 contents to the surrounding environment and extracellular DNA can bind to charged mineralogenic 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 clay-rich soil can be highly suitable for the protection of DNA against degradation, as clay particles have a 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 transformation. Although it is unlikely that this will be a large source of 'plant' DNA, it remains a possibility (Pedersen *et al.*, 2015). Whether the plant DNA in lake sediments is primarily present as extracellular DNA or as small plant parts, such as leaves, root caps, cells, stems or fruits, remains unclear.

3. Inferring taxa abundance from PCR-based *seda*DNA analyses

As with macrofossils, aquatic plants are well represented in *seda*DNA from lake sediments. The detection of a species in modern or semi-modern sediments depends on both the distance to the lake shore and its abundance in the vegetation (Alsos *et al.*, 2015; Sjögren *et al.*, 2016). For example, the dominant species growing around lakes are found with a high proportion of DNA reads and present in more PCR repeats for most cases. This also seems to be the

case for ancient samples, as all species represented from at least one macrofossil are also detected in more than one PCR replicate (Alsos *et al.*, 2016). Nevertheless, for the time being, we should be very cautious about interpreting quantities of DNA beyond rough estimates when using PCR-based methods, as several metabarcoding processes may cause bias (e.g. primer binding site, amplicon length, taxonomy and diversity of extract) (Pornon *et al.*, 2016).

VIII. Recent key findings and future methods using HTS techniques

HTS expands on traditional PCR-based Sanger sequencing techniques and has facilitated the rapid development of aDNA research during the last decade. aDNA molecules converted to sequencing libraries can now be parallel sequenced massively on HTS platforms, such as the Illumina HiSeq or Complete Genomics platforms. HTS approaches have become increasingly affordable and are now routinely used by most aDNA laboratories, allowing the screening of *seda*DNA from a wide range of complex ancient substrates. With the newest generation of desktop HTS platforms, for example NextSeq, even small laboratories can now sequence their own datasets. There are, in principle, three HTS methodological strategies for the analysis of plant aDNA in lake sediments: metabarcoding or shotgun sequencing of *seda*DNA and HTS of pollen DNA. 'Metabarcoding' is a relatively well-established method, which relies on the information on one single locus; the second method is newer and relies on shotgun sequencing, that is sequencing a mixed pool of aDNA. For clarity, it is important to define this new method. Here, we suggest the term 'shotgun metabarcoding' when shotgun sequencing of environmental DNA is used to identify taxa, and 'metagenomics' when shotgun is used for functional analyses of the environments. This will allow researchers to discriminate between studies focusing on taxon identification and studies focusing on functional and attribute analysis. In the following paragraphs, we review and discuss the progress made and the future of these three methodologies.

1. Metabarcoding on *seda*DNA

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

Metabarcoding has already been used in a variety of palaeoenvironmental studies (Willerslev *et al.*, 2003; Valentini *et al.*, 2009; Taberlet *et al.*, 2012b), including permafrost (Lydolph *et al.*, 2005; Jørgensen *et al.*, 2011, 2012; Willerslev *et al.*, 2014), mid- to high-latitude/altitude lake sediments (Coolen & Gibson, 2009; Parnucci *et al.*, 2012; Alsos *et al.*, 2015, 2016; Epp *et al.*, 2015; Pansu *et al.*, 2015; Paus *et al.*, 2015), tropical lake sediments (Boessenkool *et al.*, 2013; G. F. Ficetola, unpublished) and deposits preserved

under ice-sheets, even in the absence of visible macrofossil remains (Willerslev *et al.*, 1999, 2007).

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

The performance of the metabarcoding approach is often limited if markers are not able to amplify all the present taxa (universality), and by the capacity to differentiate and identify closely related species (resolution) (Ficetola *et al.*, 2010; Sønstebo *et al.*, 2010; Furlan *et al.*, 2016). Therefore, the use of primers with appropriate features (i.e. 'universal' primers amplifying with high specificity all plants, and achieving a high taxonomic resolution) (Ficetola *et al.*, 2010; Furlan *et al.*, 2016) is essential for the success of metabarcoding studies. In aDNA studies, the choice of the marker is particularly difficult, as prerequisites, such as minimal bias in amplification of different taxa and short sequence length, drastically reduce the ability to resolve taxa (Taberlet *et al.*, 2007). Plant metabarcoding studies can use either a single marker or a combination of multiple primers to resolve taxa. For instance, the *trnL g/h* primers (Taberlet *et al.*, 2007) can first be used to obtain an overall analysis of plant diversity with a good resolution for most plant families, and additional primers (e.g. ITS1-F/ITS1Poa-R for Poaceae and ITS1-F/ITS1Ast-R for Asteraceae) may be added later to increase resolution within families (Baamrane *et al.*, 2012). We should, however, remember that different primers might favour the amplification of different taxa, which may lead to biases in the final results (Yoccoz, 2012).

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

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

2. Shotgun metabarcoding on *sedaDNA*

Shotgun metabarcoding analysis is a newer alternative to the traditional metabarcoding approach. It relies on shotgun sequencing DNA (Orlando *et al.*, 2015) from a mixed genetic pool obtained from environmental samples; computational investigation, for example bioinformatic analysis, is then used to decipher the taxonomic composition. The shotgun sequencing approach traditionally was used for the study of the fraction of un-cultivable microbes in modern environmental samples (Sharon & Banfield, 2013; Vos *et al.*, 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 appears to give a more comprehensive insight into the community composition from all trophic layers represented by the environmental sample (Pedersen *et al.*, 2016). Although this approach has played an expanding role in the study of ancient prokaryotic communities in soil and sediments (> 90% of DNA molecules in a sample are likely to be prokaryotic), but also teeth, coprolites, gut and ice (Warinner *et al.*, 2015; Wood & Wilmschurst, 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). There exists a considerable potential for the application of low-coverage shotgun sequencing of genomic DNA (genome skimming) for the study of plant aDNA (Coissac *et al.*, 2016), but the absence of consensus data processing and lack of experience of interpreting such datasets make it prone to misinterpretations and false positives (Bennett, 2015; Weiß *et al.*, 2015) (see section X). Species identification using genomic plant data outside chloroplast barcode regions can be particularly problematic. For instance, of the *c.* 391 000 vascular plant species existing on earth (number constantly changing through new discoveries and taxonomic revisions), only 1092 are represented by complete chloroplast genomes (National Center for Biotechnological Information (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 over-represented 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 20 most sequenced organisms (number of bases, NCBI, <ftp://ftp.ncbi.nlm.nih.gov/genbank/gbrel.txt> – accessed 29 March 2016). Although we can limit false positives and confirm aDNA authenticity using bioinformatic techniques (see section X), the lack of genomic references and the limited experience in interpreting such datasets make proper contextualized biological interpretation a necessity. However, reference databases are quickly improving and on-going projects (e.g. PhyloAlps, <https://www.france-genomique.org/spip/spip.php?article112&lang=fr>, and NorBol, <http://norbol.org/>) are currently assembling the entire chloroplast and nuclear ribosomal genomes of the whole floras of the Alps, Norway and parts of the Arctic (Coissac *et al.*,

2016). Shotgun metabarcoding will therefore become an important tool in future years, as it will allow the detection of organismal diversity and, potentially, the differentiation of population structures, and will further help to bridge the gap between different scientific disciplines in palaeoecology.

An important advantage of shotgun sequencing is the possibility to quantify the degree of DNA degradation exerted by the sample, for example, fragmentation and hydrolysis of cytosines (C), which results in the accumulating misincorporation frequencies of thymine (T) instead of C at the ends of the DNA molecules during sequencing (Jónsson *et al.*, 2013). Lastly, it has been shown recently that DNA damage correlates with the age of the lake sediments (Pedersen *et al.*, 2016) and should therefore be used as an independent tool for aDNA authentication.

3. HTS on pollen

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

A more efficient way of analysing ancient pollen is to construct HTS libraries directly from single pollen grains isolated from sediments, employing the SCS technology. SCS is a powerful new set of technologies for the parallel sequencing of hundreds of single cells isolated from complex populations (Wang & Navin, 2015). SCS methodology on fossil pollen seems at the moment the best available approach for the efficient investigation of large numbers of grains, and a good alternative to the more time-consuming single-pollen genotyping technique (Parducci *et al.*, 2005). Because pollen is normally abundant in lake sediments, this method offers a unique opportunity for analysing the genetic structures of large numbers of single plant individuals simultaneously, over different spatial and temporal time scales (from decades to millennia). In order to sequence a single cell from pollen, single grains must first be isolated from sediments and cleaned carefully. Many different approaches can be used for the capture and cleaning of single grains taken at random from the abundant pollen populations usually present in sediments: mouth or hand pipetting under a microscope (see Box 1 and Parducci & Suyama, 2011), serial dilution, flow-assisted cell sorting (FACS), microfluidic manipulation (Wang & Navin, 2015), flow sorting or micro-manipulation (Kron & Husband, 2012). Pollen walls must be disrupted to allow DNA release for successive PCR amplification and HTS library construction. Before

pollen disruption, pollen grains can also be screened for DNA content by staining pollen suspensions with DAPI (4',6-diamidino-2-phenylindole), so that the grains can be selected based on DNA content using fluorescence-based flow cytometry (L. Parducci, pers. comm.). To break down the highly resistant pollen walls, bead beating with glass or ceramic beads can be used (Roberts, 2007) to avoid the use of chemicals or enzymes that may later interfere during amplification. Alternative methods 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 previously been used successfully to extract DNA from manually-crushed fossil pollen grains (Parducci & Suyama, 2011). After breaking the pollen wall the SCS approach may involve first whole genome amplification using multiple displacement amplification (MDA) to obtain a reasonable quantity of DNA for genomic analysis. Successive downstream analyses involve library construction for HTS sequencing through the direct PCR amplification of short regions of interest (e.g. the *trnL* g per hour cpDNA region, 10–143 bp (Taberlet *et al.*, 2007; Isagi & Suyama, 2010)) or whole genome sequencing (Troell *et al.*, 2016) with the incorporation of indexing adapter tags and sequencing Illumina adaptors on each pollen grain.

IX. Challenges when studying aDNA from lake sediments

1. 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 no single strategy valid for the avoidance of all possible contamination sources occurring from: (1) laboratory facilities; (2) cross-contamination; and (3) reagents used during extraction and downstream preparation before sequencing. In general, multi-strategy procedures should be employed (Champlot *et al.*, 2010) to avoid contamination. However, although contamination from laboratory facilities and cross-contamination are strictly related to the 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. First, not all reagents can be filtered, UV exposed, bleached or DNase treated (the preferred decontaminating method). Second, 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 of always using experimental controls covering all reagents and all steps during handling. It is equally important that the experimental setup is properly designed, thus enabling an understanding of pre-analysis workflows and results for better and future-proof data production. This involves steps from sample collection and laboratory work to eventual multiplexing of 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 (Gilbert *et al.*, 2005; Hebsgaard *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: replication and the use of controls, and DNA damage estimates.

2. Replication and use of controls

The 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 within controls (usually from known food or exotic plants) are usually easy to recognize 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 the 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 samples 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, independent of sample age. However, increasing the number of extractions or PCR repeats increased the chances of detecting rare species. When the probability of detecting the species of interest is low because of either low biomass in the local environment or high degradation caused by age or temperature, multiple PCR replicates are needed for a more complete description of communities. Simulation studies have shown that performing 6–12 PCRs from the same extract may produce robust results (Ficetola *et al.*, 2015). Unfortunately, increasing repeats can have the drawback of increasing the probability of detecting 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 the 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 the 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).

3. Authentication of ancient origin (DNA damage)

In all aDNA sequences, an excess of C-to-T transitions is observed at the 5' and 3' ends 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 to distinguish 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 or not a sequence was of ancient origin. The method may be particularly useful for the validation of future metagenomic studies.

X. Bioinformatic processing

1. Metabarcoding bioinformatics

HTS generates large DNA datasets 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 involve the assembly of paired-end reads, the assignment of HTS reads to samples or demultiplexing (metabarcoding studies generally analyse multiple samples within each HTS run) (Coissac, 2012 and subsequent papers) and the filtering of erroneous sequences originating from nonspecific amplification and PCR/sequencing errors. The latter task is often critical to ensure 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 probably 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, the 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 (MOTUs). The fifth and final step is the taxonomic assignment of sequences to currently known taxa. This is generally performed by comparing the retrieved sequences with reference databases, which can either be broad databases, such as GenBank (Benson *et al.*, 2013) or ENA-EMBL (Leinonen *et al.*, 2011), or high-quality reference databases containing the verified and curated sequences of potentially present species, such as the Arctic plant database (Sønstebo *et al.*, 2010) or the on-going PhyloAlps and NorBol projects (Coissac *et al.*, 2016). Multiple bioinformatics

tools are currently available, and a more detailed description can be found in Notes S1 in the Supporting Information.

2. Shotgun metabarcode bioinformatics

The idea of processing shotgun metabarcode data is to take all DNA sequences present in a sample dataset, align them against a reference database, parse the alignment information, assign a taxonomic label to each read and thereby generate the taxonomic profile of the metagenome/sample. For short-read alignment, standard programs, such as BWA (Li & Durbin, 2009) and BOWTIE2 (Langmead & Salzberg, 2012), are often used to align reads to reference sequences; however, they offer no tools to decipher between the alignments or for downstream handling, and additional tools and further analyses are therefore required. Over recent years, several tools for taxonomic profiling of shotgun sequence data have therefore been developed. Overall, these can be categorized as tools using all available sequences (GenBank), such as MEGAN (Huson *et al.*, 2007), which can also be used for metabarcoding, the metagenomic MG-RAST server (Wilke *et al.*, 2015), EBI-METAGENOMICS (Mitchell *et al.*, 2016), CLARK (Ounit *et al.*, 2015), HOLI (Pedersen *et al.*, 2016), KRAKEN (Wood & Salzberg, 2014) and KAIJU (Menzel *et al.*, 2016), and tools using selected marker genes, such as METAPHLAN (Segata *et al.*, 2012), MOTU (Sunagawa *et al.*, 2013) and METABIT (Louvel *et al.*, 2016). Pipelines, such as MG-RAST and EBI-METAGENOMICS, accept raw and untrimmed datasets, and are able to parse these through piped-programs aligning against in-house databases and, eventually, DNA, performing sequence classification and graphical presentation. The majority of the tools, however, are designed and optimized for fast and accurate alignments (e.g. BWA and BOWTIE2) or for alignment and eventual DNA sequence classification (e.g. KRAKEN, HOLI, CLARK and KAIJU) against custom-built databases. Alternatively, unique marker-gene databases are used to lower computational time and for robust abundance estimates, taxon classification (e.g. METAPHLAN and MOTU) and graphical presentation (METABIT).

Although the accuracy and speed of these tools make shotgun metabarcode dataset analysis fast and easy to perform (for, more

details, see Lindgreen *et al.*, 2016), a common issue is that they are designed for modern sequencing datasets and for analysis of the microbial 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 built. However, none of these tools is designed for aDNA analysis and they therefore lack the stringent and robust criteria required for the taxonomic consideration of short and damaged reads. Schubert *et al.* (2012) compared different computational methods for improving the accuracy and sensitivity of aDNA sequence identification, and showed that the use of reads ≥ 30 bp increases the quality of alignments to modern reference genomes and lowers the number of false positives. Currently, from the limited number of studies of ancient plant shotgun metabarcoding, the data indicate 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 'background' 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 remain unknown and probably vary with the genomic composition of the sample. No clear-cut choice exists for the metagenomic analysis tool, but with the criteria suggested above 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 the 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 could eventually use a lowest common ancestor algorithm to resolve reads with alignment against multiple species. Such a method, with no 'a priori' assumption about the environment or species composition in a sample, is objective and would be the preferred methodology.

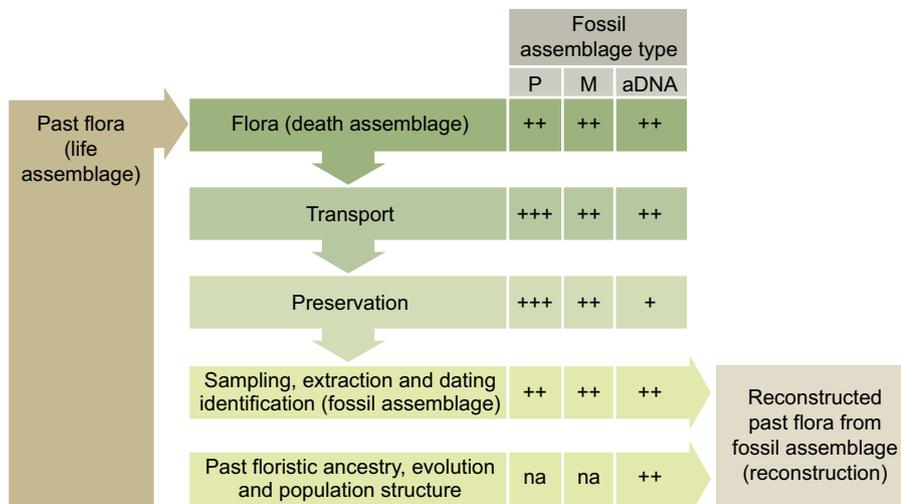


Fig. 7 Representation of the chain of processes involved in the transformation of plant information present in the three lake sediment assemblage types: pollen (P), macrofossils (M), ancient DNA (aDNA). Current understanding of the processes is indicated as good (+++), reasonable (++) or poor (+). The publications on which the levels of understanding are based are shown in Table 1. Modified from Birks & Birks (2016) and based on Jackson's (2012) general conceptual model for the representation of floristic material in palaeoecological assemblages. na, not applicable.

Pedersen *et al.* (2016) employed this methodology to data from ancient lake sediments using NCBI's nucleotide database, and showed that, using stringent assignment criteria even to a non-complete database, the plant metagenomic profile at the genus level was in line with taxa found from pollen, macrofossils and the 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, as a result of the nature of an environmental DNA sample, the majority of the sequenced DNA obtained using a shotgun metabarcoding approach cannot usually be identified with the databases currently available (Pedersen *et al.*, 2016). Often, >90% of the reads produced cannot be aligned to a reference and, in most cases, <2% of the reads are unique to any taxon. However, as more genomes become available, these proportions will improve.

XI. Conclusions and directions for future research

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

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

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

(3) Ancient plant DNA from lake sediments will, in time, become more precise at determining the local vegetation relative to macrofossil and pollen analysis.

(4) An improved understanding of DNA taphonomy from lake sediments now allows 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) to improve the determination of the power and limitations of the new tools presented in this review.

(5) SCS profiling of pollen from lake sediments will probably increase in the coming years and become an important tool for the investigation of histories and dynamics of plants at the population level.

Acknowledgements

The authors thank Sebastian Sobek for constructive comments on an earlier version of the manuscript and Pierre Taberlet for useful discussions on ancient DNA terminology. This work was supported by the Swedish Research Council (grant no. 2013-D0568401), SciLifeLab Stockholm and the Carl Triggers' Foundation (grant no. 14:371) to L.P., and the Research Council of Norway to I.G.A. (grant no. 213692/F20).

Author contributions

L.P. conceived and designed the work with significant contributions from K.D.B. and M.W.P. L.P., K.D.B., G.F.F., I.G.A., J.R.W. and M.W.P. contributed to: data collection, data analysis, data interpretation, writing and final critical revision of the article. Y.S. partially contributed to the writing. All authors approved the final version of the article.

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, 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.
- 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.
- 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 *trnL* approach. *PLoS ONE* 7: e35643.
- Barnes MA, Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* 17: 1–17.
- Belle S, Parent C, Frossard V, Verneaux V, Millet L, Chronopoulou P-M, Sabatier P, Magny M. 2014. Temporal changes in the contribution of methane-oxidizing bacteria to the biomass of chironomid larvae determined using stable carbon isotopes and ancient DNA. *Journal of Paleolimnology* 52: 215–228.
- Bennett KD. 2015. Comment on 'Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago'. *Science* 349: 247.

- Bennett KD, Willis KJ. 2001. Pollen. In: Smol P, Birks HJB, Last WM, eds. *Tracking environmental change using lake sediments*. Dordrecht, the Netherlands: Kluwer Academic, 5–32.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. *Nucleic Acids Research* 41: D36–D42.
- 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. In: Mock HP, Elias SA, eds. *Encyclopedia of quaternary science*. Amsterdam, the Netherlands: Elsevier, 593–612.
- Birks HJB, Birks HH. 2016. How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? *New Phytologist* 209: 499–506.
- Bissett A, Gibson J, Jarman SN. 2005. Isolation, amplification, and identification of ancient copepod DNA from lake sediments. *Limnology and Oceanography: Methods* 3: 533–542.
- 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.
- 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.
- Capo E, Debroas D, Arnaud F, Domaizon I. 2015. Is planktonic diversity well recorded in sedimentary DNA? Toward the reconstruction of past protistan diversity. *Microbial Ecology* 70: 865–875.
- 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. In: Pompanon F, Bonin A, eds. *Methods in molecular biology. Data production and analysis in population genomics*. New York, NY, USA, Dordrecht, the Netherlands, Heidelberg, Germany, London, UK: Springer, 13–31.
- Coissac E, Hollingsworth PM, Laverigne 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* 17: 104–106.
- Coolen M, Muyzer G, Rijpstra W, Schouten S, Volkman JK, Sinninghe Damsté JS. 2004. Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. *Earth and Planetary Science Letters* 223: 225–239.
- Coolen M, Overmann J. 1998. Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment. *Applied and Environmental Microbiology* 64: 4513–4521.
- Coolen M, Talbot HM, Abbas BA, Ward C, Schouten S, Volkman JK, Damsté JSS. 2008. Sources for sedimentary bacteriohopanepolyols as revealed by 16S rDNA stratigraphy. *Environmental Microbiology* 10: 1783–1803.
- 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* 5: a012567.
- D'Andrea WJ, Lage M, Martiny J. 2006. Alkenone producers inferred from well-preserved 18S rDNA in Greenland lake sediments. *Journal of Geographical Research* 111: G03013.
- 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.
- Domaizon I, Savichtcheva O, Debroas D, Arnaud F, Villar C, Pignol C, Alric B, Perga ME. 2013. DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus* assemblages. *Biogeosciences* 10: 3817–3838.
- 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, Schröder-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.
- Epp LS, Stooft-Leichsenring KR, Trauth MH, Tiedelman R. 2011. Molecular profiling of diatom assemblages in tropical lake sediments using taxon-specific PCR and Denaturing High-Performance Liquid Chromatography (PCR-DHPLC). *Molecular Ecology Resources* 11: 842–853.
- Epp LS, Stooft-Leichsenring KR, Trauth MH, Tiedemann R. 2010. Historical genetics on a sediment core from a Kenyan lake: intraspecific genotype turnover in a tropical rotifer is related to past environmental changes. *Journal of Paleolimnology* 43: 939–954.
- Etienne D, Destas M, Lyautey E, Marti R, Ruffaldi P, Georges-Leroy M, Dambrine E, Topp E. 2015. Two thousand-year reconstruction of livestock production intensity in France using sediment-archived fecal Bacteroidales and source-specific mitochondrial markers. *The Holocene* 25: 1384–1393.
- Faegri K, Iversen J, Kaland PE, Krzywinski K. 1989. *Textbook of pollen analysis, 4th edn*. Chichester, UK: Wiley.
- 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: 434–457.
- Ficetola GF, Pansu J, Bonin A, Coissac E, Giguët-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.
- da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Caroe C, Ávila-Arcos MAC, Huftnagel DE *et al.* 2015. The origin and evolution of maize in the Southwestern United States. *Nature Plants* 1: 14003–14008.
- Fulton TL. 2012. Setting up an ancient DNA laboratory. In: Shapiro B, Hofreiter M, eds. *Methods in molecular biology. Ancient DNA: methods and protocols*. New York, NY, USA, Dordrecht, the Netherlands, Heidelberg, Germany, London, UK: 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.
- Giguët-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 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.
- Hou W, Dong H, Li G, Yang J, Coolen MJL, Liu X, Wang S, Jiang H, Wu X, Xiao H *et al.* 2014. Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Scientific Reports* 4: 6648–6649.
- Hu FS, Hampe A, Petit RJ. 2009. Paleoeecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment* 7: 371–379.
- Huang YT. 2014. *Studies on carbon and DNA preservation in allophanic soils and paleosols on Holocene tephra in New Zealand*. PhD thesis, University of Waikato, Hamilton, New Zealand.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17: 377–386.
- Hutchinson GE. 1957. *A treatise on limnology, vol. 1. Geography, physics and chemistry*. New York, NY, USA: Wiley.
- Hyvärinen H. 1970. Flandrian pollen diagrams from Svalbard. *Geografiska Annaler* 52: 213–222.
- Isagi Y, Suyama Y. 2010. *Single-pollen genotyping*. Tokyo, Japan: 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.
- Jiang H, Dong H, Yu B, Liu X, Li Y, Ji S. 2007. Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environmental Microbiology* 9: 2603–2621.
- 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, Møller P, Andreev VA, 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 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.
- Kojima H, Iwata T, Fikui M. 2009. DNA-based analysis of planktonic methanotrophs in a stratified lake. *Freshwater Biology* 54: 1501–1509.
- Kron P, Husband BC. 2012. Using flow cytometry to estimate pollen DNA content: improved methodology and applications. *Annals of Botany* 110: 1067–1078.
- Kyle M, Haande S, Sonstebo JH, Rohrlack T. 2015. Amplification of DNA in sediment cores to detect historic *Planktothrix* occurrence in three Norwegian lakes. *Journal of Paleolimnology* 53: 61–72.
- 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–D31.
- Li G, Dong H, Hou W, Wang S, Jiang H, Yang J, Wu G. 2016. Temporal succession of ancient phytoplankton community in Qinghai Lake and implication for paleo-environmental change. *Scientific Reports* 6: 19769.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Limburg PA, Weider LJ. 2002. ‘Ancient’ DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society B: Biological Sciences* 269: 281–287.
- 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. *Nature* 362: 709–715.
- Louvel G, Der Sarkissian 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* 16: 1415–1427.
- 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.
- Madeja J. 2015. A new tool to trace past human presence from lake sediments: the human-specific molecular marker *Bacteroides* strain HF 183. *Journal of Quaternary Science* 30: 349–354.
- Madeja J, Wacnik A, Wypasek E, Chandran A, Stankiewicz E. 2010. Integrated palynological and molecular analyses of late Holocene deposits from Lake Miłkowskie (NE Poland): verification of local human impact on environment. *Quaternary International* 220: 147–152.
- Madeja J, Wacnik A, Zyga A, Stankiewicz E, Wypasek E, Guminski W, Harmata K. 2009. Bacterial ancient DNA as an indicator of human presence in the past: its correlation with palynological and archaeological data. *Journal of Quaternary Science* 24: 317–321.
- Magyari EK, Major A, Balint M, Nedli J, Braun M, Racz I, Parducci L. 2011. Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evolutionary Biology* 11: 66.
- Marková S, Černý M, Rees D, Stuchlík E. 2006. Are they still viable? Physical conditions and abundance of *Daphnia pulicaria* resting eggs in sediment cores from lakes in the Tatra Mountains. *Biologia* 61: 135–146.
- 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.
- Matisoo-Smith E, Roberts K, Welikala N, Tannock G, Chester P, Feek D, Flenley J. 2008. Recovery of DNA and pollen from New Zealand lake sediments. *Quaternary International* 184: 139–149.
- 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.
- 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.
- Mergeay J, Vanoverbeke J, Verschuren D, De Meester L. 2007. Extinction, recolonization, and dispersal through time in a planktonic crustacean. *Ecology* 88: 3032–3043.

- Mitchell A, Bucchini F, Cochrane G, Denise H, Ten Hoopen 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–D603.
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLoS ONE* 10: 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 genome sequence 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.
- Pääbo S. 1984. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314: 644–645.
- 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.
- Pal S, Gregory-Eaves I, Pick FR. 2015. Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. *Journal of Paleolimnology* 54: 87–101.
- 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, 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 sediments from central Scandinavia. *Molecular Ecology* 22: 3511–3524.
- Parducci L, Suyama Y. 2011. Single-pollen genotyping of Holocene lake sediments. In: Isagi Y, Suyama Y, eds. *Ecological research monographs. Single-pollen genotyping*. Tokyo, Japan: Springer, 101–109.
- 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, Väiliranta 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, Hafliadason 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.
- 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.
- Pedersen 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, Kjeldsen KK, Mendoza MLZ, Beaudoin AB, Zutter C, Larsen NK *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 from a single 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.
- Pollard HG, Colbourne JK, Keller W. 2003. Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination. *AMBIO: A Journal of the Human Environment* 32: 214–218.
- 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.
- Poullain AJ, Aris-Brosou SEP, Blais JM, Brazeau M, Keller WB, Paterson AM. 2015. Microbial DNA records historical delivery of anthropogenic mercury. *The ISME Journal* 9: 2541–2550.
- Randlett M-È, Coolen M, Stockhecke M, Pickarski N, Litt T, Balkema C, Kwiecien O, Tomonaga Y, Wehrli B, Schubert CJ. 2014. Alkenone distribution in Lake Van sediment over the last 270 ka: influence of temperature and haptophyte species composition. *Quaternary Science Reviews* 104: 53–62.
- Ravasi DF, Peduzzi S, Guidi V, Peduzzi R, Wirth SB, Gilli A, Tonolla M. 2012. 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. *Geobiology* 10: 196–204.
- 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.
- Reid VA, Carvalho GR, George DG, Griffiths HI. 2002. A technique for the molecular genetic analysis of *Daphnia* resting eggs from sub-recent lake sediments. *Journal of Paleolimnology* 27: 481–486.
- 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.
- Savichtcheva O, Debros D, Kurmayer R, Villar C, Jenny JP, Arnaud F, Perga ME, Domaizon I. 2011. Quantitative PCR enumeration of total/toxic *Planktothrix rubescens* and total cyanobacteria in preserved DNA isolated from lake sediments. *Applied and Environmental Microbiology* 77: 8744–8753.
- 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 IB, Bohmann K, Gilbert MTP. 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*. New York, NY, USA, Dordrecht, the Netherlands, Heidelberg, Germany, London, UK: 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. 2017. Lake sedimentary DNA accurately records 20th century introductions of exotic conifers in Scotland. *New Phytologist* 213: 929–941.

- 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, Sønstebo 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.
- Sønstebo 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.
- Southworth D. 1974. Solubility of pollen exines. *American Journal of Botany* 61: 36–44.
- Stager JC, Sporn LA, Johnson M, Regalado S. 2015. Of paleo-genes and perch: what if an 'alien' is actually a native? *PLoS ONE* 10: e0119071.
- 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.
- Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedelman R. 2012. Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular Ecology* 21: 1918–1930.
- Stoof-Leichsenring KR, Herzschuh U, Pestryakova LA, Klemm J, Epp LS, Tiedelman R. 2015. Genetic data from algae sedimentary DNA reflect the influence of environment over geography. *Scientific Reports* 5: 12924.
- 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 phylogenetic marker genes. *Nature Methods* 10: 1196–1199.
- 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.
- 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.
- 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, Vermet T, Cortthier G, Brochmann C, Willerslev E. 2007. Power and limitations of the chloroplast *trnL* (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.
- Troell K, Hallström B, Divne A-M, Alsmark C, Arrighi R, Huss M, Beser J, Bertilsson S. 2016. Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes. *BMC Genomics* 17: 471.
- 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.
- Vuillemin A, Ariztegui D, Leavitt PR, Bunting L, the PASADO Science Team. 2016. Recording of climate and diagenesis through sedimentary DNA and fossil pigments at Laguna Potrok Aike, Argentina. *Biogeosciences* 13: 2475–2492.
- 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. *Philosophical 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.
- Wetzel RG. 2001. *Limnology. Lake and river ecosystems, 3rd edn.* San Diego, CA, USA: Academic Press.
- 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.
- 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.
- 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.
- Wooller MJ, Gaglioti B, Fulton TL, Lopez A, Shapiro B. 2015. Post-glacial dispersal patterns of Northern pike inferred from an 8800 year old pike (*Esox cf. lucius*) skull from interior Alaska. *Quaternary Science Reviews* 120: 118–125.
- Wright HE, Mann DH, Glaser PH. 1984. Piston corers for peat and lake sediments. *Ecology* 65: 657–659.
- Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J, Yang Y, Yao T, Knight R, Chu H. 2012. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environmental Microbiology* 14: 2457–2466.
- Xu Z-H, Jiang X-D, Wang G-Z, He J-F, Cai M-H, Wu L-S, Jiang J-L, Chen X-L. 2011. DNA extraction, amplification and analysis of the 28S rRNA portion in sediment-buried copepod DNA in the Great Wall Bay and Xihu Lake, Antarctica. *Journal of Plankton Research* 33: 917–925.
- Yang J, Jiang H, Dong H, Hou W, Li G, Wu G. 2015. Sedimentary archaeal amoA gene abundance reflects historic nutrient level and salinity fluctuations in Qinghai Lake, Tibetan Plateau. *Scientific Reports* 5: 1–9.
- Yoccoz N. 2012. The future of environmental DNA in ecology. *Molecular Ecology* 21: 2031–2038.
- Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Von Stedingk 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.
- 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.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Notes S1 Bioinformatic tools for metabarcoding datasets.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.