Five thousand years of tropical lake sediment DNA records from Benin

Bremond L.[1,2], Favier C.[1], Ficetola G.F.[3,4,5], Tossou M.G.[6], Akouégninou A.[6], Gielly L.[3,4], Giguet-Covex C.[7], , Oslisly. R.[8], Salzmann U.[9]

1 Institut des Sciences de l'Évolution - Montpellier, UMR 5554 CNRS-IRD-Université Montpellier-EPHE, Montpellier, France.

2 École Pratique des Hautes Études, PSL Research University, 4-14 rue Ferrus, 75014 Paris, France.

3 University Grenoble Alpes, LECA, F-38000 Grenoble, France

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

5 Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, 20133, Milan, Italy.

6 National Herbarium, Department of Botany and Plant Biology, Faculty of Sciences and Technology (FAST) University of Abomey-Calavi (UAC) Cotonou, Benin.

7 EDYTEM, Université de Savoie, CNRS, Pôle Montagne, 73376 Le Bourget du Lac, France.

8 UMR PALOC Institut de Recherche pour le Développement et Agence Nationale des Parcs Nationaux, BP 20379 Libreville, Gabon.

9. Department of Geography, Faculty of Engineering and Environment, Northumbria University, Ellison Building, Newcastle upon Tyne, NE1 8ST, UK

Corresponding author: Laurent Bremond

E-mail address: laurent.bremond@umontpellier.fr
Postal address: Institut des Sciences de l'Evolution - Montpellier
UMR5554 UM – CNRS – IRD - EPHE
Université de Montpellier
CC 061 34095 MONTPELLIER Cedex 05

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.

Abstract

Until now, sedimentary DNA (*sed*DNA) studies have only focused on cold and temperate regions were 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 *sed*DNA from Lake Sele, localized in the tropical lowlands of Benin (Africa), and compared the taxonomic diversity detected by DNA analyses with pollen assemblages. Plant *sed*DNA was successfully amplified from 33 of the 34 successfully extracted samples. In total, 43 taxa were identified along the 5,000 years spanned by the sediment: 22 taxa were identified at the family level and 21 at the genus level. The plant diversity recovered through *sed*DNA 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 *sed*DNA. 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.

Keywords: *sed*DNA, metabarcoding, tropical, Africa, vegetation dynamics, plant diversity, pollen

Introduction

Plant sedimentary DNA (*sed*DNA) 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 *sed*DNA 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; Giguet-Covex et al., 2014; Parducci et al., 2017). The oldest *sed*DNA 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 *sed*DNA 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 *sed*DNA 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 et al., 2015).

Few studies have explored *sed*DNA 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 *sed*DNA 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 *sed*DNA 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) *sed*DNA sequences were obtain from 4,500-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. *Sed*DNA 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 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-Sanchez 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 5,000 years, and to test the pertinence of *sed*DNA 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 *sed*DNA.

Material and Method

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; Figure 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 <1.5m 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 ~200-300m 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 guinensis*, 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, namely *Triplochiton scleroxylon* (abachi, Sterculiaceae), *Celtis* spp. (Cannabaceae), and Ulmaceae*, Milicia excelsa* (iroko, Moraceae) are presented in Salzmann and Hoelzmann (2005).

Sampling

We collected five meters of sediment cores from the deepest part of the lake (N7.15537° E2.44106°). 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 lowers 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 ^{14}C analyses were carried out on terrestrial macroremains (seeds and piece of plant leafs) by the Poznan Radiocarbon Laboratory. The ^{14}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 4,400 years cal. BP and 3,100 years cal. BP, as previously indicated by Salzmann and Hoelzmann (2005) (AMS ^{14}C dates and age-depth model in supplementary data: Appendix S3).

158 **Samples for DNA analysis**

159 From the sediment core, we sampled 50 slices at 2 cm thick. Sampling of core slices was
160 carried out at the University of Savoie. To avoid DNA contamination, particularly with
161 regards to water circulation along the coring tube, 10 mm was removed from the edge of
162 each sediment slice. The tools (knives, spoons, pliers) were cleaned after each sampling
163 with distilled water, then alcohol and inflamed. DNA extraction, PCR set-up and PCR
164 amplification steps were later performed in three separate rooms at the University
165 Grenoble Alpes, which is specifically dedicated to ancient DNA studies. *Sed*DNA
166 extraction targeted extracellular DNA (Pansu et al., 2015). For each sediment slice, we
167 mixed approximately 15 g of sediment with 15 ml of saturated phosphate buffer
168 (Na2HPO4; 0.12 M; pH ≈ 8) for 15 minutes. The mixture was then centrifuged (10 min at
169 10000 g). The resulting supernatant (12 ml) was transferred to Amicon ® Ultra-15 10K
170 Centrifugal Filter Devices (Millipore) and centrifuged (20 min at 4000 g) to concentrate
171 *sed*DNA. Of the resulting concentrate, 400µl were kept as a starting material for the
172 following extraction steps, using the NucleoSpin® Soil kit (Macherey-Nagel, Düren,
173 Germany), skipping the cell lysis step and following the manufacturer's instructions
174 (Taberlet et al., 2012b). Three extraction controls were performed.

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

177

178 **DNA amplification and high-throughput sequencing**

179 Plant DNA was amplified with the universal primers "g" and "h" (Taberlet et al., 2007),
180 targeting the short and variable P6 loop region of the chloroplast trnL (UAA) intron.
181 These primers are highly conserved in angiosperms and gymnosperms, and are thus of
182 interest for plant metabarcoding (Taberlet et al., 2007). To assign sequence reads to the

183 relevant sample, 8 bp tags (with at least 5 differences between them) were added to the
184 5' end of primers (Binladen et al., 2007; Valentini et al., 2009). DNA amplifications were
185 carried out in a final volume of 30 µl containing 3 µl of diluted DNA extract. The
186 amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied
187 Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM of MgCl2, 0.2 mM of each dNTP, 0.2 µM
188 of each primer, and 4.8 µg of bovine serum albumin (BSA, Roche Diagnostic). After 10
189 minutes at 95° C for polymerase activation, the PCR mixture underwent 45 cycles of 30 s
190 at 95 °C, 30 s at 50 °C, and 1 min at 72 °C, followed by a final elongation step (7 min at
191 72 °C). In order to reduce the rate of false negatives (Ficetola et al., 2015) PCR products
192 were then purified. All the samples were identified by unique 8 bp tags at the 5' end of
193 the primers, which allowed the identification of the products of each PCR. This allowed
194 the pooling of all the different PCR products before library preparation. To summarize,
195 we performed PCRs on 51 samples, plus 3 extraction controls, plus 5 negative and 2
196 positive PCR controls (total: 61 samples including controls, arranged in one 96-wells
197 micrititer plate; 61 samples X 8 PCRs = 488 PCR reactions). Sequencing was performed
198 by 2 x 125 bp pair-end sequencing on an Illumina HiSeq 2500 platform, which returned
199 a total of 8'321'810 reads.
200
201 **Sequences filtering and taxa assignment**
202 DNA sequences were filtered using OBITools software (Boyer et al., 2016), following the
203 protocol described in Pansu et al. (2015). The obtained sequences were then assigned to
204 the relevant taxa using the *ecotag* program that locates highly similar sequence(s) in a
205 suitable database. Here, sequences were compared to a global plant database generated
206 using silico PCR from EMBL with the *ecoPCR* program (Ficetola et al., 2010) and a BLAST
207 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).

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 *sed*DNA was extracted, Loss On Ignition (LOI) analyses were performed on 1 cm^3 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 *sed*DNA 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 *sed*DNA. 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.

Pollen assemblages

The pollen counts used for comparison with *sed*DNA 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).

Results

Plant DNA *extraction and identification*

*Sed*DNA 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 5,365 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 (compatible 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, Poaceae, *Musa*,

Apiaceae, *Crassula*, Polygonaceae, Amaranthaceae, Rubiaceae, and Fabaceae, and were therefore removed from analyses.

The number of taxa and reads replicates decreases with the sediment's age (Figure 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 (figure 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 (Figure 2). Organic matter production induces modifications in the oxygenation conditions at the lake bottom, which also affects the conditions of *sed*DNA 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 2,100 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 3,000 years cal. BP and 1,000 years cal. BP.

The comparison with the pollen spectra showed important differences among the diversity and the presence of the main taxa (Figure 3). Some taxa, such as *Nymphaea* and Asteraceae, presented coherent signals. More taxa were recovered from pollen data than from *sed*DNA. However, 13 DNA taxa have no corresponding pollen and spore taxa (Figure 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 *sed*DNA taxa cannot be directly compared with pollen assemblages dynamic because of the scarcity of DNA taxa in the deep part of the core.

Discussion

1- Lessons for *sed*DNA analyses in lowland tropical lakes

For the first time, we obtained amplification of up to 5,000-year-old plant *sed*DNA from lowland tropical lake sediments. While the use of plant *sed*DNA 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 *sed*DNA 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 *sed*DNA in general, as well as in the specific context of lowland tropical lakes, namely: i) deposition 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 *sed*DNA 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 *sed*DNA and the number of reads tended to decline along the core. Nevertheless, some taxa were detected in sediments even older than 4,500 years. Despite the high tropical temperatures, *sed*DNA 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 1.5m 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 3,000 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 to 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 *sed*DNA, 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 *sed*DNA.

The specific alluvial nature of sediments (after the hiatus to 900 years. cal. BP) with coarse detrital input during a high lake lev@el (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 *sed*DNA (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.

2- Taxonomic resolution of taxa identified by DNA

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 *sed*DNA from trees is extremely limited. More taxa were recovered from pollen data than from *sed*DNA; 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 *sed*DNA records often show low overlapping (Jørgensen et al., 2012; Parducci et al., 2013; Parducci et al., 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 pollens can also represent regional vegetation, while *sed*DNA provides a more local signal (Boessenkool et al., 2014; Parducci et al., 2017). For the moment, *sed*DNA data are more related to the presence/absence of taxa, as macroremain analysis are interpreted in paleoecological studies (Birks, 2001). But some

383 recent studies on soils suggest that the number of reads can be related to biomass plant
384 production (Yoccoz et al., 2012; Pansu et al., 2015). In our data, it appears that for some
385 taxa (e.g. *Nymphaea*) the number of reads and pollen percentages actually correlate.
386 Combining pollen and *sed*DNA should help improve vegetation reconstruction, notably
387 in tropical forests where a large amount of trees do not have an entomophilous pollen
388 dispersal mode.
389
390 *Taxonomic reference library*
391 The lack of a complete DNA reference library of local species is a major issue for the
392 application of *sed*DNA for vegetation reconstruction. For instance, one sequence
393 detected matched well (97% identity) with the GenBank sequences of the genus
394 *Pontederia*. As this genus is not currently present in Benin, it is likely that the DNA
395 sequence belongs to *Eichhornia crassipes* (Pontederiaceae), for which no sequences are
396 available in GenBank. This plant was introduced from South America several decades
397 ago and now completely covers the periphery of the lake, which confirms that the DNA
398 signal is essentially coming from the vegetation that is developing on the shoreline and
399 in the shallow parts of the lake.
400 It is also possible that some locally present species, for which no sequences are available
401 in GenBank, were excluded because they matched related taxa with similar sequences
402 that are absent in Benin. Indeed, a large number of exotic taxa exotic from the study area
403 that were found in the lake samples were not from the control samples. The
404 identification of taxa through metabarcoding is severely limited by the availability of
405 sequences within reference databases; and if no sequences of a given taxon are
406 available, taxonomic assignment tools tend to assign sequences to the closest taxon that
407 can be available further apart. The development of extensive reference databases is,

408 therefore, a key priority to fully exploit the power of metabarcoding for environmental
409 reconstruction (Taberlet et al., 2012a). Efforts to map the genetic diversity of vegetation
410 are ongoing; unfortunately these mainly focus on Europe and North America, while
411 information on highly diverse tropical areas remains scarce (Coissac et al., 2016;
412 Miraldo et al., 2016).
413
414 3. Agriculture recorded by *sed*DNA
415 We detected the DNA of the genus *Ipomoea* (Convolvulaceae) in recent sediments.
416 Although the specific attribution is not certain, it is likely that the DNA signal was
417 produced by sweet potatoes (*Ipomoea batatas*); these were introduced to the Lake Sele
418 region from South America, probably at the same time as *Eichhornia crassipes,* in the
419 sixteenth century (O'Brien, 1972). There is little chance that another *Ipomoea* may have
420 produced this DNA signal, given that the pollen record indicates a continuous presence
421 of (non-cultivated) Convolvulaceae growing presumably on the lakeshores since the
422 level dropped after 3,000 cal. BP.
423 The original aim of this study was to test if agriculture could be recorded by *sed*DNA.
424 The detection of *Ipomoeae sp.* (sweet potato) DNA in recent sediments suggests that
425 *sed*DNA has the potential to trace the history of agricultural developments in these
426 environments. Nowadays sweet potato is cultivated closed to permanent water, on the
427 shoreline during the dry season when the lake is low. In some lakes, the DNA of plants
428 growing near the shore is more easily detected than the DNA of plants living further
429 away (Inger Alsos, pers. com.), which may explain why this species is better recorded
430 than the cassava which is grown away from the shore. However, it is surprising that
431 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 8,000 years ago in the British Isles, before the Neolithic introduction of domestic cereals, is an illustrative example.

Conclusion

Our results confirm the usefulness of *sed*DNA 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 *sed*DNA. Fine clayous sediments are more susceptible to contain extractable *sed*DNA 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. *Sed*DNA is a valuable approach in tropical areas for searching specific plants growing close to the water. The plant diversity recovered through *sed*DNA 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 *sed*DNA studies. Also, although mammal DNA was not researched

456 in our sediments it must be considered because in tropical areas, more than anywhere
457 else, permanent water is an attractive place for wild fauna and domestic stocks.
458
459 **Acknowledgment**
460 This work was supported by project PEPS-CNRS “Banana Connexion”. We would like to
461 thank Benoit Brossier for his technical support during field trip. We also thank Allowen
462 Evin for fruitful discussions and Frederic Boyer for raw data handling. We thank Jill
463 Cucchi for professional English editing.
464
465 **Conflict of interest**
466 The authors would like to mention that LG is one of the co-inventors of patents related
467 to g-h primers and the subsequent use of the P6 loop of the chloroplast *trn*L (UAA)
468 intron for plant identification using degraded template DNA. These patents only restrict
469 commercial applications and have no impact on the use of this locus by academic
470 researchers.
471
472
473
474
475
476
477
478
479
480

References

1 Adam, K.S., Boko, M., 1983. Le Bénin. les Ed. du Flamboyant EDICEF, Paris.

2 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. The Royal Society, pp. 4724-4733.

3 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. The Holocene 26, 627-642.

4 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 Evolutionary biology 11:30, 1-15.

5 Binladen, 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.

6 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.

7 Blaauw, M., 2010. Methods and code for 'classical'age-modelling of radiocarbon sequences. Quaternary Geochronology 5, 512-518.

8 Blaauw, M., Christen, J.A., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Analysis 6, 457-474.

9 Blum, S.A.E., Lorenz, M.G., Wackernagel, W., 1997. Mechanism of Retarded DNA Degradation and Prokaryotic Origin of DNases in Nonsterile Soils. Systematic and Applied Microbiology 20, 513-521.

10 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. Conservation Biology 28, 446-455.

11 Born, C., Kjellberg, F., Chevallier, M.-H., Vignes, H., Dikangadissi, J.-T., Sanguié, J., Wickings, E.J., Hossaert-McKey, M., 2008. Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, Aucoumea klaineana. Proceedings of the Royal Society B: Biological Sciences 275, 2171-2179.

12 Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: a unix‐inspired software package for DNA metabarcoding. Molecular Ecology Resources 16, 176-182.

13 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. Journal of Fish Biology 71, 239-257.

14 Coissac, E., Hollingsworth, P.M., Lavergne, S., Taberlet, P., 2016. From barcodes to genomes: extending the concept of DNA barcoding. Molecular Ecology.

15 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.

16 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.

17 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.

18 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. Journal of Paleolimnology 43, 939-954.

19 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, 1.

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

21 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. Proceedings of the National Academy of Sciences 111, 6147-6152.

22 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.

576 23 Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J.r.m., Griggo, C., Gielly, L., Domaizon,
577 I., Coissac, E., David, F., Choler, P., Poulenard, J.r.m., Taberlet, P., 2014. Long
578 livestock farming history and human landscape shaping revealed by lake
579 sediment DNA. Nat Commun 5.
580
581 24 Haile, J., Froese, D.G., MacPhee, R.D.E., Roberts, R.G., Arnold, L.J., Reyes, A.V.,
582 Rasmussen, M., Nielsen, R., Brook, B.W., Robinson, S., Demuro, M., Gilbert, M.T.P.,
583 Munch, K., Austin, J.J., Cooper, A., Barnes, I., Möller, P., Willerslev, E., 2009. Ancient
584 DNA reveals late survival of mammoth and horse in interior Alaska. Proceedings
585 of the National Academy of Sciences 106, 22352-22357.
586
587 25 Heiri, O., Lotter, A.F., Lemcke, G., 2001. Loss on ignition as a method for
588 estimating organic and carbonate content in sediments: reproducibility and
589 comparability of results. Journal of Paleolimnology 25, 101-110.
590
591 26 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2004. The WorldClim
592 interpolated global terrestrial climate surfaces. Version 1.3.
593
594 27 Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M.,
595 Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., 2012. A comparative study of
596 ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of
597 northern Siberia reveals long‐term vegetational stability. Molecular Ecology 21,
598 1989-2003.
599
600 28 Matisoo-Smith, E., Roberts, K., Welikala, N., Tannock, G., Chester, P., Feek, D.,
601 Flenley, J., 2008. Recovery of DNA and pollen from New Zealand lake sediments.
602 Quaternary International 184, 139-149.
603
604 29 Mergeay, J., Verschuren, D., Kerckhoven, L.V., Meester, L.D., 2004. Two hundred
605 years of a diverse Daphnia community in Lake Naivasha (Kenya): effects of
606 natural and human-induced environmental changes. Freshwater Biology 49, 998-
607 1013.
608
609 30 Metcalfe, S.E., Jones, M.D., Davies, S.J., Noren, A., MacKenzie, A., 2010. Climate
610 variability over the last two millennia in the North American Monsoon region,
611 recorded in laminated lake sediments from Laguna de Juanacatlán, Mexico. The
612 Holocene 20, 1195-1206.
613
614 31 Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S.,
615 Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., Nogués-Bravo, D., 2016. An
616 Anthropocene map of genetic diversity. Science 353, 1532-1535.
617
618 32 O'Brien, P.J., 1972. The sweet potato: its origin and dispersal. American
619 Anthropologist 74, 342-365.
620
621 33 Pansu, J., Giguet‐Covex, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F.,
622 Poulenard, J., Taberlet, P., Choler, P., 2015. Reconstructing long‐term human
623 impacts on plant communities: an ecological approach based on lake sediment
624 DNA. Molecular Ecology 24, 1485-1498.

34 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 Phytologist . DOI: 10.1111/nph.14470.

35 Parducci, L., Jørgensen, T., Tollefsrud, 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.

36 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. Molecular Ecology 22, 3511-3524.

37 Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., Willerslev, E., Kjaer, K.H., 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. Quaternary Science Reviews 75, 161-168.

38 Salzmann, U., Hoelzmann, P., 2005. The Dahomey Gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. The Holocene 15, 190-199.

39 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.

40 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. Molecular Ecology 21, 1918-1930.

41 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. Journal of Paleolimnology 45, 353-367.

42 Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012a. Environmental DNA. Molecular Ecology 21, 1789-1793.

43 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 research 35, e14.

44 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. Molecular Ecology 21, 1816-1820.

45 Thomsen, P.F., Willerslev, E., 2015. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation 183, 4-18.

46 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. Molecular Ecology Resources 9, 51-60.

47 Weiß, 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.

48 White, F., 1983. The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. Unesco, Paris.

49 Willerslev, E., Cooper, A., 2005. Ancient DNA. Proc Biol Sci 272, 3-16.

50 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.

51 Yoccoz, N., Bråthen, 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. Molecular Ecology 21, 3647-3655.

Figure captions

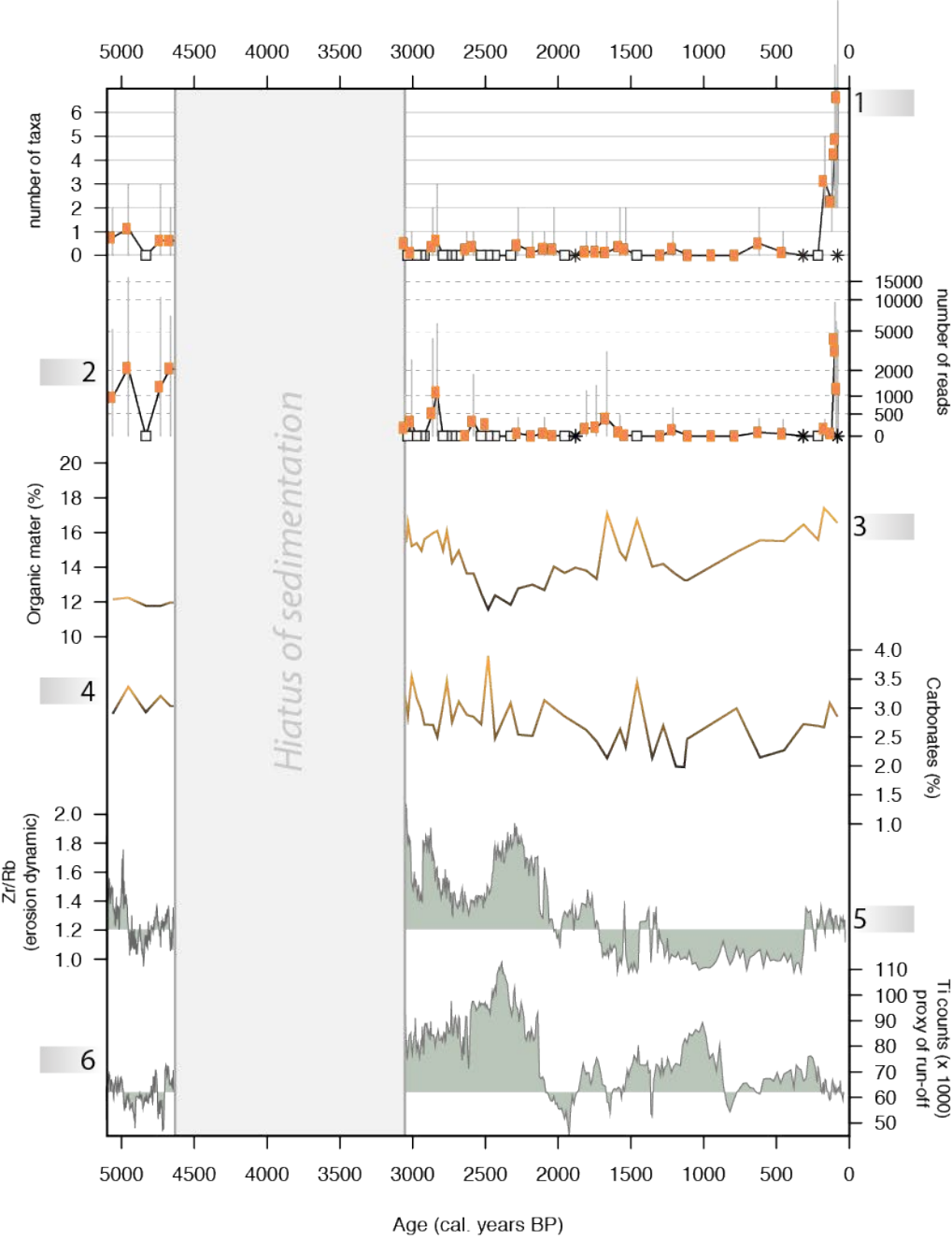
Figure 1. Localization of the coring site in Lake Sele (Benin). The annually submerged shorelines are identifiable by the greenness of the aquatic plants and the agricultural parcels which present geometric forms due to the shapes of the fields.

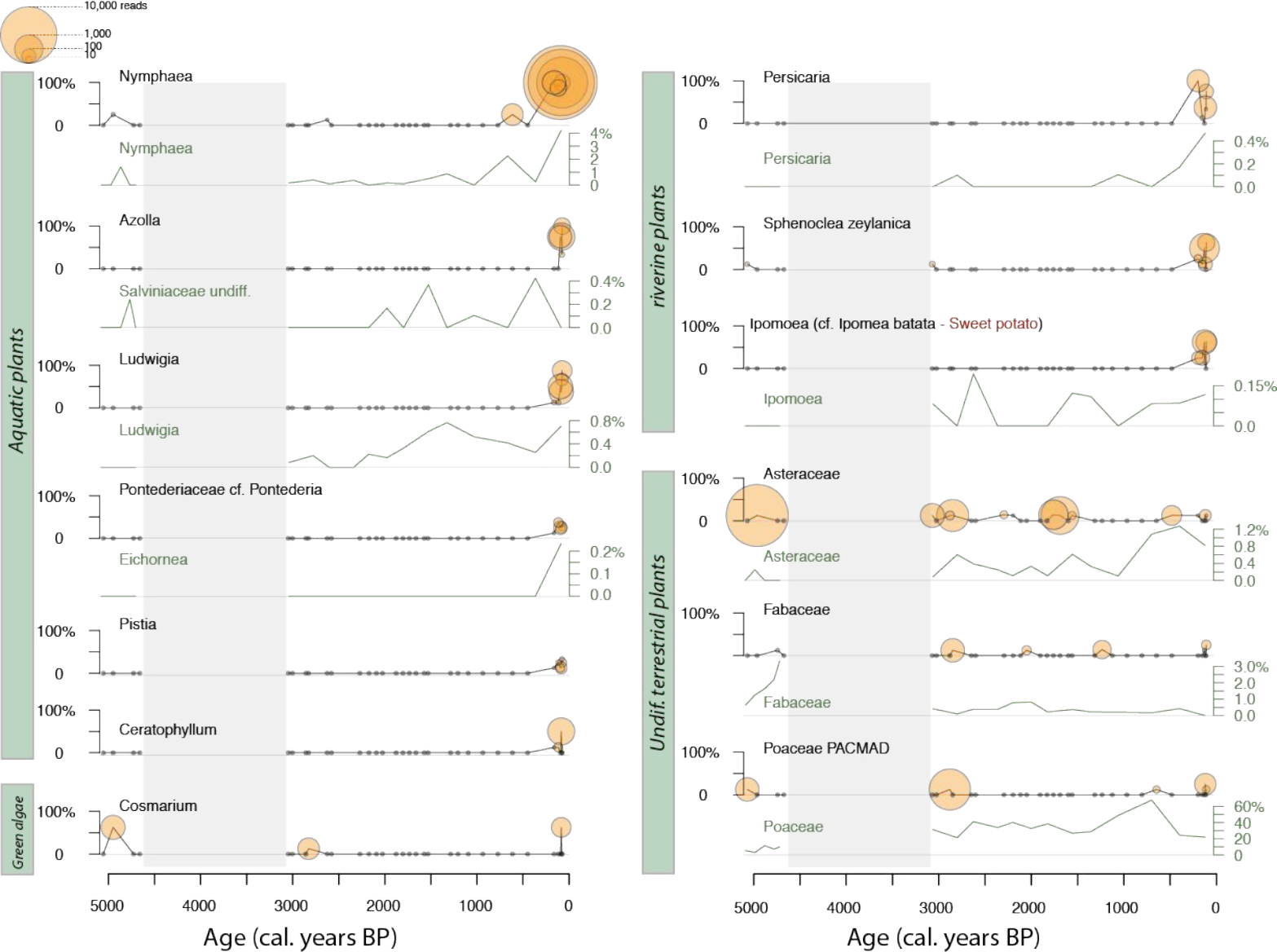
Figure 2. Comparison between *sed*DNA and sedimentology during the last 5,000 years in Lake Sele: 1) Number of identified taxa (mean values and standard deviations of the replicates) and 2) number of reads (mean values and standard deviations of the replicates). The stars represent the samples where 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 ($LOI_{550°C}$). 4) Carbonate content ($LOI_{950°C}$). 5) Zr/R–b as proxy of erosion dynamic (high values = coarser grain-size), and 6) Ti related to increased run-off/rainfall.

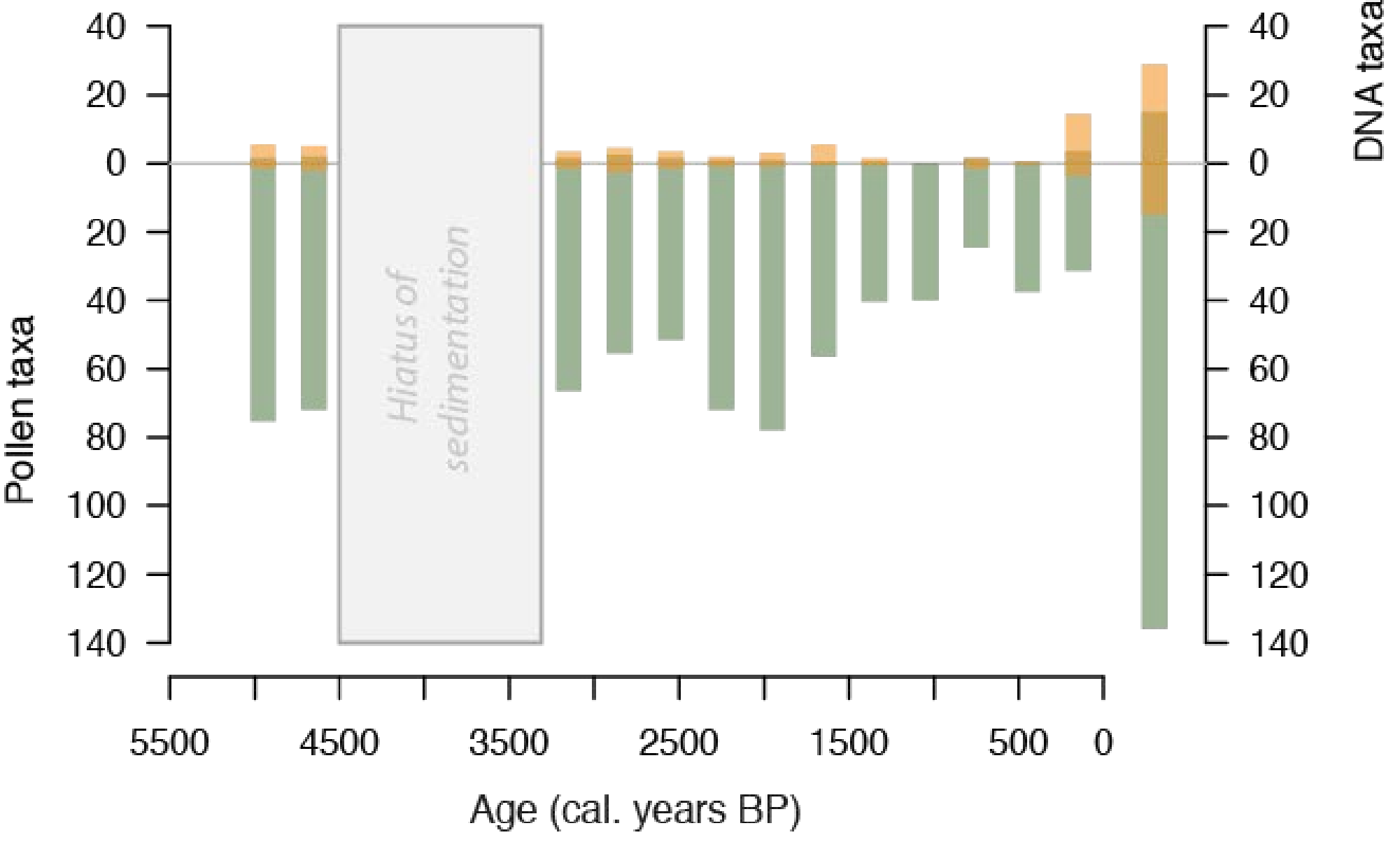
Figure 3. Comparison between the *sed*DNA of plants (in orange) and pollen data (in green) from Lake Sele. The DNA signal is expressed as a percentage of the total number of replicates per level. The size of the circles indicates the number of reads for the taxon. The pollen percentage is calculated using the total pollen sum excluding undetermined, modified from Salzmann and Hoelzmann (2005).

Figure 4. Number of taxa identified, per 200 years intervals, by *sed*DNA (orange) and pollen (green). Common taxa are in dark orange.


N
SELE-1&2
500m
Image © 2016 CNES / Astrium
© 2016 Google
Google earth


5000 4500 4000 3500 3000 2500 2000 1500 1000 500 0
number of taxa
0 1 2 3 4 5 6
1
2
number of reads
15000 10000 5000 2000 1000 500 0
Hiatus of sedimentation
Organic mater (%)
10 12 14 16 18 20
3
4
Carbonates (%)
1.0 1.5 2.0 2.5 3.0 3.5 4.0
Zr/Rb
(erosion dynamic)
1.0 1.2 1.4 1.6 1.8 2.0
5
6
Ti counts (x 1000)
proxy of run-off
50 60 70 80 90 100 110
5000 4500 4000 3500 3000 2500 2000 1500 1000 500 0
Age (cal. years BP)


10,000 reads
1,000
100
10
Aquatic plants
Green algae
riverine plants
Undif. terrestrial plants
Nymphaea
Nymphaea
Azolla
Salviniaceae undiff.
Ludwigia
Ludwigia
Pontederiaceae cf. Pontederia
Eichornea
Pistia
Ceratophyllum
Cosmarium
Persicaria
Persicaria
Sphenoclea zeylanica
Ipomoea (cf. Ipomea batata - Sweet potato)
Ipomoea
Asteraceae
Asteraceae
Fabaceae
Fabaceae
Poaceae PACMAD
Poaceae
100%
0
4%
3
2
1
0.4%
0.2
0.8%
0.4
0.2%
0.1
1.2%
0.8
3.0%
2.0
1.0
60%
40
20
5000
4000
3000
2000
1000
0
Age (cal. years BP)


Pollen taxa
DNA taxa
40
20
0
20
40
60
80
100
120
140
Hiatus of sedimentation
5500
4500
3500
2500
1500
500
0
Age (cal. years BP)