



## Five thousand years of tropical lake sediment DNA records from Benin

L. Bremond <sup>a, b, \*</sup>, C. Favier <sup>a</sup>, G.F. Ficetola <sup>c, d, e</sup>, M.G. Tossou <sup>f</sup>, A. Akouégninou <sup>f</sup>, L. Gielly <sup>c, d</sup>, C. Giguet-Covex <sup>g</sup>, R. Oslisly <sup>h</sup>, U. Salzmann <sup>i</sup>

<sup>a</sup> Institut des Sciences de l'Évolution - Montpellier, UMR 5554 CNRS-IRD-Université Montpellier-EPHE, Montpellier, France

<sup>b</sup> École Pratique des Hautes Études, PSL Research University, 4-14 rue Ferrus, 75014 Paris, France

<sup>c</sup> Université Grenoble Alpes, LECA, F-38000 Grenoble, France

<sup>d</sup> CNRS, LECA, F-38000 Grenoble, France

<sup>e</sup> Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, 20133, Milan, Italy

<sup>f</sup> National Herbarium, Department of Botany and Plant Biology, Faculty of Sciences and Technology (FAST), University of Abomey-Calavi (UAC) Cotonou, Benin

<sup>g</sup> EDYTEM, Université de Savoie, CNRS, Pole Montagne, 73376 Le Bourget du Lac, France

<sup>h</sup> UMR PALOC, Institut de Recherche pour le Développement et Agence Nationale des Parcs Nationaux, BP 20379 Libreville, Gabon

<sup>i</sup> Department of Geography, Faculty of Engineering and Environment, Northumbria University, Ellison Building, Newcastle upon Tyne, NE1 8ST, UK



### ARTICLE INFO

#### Article history:

Received 21 March 2017

Received in revised form

23 June 2017

Accepted 23 June 2017

Available online 8 July 2017

#### Keywords:

sedDNA

Metabarcoding

Tropical

Africa

Vegetation dynamics

Plant diversity

Pollen

### ABSTRACT

Until now, sedimentary DNA (sedDNA) studies have only focused on cold and temperate regions where DNA is relatively well preserved. Consequently, the tropics, where vegetation is hyperdiverse and natural archives are rare, have been neglected and deserve attention. In this study, we used next-generation sequencing to barcode sedDNA from Lake Sele, localized in the tropical lowlands of Benin (Africa), and compared the taxonomic diversity detected by DNA analyses with pollen assemblages. Plant sedDNA was successfully amplified from 33 of the 34 successfully extracted samples. In total, 43 taxa were identified along the 5000 years spanned by the sediment: 22 taxa were identified at the family level and 21 at the genus level. The plant diversity recovered through sedDNA from Lake Sele showed a specific local signal and limited overlapping with pollen. Introduced plants, grown and cultivated close to the water, such as sweet potato, were also well recorded by sedDNA. It appears, therefore, to be a promising approach to studying past diversity in tropical regions, and could help in tracking the introduction and history of agriculture. This is the first time this method has been used in the field of domestication and dissemination of several specific crops, and the results are very encouraging.

© 2017 Elsevier Ltd. All rights reserved.

### 1. Introduction

Plant sedimentary DNA (sedDNA) has recently emerged as a promising proxy for paleoecological reconstructions over thousands of years, complementary to pollen and macrofossils analysis (Thomsen and Willerslev, 2015; Parducci et al., 2017). However, until now, paleoenvironmental plant sedDNA analyses have essentially been conducted in temperate and boreal zones (Willerslev et al., 2003; Haile et al., 2009; Jørgensen et al., 2012; Parducci et al., 2013, 2017; Giguet-Covex et al., 2014). The oldest sedDNA records were recovered from the permafrost of Siberia

(Willerslev et al., 2003) and Northern Scandinavia (Parducci et al., 2012). In more temperate zones, the published results generally span over the last 10,000 years (Anderson-Carpenter et al., 2011; Giguet-Covex et al., 2014; Pansu et al., 2015; Alsos et al., 2016), as the studied lakes have accumulated sediments since the last deglaciation. While the taxonomic resolution of sedDNA is not necessarily better than pollen or macroremain counts (Matisoo-Smith et al., 2008; Parducci et al., 2013; Pedersen et al., 2013), they nevertheless allow the detection of specific taxa either poorly or not recorded by classical paleoecological methods. The analysis of plant sedDNA has enabled multiple improvements to our paleoecological knowledge, such as the discovery that coniferous trees survived the last glaciation in Northern Scandinavia despite this not being detected by pollen analysis (Parducci et al., 2012), and the description of long-term vegetation changes determined by human activities and pasture (Giguet-Covex et al., 2014; Pansu

\* Corresponding author. Institut des Sciences de l'Evolution - Montpellier, UMR5554 UM – CNRS – IRD – EPHE, Université de Montpellier, CC 061 34095, Montpellier, Cedex 05, France.

E-mail address: [laurent.bremond@umontpellier.fr](mailto:laurent.bremond@umontpellier.fr) (L. Bremond).

et al., 2015).

Few studies have explored sedDNA in tropical regions (Epp et al., 2010; Stoof-Leichsenring et al., 2012), and only one, from a high-altitude lake, analyzed long-term variations in terrestrial plant DNA (Boessenkool et al., 2014). The initial focus on high-latitude or high-altitude areas is related to the fact that DNA degradation increases at high temperatures (Allentoft et al., 2012), so the absence of publications on plant sedDNA analyses in low-altitude and latitude sites might merely be due to the absence of trials or a failure of these trials (Parducci et al., 2017). Despite the relatively short time period examined, the recovery of diatom sedDNA dating back 200 years by Stoof-Leichsenring et al. (2011) in Lake Naivasha, Kenya, is particularly interesting as water temperatures rarely fall below 20 °C (in Britton et al., 2007). In the two high-altitude lakes studied by Boessenkool et al. (2014) sedDNA sequences were obtained from 4500-year-old samples. With the help of a reference library covering the majority of the afro-alpine flora, the authors were able to identify several taxa at a better/more precise taxonomic level than pollen analyses, but with fewer species detected. If DNA can be preserved in sediments from warm climates, the method could be an important proxy to identify early agriculture and farming in the tropics. In Africa, for example, the onset of the cultivation of pearl millet (*Pennisetum glaucum*) or sorghum (*Sorghum bicolor*) is poorly understood (Fuller et al., 2014) because traditional bioproxies such as pollen, and phytoliths are inadequate (phytolith and pollen of African grasses cannot be differentiated into cereals and wild grasses), and charred grains are rare. SedDNA may also be a valuable tool to complement and alleviate the time constraints of phylogeography studies, the study of the geographic distribution of genetic lineages within species (e.g. for *Aucoumea klaineana* in Born et al., 2008), or to track the history of specific vegetation types badly recovered by more conventional methods, for instance the Marantaceae forests (Cuni-Sánchez et al., 2016).

The objective of this study is to explore the preservation of plant DNA in the sediments of a low-elevation tropical lake over the last 5000 years, and to test the pertinence of sedDNA analysis to record past changes in plant diversity and agriculture practices on the shoreline of the lake. The DNA of wild and cultivated plants is generally incorporated into the soils (Yoccoz et al., 2012), and may be transported to the lake by erosion during rainfall. We chose Lake Sele (south of Benin) because the shorelines are flat, with large parts being used for seasonal agriculture during the dry season when the lake level is low. Furthermore, its sediments have already been studied for fossil pollen content (Salzmann and Hoelzmann, 2005), allowing comparison between these records and the sedDNA.

## 2. Material and method

### 2.1. Study site

Lake Sele in Benin is located about 1 km east of the Ouémé River (7° 9'19.29"N, 2°26'25.57"E; Fig. 1) at an altitude of less than 10 m above sea level, with a mean annual temperature of 28 °C (Hijmans et al., 2004); and though it is not directly fed by the river it is within the same riverbed. The lake is elongated and shallow at around 2.5 km in length, 1 km wide (at its maximum point), and with a maximum water depth of <1.5 m during the dry season (Salzmann and Hoelzmann, 2005). Nowadays, during the dry season, the shorelines are partially covered with Cyperaceae and the water hyacinth *Eichhornia crassipes* (Pontederiaceae; introduced from South America) (Salzmann and Hoelzmann, 2005). The shorelines are generally cultivated when the lake level is low. Agricultural parcels are present all around the lake forming a belt ~200–300 m wide, which are partially or totally submerged during the wet

season. The main cultures are sweet potatoes (*Ipomoea batatas*, Convolvulaceae), the red variety planted close to the water and the white variety further away; also cultivated away from the shoreline are cassava (*Manihot esculenta*, Euphorbiaceae), maize (*Zea mays*, Poaceae), and peanuts (*Arachis hypogaea*, Fabaceae). Pigeon pea (*Cajanus cajan*, Fabaceae), bananas (*Musa sp.*, Musaceae), tomatoes (*Solanum lycopersicum*, Solanaceae), and other legumes are cultivated in the upper part of the shoreline, along with coconut (*Cocos nucifera*, Arecaceae) and oil palms (*Elaeis guineensis*, Arecaceae). The natural vegetation belongs to the Guinean transition zone with its mosaic of forests and savannas (White, 1983). The main trees of the residual forest stands, include *Triplochiton scleroxylon* (abachi, Sterculiaceae), *Celtis* spp. (Cannabaceae), and Ulmaceae, *Milicia excelsa* (iroko, Moraceae).

### 2.2. Sampling

We collected five meters of sediment cores from the deepest part of the lake (N7.15,537° E2.44,106°). The two upper meters of the sediment were collected with an Uwitec Gravity corer (63 mm inner diameter), in order to get the longest possible section in one tube to minimize contamination; the three lower meters were taken with a modified Livingston piston corer (47 mm inner diameter), which permitted us to reach the deepest sediments in one meter sections. The whole sediment core was kept closed in either plastic (Uwitec) or aluminum (Livingston) tubes and stored for several days in an air-conditioned room in Benin, before being moved to a cold chamber (4 °C) in Montpellier, France. Sampling for DNA extraction was performed one month later after arrival in France. Seven AMS <sup>14</sup>C analyses were carried out on terrestrial macroremains (seeds and piece of plant leafs) by the Poznan Radiocarbon Laboratory. The <sup>14</sup>C ages were converted to calendar years using CLAM software (Blaauw, 2010) and the age-depth model was generated using BACON software (Blaauw and Christen, 2011). A hiatus of sedimentation was detected between 4400 years cal. BP and 3100 years cal. BP, as previously indicated by Salzmann and Hoelzmann (2005) (AMS <sup>14</sup>C dates and age-depth model in supplementary data: Appendix S3).

### 2.3. Samples for DNA analysis

From the sediment core, we sampled 50 slices each 2 cm thick. Sampling of core slices was carried out at the University of Savoie. To avoid DNA contamination, particularly with regards to water circulation along the coring tube, 10 mm was removed from the edge of each sediment slice. The tools (knives, spoons, pliers) were cleaned after each sampling with distilled water, then alcohol and inflamed. DNA extraction, PCR set-up and PCR amplification steps were later performed in three separate rooms at the University Grenoble Alpes, which is specifically dedicated to ancient DNA studies. SedDNA extraction targeted extracellular DNA (Pansu et al., 2015). For each sediment slice, we mixed approximately 15 g of sediment with 15 ml of saturated phosphate buffer (Na<sub>2</sub>HPO<sub>4</sub>; 0.12 M; pH ≈ 8) for 15 min. The mixture was then centrifuged (10 min at 10,000 g). The resulting supernatant (12 ml) was transferred to Amicon® Ultra-15 10 K Centrifugal Filter Devices (Millipore) and centrifuged (20 min at 4000 g) to concentrate sedDNA. Of the resulting concentrate, 400 µl were kept as a starting material for the following extraction steps, using the NucleoSpin® Soil kit (Macherey-Nagel, Düren, Germany), skipping the cell lysis step and following the manufacturer's instructions (Taberlet et al., 2012b). Three extraction controls were performed.

The sedDNA extractions and amplifications were performed in laboratories specifically dedicated to ancient DNA studies in the University Grenoble-Alpes, France.



**Fig. 1.** Localization of the coring site in Lake Sele (Benin). The annually submerged shorelines are identifiable by the greenness of the aquatic plants and the agricultural parcels which present geometric forms due to the shapes of the fields.

#### 2.4. DNA amplification and high-throughput sequencing

Plant DNA was amplified with the universal primers “g” and “h” (Taberlet et al., 2007), targeting the short and variable P6 loop region of the chloroplast *trnL* (UAA) intron. These primers are highly conserved in angiosperms and gymnosperms, and are thus of interest for plant metabarcoding (Taberlet et al., 2007). To assign sequence reads to the relevant sample, 8 bp tags (with at least 5 differences between them) were added to the 5' end of primers (Binladen et al., 2007; Valentini et al., 2009). DNA amplifications were carried out in a final volume of 30  $\mu$ l containing 3  $\mu$ l of diluted DNA extract. The amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM of MgCl<sub>2</sub>, 0.2 mM of each dNTP, 0.2  $\mu$ M of each primer, and 4.8  $\mu$ g of bovine serum albumin (BSA, Roche Diagnostic). After 10 min at 95 °C for polymerase activation, the PCR mixture underwent 45 cycles of 30 s at 95 °C, 30 s at 50 °C, and 1 min at 72 °C, followed by a final elongation step (7 min at 72 °C). PCR products were then purified. All the samples were identified by unique 8 bp tags at the 5' end of the primers, which allowed the identification of the products of each PCR. This allowed the pooling of all the different PCR products before library preparation. To summarize, we performed PCRs on 51 samples, plus 3 extraction

controls, plus 5 negative and 2 positive PCR controls (total: 61 samples including controls, arranged in one 96-wells micrititer plate; 61 samples X 8 PCRs = 488 PCR reactions). Sequencing was performed by 2  $\times$  125 bp pair-end sequencing on an Illumina HiSeq 2500 platform, which returned a total of 8,321,810 reads.

#### 2.5. Sequences filtering and taxa assignment

DNA sequences were filtered using OBITools software (Boyer et al., 2016), following the protocol described in Pansu et al. (2015). The obtained sequences were then assigned to the relevant taxa using the *ecotag* program that locates highly similar sequence(s) in a suitable database. Here, sequences were compared to a global plant database generated using silico PCR from EMBL with the *ecoPCR* program (Ficetola et al., 2010) and a BLAST search was conducted using NCBI (<http://blast.ncbi.nlm.nih.gov>) for taxon attribution. Only sequences with a similarity of >90% to a known taxon were kept for subsequent analyses.

In order to remove potential contaminants and PCR errors, we adopted the following conservative selection procedure. Firstly, all sequences identified to a taxa (species, group of species or family) currently absent from Benin were discarded. Secondly, we considered a sequence genuine in a PCR product if its count was >10 reads.

Thirdly, we only kept sequences that reached this threshold in at least one replica in the core samples, and that did not reach this threshold in any of the control replicas; we then grouped sequences with the same identification (Parducci et al., 2017).

## 2.6. Geochemical and sedimentological analyses

All geochemical and sedimentological analyses were performed on the half intact cores collected with the Livingston corer. In order to characterize the sediment from which sedDNA was extracted, Loss On Ignition (LOI) analyses were performed on 1 cm<sup>3</sup> of sediment, following the procedure described by Heiri et al. (2001). In total, 48 samples were collected from the Livingston core at the depth levels of the sedDNA analysis excluding the very upper layers, where only one sample was collected at 8.5 cm, which almost correspond to samples collected at a 8.8, 10.4, 11.7 cm depth for sedDNA. Continuous X-ray fluorescence (XRF) analyses were performed with a core scanner Avaatech (X-Ray beam generated with a rhodium anode) at EDYTEM Laboratory (CNRS-University Savoie Mont Blanc, France). Two runs of analyses were performed with a 5 mm resolution. For the first run a voltage of 10 kV, an intensity of 1.2 mA and a counting time of 15 s were applied; for the second run, a voltage of 30 kV, an intensity of 0.75 mA and a counting time of 30 s were applied.

## 2.7. Pollen assemblages

The pollen counts used for comparison with sedDNA were previously published by Salzmann and Hoelzmann (2005). Our sediment core was collected from the same part of the lake. The age-depth model was recalculated from Salzmann and Hoelzmann (2005) using Bacon software (Blaauw and Christen, 2011).

## 3. Results

### 3.1. Plant DNA extraction and identification

SedDNA was successfully amplified and sequenced from 33 of the initial 50 samples. Sixteen samples were lost during extraction using the NucleoSpin® Soil kit because fine sediments clogged the kit filters, hampering the washing and elution steps of the extraction. One sample was sequenced but without DNA being recovered.

A total of 5,560,355 raw reads could be assigned to the 33 samples (Appendix S1). After cleaning with obiclean, 4,971,763 reads and 5365 unique sequences remained. When we retained only those sequences with a similarity of >90% to a known taxon, we obtained 1,443,822 reads and 271 unique sequences (Appendix S2). From among these sequences, we aggregated those attributable to the same high-level taxon (e.g. family); these sequences were checked again by a BLAST search. After aggregation, 148 taxa were identified, of which 74 were considered as exotic (i.e. with no species present today in Benin) and 30 were discarded because they were either detected in controls or not significantly present (<10 reads) in only one replica. This left 43 single taxa. The exotic sequences not in the control but removed were assigned to Betulaceae, *Alnus*, *Cedrus*, *Pelargonium*, *Cardamine*, *Holcus*, *Prunus*, and *Pinaceae*. None of these taxa are native to tropical lowlands and their presence in the samples might be due to contamination. Taxa present in controls were Solanaceae, Lamiaceae, some Poaceae, *Musa*, Apiaceae, *Crassula*, Polygonaceae, Amaranthaceae, Rubiaceae, and some Fabaceae, and were therefore removed from analyses.

The number of taxa and reads replicates decreases with the sediment's age (Fig. 2). The first six samples, corresponding to the first 25 cm depth (probably less than 300 years old), presented the

clearest evidence of PCR success with several taxa detected, each in seven or eight PCR replicates. In the deepest sections of the core the success of DNA amplification was much lower, with the majority of samples being given a taxon detected in only one PCR (Fig. 3).

Among the 43 taxa, 22 were identified at the family level and 21 at the genus level. The species level was not considered because of the absence of an accurate local reference library for taxonomic assignation. However, the herb *Sphenoclea zeylanica*, which occurs in damp habitats throughout the tropics, is an exception because the sequences of two species of the genus were available in the GenBank.

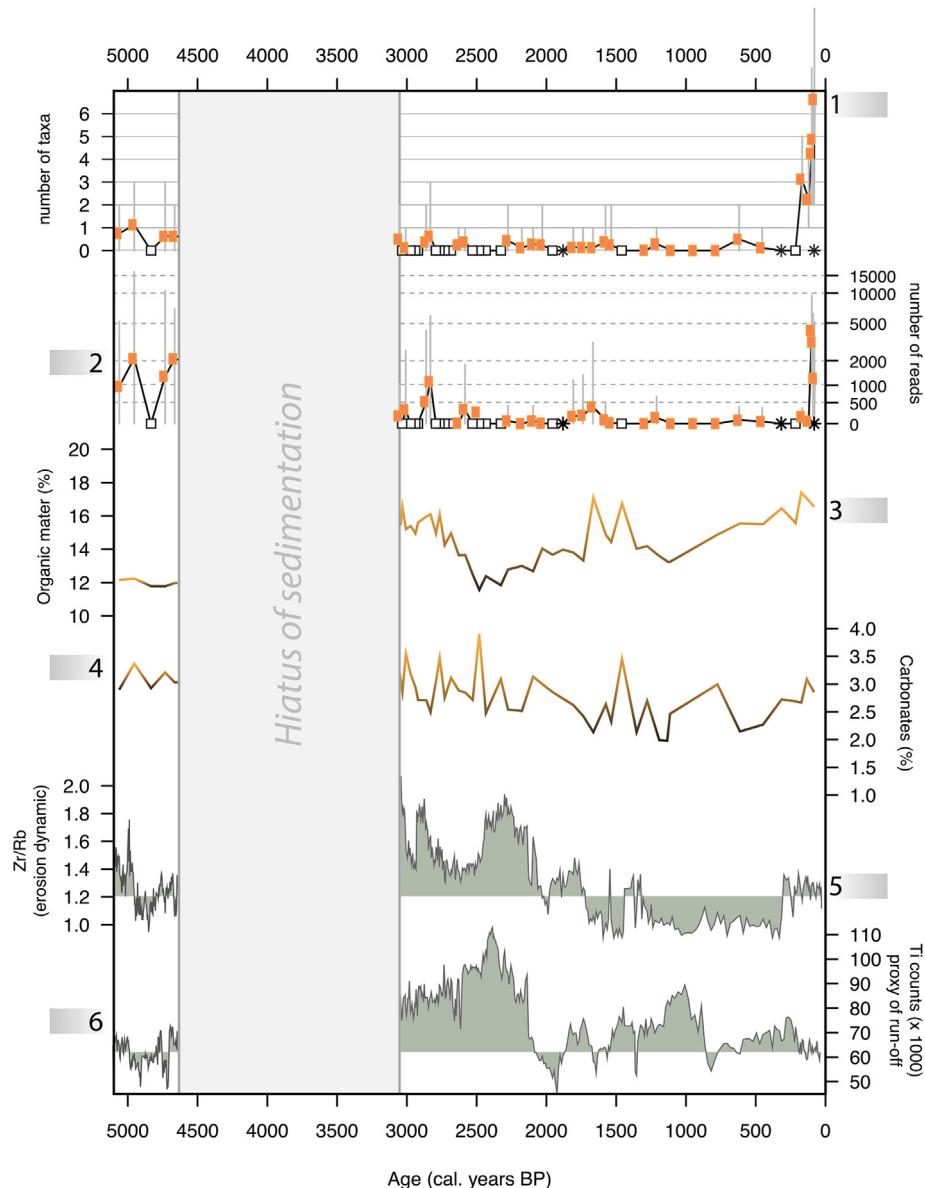
The LOI550 represents the quantity of organic matter in the sediment (% of the dry weight) that was of aquatic and/or terrestrial origin (Fig. 2). Organic matter production induces modifications in the oxygenation conditions at the lake bottom, which also affects the conditions of sedDNA preservation. In fact, low oxygen content limits the DNA damage by oxidation and microbial activity, supposedly the primary factor of extracellular DNA degradation through the production of DNase (Blum et al., 1997; Corinaldesi et al., 2005; Willerslev and Cooper, 2005; Parducci et al., 2017). Along the core, the organic matter concentration varied from 11% to 17%. The zirconium coming from zircon (Zr) was generally more concentrated in the sand sediment fraction, and the rubidium (Rb) more enriched in clays (Davies et al., 2015). The ratio Zr/Rb is thus used to trace the changes in the contribution of clay-size particles relative to sand, which can be interpreted as qualitative changes in the erosion dynamic. Titanium (Ti) can also be interpreted as a proxy of runoff and, therefore, rainfall (Metcalfe et al., 2010). The results show short phases with low-erosion rates after the hiatus, during a long period of high-erosive sediment influx from the hiatus of sedimentation until 2100 years cal. BP. Rainfall appears to have decreased until 900 years cal. BP, evident in a more or less constant fine sedimentation due to this lower rainfall. The pollen assemblages presented in Salzmann and Hoelzmann (2005) also confirm a high lake level between 3000 years cal. BP and 1000 years cal. BP.

The comparison with the pollen spectra showed important differences among the diversity and the presence of the main taxa (Fig. 3). Some taxa, such as *Nymphaea* and Asteraceae, presented coherent signals. More taxa were recovered from pollen data than from sedDNA. However, 13 DNA taxa have no corresponding pollen and spore taxa (Fig. 4) such as *Sphenoclea zeylanica*, *Pistia*, or *Ceratophyllum*. *Azolla* is a fern without identifiable spores and *Cosmarium* a green alga without pollen or spores. The dynamic of the sedDNA taxa cannot be directly compared with pollen assemblages dynamic because of the scarcity of DNA taxa in the deep part of the core.

## 4. Discussion

### 4.1. Lessons for sedDNA analyses in lowland tropical lakes

For the first time, we obtained amplification of up to 5000-year-old plant sedDNA from lowland tropical lake sediments. While the use of plant sedDNA as a palaeoecological proxy has created increasing interest over the last few years – with a focus on mid-to high-latitude lake sediments – its application in the humid tropics has remained limited and focused on high altitudes (Boessenkool et al., 2014) where the temperatures are low, or on very recent sediments (Stoof-Leichsenring et al., 2012). The successful amplification of plant sedDNA in lowland tropical lake sediments presented here, therefore, demonstrates the usefulness of the method in low latitude and altitude sites. Our results also emphasize a series of insights on the requirements to retrieve sedDNA in general, as well as in the specific context of lowland tropical lakes, namely:

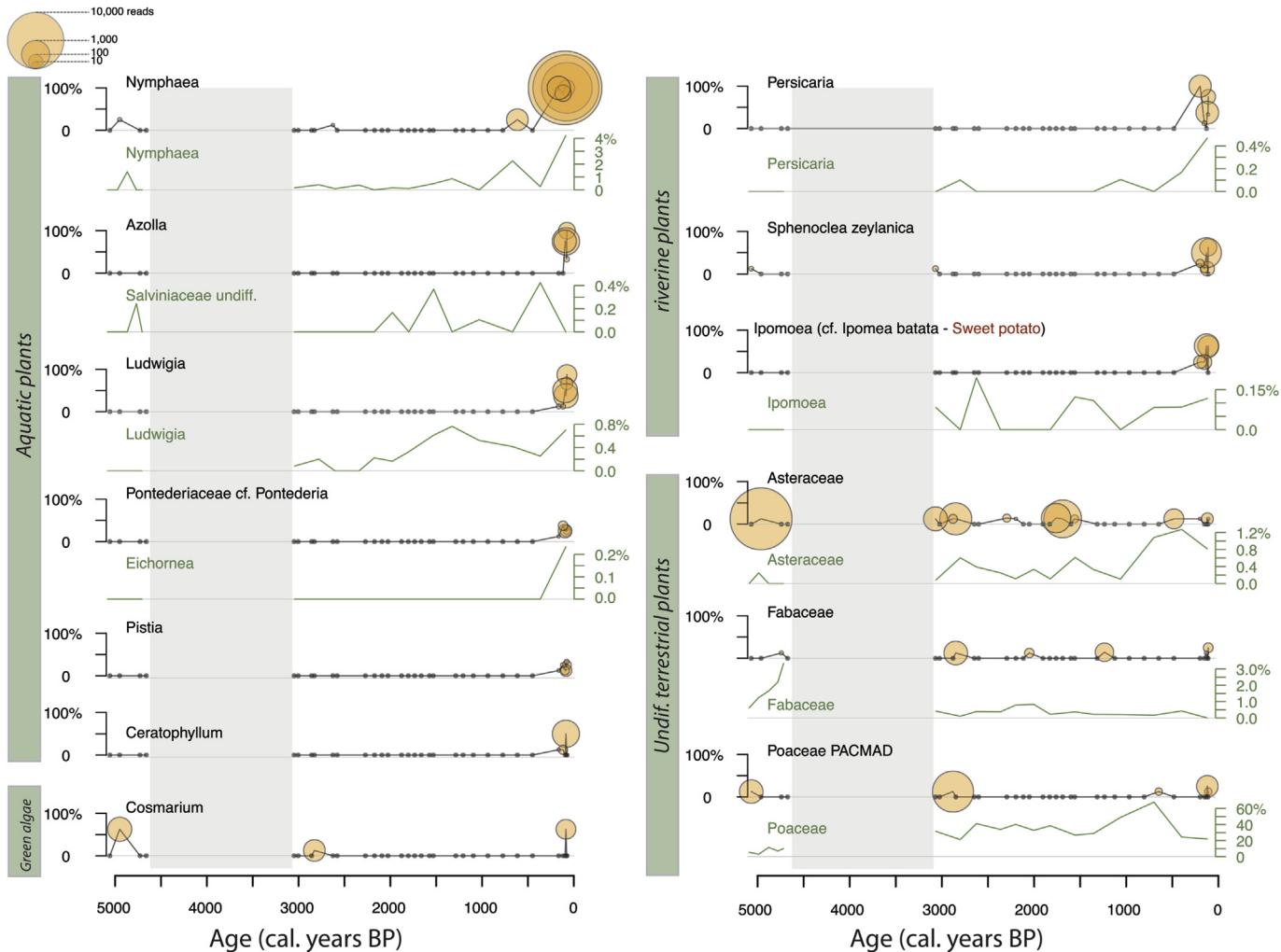


**Fig. 2.** Comparison between sedDNA and sedimentology during the last 5000 years in Lake Sele: 1) Number of identified taxa (mean values and standard deviations of the replicates) and 2) number of reads (mean values and standard deviations of the replicates). The stars represent the samples where PCR did not permit extraction of DNA, and the white squares are the samples that clogged the kit filters and hampered extraction. 3) Organic matter content ( $\text{LOI}_{550^\circ\text{C}}$ ). 4) Carbonate content ( $\text{LOI}_{550^\circ\text{C}}$ ). 5)  $\text{Zr}/\text{Rb}$  as proxy of erosion dynamic (high values = coarser grain-size), and 6) Ti related to increased run-off/rainfall.

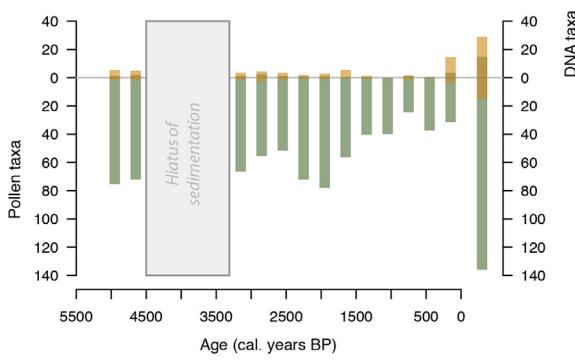
i) deposition of plant DNA in lake sediments, through taphonomical processes; ii) DNA preservation through time in lake sediments; and iii) technical methods to extract DNA from sediments.

Plant sedDNA can originate from lacustrine/shoreline plant remains, which readily deposit in the lake, or from terrestrial plant tissues. In some lakes, plants nearby the lakeshore are the main source of a terrestrial plant's DNA (Anderson-Carpenter et al., 2011; Inger Alsos, pers. com.). Extracellular DNA from terrestrial plants is, supposedly, bound mainly to soil particles and then transferred and deposited in the lake during soil erosion events (Giguet-Covex et al.; Parducci et al., 2017). In Lake Sele, lacustrine and shoreline sources also dominate the records, even in the most recent sediments. Very few DNA taxa were from terrestrial plants, probably because the majority of DNA from living plant tissue is lost before it reaches the lake sediments. Rapid DNA degradation in a warm tropical climate is a tentative explanation for the limited amount of

terrestrial plant DNA successfully amplified. Studies on the soil surfaces of tropical regions have proved that DNA can be preserved in these type of soils (Yoccoz et al., 2012), and in high enough rates that they could be transported to the lake during rainfalls events; however, measures of the degradation rate of this environmental DNA are lacking. An alternative explanation for the absence of DNA preservation is related to the lake-catchment features. Its flat topography and the riverine vegetation likely limit the flow of terrestrial plant DNA from the surrounding soils to the sediments in the middle of the lake. The question of whether DNA is preserved over long periods in sediments from warm climates is subject to controversy. As commonly observed (Anderson-Carpenter et al., 2011; Boessenkool et al., 2014), the number of taxa detected through sedDNA and the number of reads tended to decline along the core. Nevertheless, some taxa were detected in sediments even older than 4500 years. Despite the high tropical temperatures,



**Fig. 3.** Comparison between the sedDNA of plants (in orange) and pollen data (in green) from Lake Sele. The DNA signal is expressed as a percentage of the total number of replicates per level. The size of the circles indicates the number of reads for the taxon. The pollen percentage is calculated using the total pollen sum excluding undetermined, modified from Salzmann and Hoelzmann (2005). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Number of taxa identified, per 200 years intervals, by sedDNA (orange) and pollen (green). Common taxa are in dark orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sedDNA was detected. This success reflects the presence of favorable conditions for DNA preservation, which might be due to low ventilation minimizing the oxidation as well as the bacterial activity. Lake Sele is a shallow lake with a depth of less than 1.5 m during the dry season, but it is less susceptible to desiccation

because of its link to the large groundwater table of the Ouémé River (Adam and Boko, 1983). This would explain why the lake has never experienced any dry out during the last 3000 years. In the Kenyan Lake Naivasha (Stoof-Leichsenring et al., 2012), with a comparable sediment and mean annual temperature, the organic matter content was higher: 20–50% in recent sediments (Mergeay et al., 2004). A high quantity of organic matter can induce anoxic conditions, with a low pH, creating unfavorable conditions for the preservation of sedDNA, even though it can contain more plant DNA susceptible to preservation over longer time periods. At Lake Sele, it appears that more than 14% organic matter in the sediment is required to allow extraction of sedDNA.

The specific alluvial nature of sediments (after the hiatus to 900 years cal. BP) with coarse detrital input during a high lake level (Salzmann and Hoelzmann, 2005), might explain the amount of failures during the purification and extraction processes. Out of 51 samples, 16 clogged the kit filters and most replicates of 3 additional samples were found with no, or negligible, DNA. This emphasizes that the extraction and methodology used to purify and extract the sedDNA (Pansu et al., 2015) should be adapted to this kind of alluvium sediment. In the samples where extraction was successful, the majority of the eight amplification replicates contained DNA, reinforcing the idea that DNA preservation was good all

along the sediment core.

#### 4.2. Taxonomic resolution of taxa identified by DNA

##### 4.2.1. Taphonomy of sedDNA

As previously mentioned, the taxa identified by DNA analyses are mainly aquatics plants. This is the case for *Nymphaea* which is well recorded in the upper part of the core with the highest number of reads and PCR numbers; *Azolla*, *Ludwigia*, *Persicaria*, and *Sphenoclea zeylanica* that were recorded with the highest reads and PCR numbers are also aquatic plants. The limited depth of the lake also probably favors the abundance of aquatic plants and prevents the development of trees or shrubs nearby, due to the flooding of the shorelines during the wet season. These water oscillations may explain why the amount of sedDNA from trees is extremely limited. More taxa were recovered from pollen data than from sedDNA; however, some DNA taxa have no corresponding pollen taxa. The aquatic plant *Azolla* is a fern, which probably do not produce recognizable spores, but *Sphenoclea*, with identifiable pollen, should have been recorded. Pollen and sedDNA records often show low overlapping (Jørgensen et al., 2012; Parducci et al., 2013, 2017), which can be explained by the differences in source productions, the taxonomic resolution, and the primers used for sequencing (Parducci et al., 2013). Furthermore, the pollen of some species can be dispersed over long distances, thus representing regional vegetation, while sedDNA provides a more local signal (Boessenkool et al., 2014; Parducci et al., 2017). For the moment, sedDNA data are more related to the presence/absence of taxa, as macroremain analysis are interpreted in paleoecological studies (Birks, 2001). But some recent studies on soils suggest that the number of reads can be related to biomass plant production (Yoccoz et al., 2012; Pansu et al., 2015). In our data, it appears that for some taxa (e.g. *Nymphaea*) the number of reads and pollen percentages actually correlate. Combining pollen and sedDNA should help improve vegetation reconstruction, notably in tropical forests where many trees are entomophilous.

##### 4.2.2. Taxonomic reference library

The lack of a complete DNA reference library of local species is a major issue for the application of sedDNA for vegetation reconstruction. For instance, one sequence detected matched well (97% identity) with the GenBank sequences of the genus *Pontederia*. As this genus is not currently present in Benin, it is likely that the DNA sequence belongs to *Eichhornia crassipes* (Pontederiaceae), for which no sequences are available in GenBank. This plant was introduced from South America several decades ago and now completely covers the periphery of the lake, which confirms that the DNA signal is essentially coming from the vegetation that is developing on the shoreline and in the shallow parts of the lake.

It is also possible that some locally present species, for which no sequences are available in GenBank, were excluded because they matched related taxa with similar sequences that are absent in Benin. Indeed, a large number of exotic taxa from the study area that were found in the lake samples were not from the control samples. The identification of taxa through metabarcoding is severely limited by the availability of sequences within reference databases; and if no sequences of a given taxon are available, taxonomic assignment tools tend to assign sequences to the closest taxon that can be available further apart. The development of extensive reference databases is, therefore, a key priority to fully exploit the power of metabarcoding for environmental reconstruction (Taberlet et al., 2012a). Efforts to map the genetic diversity of vegetation are ongoing; unfortunately these mainly focus on Europe and North America, while information on highly diverse tropical areas remains scarce (Coissac et al., 2016; Miraldo et al.,

2016).

#### 4.3. Agriculture recorded by sedDNA

We detected the DNA of the genus *Ipomoea* (Convolvulaceae) in recent sediments. Although the specific attribution is not certain, it is likely that the DNA signal was produced by sweet potatoes (*Ipomoea batatas*); these were introduced to the Lake Sele region from South America, probably at the same time as *Eichhornia crassipes*, in the sixteenth century (O'Brien, 1972). There is little chance that another *Ipomoea* may have produced this DNA signal, given that the pollen record indicates a continuous presence of (non-cultivated) Convolvulaceae growing presumably on the lakeshores since the level dropped after 3000 cal. BP.

The original aim of this study was to test if agriculture could be recorded by sedDNA. The detection of *Ipomoea* sp. (sweet potato) DNA in recent sediments suggests that sedDNA has the potential to trace the history of agricultural developments in these environments. Nowadays sweet potato is cultivated close to permanent water, on the shoreline during the dry season when the lake is low. In some lakes, the DNA of plants growing near the shore is more easily detected than the DNA of plants living further away (Inger Alsos, pers. com.), which may explain why this species is better recorded than the cassava which is grown away from the shore. However, it is surprising that cassava or even Euphorbiaceae were not detected in the recent sediment. One possible explanation is that parts of the cultivated plant, after gathering tubers and fresh leaves, are taken and burnt somewhere that prevents possible transfer to the lake sediments.

We also found sporadic contamination by food crops such as the banana or cacao. DNA from *Musa* was detected in several samples along the core but also in the controls. Even more problematic are amplified sequences matching with European trees (*Cedrus*, *Betula* and *Pinaceae*), which were detected in the samples but not in the control. This highlights the complexity of using ancient DNA for the study of domestic and cultivated species (Weiß et al., 2015). The controversial study of Smith et al. (2015) concerning wheat cultivated 8000 years ago in the British Isles, before the Neolithic introduction of domestic cereals, is an illustrative example.

#### 5. Conclusion

Our results confirm the usefulness of sedDNA for the study of tropical sediments, as DNA can be preserved for thousand of years even in "hot" environmental conditions. The preservation seems to decrease with age but the quality of the sediment may be a more important factor for the successful analysis of sedDNA. Fine clayous sediments are more susceptible to contain extractable sedDNA and a minimum of organic matter in the bulk sediment is necessary. A small lake catchment with slopes limiting the development of large riverine aquatic plants is also probably a better configuration to record a DNA terrestrial signature than a flat lake catchment. SedDNA is a valuable approach in tropical areas for searching specific plants growing close to the water. The plant diversity recovered through sedDNA in Lake Sele clearly provides a local signal. Volcano crater lakes with inner ring vegetation, like most of the East African lakes, would be ideal to attempt new sedDNA studies. Also, although mammal DNA was not researched in our sediments it must be considered because in tropical areas, more than anywhere else, permanent water is an attractive place for wild fauna and domestic stocks.

#### Conflict of interest

The authors would like to mention that LG is one of the co-

inventors of patents related to g-h primers and the subsequent use of the P6 loop of the chloroplast *trnL* (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers.

## Contributors

LB and CF designed the study; RO, MGT and AA contributed to the design and permit technical support; GFF, LG, LB performed sampling and DNA extraction; GFF, LG, CF analyzed the data; LB, GFF, CF, US, CGC performed data interpretation; LB drafted the article, and all authors critically revised the article and approved the version to be published.

## Acknowledgment

This work was supported by project PEPS-CNRS “Banana Connexion”. We would like to thank Benoit Brossier for his technical support during field trip. We also thank Allowen Evin for fruitful discussions and Frederic Boyer for raw data handling. We thank Jill Cucchi for professional English editing.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.quascirev.2017.06.025>.

## References

- Adam, K.S., Boko, M., 1983. Le Bénin. les Ed. du Flamboyant EDICEF, Paris.
- Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaniego, J.A., Gilbert, M.T.P., Willerslev, E., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B. R. Soc.* 4724–4733.
- Alsos, I.G., Sjögren, P., Edwards, M.E., Landvik, J.Y., Gielly, L., Forwick, M., Coissac, E., Brown, A.G., Jakobsen, L.V., Føreid, M.K., 2016. Sedimentary ancient DNA from Lake Skartjørna, Svalbard: assessing the resilience of arctic flora to Holocene climate change. *Holocene* 26, 627–642.
- Anderson-Carpenter, L.L., McLachlan, J.S., Jackson, S.T., Kuch, M., Lumibao, C.Y., Poinar, H.N., 2011. Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol. Biol.* 11 (30), 1–15.
- Bindliden, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R., Willerslev, E., 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2, e197.
- Birks, H.H., 2001. Plant macrofossils. In: Smol, J., Birks, H., Last, W. (Eds.), *Terrestrial Algal and Siliceous Indicators. Tracking Environmental Change Using Lake Sediments*. Kluwer Academic Publishers, Dordrecht, pp. 49–74.
- Blaauw, M., 2010. Methods and code for ‘classical’ age-modelling of radiocarbon sequences. *Quat. Geochronol.* 5, 512–518.
- Blaauw, M., Christen, J.A., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal.* 6, 457–474.
- Blum, S.A.E., Lorenz, M.G., Wackernagel, W., 1997. Mechanism of retarded DNA degradation and prokaryotic origin of DNases in nonsterile soils. *Syst. Appl. Microbiol.* 20, 513–521.
- Boessenkool, S., McGlynn, G., Epp, L.S., Taylor, D., Pimentel, M., Gizaw, A., Nemomissa, S., Brochmann, C., Popp, M., 2014. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conserv. Biol.* 28, 446–455.
- Born, C., Kjellberg, F., Chevallier, M.-H., Vignes, H., Dikangadissi, J.-T., Sangué, J., Wickings, E.J., Hossaert-McKey, M., 2008. Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, *Aucoumea klaineana*. *Proc. R. Soc. B Biol. Sci.* 275, 2171–2179.
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* 16, 176–182.
- Britton, J.R., Boar, R., Grey, J., Foster, J., Lugonzo, J., Harper, D., 2007. From introduction to fishery dominance: the initial impacts of the invasive carp *Cyprinus carpio* in Lake Naivasha, Kenya, 1999 to 2006. *J. Fish Biol.* 71, 239–257.
- Coissac, E., Hollingsworth, P.M., Lavergne, S., Taberlet, P., 2016. From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.* 25 (7), 1423–1428.
- Corinaldesi, C., Danovaro, R., Dell'Anno, A., 2005. Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies from marine sediments. *Appl. Environ. Microbiol.* 71, 46–50.
- Cuni-Sanchez, A., White, L.J., Calders, K., Jeffery, K.J., Abernethy, K., Burt, A., Disney, M., Gilpin, M., Gomez-Dans, J.L., Lewis, S.L., 2016. African savanna-forest boundary dynamics: a 20-year study. *PLoS One* 11, e0156934.
- Davies, S.J., Lamb, H.F., Roberts, S.J., 2015. Micro-XRF Core Scanning in Palaeolimnology: Recent Developments, Micro-XRF Studies of Sediment Cores. Springer, pp. 189–226.
- Epp, L.S., Stoof, K.R., Trauth, M.H., 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. *J. Paleolimnol.* 43, 939–954.
- Ficetola, G.F., 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.
- Fuller, D.Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C.J., Qin, L., Allaby, R.G., Purugganan, M.D., 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci.* 111, 6147–6152.
- Giguet-Covex, C., Ficetola, G.F., Walsh, K.J., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., David, F., Taberlet, P., Arnaud, F., New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. submitted QSR.
- Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J.r.m., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J.r.m., Taberlet, P., 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5.
- Haile, J., Froese, D.G., MacPhee, R.D.E., Roberts, R.G., Arnold, L.J., Reyes, A.V., Rasmussen, M., Nielsen, R., Brook, B.W., Robinson, S., Demuro, M., Gilbert, M.T.P., Munch, K., Austin, J.J., Cooper, A., Barnes, I., Möller, P., Willerslev, E., 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl. Acad. Sci.* 106, 22352–22357.
- Heiri, O., Lotter, A.F., Lemcke, G., 2001. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *J. Paleolimnol.* 25, 101–110.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2004. The WorldClim Interpolated Global Terrestrial Climate Surfaces. Version 1.3.
- Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., 2012. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* 21, 1989–2003.
- 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. *Quat. Int.* 184, 139–149.
- Mergeay, J., Verschuren, D., Kerckhoven, L.V., Meester, L.D., 2004. Two hundred years of a diverse *Daphnia* community in Lake Naivasha (Kenya): effects of natural and human-induced environmental changes. *Freshw. Biol.* 49, 998–1013.
- Metcalfe, S.E., Jones, M.D., Davies, S.J., Noren, A., MacKenzie, A., 2010. Climate variability over the last two millennia in the North American Monsoon region, recorded in laminated lake sediments from Laguna de Juanacatlán, Mexico. *Holocene* 20, 1195–1206.
- Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., Nogués-Bravo, D., 2016. An Anthropocene map of genetic diversity. *Science* 353, 1532–1535.
- O'Brien, P.J., 1972. The sweet potato: its origin and dispersal. *Am. Anthropol.* 74, 342–365.
- Pansu, J., Giguet-Covex, C., Ficetola, G.F., 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. *Mol. Ecol.* 24, 1485–1498.
- Parducci, L., Bennet, K.D., Ficetola, G.F., Alsos, I.G., Suyama, Y., Wood, J.R., Pedersen, M.W., 2017. Ancient plant DNA in lake sediments. *New Phytol.* 214 (3), 924–942.
- Parducci, L., Jørgensen, T., Tollefson, M.M., Elverland, E., Alm, T., Fontana, S.L., Bennett, K.D., Haile, J., Matetovici, I., Suyama, Y., Edwards, M.E., Andersen, K., Rasmussen, M., Boessenkool, S., Coissac, E., Brochmann, C., Taberlet, P., Houmark-Nielsen, M., Larsen, N.K., Orlando, L., Gilbert, M.T.P., Kjær, K.H., Alsos, I.G., Willerslev, E., 2012. Glacial survival of boreal trees in northern Scandinavia. *Science* 335, 1083–1086.
- Parducci, L., Matetovici, I., Fontana, S.L., Bennett, K.D., Suyama, Y., Haile, J., Kjær, K.H., Larsen, N.K., Drouzas, A.D., Willerslev, E., 2013. Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Mol. Ecol.* 22, 3511–3524.
- Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., Willerslev, E., Kjær, K.H., 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quat. Sci. Rev.* 75, 161–168.
- Salzmann, U., Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *Holocene* 15, 190–199.
- Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V., Allaby, R.G., 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* 347, 998–1001.
- Stoof-Leichsenring, K.R., Epp, L.S., Trauth, M.H., Tiedemann, R., 2012. Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Mol. Ecol.* 21, 1918–1930.

- Stoof-Leichsenring, K.R., Junginger, A., Olaka, L.A., Tiedemann, R., Trauth, M.H., 2011. Environmental variability in Lake Naivasha, Kenya, over the last two centuries. *J. Paleolimnol.* 45, 353–367.
- Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012a. Environmental DNA. *Mol. Ecol.* 21, 1789–1793.
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic acids Res.* 35, e14.
- Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., 2012b. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* 21, 1816–1820.
- Thomsen, P.F., Willerslev, E., 2015. Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18.
- Valentini, A., Miquel, C., Nawaz, M.A., Bellemain, E., Coissac, E., Pompanon, F., Gielly, L., Cruaud, C., Nascetti, G., Wincker, P., 2009. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* 9, 51–60.
- Weiβ, C.L., Dannemann, M., Prüfer, K., Burbano, H.A., 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.
- White, F., 1983. The Vegetation of Africa, a Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa. Unesco, Paris.
- Willerslev, E., Cooper, A., 2005. Ancient DNA. *Proc. Biol. Sci.* 272, 3–16.
- Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A., 2003. Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science* 300, 791–795.
- Yoccoz, N., Bräthen, K., Gielly, L., Haile, J., Edwards, M., Goslar, T., Von Stedingk, H., Brysting, A., Coissac, E., Pompanon, F., 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* 21, 3647–3655.