Reconstructing long-term human impacts on plant communities: an ecological

approach based on lake sediment DNA

Johan Pansu[1,2*], Charline Giguet-Covex[1,2,3,4,5*], Gentile Francesco Ficetola[1,2], Ludovic

Gielly[1,2], Frédéric Boyer[1,2], Lucie Zinger[6], Fabien Arnaud[3,4,5], Jérôme Poulenard[3,4,5],

Pierre Taberlet[1,2] & Philippe Choler[1,2,7,8]

[1] Univ. Grenoble Alpes, LECA, F-38000 Grenoble, France

[2] CNRS, LECA, F-38000 Grenoble, France

[3] Univ. Savoie Mont Blanc, EDYTEM, F-73000 Chambéry, France

[4] CNRS, EDYTEM, F-73000 Chambery, France

[5] Ministère de la culture et de la communication, EDYTEM, F-73000 Chambéry, France

[6] Laboratoire Evolution & Diversité Biologique, UMR CNRS-UPS 5174, Toulouse, France

[7] Univ. Grenoble Alpes, SAJF, F-38000 Grenoble, France

[8] CNRS, SAJF, F-38000 Grenoble, France

* These authors contributed equally to this work.

johan.pansu@gmail.com

charline.giguet@gmail.com

francesco.ficetola@gmail.com

ludovic.gielly@ujf-grenoble.fr

frederic.boyer@ujf-grenoble.fr

lucie@zinger.fr

fabien.arnaud@univ-savoie.fr

jerome.poulenard@univ-savoie.fr

pierre.taberlet@ujf-grenoble.fr

philippe.choler@ujf-grenoble.fr

Running title

Human impact on plant communities from *seda*DNA

First and corresponding authors: Johan Pansu and Charline Giguet-Covex.

johan.pansu@gmail.com

Tel: +33 (0)4 76 51 36 71

Fax: +33 (0)4 76 51 42 79

Abstract

Paleoenvironmental studies are essential to understand biodiversity changes over long timescales and to assess the relative importance of anthropogenic and environmental factors. Sedimentary ancient DNA (*seda*DNA) is an emerging tool in the field of paleoecology and has proven to be a complementary approach to the use of pollen and macroremains for investigating past community changes. *Seda*DNA-based reconstructions of ancient environments often rely on indicator taxa or expert knowledge, but quantitative ecological analyses might provide more objective information. Here, we analysed *seda*DNA to investigate plant community trajectories in the catchment of a high-elevation lake in the Alps over the last 6,400 years. We

combined data on past and present plant species assemblages along with sedimentological and geochemical records to assess the relative impact of human activities through pastoralism, and abiotic factors (temperature and soil evolution). Over the last 6,400 years, we identified significant variation in plant communities, mostly related to soil evolution and pastoral activities. An abrupt vegetational change corresponding to the establishment of an agro-pastoral landscape was detected during the Late Holocene, approximately 4,500 years ago, with the replacement of mountain forests and tall-herb communities by heathlands and grazed lands. Our results highlight the importance of anthropogenic activities in mountain areas for the long-term evolution of local plant assemblages. *Seda*DNA data, associated with other paleoenvironmental proxies and present plant assemblages, appears to be a relevant tool for reconstruction of plant cover history. Their integration, in conjunction with classical tools, offers interesting perspectives for a better understanding of long-term ecosystem dynamics under the influence of human-induced and environmental drivers.

Introduction

The "Anthropocene" concept was originally defined by geologists to characterise a new period in which human activities became one of the most important factors affecting geological processes at the global scale (Crutzen 2002; Steffen *et al.* 2011; Wilkinson 2005; Ruddiman 2012). However, biodiversity and ecological processes are also affected by human activities. Thus, this definition has been expanded to the earth's

biotic component, as a period in which ecosystem functioning is primarily influenced by human activities (Ellis 2011). Understanding how and when human activities began modifying ecosystems, and how ecosystems responded to these human-modified landscapes, is a key challenge (Seddon *et al.* 2014).

Because ecological monitoring rarely extends beyond a few decades, long-term studies are crucial to understand the causes and the dynamics of environmental changes (Seddon *et al.* 2014). Lakes are recognised as sentinels of environmental changes (Adrian *et al.* 2009). Their sediments accumulate information about biological communities living both within the water body (*e.g.,* cyanobacteria, diatoms, protists, invertebrates; Lotter 1998; Heiri *et al.* 2003; Millet *et al.* 2009; Coolen & Gibson 2009), and within the neighbouring terrestrial environment (*e.g.,* Birks & Birks 2000; Gaillard *et al.* 2008; Gauthier *et al.* 2010*;* Tinner *et al.* 2003). The history of vegetation and human impact in the areas surrounding lakes is usually inferred from pollen and macroremains (*e.g.,* Birks & Birks 2000; David *et al.* 2010; Anderson-Carpenter *et al.* 2011; Rey *et al.* 2013). The combination of these proxies is essential to reconstruct trajectories of plant assemblages: pollen provides information representing wide spatial scales, due to the wind transport, while macroremains mainly represent the local plant assemblages (Birks & Birks 2000; Seppä & Bennett 2003). These macrofossils are more sporadic and their identification level is variable, depending on the nature and the preservation state of plant remains (Jørgensen *et al.* 2012; Boessenkool *et al.* 2014).

A complementary approach based on sedimentary ancient DNA (*seda*DNA) is now emerging (*e.g.,* Willerslev *et al.* 2003; Sønstebø *et al.* 2010; Jørgensen *et al.* 2012). The DNA of a variety of taxa, including plants and mammals, can be directly obtained

from recent and past sediments (Chariton *et al.* 2010; Hofreiter *et al.* 2012; Pedersen *et al.* 2013; Willerslev *et al.* 2014) even in the absence of visible macrofossils. This *seda*DNA constitutes a potential source of information to assess past lake communities (Stoof-Leichenring *et al.* 2012; Ravasi *et al.* 2012; Domaizon *et al.* 2013), but also species assemblages from the surrounding terrestrial ecosystems (Jørgensen *et al.* 2012; Parducci *et al.* 2013; Pedersen *et al.* 2013). Recent developments of high-throughput sequencing now allow massive sequencing of the ancient DNA contained in sedimentary archives.

Direct sequencing of environmental DNA allows the investigation of taxonomic and functional attributes of communities. Nevertheless, complete reference databases are required for reliable taxonomic identification. In the absence of such databases, identification relies primarily on the use of standardised, well-studied and discriminative genomic fragments (*e.g.,* 16S for bacteria) (Fierer *et al*. 2013). The DNA metabarcoding approach seeks to identify multiple taxa from environmental DNA samples in a single process by targeting these short DNA fragments (Taberlet *et al.* 2012a; Deagle *et al*. 2014). Different primer pairs have been developed to focus on particular taxonomic groups (*e.g.,* Epp *et al*. 2012). Recent studies have demonstrated the technical feasibility of *seda*DNA-based metabarcoding analyses and their reliability for paleoecological studies (*e.g.,* Jørgensen *et al.* 2012; Parducci *et al.* 2013; Boessenkool *et al.* 2014; Willerslev *et al.* 2014; Giguet-Covex *et al.* 2014). Several of these studies compared DNA-based reconstruction with pollen- and/or macroremains-based reconstructions, and concluded that metabarcoding of lake sediments is a tool complementary to classical approaches for investigating past flora (Jørgensen *et al.* 2012; Parducci *et al.* 2013; Boessenkool *et al.* 2014).

Paleoecological trajectories of plant communities are often studied using indicator taxa (Mazier *et al.* 2009). Multivariate analysis of ecological communities is a powerful tool that allows an objective understanding of the processes determining environmental variation (Legendre & Legendre 2012), but they have seldom been applied to *seda*DNA data. In this study, we combined these two approaches, using metabarcoding data from a subalpine lake, along with data on present plant species assemblages. Our purposes was to infer past vegetation dynamics in the catchment during the Holocene, and to assess the relative importance of human impacts, natural evolution of communities, and variation in climatic conditions in the observed trajectories. First, we analysed time-series (chronosequences) of plant DNA preserved in lake sediments. Second, we used information on species' current habitat preferences and other proxies (grazing activity, development of soils, climate change) to reconstruct past landscape history and the potential drivers of vegetation changes.

Material and methods

Study site and sampling

The study was carried out at Lake Anterne in North French Alps (2,063 m above sea level (a.s.l.) and above the present-day treeline). Catchment vegetation is mostly composed of subalpine grasslands dominated by *Festuca nigrescens* (Chewing's fescue), *Carex sempervirens* (evergreen sedge), and *Nardus stricta* (matgrass); scattered pockets of heathland dominated by *Vaccinium uliginosum* (northern bilberry) and *Rhododendron ferrugineum* (alpenrose) and alpine meadows dominated by *Sesleria caerulea* (blue moor-grass) and *Kobresia myosuroides* (Bellard's kobresia) (Fig. S1 in Supporting Information). Three main soil types were identified in the catchment: leptosols (~50% of the catchment area), cambisols (~25%) and acid soils

(podzols/stagnosols, ∼15%) (Giguet-Covex *et al.* 2011). The catchment is also characterised by the presence of steep slopes and highly erodible rocks (shales, calcshists and black shales). Consequently, Anterne lake sediments are mainly of detrital origin, *i.e.,* made of particles from rocks and soils of the catchment. They are characterised by the alternation of coarse-grained laminae and silty clay laminae corresponding to flood deposits triggered by precipitation events (Giguet-Covex *et al.* 2012). In detrital sediments a high proportion of pollen grains can be degraded, crumpled, broken or corroded, which makes their identification more difficult (Cushing 1967; Birks 1970; Wilmshurst & McGlone, 2005; Doyen *et al.* 2013). In such a context, ancient sediment DNA could be an interesting approach to reconstruct past vegetation cover.

The lake bottom is subjected to anoxic conditions during a large part of the year due to the combination of two processes: (i) the oxygen consumption by aerobic bacteria that decompose the organic matter, and (ii) the absence of processes allowing water column mixing, and thus the renewal of the oxygen in lake bottom. Indeed, in winter, the lake is ice-covered for six to seven months. This ice cap provides protection from wind and water inputs to the lake, and thus prevents the water column from mixing (Fig. S2). In summer, thermal stratification, combined with low flood activity, limit the oxygen supply in lake bottom (Fig. S2). Temperature monitoring shows variations between 4 and 8 °C in the lake bottom (unpublished data). These cold temperatures, combined with quasi-anoxic conditions, are favourable for DNA preservation (Coolen & Overmann 2007; Corinaldesi *et al.* 2008; Coolen & Gibson 2009). Moreover, sediments are well stratified, and their detrital nature induces fast compaction, which limits the thickness of the water/sediment interface where mixing is

possible. After compaction, vertical percolation of liquids, and thus DNA leaching, are prevented by the permanent saturation of sediments with water (Giguet-Covex *et al*. 2014).

The 20.2 m long core (ANT-07) used for the plant DNA study was retrieved from the deeper part of the lake (13 m) in 2007. Core parts specifically dedicated to DNA analyses were shrink-wrapped to limit oxidation of the sediments, stored in a refrigerated area (4 °C; a temperature similar to the lake bottom) and opened only for DNA extractions. Core parts used in DNA analyses covered the last 10,165 years according to the age-depth model based on 18 ^{14}C dates of terrestrial macroremains (Giguet-Covex *et al.* 2011). Soils and erosion histories are described elsewhere (Giguet-Covex *et al.* 2011; Giguet-Covex *et al.* 2012; Fig. S3). Moreover, a previous metabarcoding study established the presence of DNA attributed to domestic animals in these sediments records corresponding to the late Iron Age/Roman period and the Middle Ages/Early Modern Period (Giguet-Covex *et al.* 2014).

Extracellular DNA extraction

From the sediment core, 47 slices of 1 cm thickness were sampled with a focus on the last 6,000 years. Sampling of core slices was carried out in a laboratory at University of Savoie where no DNA research had previously been done. DNA extraction, PCR reaction setup and PCR amplification steps were later performed in 3 separate rooms specifically dedicated to ancient DNA studies at University Grenoble Alpes. To avoid potential contamination, we removed the edges of slices, which could contain contaminating DNA due to water circulation along the coring tube. Each sample was divided into 2 subsamples. Then, 2 extractions per subsample were carried out (*i.e.,* 4

per sample). Finally, 2 different DNA amplifications per extraction were performed to obtain 8 amplification replicates per core slice (Taberlet *et al.* 2012b). This strategy allowed us to maximise species detection, and also revealed potential sporadic contamination.

Sedimentary DNA includes both intracellular and extracellular DNA. The results from total DNA extraction and sequencing can thus be biased towards sequences from macroremains (Pedersen *et al.* 2013). We targeted extracellular DNA. Adsorption of extracellular DNA on sediment particles protects it against nuclease degradation (Pietramellara *et al.* 2009). The depurination kinetics of adsorbed DNA in marine sediments has been proven to proceed at much slower rates than those theoretically predicted or estimated for terrestrial environments (Corinaldesi *et al.* 2008; Corinaldesi *et al.* 2011). Extracellular DNA was removed from approximately 15 g of sediment by mixing it for 15 min with 15 ml of saturated phosphate buffer (Na$_2$HPO$_4$; 0.12 M; pH $\approx$ 8). Two millilitres of the mixture was then centrifuged for 10 min at 10000 g. Of the resulting supernatant, 400 µl was kept as starting material for the next 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). The DNA extract was eluted in 100 µl of elution buffer (SE) and diluted 10 times before being used as a PCR template. Four extraction controls were performed.

DNA amplification and high-throughput sequencing

Degradation and/or fragmentation processes limit the length of extracellular DNA sequences that can be retrieved from environmental samples (Yoccoz *et al*. 2012; Taberlet *et al*. 2012). We amplified plant DNA with universal primers *"g" and "h"*, which

target the short and variable P6 loop region of the chloroplast *trn*L (UAA) intron (between ~7 and ~143 base pairs (bp) but <100 bp in more than 99% of cases) (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 μl containing 3 μl of diluted DNA extract. The amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Courtaboeuf, France), 15 mM of Tris-HCl (pH 8.3), 50 mM KCl, 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 μM of each primer and 4.8 μg of bovine serum albumin (BSA, Roche Diagnostic). After 10 min at 95 °C for polymerase activation, the PCR mixture underwent 45 cycles: 30 s at 95 °C, 30 s at 50 °C, and 1 min at 72 °C. A final elongation step (7 min at 72 °C) was performed. Eight PCR controls were carried out, in addition to extraction controls, and randomly distributed on PCR plates. PCR products were then purified and mixed together before sequencing. Our amplification products were expected to range in length between ~65 and ~155 bp in most cases (Yoccoz *et al*. 2012). Hence, we performed a 2 x 100 bp paired-end sequencing on an Illumina HiSeq 2500 platform (part of a single lane), using the TruSeq SBS Kit v3, which allows recovery of the entire amplicons sequence (with an overlap typically greater than 45 bp).

Sequence filtering and taxa assignment

DNA sequences were filtered using the OBITOOLS software (http://metabarcoding.org/obitools). Because we performed a paired-end sequencing, the direct and reverse reads corresponding to the same sequence were first aligned and merged using a gap-free end algorithm implemented in the *IlluminaPairEnd* program. Only merged sequences with a high alignment quality score were retained (> 40; this

corresponds to a pair of reads that can align perfectly on at least 10 bp at each read end). The *ngsfilter* program assigned each merged sequence to its original sample using tags information. Only sequences containing both primers (with a maximum of 3 mismatches per primer) and exact tag sequences were selected. Sequences shorter than 7 bp or containing ambiguous nucleotides were discarded by the *obigrep* command. We used the *obiuniq* program to merge together strictly identical sequences, keeping the information about the original samples from which they came. Afterwards, singletons (*i.e.,* sequences occurring only once in the dataset) were filtered out. The *obiclean* program allows the detection of sequences corresponding to putative amplification or sequencing errors. For each PCR, the program creates a network that connects sequences differing by a single nucleotide substitution or indel. In each network, two sequences were considered linked if they differed in frequency by more than 50%; the less frequent sequence is assumed to derive from the other sequence. In each network, each sequence was then assigned to one of three categories: "head" (*i.e.,* a sequence that is linked only to sequences with a lower count), "singleton" (*i.e.,* a sequence with no link in the network) or "internal" (*i.e.,* sequences that do not fall into either of the other categories and are therefore likely to be errors). Sequences whose status in the global dataset was more commonly "internal" than "head" or "singleton" were discarded.

Only sequences with a similarity of > 80% to a known taxon were selected for subsequent analyses. Potential contaminants, identified by having their highest occurrence in negative control samples, were removed. We then filtered PCR replicates to discard i) highly contaminated PCR replicates and ii) inconsistent PCR replicates within a sample. For this step, we used a graph-partitioning approach (*igraph* package in R, Csardi & Nepusz 2006) based on thresholded Bray-Curtis distances. First, PCR

replicates clustering with control samples (*i.e.*, highly contaminated) and PCR replicates showing no similarity to any other PCR replicate (*i.e.*, resulting primarily from biases during the amplification and sequencing processes) were removed. Additionally, because we expected the 8 PCR replicates from a same sample to cluster together, replicates that did not cluster with at least 3 others replicates from the same sample were also discarded.

Remaining sequences were considered as molecular operational taxonomic units (MOTUs) and assigned to a relevant taxon by comparing them to three reference databases: an alpine database, a boreal/Arctic database and a global database. The global database was generated by *in silico* PCR from EMBL (release 117) with the *ecoPCR* program (Ficetola *et al.* 2010). The alpine database is a subset of the global database restricted to alpine taxa. The boreal/Arctic database groups most of the Arctic and boreal plant taxa (Willerslev *et al.* 2014). Taxonomic assignment was performed using the *ecotag* program, which assigns sequences based on a reference database, along with the taxonomy. We considered only strictly identical matches with reference sequences. In the case of multiple hits, the sequence was annotated with the taxon corresponding to the last common ancestor of the reference sequences. We first assigned taxonomic information using the alpine reference database. If the assignment was not successful (*i.e.*, no reference sequence was identical to the query), the query sequence was then tested against the boreal/Arctic reference database. Again, if the assignment was not successful, the sequence was then tested against the global database. In the final dataset, we kept only sequences with a taxonomic assignment, and whose occurrence was greater than 1,000 reads over the entire dataset (*i.e.,* corresponding to less than 0.02% of the most frequent sequence).

Attesting to the presence of a DNA sequence in sediments during a given timespan requires caution and stringent filtering. Due to the low quantity of the initial template DNA and to potential sporadic contaminations, the amplification process is relatively stochastic. The detection probability of a taxon during a given timespan may be low. On the other hand, false positives can never be excluded, for instance, because of occasional contamination or sequence misidentification. To overcome these issues, we performed 8 replicates per sample. We considered a sequence as present in a replicate if its count was greater than or equal to 10 reads in that replicate. A sequence was then considered as present within a given timespan if it was detected in at least 3 replicates (Fig. S4). These criteria obviously dismiss many putative taxa present at low abundance, but also allow filtering of potential errors. Finally, a few haplotypes of exotic origin or detected as contaminants in control samples were removed. The taxonomic assignment of these sequences includes *Actinidia*, *Myriophillum*, *Trifolium*, *Capsicum*, *Acer*, *Tilia*, *Malae* and *Pooideae.* With the exception of the exotic *Actinidia* and *Myriophillum,* which is an aquatic taxon, these sequences were found only once in control samples. The conservative filtering procedure used here prevented the detection of these false positives. Thus, the sequences used for reconstructing past plant communities come from endogenous DNA, and are unlikely to originate from contemporaneous contaminations.

We used occupancy models (Miller *et al.* 2011) to evaluate the detection probability and the probability of false detection for each of the remaining MOTUs (Table S1). This analysis provides a measure of the reliability of the pattern of presence/absence. Simulation analysis showed that this approach could successfully estimate these parameters in environmental DNA studies (Ficetola *et al.* 2014). Thus,

we estimated the probability of non-detection for each MOTU when it is actually present, and the probability of obtaining three false positives in a given sample. Occupancy modelling was performed using the *unmarked* package in R (Fiske & Chandler 2011).

Data analysis

To investigate plant community trajectories over the last 6,000 years, we focused on points containing at least 10 taxa identified to the genus or species level. Redundancy analysis (RDA) based on the presence/absence of taxa was performed to identify community changes among periods, and taxa that are specifically associated with them (Mazier *et al.* 2009; Borcard *et al.* 2011). For the most contrasted periods, we selected a set of diagnostic taxa (species or genera) for which barcodes were available in the databases in order to investigate their long-term dynamics.

Using RDA we also assessed the potential roles of temperature, soil type and human activities on the evolution of plant communities over time. Available temperature variation reconstruction from the Lake Anterne, based on chironomid records (Millet *et al.* 2009), covers only the last 2,000 years. Consequently, we used the mid-European decadal isotope-climate record from Von Grafenstein *et al.* (1999), extended to the present, as a measure of long-term variation in temperature. Calcium content of sediments is used as a proxy for carbonated rocks dissolution and thus for past soil acidity (Giguet-Covex *et al.* 2011, Fig. S3). The presence of pastoralism (presence of cow and/or sheep herds), assessed from *seda*DNA (Giguet-Covex *et al.* 2014), was considered as a proxy for human impact, because pastoralism is presumed

to be the primary human activity in the past in this area. The significance of RDA was assessed through 10,000 permutations.

The ecology of taxa identified by DNA metabarcoding was derived from a multivariate analysis of 3,077 mountain vegetation relevés (*i.e.,* an exhaustive list of species with visual estimates of local species' abundance) from the northern French Alps. Relevés were extracted from a large floristic database maintained by the Conservatoire Botanique National Alpin (CBNA). Only those corresponding to mountain forests, grasslands and meadows above 1,500 m a.s.l. were retained. A non-metric multidimensional scaling (NMDS) based on Bray-Curtis distance was to summarise the current pattern of species' co-occurrence (Borcard *et al.* 2011). NMDS was limited to 982 alpine plant species, detected in at least 5 relevés. We then used a vector fitting approach to relate the ordination scores of 223 species with their habitat preferences (light, temperature, nitrogen, pH and moisture), as given by Ellenberg indicator values (Ellenberg *et al.* 1992). The vector fitting was tested for significance using 1,000 permutations. Habitat preferences of species identified by metabarcoding were estimated by the projection of their NMDS scores along habitat vectors. For taxa only identified at the genus level, we used the mean score of all species belonging to that genus (Fig. S5). Statistical analyses were performed using the *vegan* package in R (Oksanen *et al.* 2013).

Results

Plant DNA identification

High-throughput sequencing provided 28,019,267 usable merged reads out of 38,030,118 clusters, corresponding to 219,266 unique sequences. After the filtering procedure, we obtained 17,759,158 reads and 144 unique sequences. Eighty of them

(corresponding to 17,264,015 reads) met, at least once, the threshold of three replicates per core slice (Appendix S1, Fig. S4). The taxonomic assignment of these sequences allowed us to identify 30 different plant families, 38 genera and 9 species. For most of the identified taxa, detection probability was high (generally > 0.3), while the probability of obtaining 3 false positives was always < 5%, allowing reliable inference about presence/absence patterns (Table S1).

Plant community changes

Among the 47 slices of sediments, 23 contained at least 10 well-identified taxa and were used to investigate vegetation dynamics during the Holocene. Among them, 8 were between ≈6,400 to 4,600 years calibrated before present (cal. BP), 5 were between ≈3,900 to 3,550 cal. BP, 4 were between ≈2,400 and 2,000 cal. BP, 1 was approximately 1,500 cal. BP and 5 between ≈1,000 to ≈300 cal. BP. When two or more sequences were assigned to the same genus (*e.g., Alchemilla*, *Ranunculus*, *Plantago*, etc.), they were considered separately because their barcodes belonged to different species. Therefore, 49 sequences (corresponding to 39 taxa identified at the genus or the species level) were available for ecological analyses. The redundancy analysis of plant DNA chronosequences (time-series) showed that these groups of dates were well clustered (Fig. 1a). The first RDA axis discriminated the mid-Holocene climatic optimum (≈ 6,400 to 4,600 cal. BP) from the recent period, between the Iron Age and the Modern times (≈ 2,400 to 300 cal. BP), when two important phases of pastoral activities and erosion are identified (Fig. 2).

The predictors used in RDA (Fig. 1a & b) explained 11.2% of the variation (permutation P = 0.002). The pure effects of soil acidity (carbonate content of eroded soils) and pastoralism (proven presence of herds) accounted for 26% and 14% of explained variation respectively, whereas the pure effect of temperature did not significantly explained the observed variation. Furthermore, the combined effects of soil acidity and pastoralism accounted for 42% of explained variation.

Plant taxa chronosequences (Fig. 1b & Fig. 2) allowed the identification of diagnostic taxa for the two main periods. *Achillea macrophylla* and taxa from the genera *Pinus*, *Veratrum*, *Hypericum*, *Rumex*, *Pedicularis* and *Ranunculus* were characteristic of the mid-Holocene optimum. Agro-pastoral periods, the Iron Age-Roman period (≈ 2,400-1,800 cal. BP) and Middle Ages-Modern Times (≈ 1,200-150 cal. BP) were characterised by the presence of *Helianthemum nummularium*, *Saxifraga stellaris* and taxa from the genera *Plantago*, *Alnus*, *Alchemilla* and *Potentilla* (Fig. 1b & Fig. 2). The ecological characteristics of these periods were inferred by comparison of the diagnostic taxa with the current species assemblages (Fig. 3 & S5). The Holocene Climatic Optimum corresponded to vegetation currently observed in relatively warmer, shaded and nutrient-rich environments, in comparison to vegetation over for the last 2,500 years (Fig. 3). No trends were found for pH and moisture. Other taxa detected during the mid-Holocene included several Ericaceae: the *Vaccinium* genus (including *V. uliginosum*) and the Rhodoreae tribe (Fig. S4), typical of developed acidic soils (podzols), as are *Pinus* trees.

For the recent period, most identified taxa were low stature plants from grazed grasslands and meadows, some on acidic soils and others on basic soils. Some species of these genera, such as *Plantago alpina* and *P. lanceolata*, are well-known indicators of pastoralism (Wick *et al.* 2003). *Alnus* sp. was detected at the beginning of the two agro-pastoral periods but disappeared approximately 2,000 and 1,000 cal. BP (Fig. 2), respectively, whereas *Vaccinium* sp. was punctually found during periods of pastoral activity.

Discussion

Plant community evolution

Sedimentary DNA is presumed to come primarily from the surrounding environment and to have a local origin (Jørgensen *et al.* 2012; Parducci *et al.* 2013). In this study, it provided fine-scale information about past plant biodiversity and community dynamics. Although limitations on taxonomic resolution exist, it was possible to identify diagnostic taxa showing, in particular, an abrupt distribution change from the pre-human to the agro-pastoral period (Fig. 1 & 2). The integration between metabarcoding data and current information on species' distributions allowed the acquisition of reliable ecological information from certain taxa at a level other than species.

Our results demonstrated an important phase of vegetational change between the mid-Holocene and the last 2,500 years (Fig. 1 & Fig. 2). These results are congruent with palynological data obtained from two neighbouring peatbogs (David 2010a; David 2010b). Comparison of diagnostic taxa with their current distributions shows good congruence between past species' distributions and present ones: diagnostic taxa from

these two periods currently display strong ecological differences (Fig. 3). Herbaceous taxa from the mid-Holocene include large-leaved species (*e.g., Veratrum sp.* and *Achillea macrophylla*) that are characteristic of tall-herb communities usually found in mountain forest understories or fringes. We hypothesised that subalpine pine forests and tall-herb communities were the dominant vegetation in the lake catchment before human impacts. This period (≈ 6,400 to 4,600 cal. BP) corresponds to the hypsithermal, characterised by forest expansion all across Europe (*e.g.,* Nicolussi *et al.* 2005; Prentice *et al.* 2011; Marquer *et al.* 2014), which is congruent with our results. The subsequent opening of the landscape in approximately 4,600 cal. BP was also observed in the Survilly peatbog (2235 m), along with the regression of *Pinus* trees (David 2010a).

The last 2,500 years are marked by two successive waves of pastoralism followed by a decrease in human activity. These waves were characterised by the presence of heliophilous taxa from open environments, typically subalpine meadows: *Helianthemum nummularium*, *Potentilla* sp. and *Alchemilla* sp. Furthermore, the genus *Plantago* contains several taxa recognised as pastoralism indicators. A similar plant cover was reconstructed in surrounding peatbogs during this period (David 2010a; David 2010b). DNA analyses have proven the presence of cattle and/or sheep herds in the Lake Anterne catchment during the late Iron Age and the Roman period, as well as during the Middle Ages and the Modern times (Fig. 1, Giguet-Covex *et al.* 2014). *Hedysarum hedysaroides*, and *Potentilla*, some of these species being associated with pastoral activities, were found specifically during these two periods. These taxa disappear between the two phases of pastoral activity (≈1,800 to 1,200 cal. BP), and are absent from the two most recent points, which is congruent with the absence of mammalian DNA (Fig. 2). This absence suggests the agricultural abandonment of the

catchment at these times. *Alnus* was also detected at the beginning of pastoral activities but disappeared afterward, likely related to grazing intensification and shrub clearance associated with pasturing (Wick & Tinner 1997). A similar decrease in *Alnus* was also observed in the Ecuelles peatbog (1855 m) approximately 1,960 cal. BP (David 2010b). Our results suggest a landscape mosaic corresponding to open environments, such as subalpine grasslands and alpine meadows dominated by low stature plants, with remnant pockets of heathlands (*Vaccinium* and *Alnus*). The detection of taxa belonging to different communities, currently present on different sides of the lake, suggests that sediment DNA offers a good representation of the catchment as a whole. We hypothesise that this human-modified landscape resulted from a combination of logging, burning and grazing (Gobet *et al.* 2003; Blarquez *et al.* 2010; Rey *et al.* 2013).

Role of anthropogenic and abiotic factors

Previous studies have highlighted the important role of humans in shaping landscapes (Gobet *et al.* 2003; Röpke *et al.* 2011). In this study, pastoralism was significantly correlated with plant community changes during the last 6400 years (Fig. 1a). The importance of pastoralism might be underestimated in our analysis, as data used to infer past farming history only consider periods where mammalian DNA was found with certainty (Giguet-Covex *et al.* 2014). However, other data suggest that the first human impact in the catchment might have occurred 1,000 years before the first attested presence of herds (Covex *et al.* 2011; Giguet-Covex *et al.* 2014): regular detection of *Plantago* sp. since 3,500 cal. BP, presence of *Alnus* sp. at 3,230 and 2,779 cal. BP (Fig. 2) and increases in the frequency of erosive events and the thickness of deposits at approximately 3,400 cal. BP that are maintained thereafter. This increase is concomitant with those in the Rock-Eval pyrolysis Oxygen Index that reflects the

erosion of deep soil horizons (Fig. S3; Giguet-Covex *et al.* 2011), indicating an important and nearly continuous process of soil destabilisation since this period. Reconstruction of past fire activity based on charcoal remains (*e.g.,* Blarquez & Carcaillet 2010; Marlon *et al.* 2012; Rey *et al.* 2013) would provide additional information on human activities associated with landscape clearance around the lake.

The acidity of eroded soils was also strongly related to past plant communities (Fig. 1a) even if no difference was found using the ecology of diagnostic taxa. This correlation highlights the importance of soil properties in plant community changes over long timescales. Even if soil characteristics are reduced to acidity in this study, others parameters (not reconstructed from lake sediments) have probably an effect. During the Mid-Holocene, calcium content in terrigenous sediments reaches its minimum (CaO, Fig. S3). The low calcium content is concordant with the presence of a subalpine dwarf-shrub heath community under acid soils (podzols). These soils have first experienced a complete dissolution of $CaCO_3$, then a leaching of bases, leading to acidification. In addition, the strong decrease of *Pinus* (Fig. 2) and *Vaccinium* (Fig. S4) is concomitant with rises in silica content (Fig. S3), which could reflect major erosion of previously formed podzols or the formation and erosion of another type of acid soil called stagnosols (hydromorphic soils). The degradation of acid soils lasts until 3,200 cal. BP. During the primary periods of livestock farming, a progressive soil degradation with erosion of soil surface horizons (indicated by an increase in the Hydrogen Index) followed by those of deeper horizons (indicated by an increase in the Oxygen Index) is recorded (Fig. S3; Giguet-Covex *et al.* 2011). After these periods, the increase in calcium content suggests the appearance of new raw soils (carbonated bedrock), probably linked to the previous degradation of developed soils (Fig. S3). These results suggest an

indirect impact of human activities on plant communities by modifying soils characteristics.

Climate was a key element of biome structuring and played important role during deglaciation and the first vegetation processes (Von Grafenstein *et al.* 1999). In this study, we focused on temperature, a climatic parameter for which reliable proxies exist over long time periods. Although past altitudinal variations of some species were correlated with temperature (*e.g.,* Holtmeier & Broll 2005; Colombaroli *et al.* 2010) and the climatic dependence of human activities was observed in other alpine areas (Tinner *et al.* 2003; Röpke *et al.* 2011), we found no significant relationship between mean annual temperature and changes in plant communities during the last 6,000 years (Fig. 1). The paleoclimatic reconstruction used here was inferred from an alpine area, and has an appropriate resolution but it refers to a low-altitude lake and temperature changes are known to have greater amplitude in mountainous areas (Beniston 2005). This altitude difference might partially explain why we did not find any correlation between mean temperatures *per se* and vegetation changes over the last 6,000 years (Fig. 1). The role of other parameters, such as precipitation and humidity, are more difficult to infer, but may also be relevant. Nevertheless, this highlights the importance of anthropogenic factors for determining local plant assemblages.

These results illustrate the capacity of human societies to profoundly and durably modify mountain vegetation and landscapes. In Europe, the first large-scale human impacts on the environment occurred during the Neolithic with sedentarisation, development of agriculture and the associated forest clearance (Behre 1988). The onset of strong human impacts on plant cover and ecosystems at high altitude has been dated

to the Bronze Age (Schmidt *et al.* 2002; Koinig *et al.* 2003; Lotter *et al.* 2006; Walsh *et al.* 2007; Roepke & Krause 2013). Around Lake Anterne, the introduction of livestock to the system induced a dramatic shift in the plant community. Pastoralism was a severe perturbation that shifted the vegetation towards an alternative state (Miehe *et al.* 2014). Some species that disappeared were not subsequently recorded, even during periods with supposed absence of livestock (*e.g., Eryngium* and *Pinus* genera, *Achillea macrophylla, etc.*). The effects of human activities persist over long timescales because disturbance also affects physical and chemical parameters of the environment (Scheffer *et al.* 2001), preventing a possible return to the forest mountain landscape during climatically favourable periods. Ecological and historical trajectories are closely related, and the legacy of past disturbances, particularly on soils, could play an important role in ecological trajectories. The relative importance of disturbance history and soils characteristics on plant community trajectories also highlights the limits of species distribution models based only on species climate envelope ranges (Pearman *et al.* 2008). Our approach facilitates an understanding of the coupled dynamics of soils, human activities, climate and vegetation.

Perspectives

*Seda*DNA-based paleoecological reconstructions are still in their infancy (Boessenkool *et al.* 2014). Here, we showed the relevance of these data, combined with other proxies and present plant assemblages, to the reconstruction of past plant cover history and the potential drivers of changes. Multiproxy approaches are essential for the understanding of the long-term evolution of ecosystems (Birks & Birks 2000; Röpke *et al.* 2011). Thus we argue that integration of *seda*DNA in paleoecological studies, along

with pollen and macroremains, will certainly improve the understanding of past human-environment interactions.

The *seda*DNA-based approach currently has some limitations but offers, in addition to a relatively fast and easy protocol, some interesting perspectives. In this study, only sequences with a perfect assignment score (100%) were considered; consequently, data were highly dependent on the completeness of the reference database. In addition, the taxonomic resolution of the *trn*L P6 loop does not always allow identification to the species level, with barcodes often shared by several species. However, this fragment represents only a very limited part of the chloroplast DNA, less than one one-thousandth of the whole molecule. The fact that such a small part of the chloroplast DNA can be reliably amplified demonstrates that a very large number of chloroplast DNA fragments are present in the DNA extract, and thus in the core slice. The next challenge will be to more efficiently extract the taxonomic information from those DNA fragments contained in the core. One possibility might be to directly sequence the DNA extracts using a shotgun approach (Taberlet *et al.* 2012c; Zhou *et al.* 2013). An ongoing project exists to sequence the whole chloroplast genome for approximately 4,500 species of alpine flora. Shotgun sequencing combined with the availability of an exhaustive reference database containing the entire chloroplast DNA molecules creates the potential for a much more precise reconstruction of past plant communities, with identifications to the species level, and opens unprecedented avenues in paleoecology.

References

Adrian R, O'Reilly CM, Zagarese H *et al.* (2009) Lakes as sentinels of climate change. *Limnology and Oceanogr*aphy, **54**, 2283-2297.

Anderson-Carpenter LL, McLachlan JS, Jackson ST *et al.* (2011) Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biol*ogy, **11**, 30.

Behre KE (1988) The rôle of man in European vegetation history. In Huntley & Webb T (Eds.), *Vegetation history*, 633–672. Dordrecht: Springer Netherlands.

Beniston M (2005) Mountain climates and climatic change: An overview of processes focusing on the European Alps. *Pure Applied and Geophysics*, **162**, 1587-1606.

Binladen J, Gilbert MTP, Bollback JP *et al.* (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 HJB (1970) Inwashed pollen spectra at Loch Fada, Isle of Skye. *New Phytologist*, **69**, 807-820.

Birks HH, Birks HJB (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31–35.

Blarquez O, Carcaillet C. Mourier B, Bremond L, Radakovitch O (2010) Trees in the subalpine belt since 11 700 cal. BP: origin, expansion and alteration of the modern forest. *Holocene*, **20**, 139-146.

Blarquez O, Carcaillet C (2010) Fire, fuel composition and resilience threshold in subalpine ecosystem. *PloS One*, **5**, e12480.

Boessenkool S, McGlynn G, Epp LS *et al.* (2014) Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conservation Biology*, **28**, 446–455.

Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*. Springer, New York.

Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environ*ment, **8**, 233-238.

Colombaroli D, Henne PD, Kaltenrieder P, Gobet E, Tinner W (2010) Species responses to fire, climate and human impact at tree line in the Alps as evidenced by palaeo-environmental records and a dynamic simulation model. *Journal of Ecology*, **98**, 1346–1357.

Coolen MJL, Gibson JAE (2009) Ancient DNA in lake sediment records. *Pages News*, **17**,

104-106.

Coolen MJL, Overmann J (2007) 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environmental microbiology*, **9**, 238–249.

Crutzen PJ (2002) Geology of mankind. *Nature*, **415**, 23-23.

Csárdi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 5.

Cushing EJ (1964) Evidence for differential pollen preservation in late quaternary sediments in Minnesota. *Review of Palaeobotany and Palynology*, **4**, 87-101.

David F (2010a) An example of the consequences of human activities on the evolution of subalpine landscapes. *Comptes Rendus Palevol*, **9**, 229–235.

David F (2010b) Expansion of green alder (*Alnus alnobetula* [Ehrh] K. Koch) in the northern French Alps: a palaeoecological point of view. *Comptes Rendus Biologies*, **333**, 424–428.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA *metabarcoding* and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, In Press.

Domaizon I, Savichtcheva O, Debroas D *et al.* (2013). DNA from lake sediments reveals the long-term dynamics and diversity of Synechococcus assemblages. *Biogeosciences Discussions*, **10**, 2515–2564.

Doyen E, Vannière B, Berger JF *et al.* (2013) Land-use changes and environmental dynamics in the upper Rhone valley since Neolithic times inferred from sediments in Lac Moras. *The Holocene*, **24**, 1320-1335.

Ellenberg H, Weber HE, Düll R *et al.* (1992) Zeigerwerte von Pflanzen in Mitteleuropa. Scripta Geobotanica, **18.**

Ellis EC (2011) Anthropogenic transformation of the terrestrial biosphere. *Philosophical Transactions of the Royal Society of London Series A*, **369**, 1010-1035.

Epp LS, Boessenkool S, Bellemain EP *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular ecology*, **21**, 1821–1833.

Ficetola GF, Coissac E, Zundel S *et al.* (2010) An in silico approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

Ficetola GF, Pansu J, Bonin A *et al.* (2014) Replication levels, false presences, and the estimation of presence / absence from eDNA metabarcoding data. *Molecular Ecology Resources.* http://dx.doi.org/10.1111/ 1755-0998.12338.

Fierer N, Ladau J, Clemente JC *et al.* (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science*, **342**, 621–624.

Fiske I, Chandler R (2011) unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1-23.

Gaillard MJ, Sugita S, Bunting MJ *et al.* (2008) The use of modelling and simulation

approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History and Archaeobotany*, **17**, 419–443.

Gauthier E, Bichet V, Massat C *et al.* (2010) Pollen and non-pollen palynomorph evidence of Medieval farming activities in southwestern Greenland. *Vegetation History and Archaeobotany*, **19**, 427-438.

Giguet-Covex C, Arnaud F, Enters D *et al.* (2012) Frequency and intensity of high-altitude floods over the last 3.5 ka in northwestern French Alps (Lake Anterne). *Quaternary Research*, **77**, 12-22.

Giguet-Covex C, Arnaud F, Poulenard J *et al.* (2011) Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (Lake Anterne, 2063 m a.s.l., NW French Alps): The role of climate and human activities. *Holocene*, **21**, 651-665.

Giguet-Covex C, Pansu J, Arnaud F *et al.* (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications,* **5**, 3211.

Gobet E, Tinner W, Hochuli PA, van Leeuwen JFN, Ammann B (2003) Middle to Late Holocene vegetation history of the Upper Engadine (Swiss Alps): the role of man and fire. *Vegetation History and Archaeobotany*,**12**, 143-163.

Heiri O, Wick L, van Leeuwen JFN, van der Knaap WO, Lotter AF (2003) Holocene tree immigration and the chironomid fauna of a small Swiss subalpine lake

(Hinterburgsee, 1515 m asl). *Paleogeography, Paleoclimatology, Paleoecology*, **189**, 35–53.

Hofreiter M, Collins M, Stewart JR (2012) Ancient biomolecules in Quaternary palaeoecology. *Quaternary Science Review*, **33**, 1-13.

Holtmeier FK, Broll G (2005) Sensitivity and response of northern hemisphere altitudinal and polar treelines to environmental change at landscape and local scales. *Global Ecology and Biogeography*, **14**, 395–410.

Jørgensen T, Haile J, Moller P *et al.* (2012) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, **21**, 1989-2003.

Koinig KA, Shotyk W, Lotter AF, Ohlendorf C, Sturm M (2003) 9000 years of geochemical evolution of lithogenic major and trace elements in the sediment of an alpine lake-the role of climate, vegetation, and land-use history. *Journal of Paleolimnology*, **30**, 307-320.

Legendre P, Legendre L (2012) *Numerical Ecology*, 3rd Edition edn. Elsevier, Amsterdam.

Lotter AF (1998) The recent eutrophication of Baldeggersee (Switzerland) as assessed by fossil diatom assemblages. Holocene, **84**, 395–405.

Lotter AF, Heiri O, Hofmann W *et al.* (2006) Holocene timber-line dynamics at Bachalpsee, a lake at 2265 m asl in the northern Swiss Alps. *Vegetation history and archaeobotany*, **15**, 295-307.

Marlon JR, Bartlein PJ, Gavin DG *et al.* (2012) Long-term perspective on wildfires in the western USA. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E535–E543.

Marquer L, Gaillard M-J, Sugita S *et al.* (2014) Holocene changes in vegetation composition in northern Europe: why quantitative pollen-based vegetation reconstructions matter. *Quaternary Science Review*, **90**, 199-216.

Mazier F, Galop D, Gaillard MJ *et al.* (2009) Multidisciplinary approach to reconstructing local pastoral activities: an example from the Pyrenean Mountains (Pays Basque). *The Holocene*, **19**, 171–188.

Miehe G, Miehe S, Böhner J *et al.* (2014) How old is the human footprint in the world's largest alpine ecosystem? A review of multiproxy records from the Tibetan Plateau from the ecologists' viewpoint. *Quaternary Science Review*, **86**, 90-209.

Miller DA, Nichols JD, McClintock BT *et al.* (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422-1428.

Millet L, Arnaud F, Heiri O *et al.* (2009) Late-Holocene summer temperature reconstruction from chironomid assemblages of Lake Anterne, northern French Alps. *The Holocene*, **19**, 317–328.

Nicolussi K, Kaufmann M, Patzelt G, van der Plicht J, Thurner A (2005) Holocene tree-line variability in the Kauner Valley, Central Eastern Alps, indicated by dendrochronological analysis of living trees and subfossil logs. *Vegetation History and Archaeobotany*, **14**, 221-234.

Oksanen J, Blanchet FJ, Kindt R *et al.* (2013) vegan: Community Ecology Package. *R package version 2.0-7*. http://CRAN.R-project.org/package=vegan

Parducci L, Matetovici I, Fontana SL *et al.* (2013) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology*, **22**, 3511–3524.

Pearman PB, Randin CF, Broennimann O *et al.* (2008) Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**, 357-369.

Pedersen MW, Ginolhac A, Orlando L *et al.* (2013) A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Review*, **75**, 161-168.

Pietramellara G, Ascher J, Borgogni F *et al.* (2009) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, **45**, 219-235.

Prentice IC, Jolly D (2001) Mid-Holocene and glacial-maximum vegetation geography of the Northern continents and Africa. *Journal of Biogeography*, **27**, 507-519.

Ravasi DF, Peduzzi S, Guidi V *et al.* (2012) Development of a real-time PCR method for the detection of fossil 16S rDNA fragments of phototrophic sulfur bacteria in the sediments of Lake Cadagno. *Geobiology*, **10**, 196-204.

Rey F, Schworer C, Gobet E *et al.* (2013). Climatic and human impacts on mountain vegetation at Lauenensee (Bernese Alps, Switzerland) during the last 14,000 years. *The Holocene*, **23**, 1415–1427.

Roepke A, Krause R (2013) High montane–subalpine soils in the Montafon Valley (Austria, Northern Alps) and their link to land-use, fire and settlement history. *Quaternary International*, **308**, 178-189.

Röpke A, Stobbe A, Oeggl K, Kalis AJ, Tinner W (2011) Late-Holocene land-use history and environmental changes at the high altitudes of St Antonien (Switzerland, Northern Alps): Combined evidence from pollen, soil and tree-ring analyses. *The Holocene*, **21**, 485–498.

Ruddiman WF (2012) The Anthropocene. *Annual Review of Earth and Planetary Sciences*, **41**, 45-68.

Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2011) Catastrophic shifts in ecosystems. *Nature*, **413**, 591-596.

Schmidt R, Koinig KA, Thompson R, Kamenik C (2002) A multi proxy core study of the last 7000 years of climate and alpine land-use impacts on an Austrian mountain lake (Unter Landschitzsee, Niedere Tauern). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **187**, 101-120.

Seddon AWR, Mackay AW, Baker AG *et al.* (2014) Looking forward through the past: identification of 50 priority research questions in palaeoecology. *Journal of Ecology*, **102**, 256-267.

Seppä H, Bennett KD (2003) Quaternary pollen analysis: recent progress in palaeoecology and palaeoclimatology. *Progress in Physical Geography*, **27**, 548–579.

Sønstebø JH, Gielly L, Brysting AK *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, **10**, 1009–1018.

Steffen W, Grinevald J, Crutzen P, McNeill J (2011) The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society of London Series A*, **369**, 842–867.

Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R (2012) Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular Ecology*, **21**, 1918-1930.

Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, 1-8.

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

Taberlet P, Prud'homme SM, Campione E *et al.* (2012b) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology*, **21**, 1816-1820.

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

Tinner W, Lotter AF, Ammann B *et al.* (2003) Climatic change and contemporaneous land-use phases north and south of the Alps 2300 BC to 800 AD. *Quaternary Science Reviews*, **22**, 1447-1460.

Valentini A, Miquel C, Nawaz, MA *et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trn*L approach. *Molecular Ecology Resources*, **9**, 51-60.

Von Grafenstein U, Erlenkeuser H, Brauer A, Jouzel J, Johnsen SJ (1999) A mid-European decadal isotope-climate record from 15,500 to 5000 years BP. *Science*, **284**, 1654-1657.

Walsh K, Mocci F, Palet-Martinez J (2007) Nine thousand years of human/landscape dynamics in a high altitude zone in the southern French Alps (Parc National des Ecrins, Hautes-Alpes). *Preistoria Alpina*, **42**, 9-22.

Wick L, Tinner W (1997) Vegetation changes and timberline fluctuations in the central alps as indicators of Holocene climatic oscillations. *Arctic and Alpine Research*, **29**, 445-458.

Wick L, van Leeuwen JFN, van der Knaap WO, Lotter AF (2003) Holocene vegetation development in the catchment of Sägistalsee (1935 m asl), a small lake in the Swiss Alps. *Journal of Paleolimnol*ogy, **30**, 261-272.

Wilkinson BH (2005) Humans as geologic agents: A deep-time perspective. *Geology*, **33**, 161-164.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791-795.

Willerslev E, Davison J, Moora M *et al.* (2014) Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, **506**, 47-51.

Wilmshurst JM, McGlone MS (2005) Corroded pollen and spores as indicators of changing lake sediment sources and catchment disturbance. *Journal of Paleolimnology*, **34**, 503-517.

Yoccoz NG, Brathen KA, Gielly L*, et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, **21**, 3647-3655.

Zhou X, Li Y, Liu S *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, **2**, 4.

Data accessibility

- Unfiltered plant sequencing data were deposited, under fasta and table formats, in the Dryad Digital Repository doi:10.5061/dryad.6171j.

- Filtered plant sequencing data, along with the age of samples, are both provided in the Appendix S1 and in the Dryad Digital Repository doi:10.5061/dryad.6171j.

- Global and alpine databases are also available in the Dryad Digital Repository doi:10.5061/dryad.6171j while the boreal/arctic database is available from the Dryad Digital Repository doi:10.5061/dryad.ph8s5 (Willerslev *et al*. 2014).

- Mammal DNA data are available from the Dryad Digital Repository doi:10.5061/dryad.h11h7 (Giguet-Covex *et al*. 2014).

- All geochemical and sedimentological data, as well as age depth model, from Lake Anterne core are available through Pangaea database doi:10.1594/PANGAEA.842828 (Giguet-Covex *et al*. 2011 ; Giguet-Covex *et al*. 2012).

- Ammersee ostracod oxygen isotope data used to infer for the paleotemperature variations are available from the NOAA Paleolimnology Data Archive:

ftp://ftp.ncdc.noaa.gov/pub/data/paleo/europe/germany/ammersee_1999.txt   (Von Grafenstein *et al.* 1999).

- Scripts are freely available upon request from J. Pansu.

Author contributions

P.C., F.A., J. Poulenard, P.T., C.G.-C. and J. Pansu designed the study. J. Pansu, C.G.-C. and L.G. performed DNA analyses. J. Pansu, L.Z. and F.B. performed the sequence analyses and taxonomic assignment. J. Pansu, G.F.F., P.C. and C.G.-C. analysed the data. F.A., J. Poulenard and P.C. contributed their expertise on the reconstruction of past landscapes and soil changes. J. Pansu and C.G.-C. wrote the manuscript with G.F.F., and all authors contributed substantially to revisions. F.A. coordinated the PYGMALION research program. P.T coordinated the MetaBar research program. P.T., F.A., P.C. and J. Poulenard coordinated the Retromont program.

Conflict of interests

L.G. and P.T. are co-inventors of patents related to the gh primers and the use of the P6 loop of the chloroplast *trn*L (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.

Figure captions

Figure 1: Redundancy analysis (RDA) of past plant assemblages and diagnostic taxa

a. Ordination of plant DNA assemblages. Gray arrows and black arrows represent non-significant and significant explanatory variables, respectively. Pastoralism corresponds to the attested presence or absence of mammals in the Lake Anterne catchment during the corresponding period according to Giguet-Covex *et al.* (2014). CaO corresponds to carbonate content of sediments, according to Giguet-Covex *et al.* (2011), and is used as a proxy for soil acidity. Temperature corresponds to oxygen-isotope ratios of precipitation (d18OP) in Amersee Lake (Southern Germany) from Von Grafenstein *et al.* (1999), and constitutes a proxy for past temperature during the corresponding period in Central Europe. The model explains 11.2 % of the variance.

b. Ordination of identified taxa. Taxa in black on the left and on the right are, respectively, considered as "diagnostic" of the Mid-Holocene and of the recent period (the last 2,500 years).

Figure 2: Plant taxa evolution during the Holocene

Full circle: sequence found in at least in 3 replicates. Open circle: sequence found in less than 3 replicates. The diameter of the circle is proportional to the mean percentage of the taxa among the dataset for a given sample.

Mammalian data are from Giguet-Covex *et al.* (2014). Open diamonds: sediment cores sampled in the metabarcoding analysis. Colored diamonds: DNA sequence present in only one replicate. Bar: DNA sequence present in at least in two replicates. The heights of the bars represent the ranges of sequence numbers (*i.e.,* between 50 and 100; between 100 and 500; between 500 and 1,000; between 1,000 and 5,000; and more than 5,000).

Frequency of erosive events is the frequency of flooding (flood deposit > 5mm) in 100 years according to Giguet-Covex *et al.* (2014).

Figure 3: Ecological characteristics of the mid-Holocene and of the recent period (the last 2,500 years).

Boxplot reflecting ecological characteristics of the mid-Holocene and of the recent period (the last 2,500 years), inferred from current distribution of diagnostic taxa: a) Light; b) Temperature; c) Nitrogen; d) pH; e) moisture.

f: NMDS ordination of 982 alpine plants according to their current distribution in Northern French Alps and their relation to abiotic parameters (stress value=0.1305256). All fittings are significant (P<0.001).

Supporting information

The following supplementary material is available for this article:

Appendix S1: Data filtering statistics, filtered sequences and age of samples

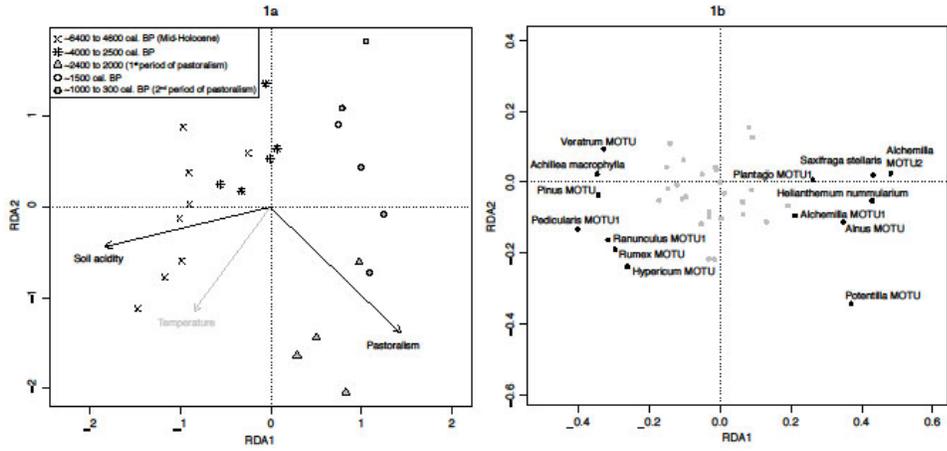**Figure S1: Land cover of the Lake Anterne catchment**

Figure S2: Oxygen concentration and turbidity in the Lake Anterne bottom
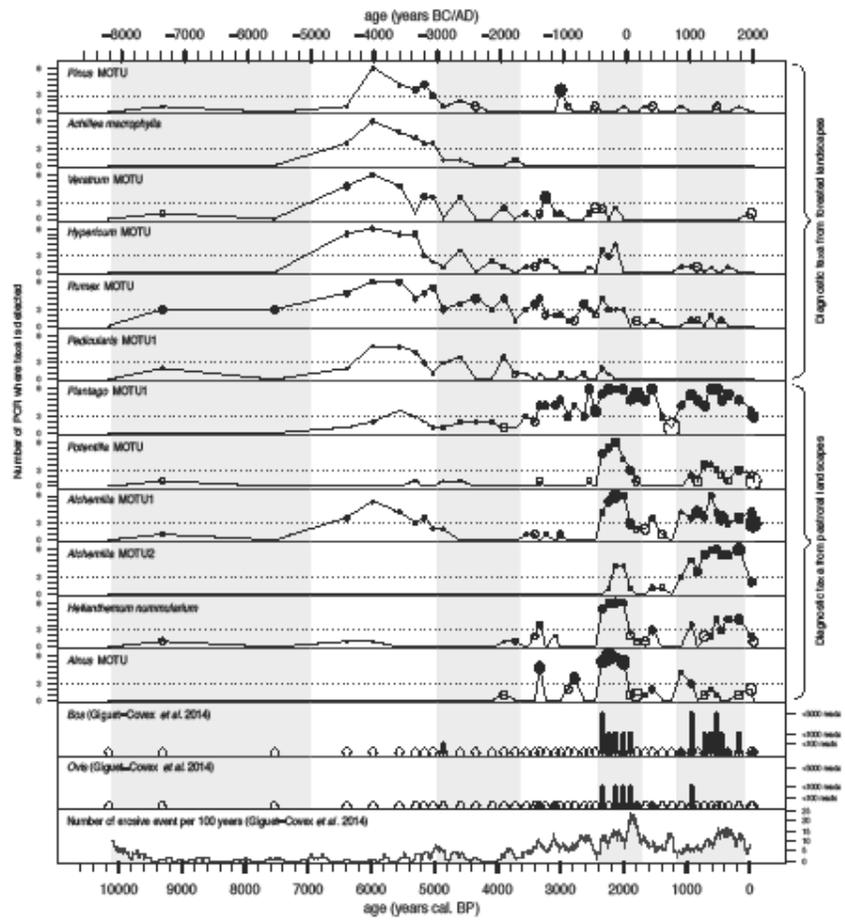
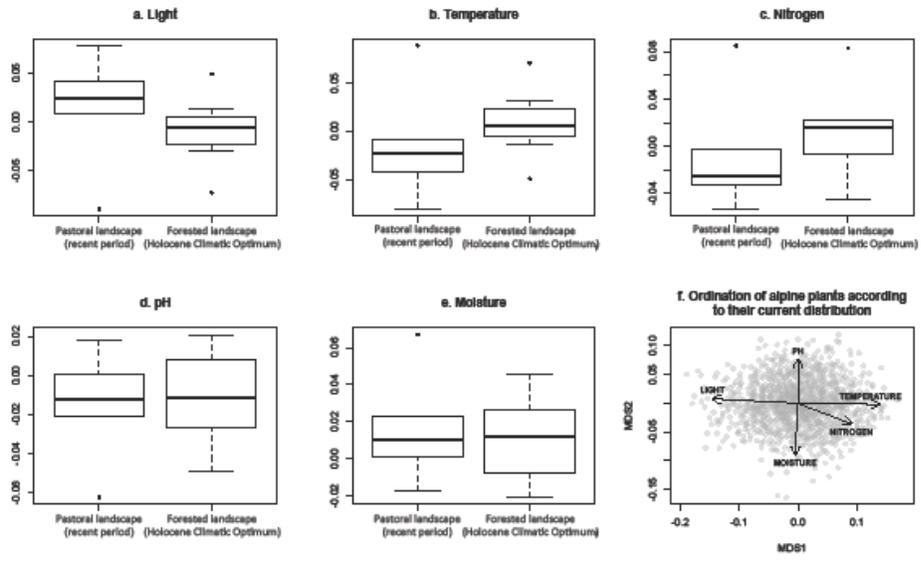**Figure S3: Geochemical analysis from sediment core**

Figure S4: DNA diagrams of filtered sequences during the Holocene

Figure S5: Ordination of taxa detected by metabarcoding according to their current co-occurrence pattern and ecological characteristics of diagnostic taxa

Table S1: Detection probability and false presence probability of filtered sequences

1a

- ✕ ~6400 to 4800 cal. BP (Mid-Holocene)
- ✳ ~4000 to 2500 cal. BP
- △ ~2400 to 2000 (1ˢᵗ period of pastoralism)
- ○ ~1500 cal. BP
- ⊕ ~1000 to 800 cal. BP (2ⁿᵈ period of pastoralism)

Soil acidity

Temperature

Pastoralism

RDA1
RDA2

1b

Veratrum MOTU
Achillea macrophylla
Pinus MOTU
Pedicularis MOTU1
Ranunculus MOTU1
Rumex MOTU
Hypericum MOTU

Plantago MOTU1
Saxifraga stellaris
Alchemilla MOTU2
Helianthemum nummularium
Alchemilla MOTU1
Alnus MOTU
Potentilla MOTU

RDA1
RDA2

a. Light

b. Temperature

c. Nitrogen

d. pH

e. Moisture

f. Ordination of alpine plants according to their current distribution