Testing plant metabarcoding on a temperate lake core from southern Italy covering thirty-one-thousand-years

Kevin Nota[1], Donatella Magri[2], Tim Temizyürek[1], Francesca Benedetti[2,] Francesco Spada[1,2], Laura Parducci[1,2]

[1] Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden, [2] Department of Environmental Biology, Sapienza University of Rome, Rome, Piazzale Aldo Moro 5, 00185, Italy

Abstract

Because DNA is likely better preserved in cold environments, the majority of ancient sedimentary (*sed*aDNA) studies to investigate past flora changes have been conducted at high latitude regions, or at lower latitudes, in cold environments such as the Alps. Temperate lakes in southern Europe are therefore underrepresented in the *sed*aDNA literature, while they are key in understanding where species survived in glacial refugia. Here we investigated a 14-meter-long sediment core taken from maar lake Lago Grande di Monticchio in southern Italy covering the last 31 thousand years and tested *trn*L and ITS1 metabarcoding. Our result shows a detailed flora in the first 7.5 thousand years of the sediment record, with DNA revealing well recorded woody taxa replacements during this period and overall, a rich herbaceous and aquatic flora. However, between 7.5 and 31thousand years the DNA signal becomes stochastic with a high number of sample dropouts, likely due to a combination of enzyme inhibitors and DNA degradation which prevented metabarcoding from working. Overall, we demonstrated that DNA is well preserved from the mid to late Holocene in this southern Italian lake, however, methodological limitations did not allow us to investigate accurately the end of the lateglacial period, and the transition into the Holocene. The large potential of metabarcoding however remains if new methodological advances will help mitigating enzyme inhibition of *sed*aDNA samples and if smaller DNA fragments are targeted.

Keywords: Ancient sedimentary DNA, metabarcoding, *tnr*L, ITS1, vegetation history, lake sediments.

Introduction

Investigating palaeoecological changes during rapid and major climatic events in the past is a key for understanding how climate might impact biological diversity in the future (Jackson and Wiliams, 2004). Vegetation is an important component and shaper of biological communities, which are heavily impacted by climate. During the Last Glacial Maximum (LGM, -21,000 years ago), most of the temperate plant species survived in glacial refugia in central and southern Europe (Bennett *et al.,* 1991; Hewitt, 1999), which acted as local biodiversity hotspots (biogeographic regions with significant levels of biodiversity maintained through time). Such regions are important locations for studying current and past changes in plant distributions in relation to climate change. One way of studying past environments is by looking at annually deposited substrates such as lake sediments, which incorporate material from local and regional flora and fauna (Rull *et al.,* 2021). Many studies have been conducted in the past century using traditional plant proxies such as pollen and macrofossils assemblies, and much is known about vegetation development and glacial refugia because of them (Gavin *et al.,* 2014).

Recently a new and emerging tool – ancient sedimentary DNA (*sed*aDNA), has received increased attention to investigate past ecosystems (Capo *et al.,* 2021). Analysis of *sed*aDNA is shown to provide a more local signal than pollen. A study by Alsos *et al.,* (2018) revealed that most plant taxa recorded in 11 lake sediments from Norway were growing within two meters from the lakeshore. Several taxa, especially aquatics, were not recorded in classical vegetation survey around the same lake. Moreover, pollen seems to contribute little to the recovery of *sed*aDNA (Parducci *et al.,* 2017), making this an excellent proxy complementary to pollen for flora reconstruction and revealing important information on local plant taxa. Efforts and ground-breaking results using *sed*aDNA have mainly been focused on alpine or high latitude environments in Europe because of the good preservation of DNA at low temperatures (Willerslev *et al.,* 2014; Rijal *et al.,* 2021, Wang *et al.,* 2021), while many temperate lakes in southern Europe remain understudied. However, low-latitude lakes, and in particular lakes in the Mediterranean basin, are very interesting for palaeoecological studies because they acted as a glacial refugia for temperate plant taxa during the last glaciation cycles (Magri *et al.,* 2017; Donders *et al.,* 2021).

This study reports the first attempt of using *sed*aDNA metabarcoding to describe local floral changes over a 31 kyr period from the long sedimentary lake record at the maar lake Lago Grande di Monticchio in southern Italy.

Material and methods

Site description

Lago Grande di Monticchio is one of the two maar lakes inside the crater of Monte Vulture formed ca. 132 ka BP in southern Italy (40°55'52.0" N 15°36'17.7" E). The lake is currently located ca 656 m above sea level facing the southwest slope of Monte Vulture (**Fig. 1**). Partly varved sediments were discovered in the lake nearly four decades ago (Watts, 1985), and since then a large body of work has been conducted on these sediments to study vegetation history and diatom assemblages (Watts, 1985; Watts *et al.,* 1996; Huntley *et al.,* 1996; Nimmergut *et al.,* 1999; Allen *et al.,* 2000, 2002; Venanzoni *et al.,* 2003; Allen and Huntley 2009, 2018), dating and chronology (Zolitschka and Negendank, 1996; Narcisi, 1996; Hajdas *et al.,* 1998; Wulf *et al.,* 2004 2008, 2012; Allen and Huntley 2009), geochemistry, climate and palaeoecology (Robinson, 1994; Ramrath *et al.,* 1999; Watts *et al.,* 2000; Brauer *et al.,* 2000, 2007, Schettler and Albéric, 2008, Tomlinson *et al.,* 2014, Wutke *et al.,* 2015; Martin-Puertas *et al.,* 2019), lake formation and other proxies (Creer and Morris, 1996; Stoppa and Principe, 1997). Multiple coring expeditions at the site revealed

that not only sediments from the Lateglacial period (20-31.2 cal. ka BP) were present, but also from the last interglacial (~115 cal. ka BP) (Brauer *et al.*, 2007) and the sediment in the maar lake records nearly ~132 kyr of past environments (Zolitschka and Negendank, 1996; Brauer *et al.*, 2007). Due to its location, Lago Grande di Monticchio has never been glaciated during the last glaciation, making this as one of the few continuous records in Europe covering more than 100 kyr (Fletcher *et al.*, 2010). Lamination-based chronology of high precision and accuracy based on multiple independent dating approaches are available (Wulf *et al.*, 2004) and previous studies revealed that the lake and its catchment responded distinctly to past rapid climate changes.

Present day woody vegetation in the catchment surrounding the lakes consist of a forested slope towards Monte Vulture with arboreal species such as *Ilex aquifolium*, *Salix alba* and *S. alba* x *fragilis* and *Prunus avium* growing close to the lake shore. Slightly further upslope, *Fraxinus ornus*, *Alnus glutinosa*, *Ulmus glabra* are common, together with *Tilia* spp. and *Acer cappadocicum*. Higher upslope, *Corylus*, *Tilia*, and *Castanea* are present. On the flatter parts of the catchment north and northeast of the lake, *Iris* spp., *S. alba*, and *Quercus cerris* are dominant. Further away from the lake, *F. ornus* together with *Alnus cordata* become common, with a few individuals of *Q. cerris*. The slope facing the eastern part of the small lake (Lago Piccolo di Monticchio), consists mainly of *Carpinus betulus* and *Fagus sylvatica*, while the steep slope facing the small lake from the north, shows mainly *Castanea sativa* mixed with *Tilia* spp.

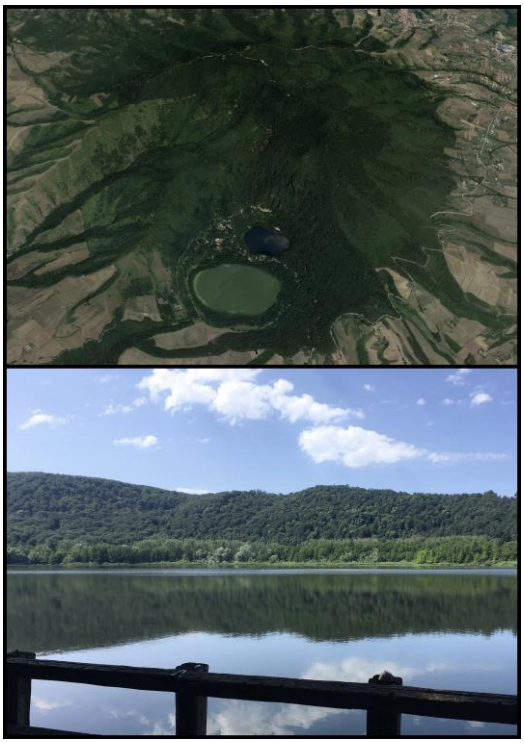

Figure 1. The top figure shows an aerial photo of Lago Grande di Monticchio. The bottom photo shows the lake surface with the local vegetation (Aerial photo from Google earth and photo by Laura Parducci).

Core subsampling and chronology

In May 2016, researchers from the German Research for Geosciences (GFZ) in Potsdam (Germany) collected five new sediment cores at Lago Grande di Monticchio (**Fig. 1**) and MON16B was selected for *sed*aDNA analysis. MON16B was13.62-meter-long and was divided into 14 sections of ca 1 m. In October 2016 we performed sub-sampling on one half of the core in a clean room at GFZ, wearing bodysuits, shoe covers, hairnets, facemasks, and sterile gloves, in a building where no people worked with DNA. We removed the top 2 cm of the sediment with sterile scalpels and used new sterile scalpels to sample the innermost part of the sediment. A total of 163 samples of ca 10 g (wet weight) were taken, immediately stored at -20 °C and transported to Uppsala university, to be further processed in a dedicated ancient DNA laboratory following established procedures for *sed*aDNA analysis (Parducci *et al.*, 2017). The sampling interval differed along the core. The first section that we thought was corresponding to the late glacial – Holocene transition was densely sampled (every 2 cm), most of the Holocene section was sampled every 8 cm, and the full glacial section every 20 cm. Due to economic constraints on sequencing,

only half of the samples were selected and sequenced.

DNA extraction

Approximately 0.25-0.35 g of sediments were extracted using the Dneasy PowerSoil extraction kit (Qiagen, Hilden, Germany) as described in case study 1 in Capo *et al.* (2021) and in Nota *et al.* (2022). In short, 94 of the 163 samples were homogenised by vortexing before DNA extraction. Bead-beating was performed on a "normal" vortex at the highest speed for 10 min by taping the tubes horizontally on the vortex. After bead-beating, 2 µL of 20 mg/mL Proteinase K and 25 µL of 1 M Dithiothreitol (DTT) were added to the tubes and incubated overnight at 37 °C in a rotating incubator. The volume of solution C3 and C4 was increased from 200 µL to 250 µL and 1200 µL to 1400 µL, respectively. Samples were incubated for 10 min at room temperature before being eluted twice in 60 µL elution buffer containing 10 mM Tris-HCL and 0.05% Tween-20. Enzyme inhibition was tested using a spiked qPCR assay as described in case study 1 of Capo *et al.*, (2021). Due to relatively high levels of inhibitors found along the core all DNA extracts were purified using the OneStep™ PCR Inhibitor Removal Kit (Zymo Research), according to the manufacturer's instructions. All extraction batches contained two extraction controls (no sediment) which were treated alongside sediment samples to monitor possible cross-contamination between sediment samples. Additional three extraction controls, not treated alongside sediment samples, were prepared to monitor for possible background contamination from the extraction kit.

TrnL P6-loop metabarcoding

PCR was performed in 96-well plates using 96 unique dual-tagged *g* and *h* primers to target the *trn*L P6-loop locus (Taberlet *et al.*, 2007) (Supporting information). Each of the 106 DNA extracts (94 samples + 12 extraction controls) was amplified eight times (950 PCR) in 25 uL PCR reactions containing 4 uL of DNA, 0.2 uM each primer, 1.75 units of Platinum™ II Taq Hot-Start DNA Polymerase (Invitrogen), 1X Platinum II PCR Buffer, 100 ng bovine serum albumin, and 1 mM $MgCl_2$ (2.5 mM in total). PCR cycling was as follows: initial denaturation 94°C for 2 min, followed by 45 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 15 sec and extension at 68°C for 7 sec. Each 96-well plate contained four PCR control samples with nuclease-free water (48 in total), four empty wells (48 in total), and sample extracts (including control samples for extraction). After PCR, 10 uL of PCR product from 24 samples (three columns in each of the 12 plates) were pooled together obtaining 42 pools in total. Of each pool, 100 uL were purified using the MinElute PCR Purification Kit (Qiagen) and eluted in 30 uL (EB-buffer). Of each cleaned pool, 16 uL were used to perform library preparation using the single-tube library prep described in Carøe *et al.*, (2018) with the following modifications: the T4-DNA polymerase was omitted from the end-repair step, we added a MinElute purification step after the ligation step according to the manufacture's recommendations, no purification step was done before indexing. All adapter ligation pools were quantified using qPCR and each reaction contained: 1 uL of 1:10 diluted adapter-ligated DNA, 0.5 uM forward and reverse primers [IS5 and IS6 primers (Meyer and Kircher, 2010)], 1X TATAA SYBR® GrandMaster Mix. The thermocycling protocol was as follow: 95°C for 30 sec, 40 cycles of 95°C for 5 sec, 55°C for 30 sec, and 72 for 10 sec.

Indexing was performed using the KAPA HiFi HotStart Library Amp Kit (Roche Sequencing Solutions) using 10 uL of adapter-ligated DNA, 0.08 µM of both index primers (15 unique, see supporting information) in a total volume of 25 ul. All 42 pools were indexed for eight cycles using the following thermocycling protocol: 95°C for 45 sec, 8 cycles of 95°C for 15 sec, 60°C for 30°C, and 72°C for 30 sec. Based on qPCR

quantification results, we pooled together samples in equal concentration and created eight indexed libraries using the same protocol as described for the adapter ligation, except using primers I7 and I8 from Meyer and Kircher (2010). Between 33 and 85 uL of the eight indexed pools, were purified using the MinElute purification kit and eluted in 15 uL (Supporting information). The pooled libraries were quantified again with qPCR and pooled together in equal concentrations, purified, and concentrated with AMPure XP beads (About Beckman Coulter Life Sciences) following the manufacturer's instructions (1.8x bead ratio). The final library was quantified using the Qubit™ 1X dsDNA HS Assay Kit on a Qubit 2.0 Fluorometer and sequenced on ~70% of one lane on Illumina NovaSeq 6000 system (v1.5 sequencing chemistry) together with samples from another project.

ITS1 metabarcoding

To increase species resolution for the Asteraceae and Poaceae, we tested two ITS1 primer pairs (Aste01 and Poac01) (Taberlet *et al.*, 2018, modified from Baamrane *et al.*, 2012). Because these primers have so far not been empirically tested on lake sediments, we used only a subset (21) of the samples that were used for the *trn*L metabarcoding. Only the Poaceae primers showed amplification and were used for further analyses. We used 16 unique dual tagged Poales ITS1 primers to amplify each sample eight times, including four extraction controls, two PCR controls and one empty sample. The PCR cycling and library preparations were as described for the *trn*L, except that the PCR annealing temperature was 57°C. The pooled library was sequenced on one Illumina MiSeq lane (v3 chemistry, 300 paired-end).

Bioinformatics analysis and data filtering

Reads were received demultiplexed based on 15 indexes by the SciLifeLab sequencing facility in Uppsala. Reads for each index were merged using the default settings of the paired-end merger *PEAR* (Zhang *et al.*, 2014). *Barcode_spitter* (Leach and Parsons, 2019) was used to subset all reads into fastq files based on the tag in the forward primer allowing for maximal one mismatch. The sample identifiers were annotated to the fatsq header before being merged. The merged files were run through *Barcode_spitter* a second time based on the tag present in the reverse primer. An inhouse python script [https://github.com/Kevinnota/Metabarcoding_scripts] was used to remove primer sequences (maximal two mismatches) and reverse complement reads which started with the reverse primer. All reads with an insert size of 10 or fewer base pairs (bp) were removed. A second inhouse python script [https://github.com/Kevinnota/Metabarcoding_scripts] was used to dereplicate (merging identical reads) and count reads with correct tag in both forward and reverse primers as well as reads with incorrect tag combinations (chimeric sequences). Singletons and reads with 50 bp or less per index were removed. All reads containing the correct tag sequences, were subset, and clustered using *swarm* (*fastidious* mode, Mahé *et al.*, 2014). All mOTUs recovered by *swarm* were merged over the whole dataset, before being assigned to taxonomy using *ecotag* (Boyer *et al.*, 2016). For mapping we used the ArctBorBryo reference database (Sønstebø *et al.*, 2010; Willerslev *et al*, 2014; Soininen *et al.*, 2015) and all EMBL sequences (release 133). The software ecoPCR was used for silicon amplification allowing for max five mismatches for the primers and maximal amplicon length of 300 bp (Ficetola *et al.*, 2010). Taxonomic assignments were checked manually for all reads with at least 90% similarity to one of the reference databases. Reads with less than 90% similarity were removed. All unique reads assigned to the same taxa were summed. These sequences likely represent sequencing errors, that are not clustered because the clustering was performed per sample rather than over the whole dataset. Due to the large number of individual PCRs performed in some

reactions an error might occur in the first few cycles creating a highly abundant sequence that differs slightly between PCRs. Therefore, genera with multiple species with reads that differed by a single nucleotide were all classified at the genus level, such as *Q. ilex*-type, *Q. robur*-type and *Q. cerris* type were all identified as *Quercus*.

Reads which contained unused index-tag combinations were removed (tag jumping, see Supporting information) and this filtering criterion was applied to all the samples, including PCR and extraction controls. The maximum number of reads detected in the PCR controls in each PCR plate was subtracted from all samples detected in the same plate. Finally, only taxa recorded in at least three PCR replicas were retained and all taxa with less than 100 reads were removed. The extraction controls were not removed from the dataset set but used in the analyses and plotted together with samples. All taxa with less than 5% of the total number of reads in the control samples (PCR and extraction controls) were removed and we omitted therefore cultivated taxa such as *Cucumis melo* (melon), *Glycine max* (soybean) and *Capsicum* (paprika). Non-native and cultivated taxa which were not removed after filtering were plotted together with true positives. All results were visualised using the *ggplot2* package in R. Relative abundance data for each of the eight PCR repeats was used to create a Bray-Curtis dissimilarity matrix using *vegdist* from R package *Vegan* (Dixon, 2003). A constrained hierarchical clustering using the *rioja* package in R (Juggins, 2020) and the CONISS method was performed and used to denote zones for a qualitative description of the detected flora.

The bioinformatic pipeline used for the analysis of the ITS1 sequences was identical to *trn*L, except for the taxonomy assignment to the mOTUs detected after the *swarm* clustering, which was done using *blastn* (Camacho *et al.*, 2009). All mOTUS's were blasted to the partially non-redundant nucleotide sequences using the default settings for Blastn, except for *word_size* set to 20, and *staxids* which was added to the tabulate output. No specific database was created in this case because it was not clear which plant taxa would amplify and if non-plant taxa could increase false positives identifications. Last common ancestor was assigned to each sequence using MEGAN6 (Huson *et al.*, 2016) using the default settings, except for the *Min support* which was set to one sequence, and *Min support percent* set to zero. Only sequences mapping to plants with a query cover above 95% were used for downstream analysis. All plant assignments were manually checked. Downstream filtering was done similarly as for *trn*L.

Results

TrnL and ITS1 metabarcoding results

In total, for the *trn*L library pools, we obtained 134,679,451 raw reads with correct unique index pairs. Of these, 42,905,737 were assigned to samples, 49,596,717 consisted of chimeric sequence with unused tag-combinations, and 42,176,997 could not be assigned to samples. The average number of reads per sample obtained after demultiplexing was 39,253±123,483 of which 30,503,454 could be taxonomically assigned. There was a clear and consistent DNA signal in eight replicas in all the samples from 1.4 cal ka BP until 7.5 cal. ka BP. Before 7.5 cal.

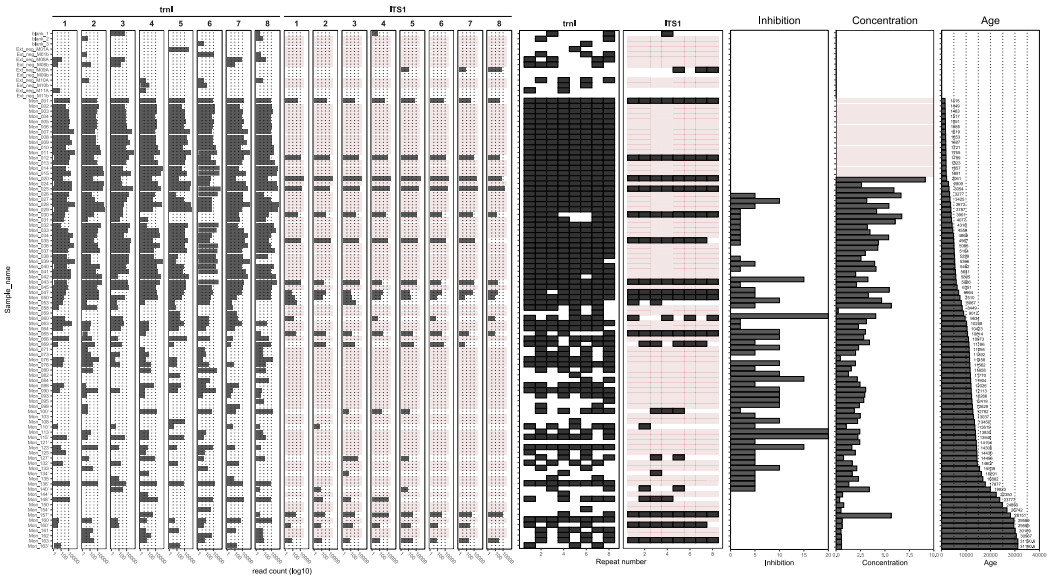

Figure 2. Overview of *trn*L and ITS1 metabarcoding results on Lago Grande di Monticchio. Left: read counts after filtering for the eight PCR repeats. Right: PCR repeats with reads, level of inhibition before Zymo purification, DNA concentration in ng/gram sediment and calibrated ages. Red shaded areas show sample not used for ITS1 metabarcoding. Most extraction controls contain few reads, with maximum four out of eight PCR repeats showing reads. A clear and consistent DNA signal is observed between 1.4 and 7.5 cal. ka BP. The signal becomes stochastic from 7.5 cal. ka to the end of the core at 31 cal. ka. with only few samples having reads in four or more PCR repeats. Enzyme inhibition is high during the period with high sample dropout, intermediate in younger samples and low in the oldest samples.

ka BP the signal stochastically disappears (**Fig. 2**). Between 8 and 11.5 cal. ka BP only six samples show more than half of the PCR replicas with more reads remaining after filtering and only one sample at ~17.9 cal. ka BP showed a consistent signal in all eight PCR repeats. Only a small number of the extraction controls showed amplification in maximal three out of eight replicas with more than 250 reads. All the reads obtained from the PCR controls were removed during the filtering.

A total of 15,223,297 raw reads were obtained from the ITS1 pool using the Poaceae primers (grasses), of which 4,313,627 were correctly merged and assigned to samples (mean 2739±8158 reads per sample). Only 15% of the reads (657,406) were assigned to plants, 2.2% to bacteria (95,699 reads) with query coverage and identity score >95%. The remaining reads (82.5%) did not map to any sequence in the partially non-redundant nucleotide database within >95% identity score. No reads mapping to plants were recovered among index-tag combinations not used during PCR set up as well as among PCR controls. Only one extraction control contained three repeats mapping to *Zea* (corn) after filtering and, similarly to *trn*L**,** we observed several sample dropouts before 7.5 cal. ka BP (**Fig. 2**).

Overall floral composition

Overall, a total of 95 taxa were recovered with *sed*aDNA analysis (*trn*L and ITS) at Lago Grande di Monticchio, with *Alnus, Quercus, Nymphaea/Nuphar, and Ceratophyllum demersum* being the most abundant. Based on the constrained hierarchical clustering of the

metabarcoding results including all recovered taxa, we recovered four general zones (CONISS in **Fig. 3**). Zone 1 (1.4 – 2.1 cal. ka BP) and zone 2 (2.8 – 6.4 cal. ka BP) are based on high quality samples, while zone 3 (6.9 – 13.5 cal. ka BP) and zone 4 (13.6 – 31.1 cal. ka BP) show a large sample dropout.

Trees, shrubs and vines

Alnus, Quercus, Carpinus/Corylus, Fraxinus/Olea, Viburnum and *Vitis vinifera* are the most detected woody taxa and occur in most samples (**Fig. 3a**). Less common taxa with consistent signals are *Acer, Abies, Fagus, Ilex, Sambucus,* and *Hedera helix*. Rare taxa with signals from only few samples include *Taxus, Cornus, Cupressus, Pinus,* and *Tilia*. Zone 1 (1.4 - 2.1 cal. ka BP) show an abundance of *Quercus* and *Alnus*, with *Quercus* representing most of the reads at 1.7-2.1 cal. ka BP and *Alnus* being more abundant at 1.4-1.7 cal. ka BP. Interestingly, at 2.1 cal. ka BP, no other tree taxa are recovered besides *Quercus* and *Alnus.* In the same zone, *Carpinus/Corylus, Salix/Populus*, and *Fagus* are also common but with lower relative abundance, and there is a stochastic presence of *Fraxinus/Olea, Acer, Sambucus, Viburnum, V. vinifera,* and *H. helix.* The first part of zone 2 (2.1-4.3 cal. ka BP) shows a lower relative abundance of *Quercus* reads and a higher relative abundance of *Fraxinus/Olea* and *Viburnum* reads. At the time the signals from *Salix/Populus* and *Fagus* disappear, *Ilex* and *Abies* become common, but are detected at low abundance, and few PCR repeats with *Taxus* appear. The second part of zone 2 (4.3-6.4 cal. ka BP) shows an increase in *Alnus*, and slightly lower relative abundance of *Fraxinus/Olea* and *Viburnum*. Before 7.5 cal. ka BP the signal becomes sporadic, although two samples at ~10 cal. ka BP shows a similar composition as zone 2, except for the absence of *H. helix, Acer, Ilex, Abies,* and *Sambucus*. Between 10.8 and 31 cal. ka BP (zone 4) *Alnus, Quercus,* and *Salix/Populus* are occasionally detected. Surprisingly, we detected the non-native cultivated taxa *Gossypium* (cotton) with high relative abundance and read count in sample Mon_050 and in two consecutive younger samples (Mon_047, and Mon_045) at lower relative abundance. In sample Mon_050, very few other taxa were recovered, besides *Gossypium.*

Finally, the *trn*L sequence could distinguish three types of *Quercus*, though with little difference between sequences. **Fig. 4** shows the relative abundance of the three oak types (without clustering PCR errors). *Quercus cerris* sequence is the most common in all four zones and interestingly the signal disappears between 5.9 and 6.9 cal. ka BP when *Q. ilex* becomes the only species detected. The same happens for the samples between 7.5 and 31 cal. ka. The third type, *Q. robur* is only present at low abundances between 1.4 and 5.4 cal. ka BP.

Herbs

Many herbaceous taxa are recorded with taxa such as *Mentha,* Nepetoideae, *Ranunculus,* Solanaceae and Asterales being very common, but not particularly abundant (relative abundance), in all zones. *Veronica* and *Urtica dioica* are the only herbaceous taxa present exclusively in zone 1, while *Cicuta virosa* and *Clematis* are mainly present in zone 1 but appear stochastically also in zones 2 and 3 (**Fig. 3a, b**). Maleae, *Filipendula,* and Apiaceae are mainly present in zones 2 and 3. Many taxa occur in only a single sample. The family Brassicaceae, *Trifolium,* and *Filipendula ulmaria* are the only taxa that are still present in the extraction controls after filtering with presence in at least three repeats.

Aquatics and grasses

The most common aquatic and grasses detected are *Phragmites australis¸ Nymphaea/Nuphar, C. demersum, Typha, Stuckenia,* and *Cladium* (**Fig 2c**). Taxa such as *Utricularia, Sparganium Potamogetonaceae, Potamogeton, Triticeae,* are detected at low relative abundance.

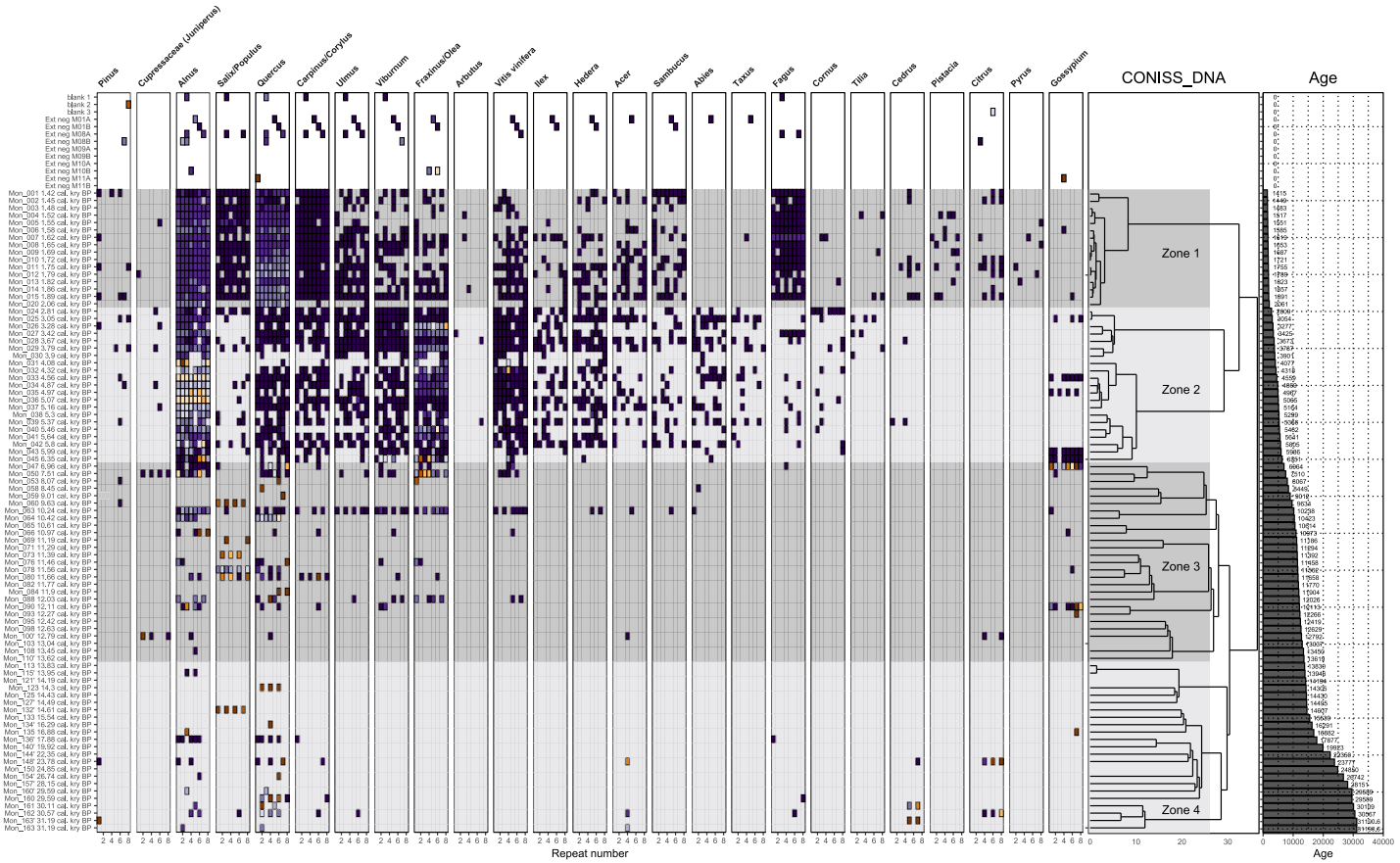

Figure 3a. Overview of *trn*L metabarcoding results for selected woody taxa. The x-axis shows the number of PCR replicas and cells are coloured based on the relative abundance of assigned reads. Samples are ordered on the y-axes with increasing age and with controls samples at the top. The grey shading represents the zones identified by the CONISS tree based on this DNA data.

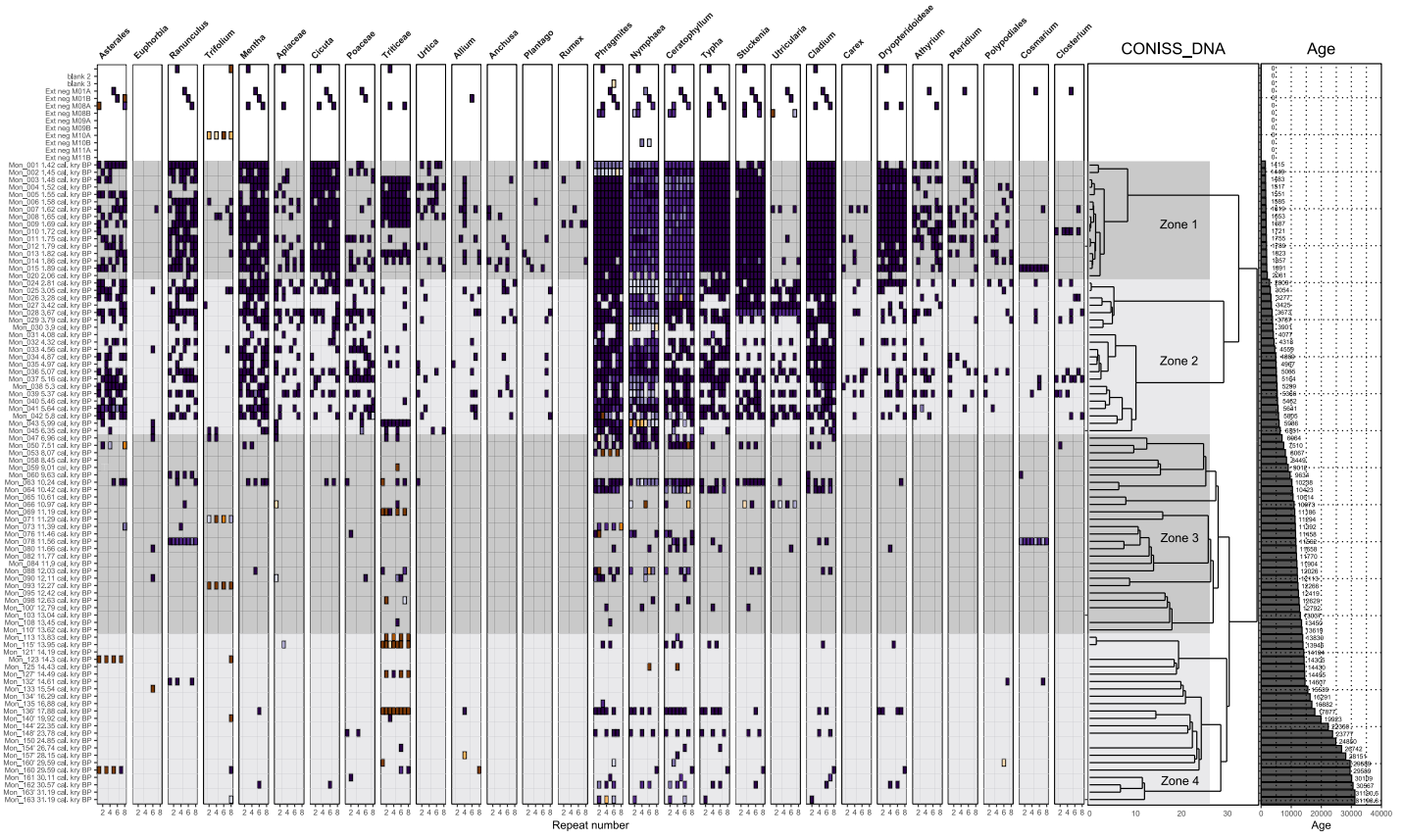

Figure 3b. Overview of *trn*L metabarcoding results for selected herbaceous, aquatic and ferns taxa (see legend from Fig. 3a for details).

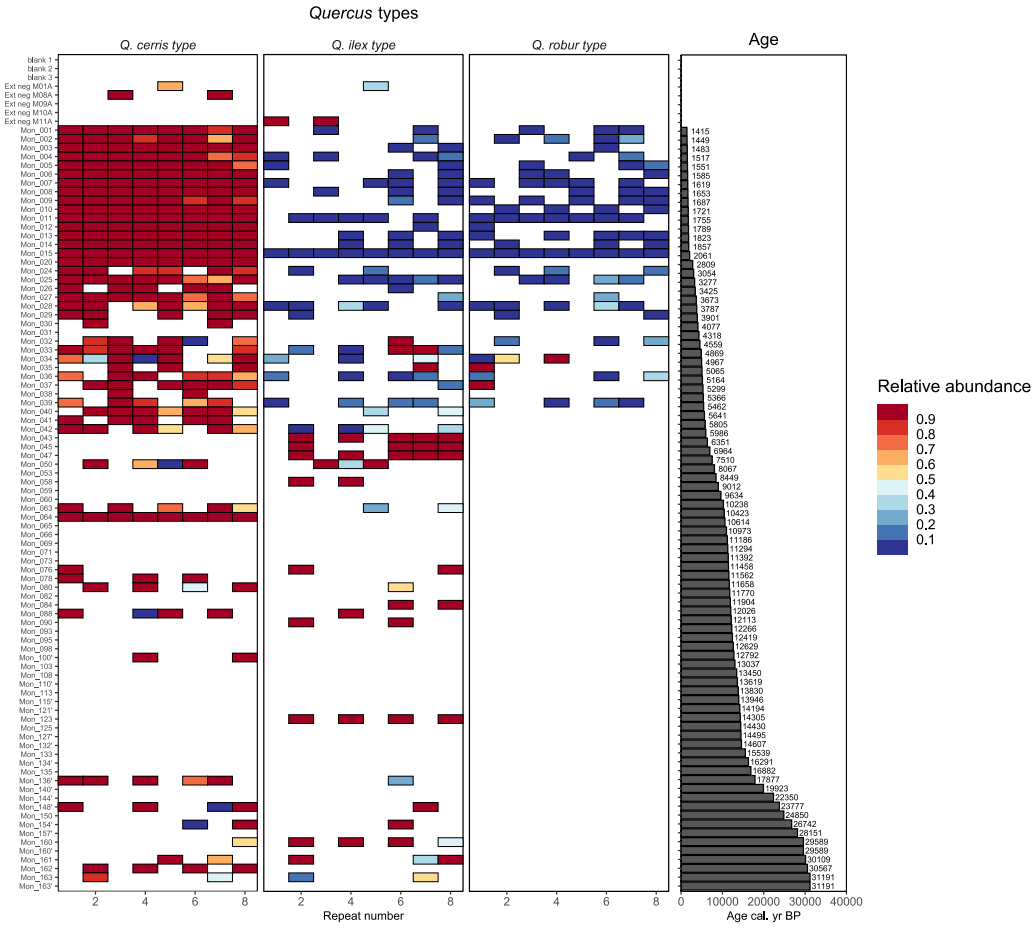

Figure 4. Overview of *trn*L metabarcoding results for the three *Quercus* types detected (see legend from Fig. 3a for details). The plot shows the raw unclustered reads counts for the three *Quercus* groups.

Zone 1 has a high abundance of *Nymphaea/Nuphar* and *C. demersum*, with consistent detection at relatively low read abundance of *Typha*, *Stuckenia* and *Cladium*. *P. australis* is very abundant in the two youngest samples (Mon_001, and Mon_002), where there is no signal of Triticeae. On the other hand, Triticeae is present consistently on part of zone 1, together with *Sparganium*, which is present when Triticeae is absent. The beginning of zone 2 has the same composition as zone 1, with increased *Nymphaea/Nuphar*, appearance of *Utricularia*, Potamogetonaceae and disappearance of Triticeae. The second part of zone 2 shows a more stochastic detection of *Stuckenia*, *Utricularia* and *Cladium*. The few high-quality samples in zone 3 and 4 contain a high abundance of Triticeae, with low read abundance of *P. australis*¸ *Nymphaea/Nuphar*, *C. demersum*, and *Typha*.

Ferns, Bryophytes, Parasitic and algae

Nearly all ferns are recovered in zone 1, with Dryopteridaceae being the most abundant taxa and stochastic detection of *Athyrium*, *Pteridium*

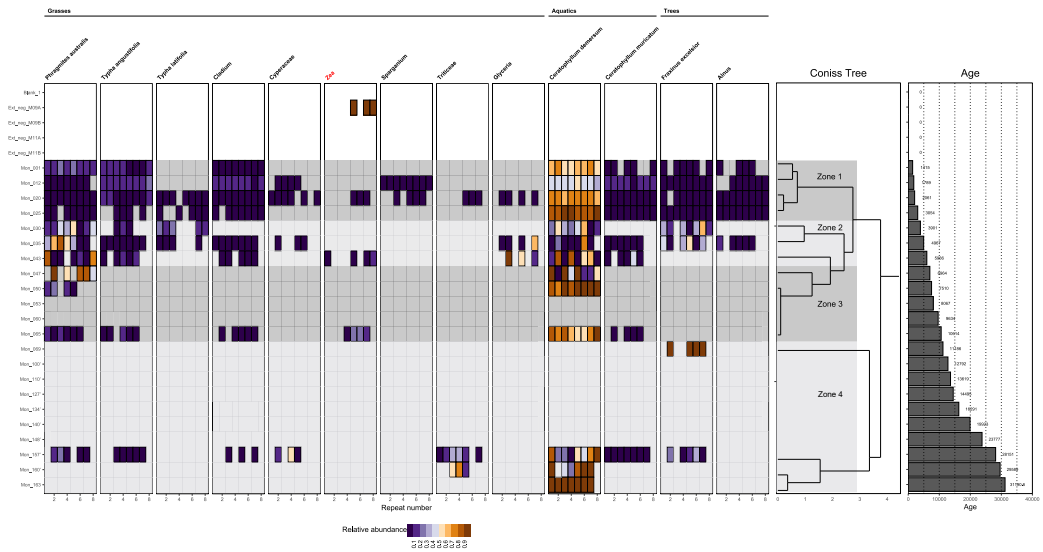

Figure 5. Overview of the ITS1 metabarcoding results (see legend from Fig. 3a for details).

and *Polystichum* (**Fig. 3c**). Zones 2 and 3 have stochastic presences of Dryopteridaceae and *Athyrium*. Very few bryophyte taxa were recovered, with *Platydictya* being detected in two samples – interestingly these were the most abundant taxa at 11.5 cal. ka BP (Mon_076). Other bryophyte taxa detected in few samples of zones 2 and 3 are *Funaria*, Hypnales, and *Antitrichia curtipendula* and the two parasitic taxa Loranthaceae and *Lathraea squamari*. Some algae were also found with *Nannochloropsis* detected in seven samples in all zones except zone 3. In two samples *Nannochloropsis* was detected together with *Cosmarium botrytis* at the beginning of zone 1 and 4. *Neglectella solitaria* is present in three samples in zones 2 and 3, *and Closterium ehrenbergii* in zones 1 and 3.

ITS1 metabarcoding Poaceae

Overall, only 12 taxa (excluding the contaminant *Zea*) were recovered with ITS1 Poaceae primers (**Fig. 5**). Of these, four taxa did not belong to the Poaceae family or Poales order: the trees *Alnus* (Betulaceae) and *Fraxinus excelsior* (Oleaceae) and the aquatics *C. demersu* and *C. muricatum* (*Ceratophyllaceae*). The CONISS tree identifies four zones: zone 1 (1.4 – 3.1 cal. ka BP), zone 2 (3.9 – 6.0 cal. ka BP), zone 3 (7.0 – 10.6 cal. ka BP) and zone 4 (11.2 – 31.2 cal. ka BP). Both zone 3 and 4 show substantial sample dropout.

Based on relative read abundance, the aquatic *C. demersum* is the most abundant in zone 1, 3, and 4 while *C. muricatum* reads are present in most samples at lower relative abundance (except for sample Mon_012). Interestingly, in sample Mon_012 *T. angustifolia* and *Cladium* are also more abundant and *Sparganium* is detected in all eight replicas. The aquatic grass *P. australis* was consistently detected in most samples and repeats and in zone 2 is the taxa with the highest relative abundance. Two species of *Typha* were identified, of which *Typha angustifolia* was the most common and *Typha latifolia* was present in the first two samples of zone 1 and in most of the samples of zone 2. Zone 4 show a stochastic presence of most taxa recovered in the previous zones (1, 2, and 3), although Triticaea is much more common in zone 4.

Alnus and *F. excelsior* are common but at low relative read abundance in zone 1 and 2, while in zone 4, in sample Mon_069, *Fraxinus* represents 100% of the plant reads recovered. In zone 3 and 4 *Alnus* is no longer detected.

Discussion

Overall metabarcoding success and constrains

The *trn*L and ITS1 metabarcoding analysis of the long sedimentary record of Lago Grande di Monticchio produced overall a low number of high quanlity reads. We plotted the extraction controls alongside the samples, and we removed known contaminants such as *Hevea* (rubber tree) or *Zea* (corn). We acknowledge that our filtering criteria were not strict, but we wanted to be inclusive as this was the first *sed*aDNA study conducted at low latitudes in southern Europe and we wanted to detect low abundant and rare taxa.

The ITS1 primers have not been experimentally tested before, and our results show a large proportion of reads matching to off-target taxa (mainly bacteria with less than 95% similarity). There are many possible causes for this mismatch, and the most likely is low primer specificity. Similarly, a high number of reads, consisting of library tag-index chimeric sequences, could not be used in the *trn*L analysis. Formation of chimeric sequences is likely due to the similarity existing reads belonging to the most abundant taxa in high-quality samples. Tag jumping is expected to occur more often in abundant taxa, however using unique tag and index combinations and the incorporation of unused index-tag controls (Taberlet *et al.*, 2018) allowed us to detect this phenomenon.

We made all efforts to minimize contamination following standard procedures for ancient DNA studies, however contamination can never be completely avoided, and control samples were excellent guides for filtering and building confidence in our data. We used eight PCR repeats per sample (Ficetola *et al.*, 2015) and included PCR and extraction controls in all runs. Overall PCR controls showed few reads, which indicates a low amount of background cross-contamination between samples. Also, extraction controls rarely showed a signal in more than two repeats except for Brassicaceae, *Trifolium*, and *F. ulmaria*. Many members of the Brassicaceae family are native in Italy, but the family include several cultivated crops. We noticed that both *Trifolium*, and *F. ulmaria* were commonly observed in samples from another project pooled and sequenced together with Monticchio samples, indicating that some cross-over between samples did occur with little impact on the quality of results. Crop species such as *Capsicum* and *Cucumis* were also detected in our samples but in larger proportion than in the controls (**Table S1**) and were therefore remove from the dataset.

Sample dropout in our dataset was consistently occurring in samples older than 7.5 cal. ka BP, with both *trn*L and ITS1, with only few exceptions. It is possible that the dropout was related to different levels of DNA degradation, DNA extraction bias due to sediment type, amplification bias linked to the presence of specific taxa, or more likely to a combination of these factors (Capo *et al.*, 2021). PCR inhibition may explain sample dropout between 7.5 cal. ka BP to 15.5 cal. ka. because high inhibition was observed before Zymo purification – while the DNA concentration remained similar (**Fig. 2**).

Comparing results from two different plant metabarcoding markers has rarely been done in lake sediments. One main difference between markers in this study was the longer amplicon size of ITS1 (112-128 bp including primer binding sites) compared to *trn*L (49-259 bp, average ~87 pb). Longer fragments are less likely amplified when DNA is degraded, however in our case we noticed similar amplification success between markers. On the other hand, ITS1 showed a better representation among PCR repeats than *trn*L, suggesting that DNA degradation is not the

main factor causing sample dropout (similar taxa were recorded with both markers). It is possible however, that ITS1 primers have a lower detection threshold, or that nuclear DNA is more common in sediment samples than multicopy chloroplast regions, as recently proposed by Wang *et al.* (2021). Likely the reason is not ecological because samples that worked poorly with *trn*L (Mon_065, Mon_157') showed similar taxon composition with ITS1. On the other hand, most taxa detected with ITS1 could also be identified with *trn*L, which allows for a direct comparison of taxon presence. The most common Poales taxa, *P. australis, Typha,* and *Cladium* are detected in the same samples with both primers but with differences in relative abundance. This suggests that changes in relative abundance should not be directly translated into quantitative changes because different markers experience different biases (Nichols et al., 2018). The tree species *Alnus* and *F. excelsior* were identified with ITS1 in samples were *trn*L failed, and vice versa. *Alnus* is absent before 6.0 cal. kyr BP with ITS, while *trn*L records *Alnus* in nearly all samples. *Fraxinus* is present in more repeats in the first samples with ITS1 than with *trn*L. In conclusion, ITS1 did not increase, as expected, the number of Poales taxa recovered, instead it increased taxonomic resolution for *Typha* revealing that *T. angustifolia* was more common than *T. latifolia and*detected a second *Ceratophyllum* species (*C. muricatum*).

Attempts were also made to create shotgun libraries from the less inhibited lateglacial samples (20-31.2 cal. ka BP) but with little success (results not shown). It is possible that either the DNA fragments were too short for the cut-off of 100 nucleotides used in the extraction kit (Qiagen PowerSoil), or that the extraction method was not efficient in extracting DNA from samples of this sediment type (minerogenic fine silt).

7.5 kyr of flora reconstruction at Lago Grand di Monticchio

We discuss here only the *sed*aDNA results for the first 7.5 cal. kyr of this unique maar lake in southern Italy as results from older samples provided little information.

The flora detected shows good similarities with the previous published pollen records from Monticchio (Watts *et al.,*1996; Allen *et al.,* 2002). It is not surprising that *Alnus* is commonly detected with *sed*DNA because this taxon (likely *A. glutinosa*) is confined to wet ground at lake margins (Allen *at al.,* 2002). Two species of *Alnus* are found today at Monticchio, *A. glutinosa* that grows at the lake margins and *A. cordata* that grows upslope together with *Fagus*. Similar to pollen however, the *trn*L marker could not differentiate between the two species. We found also a high abundance of *Quercus* DNA, as today this taxon grows also close to the lake, and abundant pollen has continually been recovered in Holocene cores from the lake.

As also observed in Parducci *et al*. (2017) *sed*aDNA recovered many aquatics such as *Nymphaea/Nuphar*, *Ceratophyllum*, *Stuckenia*, *Utricularia*, and *Potamogeton,* but also taxa usually constrained to wet or submerged environments at the margin of temperate Mediterranean lakes, such are *Salix*, *Cladium*, *Phragmites, Typha, Glyceria, Schoenoplectus,* and *Equisetum*. These taxa are not sensitive to macro-mesoclimatic fluctuations but instead to changes in water level of the lake. Variations in their abundance in term of repeat counts might indicate water level changes occurred between zone 1, and 2.

The CONISS tree based on the *sed*aDNA (CONISS_DNA), agrees well with the previous CONISS2 tree based on terrestrial and aquatic taxa (Allen *et al.,* 2002). CONISS_DNA zone 1, correspond well to CONISS2 zone 1b ending approximately 3 cal. kyr BP but it ends earlier at 2.1 cal. ka BP likely because there is a short gap in sampling between 2.1 – 2.8 cal. kyr BP.

CONISS_DNA zone 2 corresponds well with pollen zone 1C, but again, the start and end points are slightly different due to sampling resolution. CONISS_DNA zone 3 and 4 do not correspond well with pollen due to the high number of sample dropout and stochasticity of the DNA data in this period.

The detection and chronology of the local disappearance of *Abies* at the onset of the Christian era (~2 cal. ka BP) is worth being noticed. The pollen record confirms local presence of *Abies* between 8 and 3 cal. ka BP and a successive rapid reduction and disappearance coeval with a *Fagus* increase. A reduction of *Abies* is also recorded over the whole Italian peninsula, probably following the combined effects of increased human activity and climate dryness (Magri *et al.*, 2015).

The disappearance of *Abies* chronologically corresponds with the appearance fin the last two thousand years of taxa that relate to human presence like *Citrus*, *Pyrus* and *Rumex* (the latter related to stockbreeding). These taxa are not detected by pollen as this is poorly dispersed.

Tree different types of *Quercus* were recovered with *trn*L. *Q. cerris* is by far the most common type, which is in line with the pollen data (Allen *et al.*, 2002). *Q. ilex* was recovered mainly between 5.8-6 cal. kyr probably because the climate around the lake was favorable for this taxon. Finally, it was difficult to conclude about the presence of *Q. robur* type due to few differences found in the sequence of this taxa.

Unexpected finds

A find that needs critical assessment and further investigations is *Cedrus.* Palynological data reveal ´uncertain presence of this taxon in the Italian peninsula during the middle Pleistocene (~0.8 Ma, Magri *et al.*, 2017). Few pollen grains recovered in lakes from southern Italy were found to correlate with Saharan dust intensity and were therefore explained by long distance transport from north Africa (Magri and Parra, 2002). Although *Cedrus* pollen is produced is great amounts, and wind-transported over long distances, Postigo-Mijarra *et al.* (2010) concluded instead that low pollen count in the Iberian Peninsula could be interpreted as local presence, and that *Cedrus* could therefore have persisted locally until the onset of the Holocene. Detecting *Cedrus* in Italy with DNA might, therefore indicate continues presence of this taxon in southern Italy. However, like *Pinus*, *Cedrus* has been also considered a contaminant in *sed*aDNA studies when its presence could be excluded based on geography (Voldstad et al., 2020; Bremond et al., 2017; Willerslev et al, 2014). An alternative hypothesis, to be tested with further analyses, is that *Cedrus* DNA refers to trees planted for ornamental purposes, as it is mostly recorded in the last two thousand years.

Another unexpected find was *Gossypium* (cotton). This taxon occurs in all but one repeat, in three subsequent samples – but also at lower frequency in other samples, although not following a pattern typical of contaminants. *Gossypium* is not native to Italy but is a casual archaeophyte (Viot, 2019) and the recovered sequence do not belong to other native members in the Malvaceae family. Cotton on the other hand, is a common textile, and fibres may have contaminated parts of the sediment core.

Was Pinus present in the Lateglacial?

The question if *Pinus* was locally present during the Lateglacial at Monticchio could not be answered in this study. *Pinus* was detected in lateglacial sediments and during the onset of the Holocene but finding were based on low frequency reads and we cannot exclude background contamination. Failure to detect a taxon however is not prove of absence and we did detect few *Pinus reads* in zone 1 and 2. *Pinus* has been recorded as background contamination in a recent metabarcoding study (Alsos *et al.*, 2020). In our study, two out of 14 extractions controls showed one repeat with *Pinus* (two out of 112 PCR

reactions). Due to our sporadic findings and to the limits imposed by our markers we could not resolve which *Pinus* species was present in the Holocene. The pollen records from Lago Grande di Monticchio (Watts *et al.*, 1996; Allen *at al.*, 2002) supports the evidence that *Pinus* was very sparse as its pollen never attains frequencies >10%, although it is produced in huge amounts by pine trees and is easily dispersed by wind. The reduced presence of *Pinus* in the landscape during the Lateglacial and the Holocene is also confirmed by other pollen records available from southern Italy, including Lake Trifoglietti (de Beaulieu *et al.*, 2017; Joannin *et al.*, 2012), Lago Alimini Piccolo (Di Rita and Magri, 2009), and Sant'Eufemia Plain (Russo Ermolli *et al.*, 2018). Only at Lake Battaglia (Caroli and Caldara, 2007), a site located within stands of *P. halepensis*, pine pollen is recorded with values exceeding 20%.

Conclusion

Our study demonstrates how metabarcoding analysis of *sed*DNA extracted from ancient sediments from low-latitude lakes provide detailed records of past flora and, together with classical palaeobotanical data, help in reconstructing vegetation history. The potential offered by the 14-meter-long core from Monticchio however, could not been fully explored due to a lack of signal in samples older than 7.5 kyr BP probably due to a combination of DNA degradation and PCR inhibition. For these reasons, key questions related to transition from full glacial into Holocene could not be explored in full. Some contaminant taxa were excluded from our dataset with the help of control samples (*Capsicum*, *Cucumis*, and *Musa*), while uncertain taxa such as *Gossypium* (Malvaceae) and *Cedrus* (Pinaceae) remained and should be further investigated. Our results show how methodological advances in *sed*aDNA protocols as well as new bioinformatics methods are necessary to fully understand the potential of the metabarcoding technique applied to low-latitudes lakes. It has been recently shown that shotgun sequencing data mapping to nuclear DNA databases (Whang *et al.*, 2021) provide better result with degraded DNA. Unfortunately, however, to date a high-quality nuclear DNA database does not exist for the rich Mediterranean flora.

References

Allen, J. R. M., & Huntley, B. (2009). Last Interglacial palaeovegetation, palaeoenvironments and chronology: a new record from Lago Grande di Monticchio, southern Italy. Quaternary Science Reviews, 28(15–16), 1521–1538.

Allen, J. R. M., & Huntley, B. (2018). Effects of tephra falls on vegetation: A Late-Quaternary record from southern Italy. Journal of Ecology, 106(6), 2456–2472.

Allen, J. R. M., Watts, W. A., & Huntley, B. (2000). Weichselian palynostratigraphy, palaeovegetation and palaeoenvironment; the record from Lago Grande di Monticchio, southern Italy. Quaternary International, 73–74, 91–110.

Allen, J. R. M., Watts, W. A., McGee, E., & Huntley, B. (2002). Holecene environmental variability - The record from Lago Grande di Monticchio, Italy. Quaternary International, 88.

Alsos, I. G., Sjögren, P., Brown, A. G., Gielly, L., Merkel, M. K. F., Paus, A., Lammers, Y., Edwards, M. E., Alm, T., Leng, M., Goslar, T., Langdon, C. T., Bakke, J., & van der Bilt, W. G. M. (2020). Last Glacial Maximum environmental conditions at Andøya, northern Norway; evidence for a northern ice-edge ecological "hotspot." Quaternary Science Reviews, 239, 106364.

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. PLoS ONE, 13(4), e0195403.

Baamrane, M. A. A., Shehzad, W., Ouhammou, A., Abbad, A., Naimi, M., Coissac, E., Taberlet, P., & Znari, M. (2012). Assessment of the Food Habits of the Moroccan Dorcas Gazelle in M'Sabih Talaa, West Central Morocco, Using the trnL Approach. PLOS ONE, 7(4), e35643.

Bennett, K. D., Tzedakis, P. C., & Willis, K. J. (1991). Quaternary Refugia of North European Trees. Journal of Biogeography, 18(1), 103.

Beaulieu, J.-L. de, Brugiapaglia, E., Joannin, S., Guiter, F., Zanchetta, G., Wulf, S., Peyron, O., Bernardo, L., Didier, J., Stock, A., Rius, D.,

Magny, M., 2017. Lateglacial-Holocene abrupt vegetation changes at Lago Trifoglietti in Calabria, Southern Italy: The setting of ecosystems in a refugial zone. Quaternary Science Reviews 158, 44–57

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(1), 176–182.

Brauer, A., Mingram, J., Frank, U., Günter, C., Schettler, G., Wulf, S., Zolitschka, B., & Negendank, J. F. W. (2000). Abrupt environmental oscillations during the Early Weichselian recorded at Lago Grande di Monticchio, southern Italy. Quaternary International, 73–74, 79–90.

Brauer, A., Allen, J. R. M., Mingram, J., Dulski, P., Wulf, S., & Huntley, B. (2007). Evidence for last interglacial chronology and environmental change from Southern Europe. Proceedings of the National Academy of Sciences of the United States of America, 104(2), 450–455.

Bremond, L., Favier, C., Ficetola, G. F., Tossou, M. G., Akouégninou, A., Gielly, L., Giguet-Covex, C., Oslisly, R., & Salzmann, U. (2017). Five thousand years of tropical lake sediment DNA records from Benin. Quaternary Science Reviews, 170, 203–211.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10(1), 421.

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A. G., Clarke, C. L., Crump, S. E., Debroas, D., Englund, G., Ficetola, G. F., Garner, R. E., … Parducci, L. (2021). Lake Sedimentary DNA Research on Past Terrestrial and Aquatic Biodiversity: Overview and Recommendations. Quaternary, 4(1), 6.

Caroli, I., Caldara, M., 2007. Vegetation history of Lago Battaglia (eastern Gargano coast, Apulia, Italy) during the middle-late Holocene. Vegetation History and Archaeobotany 16, 317–327.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution, 9(2), 410–419.

Creer, K. M., & Morris, A. (1996). Proxy-climate and geomagnetic palaeointensity records extending back to Ca. 75,000 BP derived from sediments cored from Lago Grande di Monticchio, Southern Italy. Quaternary Science Reviews, 15(2–3), 167–188.

Di Rita, F., Magri, D., 2009. Holocene drought, deforestation and evergreen vegetation development in the central Mediterranean: a 5500 year record from Lago Alimini Piccolo, Apulia, southeast Italy. The Holocene 19

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science, 14(6), 927–930.

Donders, T., Panagiotopoulos, K., Koutsodendris, A., Bertini, A., Mercuri, A. M., Masi, A., Combourieu-Nebout, N., Joannin, S., Kouli, K., Kousis, I., Peyron, O., Torri, P., Florenzano, A., Francke, A., Wagner, B., & Sadori, L. (2021). 1.36 million years of Mediterranean forest refugium dynamics in response to glacial–interglacial cycle strength. Proceedings of the National Academy of Sciences, 118(34), e2026111118.

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), 1–10.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15(3), 543–556.

Fletcher, W. J., Sánchez Goñi, M. F., Allen, J. R. M., Cheddadi, R., Combourieu-Nebout, N., Huntley, B., Lawson, I., Londeix, L., Magri, D., Margari, V., Müller, U. C., Naughton, F., Novenko, E., Roucoux, K., & Tzedakis, P. C. (2010). Millennial-scale variability during the last glacial in vegetation records from Europe. Quaternary Science Reviews, 29(21–22), 2839–2864.

Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodríguez-Sánchez, F., Dobrowski, S. Z., Hampe, A., Hu, F. S., Ashcroft, M. B., Bartlein, P. J., Blois, J. L., Carstens, B. C., Davis, E. B., Lafontaine, G.,

Hajdas, I., Bonani, G., Zolitschka, B., Brauer, A., & Negendank, J. (1998). 14C ages of terrestrial macrofossils from Lago Grande di Monticchio (Italy). Radiocarbon, 40(2), 803–807.

Hewitt, G. (1999). Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society, 68(1–2), 87–112.

Huntley, B., Allen, J. R. M., & Watts, W. A. (1996). Weichselian late-glacial paiseoecology and palaaoenvironment at Lago Grande di Monticchio

(Basilicata, S Italy). Alpine and Mediterranean Quaternary, 9(2).

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., & Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLOS Computational Biology, 12(6), e1004957.

Jackson, S. T., & Williams, J. W. (2004). Modern Analogs In Quaternary Paleoecology: Here Today, Gone Yesterday, Gone Tomorrow? Annual Review of Earth and Planetary Sciences, 32(1), 495–537.

Juggins, S. (2020). rioja: Analysis of Quaternary Science Data (R package version 0.9-26). https://cran.r-project.org/package=rioja

Leach, R., & Parsons, L. (2019). Barcode Splitter, version 0.18.5. https://bitbucket.org/princeton_genomics/barcode_splitter

Magri, D., & Parra, I. (2002). Late Quaternary western Mediterranean pollen records and African winds. Earth and Planetary Science Letters, 200(3–4), 401–408.

Magri, D., Agrillo, E., Di Rita, F., Furlanetto, G., Pini, R., Ravazzi, C., Spada, F., 2015. Holocene dynamics of tree taxa populations in Italy. Review of Palaeobotany and Palynology 218, 267–284.

Magri, D., Di Rita, F., Aranbarri, J., Fletcher, W., & González-Sampériz, P. (2017). Quaternary disappearance of tree taxa from Southern Europe: Timing and trends. In Quaternary Science Reviews (Vol. 163, pp. 23–55). Elsevier Ltd.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. PeerJ, 2014(1), e593.

Martin-Puertas, C., Lauterbach, S., Allen, J. R. M., Perez, M., Blockley, S., Wulf, S., Huntley, B., & Brauer, A. (2019). Initial Mediterranean response to major climate reorganization during the last interglacial-glacial transition. Quaternary Science Reviews, 215, 232–241.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 5(6), pdb.prot5448.

Narcisi, B. (1996). Tephrochronology of a late quaternary lacustrine record from the Monticchio maar (Vulture volcano, southern Italy). Quaternary Science Reviews, 15(2–3), 155–165.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E., & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. Molecular Ecology Resources, 18(5), 927–939

Nimmergut, A. P., Allen, J. R. M., Jones, V. J., Huntley, B., & Battarbee, R. W. (1999). Submillennial environmental fluctuations during marine Oxygen Isotope Stage 2: a comparative analysis of diatom and pollen evidence from Lago Grande di Monticchio, South Italy. Journal of Quaternary Science, 14(2), 111–123.

Nota, K., Klaminder, J., Melesi, P., Bindler, R., Nobile, A., van Steijn, T., Bertilsson, S., Svensson, B., Hirota, S.K., Matsuo, A., Gunnarsson, U., Seppä, H., Väliranta, M,M., Wohlfarth, B., Suyama, Y., Parducci, P. (2022). Norway spruce postglacial recolonization of Fennoscandia, *in press*.

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. Ancient plant DNA in lake sediments. New Phytologist 214 (3), 924–942 (2017).

Ramrath, A., Zolitschka, B., Wulf, S., & Negendank, J. F. W. (1999). Late pleistocene climatic variations as recorded in two Italian maar lakes (Lago di Mezzano, Lago Grande di Monticchio). Quaternary Science Reviews, 18(7), 977–992.

Postigo-Mijarra, J. M., Morla, C., Barrón, E., Morales-Molino, C., & García, S. (2010). Patterns of extinction and persistence of Arctotertiary flora in Iberia during the Quaternary. Review of Palaeobotany and Palynology, 162(3), 416–426.

Rijal, D. P., Heintzman, P. D., Lammers, Y., Yoccoz, N. G., Lorberau, K. E., Pitelkova, I., Goslar, T., Murguzur, F. J. A., Salonen, J. S., Helmens, K. F., Bakke, J., Edwards, M. E., Alm, T., Bråthen, K. A., Brown, A. G., & Alsos, I. G. (2021). Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. Science Advances, 7(31), eabf9557.

Russo Ermolli, E., Ruello, M.R., Cicala, L., Di Lorenzo, H., Molisso, F., Pacciarelli, M., 2018. An 8300-yr record of environmental and cultural changes in the Sant'Eufemia Plain (Calabria, Italy). Quaternary International 483, 39–56.

Rull, V., Vegas-Vilarrúbia, T., Corella, J. P., Trapote, M. C., Montoya, E., & Valero-Garcés, B. (2021). A unique Pyrenean varved record provides a detailed reconstruction of Mediterranean vegetation and land-use dynamics over the last three millennia. Quaternary Science Reviews, 268, 107128.

Schettler, G., & Albéric, P. (2008). Laghi di Monticchio (Southern Italy, Region Basilicata): Genesis of sediments - A geochemical study. Journal of Paleolimnology, 40(1), 529–556

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., Gussarova, G., Bellemain, E., Hassel, K., Stenøien, H. K., Epp, L., Schrøder-Nielsen, A., Brochmann, C., & Yoccoz, N. G. (2015). Highly Overlapping Winter Diet in Two Sympatric Lemming Species Revealed by DNA Metabarcoding. PLOS ONE, 10(1), e0115335.

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P., & Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Molecular Ecology Resources, 10(6), 1009–1018.

Stoppa, F., & Principe, C. (1997). Eruption style and petrology of a new carbonatitic suite from the Mt. Vulture Southern Italy /: The Monticchio Lakes Formation. Journal of Volcanology and Geothermal Research, 78(3–4), 251–265.

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(3), e14–e14.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For biodiversity research and monitoring. In Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press.

Tomlinson, E. L., Albert, P. G., Wulf, S., Brown, R. J., Smith, V. C., Keller, J., Orsi, G., Bourne, A. J., & Menzies, M. A. (2014). Age and geochemistry of tephra layers from Ischia, Italy: Constraints from proximal-distal correlations with Lago Grande di Monticchio. Journal of Volcanology and Geothermal Research, 287, 22–39.

Venanzoni, R., Apruzzese, A., Gigante, D., Suanno, G., & Vale, E. (2003). Contributo alla conoscenza della vegetazione acquatica e igrofitica dei Laghi di Monticchio. Informatore Botanico Italiano, 35(1), 69–80.

Viot, C. (2019). Domestication and varietal diversification of Old World cultivated cottons (Gossypium sp.) in the Antiquity. Revue d'ethnoécologie, 15.

Voldstad, L. H., Alsos, I. G., Farnsworth, W. R., Heintzman, P. D., Håkansson, L., Kjellman, S. E., Rouillard, A., Schomacker, A., & Eidesen, P. B. (2020). A complete Holocene lake sediment ancient DNA record reveals long-standing high Arctic plant diversity hotspot in northern Svalbard. Quaternary Science Reviews, 234, 106207.

Wang, Y., Pedersen, M. W., Alsos, I. G., De Sanctis, B., Racimo, F., Prohaska, A., Coissac, E., Owens, H. L., Merkel, M. K. F., Fernandez-Guerra, A., Rouillard, A., Lammers, Y., Alberti, A., Denoeud, F., Money, D., Ruter, A. H., McColl, H., Larsen, N. K., Cherezova, A. A., … Willerslev, E. (2021). Late Quaternary dynamics of Arctic biota from ancient environmental genomics. Nature, 1–7.

Watts, W. A., Allen, J. R. M., Huntley, B., & Fritz, S. C. (1996). Vegetation history and climate of the last 15,000 years at Laghi di Monticchio, southern Italy. Quaternary Science Reviews, 15(2–3), 113–132.

Watts, W. A., Allen, J. R. M., & Huntley, B. (2000). Palaeoecology of three interstadial events during oxygen-isotope Stages 3 and 4: A lacustrine record from Lago Grande di Monticchio, southern Italy. Palaeogeography, Palaeoclimatology, Palaeoecology, 155(1–2), 83–93.

Watts, W. A. (1985). A long pollen record from Laghi di Monticchio, southern Italy: a preliminary account. Journal - Geological Society (London), 142(3), 491–499.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. Nature, 506(7486), 47–51.

Wulf, S., Kraml, M., Brauer, A., Keller, J., & Negendank, J. F. W. (2004). Tephrochronology of the 100ka lacustrine sediment record of Lago Grande di Monticchio (southern Italy). Quaternary International, 122(1 SPEC. ISS.), 7–30.

Wulf, S., Kraml, M., & Keller, J. (2008). Towards a detailed distal tephrostratigraphy in the Central Mediterranean: The last 20,000 yrs record of Lago Grande di Monticchio. Journal of Volcanology and Geothermal Research, 177(1), 118–132.

Wulf, S., Keller, J., Paterne, M., Mingram, J., Lauterbach, S., Opitz, S., Sottili, G., Giaccio, B., Albert, P. G., Satow, C., Tomlinson, E. L., Viccaro, M., & Brauer, A. (2012). The 100–133 ka record of Italian explosive volcanism and revised tephrochronology of Lago Grande di Monticchio. Quaternary Science Reviews, 58, 104–123.

Wutke, K., Wulf, S., Tomlinson, E. L., Hardiman, M., Dulski, P., Luterbacher, J., & Brauer, A. (2015). Geochemical properties and environmental impacts of seven Campanian tephra layers deposited

between 40 and 38 ka BP in the varved lake sediments of Lago Grande di Monticchio, southern Italy. Quaternary Science Reviews, 118, 67–83.
Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5), 614–620.
Zolitschka, B., & Negendank, J. F. W. (1996). Sedimentology, dating and palaeoclimatic interpretation of a 76.3 ka record from Lago Grande di Monticchio, southern Italy. Quaternary Science Reviews, 15(2–3), 101–112.

Acknowledgements

We thank Achim Brauer and Xueru Zhao for providing the chronology, and comments on the manuscript. We also thank Luca Turolo and Katerina Tasiouli for assisting in some of the ancient DNA labwork. the computations and data handling were enabled by resources provided by the Swedish National In-frastructure for Computing (SNIC) at Uppmax (Uppsala) and partially funded by the Swedish Research Council (nr. 2018-05973). The ancient DNA work was financed with grants obtained from the Swedish Phytogeographical Society.

Author contribution

L.P. and K.N. conceived the study. K.N., T.T., and F.B. preformed the molecular analysis. K.N. preformed and analysed the metabarcoding data. K.N. and L.P. wrote the manuscript with significant contribution for palaeoecological interpretation by D.M. and F.S.