Quaternary Research (2023), **112**, 180–188
doi:10.1017/qua.2022.46

Contributions to the QR Forum

Ancient DNA from speleothems: opportunity or challenge?

Alexis Marchesini[a,f]* , Daniela Festi[b], Matteo Girardi[c], Diego Micheletti[d], Cristiano Vernesi[a]*[§] and Marc Luetscher[e§]

[a]Forest Ecology Unit, Research and Innovation Centre- Fondazione Edmund Mach, via E. Mach 1-S. Michele all'Adige (Trento), 38010 Italy; [b]Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, Innrain 25, A-6020 Innsbruck Austria; [c]Conservation Genomics Unit, Research and Innovation Centre- Fondazione Edmund Mach, via E. Mach 1-S. Michele all'Adige (Trento), 38010 Italy; [d]Computational Biology Unit, Research and Innovation Centre- Fondazione Edmund Mach, via E. Mach 1-S. Michele all'Adige (Trento), 38010 Italy; [e]Swiss Institute for Speleology and Karst Studies (SISKA), Serre 68, CH-2300 La Chaux-de-Fonds and [f]Institute for Sustainable Plant Protection (IPSP), The National Research Council of Italy (CNR), 50019 Sesto Fiorentino (Florence), Italy

Abstract

Ancient environmental DNA retrieved from sedimentary records (sedaDNA) can complement fossil-morphological approaches for characterizing Quaternary biodiversity changes. PCR-based DNA metabarcoding is so far the most widely used method in environmental DNA studies, including sedaDNA. However, degradation of ancient DNA and potential contamination, together with the PCR amplification drawbacks, have to be carefully considered. Here we tested this approach on speleothems from an Alpine cave that, according to a previous palynomorphological study, have shown to contain abundant pollen grains. This offers a unique opportunity for comparing the two methods and, indirectly, trying to validate DNA-based results. The plant taxa identified by sedaDNA are fewer than those by pollen analysis, and success rate of PCR replicates is low. Despite extensive work performed following best practice for sedaDNA, our results are suboptimal and accompanied by a non-negligible uncertainty. Our preliminary data seem to indicate that paleoenvironmental DNA may be isolated from speleothems, but the intrinsic weakness of PCR-based metabarcoding poses a challenge to its exploitation. We suggest that newly developed methods such as hybridization capture, being free from PCR drawbacks and offering the opportunity to directly assess aDNA authenticity, may overcome these limitations, allowing a proper exploitation of speleothems as biological archives.

Keywords: Speleothems, SedaDNA, PCR-based eDNA metabarcoding, Contamination, Authenticity

(Received 9 December 2021; accepted 15 August 2022)

INTRODUCTION

Speleothems, secondary cave carbonates such as stalagmites and flowstones, are among the most complete terrestrial paleoenvironmental archives that can be accurately dated back to ca. 600 ka (Cheng et al., 2013). While speleothem-based inferences commonly include the reconstruction of hydrological, paleoclimatic, and biogeochemical cycles from seasonal to millennial time scales (Fairchild and Baker, 2012), the use of speleothems as bio-archives remains largely unexplored (Blyth and Frisia, 2008; Lechleitner et al., 2017; Luetscher et al., 2021).

In ventilated caves, the transport of aerosols favors incorporation of pollen grains and other particles into speleothems (Bastin, 1979; Dredge et al., 2013), and the remarkably resistant pollen walls consistently slow down degradation processes (Li et al., 2019). Although the analysis of pollen, spores, and other palynomorphs (organic-walled microfossils) encapsulated in sedimentary deposits remains the principal technique used to reconstruct long-term vegetation changes (Faegri et al., 1989), species-level identification remains difficult (Richardson et al., 2015) because several morphological features are shared within genera and sometimes even families. Moreover, pollen concentrations in these sedimentary archives are often low (Matley et al., 2020, Luetscher et al., 2021).

Thus, alternative methods based on DNA analysis, allowing the investigation of small amounts of biological particles and providing good to outstanding taxonomic resolution, hold great promise. Environmental DNA (eDNA) metabarcoding (Taberlet et al., 2018) has been successfully applied for pollen identification in different environments (e.g., Bell et al., 2016; Leontidou et al., 2021; Varotto et al., 2021), but the exploitation of terrestrial archives has been restricted mainly to lake sediments, permafrozen soils, and ice cores (Jørgensen et al., 2012; Giguet-Covex et al., 2014; ter Schure et al., 2021). These sedimentary archives have given access to relevant paleovegetation records, but dating beyond the radiocarbon age-span remains problematic (Willerslev et al., 2004). Occasionally, speleothems have also been used for molecular-based biological investigations, but limited to very specific applications (e.g., the study of microbial communities associated with biogeochemical processes; Dhami et al., 2018) or resulting in a very low number of detected organisms, with taxonomic complications due to the high degradation of DNA (Zepeda Mendoza et al., 2016; Stahlschmidt et al., 2019; see Discussion).

Here, we applied sedimentary ancient DNA (sedaDNA) metabarcoding (Pedersen et al., 2015) for identifying plant taxa in three 7600-year-old stalagmite sections from the Milchbach cave system (Switzerland, 1840 m asl; Luetscher et al., 2011). The

*Corresponding author email address: alexis.marchesini@ipsp.cnr.it
[§]these authors contributed equally

Cite this article: Marchesini A, Festi D, Girardi M, Micheletti D, Vernesi C, Luetscher M (2023). Ancient DNA from speleothems: opportunity or challenge? *Quaternary Research* **112**, 180–188. https://doi.org/10.1017/qua.2022.46

© University of Washington. Published by Cambridge University Press, 2022. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

periglacial Alpine environment and the reasonably fast calcite deposition rate (ca. 40 μm/yr) strictly limit DNA leaching, while a strong ventilation favors the incorporation of biological particles. Indeed, a remarkably high pollen concentration was found in Milchbach speleothem sections in a previous palynological analysis (564 palynomorphs/g; Festi et al., 2016). In this study, we applied a rigorous approach following a set of recently defined recommendations (Capo et al., 2021) to test if results from PCR-based sedaDNA metabarcoding are reliable and comparable to those of pollen spectra derived from the same speleothem sections by means of classical morphological analyses (Festi et al., 2016). We highlight several intrinsic limitations of the PCR-based amplicon sequencing technology, which can make mean its application to speleothem samples (and other sedaDNA matrices) incurs considerable uncertainties, and we suggest potentially more effective alternative strategies.

MATERIALS AND METHODS

Samples

We investigated for the presence of plant DNA in three coeval stalagmites (MB3, MB4, MB5-D) from Milchbach Cave, a periglacial cave system that opens along the western flank of the Upper Grindelwald Glacier valley, Switzerland (1840 m asl; see Fig. 1). U-series dating previously confirmed deposition 7000–8000 yr BP (Luetscher et al., 2011), and palynological analysis highlighted the presence of pollen entrapped in the calcite fabrics (Festi et al., 2016). Carbonate sub-samples of ∼2–4 cm^3 in size were cut along the stalagmite growth axis using a diamond wire saw. The surface of each section was cleaned in an ultrasonic bath using deionized water and stored in a sealed sterile bag until further processing. The taphonomy of Alpine speleothems, with peculiar chemism and steady deposition of calcite, together with the temperature conditions of Milchbach Cave (0–2°C; Luetscher et al., 2011) are potentially favorable to the long-term preservation of DNA (see Appendix S1 for more details on the cave system and speleothem samples).

DNA extraction, PCR amplification and sequencing

All laboratory work was carried out at an ancient DNA-dedicated facility at Fondazione Edmund Mach, following the strict procedures required for ancient DNA studies to avoid contamination with modern samples (see Appendix S1: Sample preparation and pre-processing).

Speleothem specimens were made into a fine powder with a cryogenic grinding mill (using liquid nitrogen; 6770 Freezer/Mill, SPEX SamplePrep, Metuchen, New Jersey). During the procedure, each sample was placed in a closed grinding vial. DNA was extracted from the speleothem powder using the DNeasy PowerSoil DNA extraction kit (Qiagen, Hilden, Germany). In this protocol, cell lysis occurs by mechanical (beads-beating) and chemical processes, allowing extraction of DNA even from the toughest and most difficult samples. The final DNA extract was eluted in 100 μl. We processed one specimen for each speleothem section, except for MB5-D, for which two distinct specimens were available and were therefore processed separately. For each specimen, the obtained powder was split into two samples, from which DNA was extracted in two temporally separated extractions (except for MB3: the low quantity of starting material resulted in only one sample for extraction). Moreover, for each powder sample (i.e., for each single extraction), two replicates were performed using a different amount of starting material (speleothem powder): 250 mg and 500 mg. A scheme of processed samples, DNA extraction, and extraction replicates is reported in Table 1. Two negative controls (blanks) were included in each extraction in order to check for the presence of contaminants in reagents and/or in the laboratory environment.

For identifying plant DNA, a short fragment of the chloroplast gene trnL was amplified using the universal primers c-A49325 and h-B49466 [5′-CGAAATCGGTAGACGCTACG-3′ and 5′-CCATTGAGTCTCTGCACCTATC-3′] (Taberlet et al., 2007) produced by Sigma-Aldrich (Milan, Italy). This trnL fragment has been previously tested in a methodological paper on DNA metabarcoding of airborne pollen (Leontidou et al., 2018) and subsequently used in more recent applications of eDNA analysis of pollen in different Italian Alpine habitats (Leontidou et al., 2021) and Alpine ice cores (Varotto et al., 2021). PCR amplification mixture and thermal conditions are reported in Appendix S2. Band sizes and concentrations of PCR products were checked on a gel electrophoresis using QIAxcel with DNA High Sensitive cartridge (Qiagen, GmbH, Hilden, Germany) and analyzed by QIAxcel ScreenGel Software, in a separated post-PCR room. Multiple PCR replicates were performed for each specimen/extraction, according to the availability of extracted DNA after the PCR set-up (see Table 1). A maximum number of 16 samples was amplified for each separate PCR reaction, including extraction negative controls (blanks) and at least one PCR blank every eight samples, to track potential contaminations. PCR reaction products were gel purified using the MinElute Gel Purification kit (Qiagen, GmbH, Hilden, Germany) and eluted in 50 μl. Each individual PCR replicate was treated separately and processed as an independent sample for sequencing and bioinformatics analysis. Approximately 100 ng of the purified PCR product were used for library preparation using TruSeq DNA sample preparation kit V2 (Illumina Inc., San Diego, CA, USA), pooled in equimolar ratio, and sequenced using MiSeq Reagent Kit v3 in Illumina MiSeq platform.

Bioinformatics analysis and post-processing of metabarcoding data

Raw Illumina data were processed using the MICCA v. 1.7 (Albanese et al., 2015) pipeline and the derived operational taxonomic units (OTUs) were classified using the Sintax algorithm (Edgar, 2016) as implemented in VSEARCH (Rognes et al., 2016; for the parameters used, see Appendix S3). Taxonomic assignment of OTUs was performed by comparison with a custom-made reference database including trnL DNA sequences for most of the Alpine seed plants (Spermatophyta). This local database was built starting from a previously developed trnL database (Leontidou et al., 2018), which was focused on the flora of a southeastern Italian Alpine region, subsequently complemented with public sequences (NCBI GenBank database; National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/genbank) for the other most frequent species of the Alpine flora (Aeschimann et al., 2004).

After the bioinformatic analysis, a strict, multistep post-processing procedure was implemented on the dataset in order to remove negligible OTUs (choosing a stringent threshold: number of reads <100) and identify major and minor (potential) contaminants. Details on the post-sequencing removal of potential contaminations are described in Appendix S3 (Supplementary Information).

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

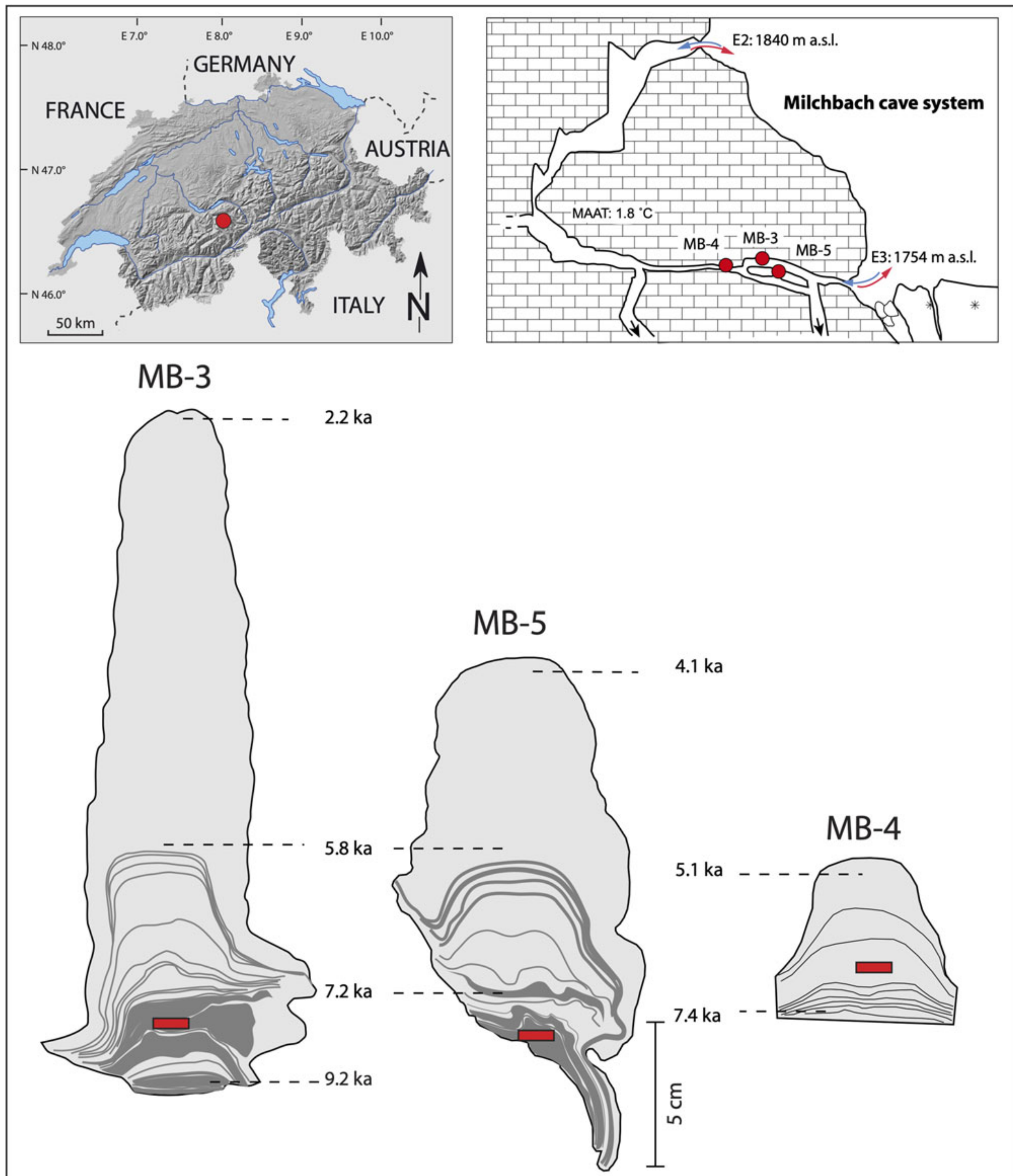

Figure 1. Upper left: location of the sampling site. Upper right: schematic representation of the Milchbach cave system. Bottom: precise location of each speleothem sample (red) from where DNA has been extracted.

RESULTS

Eighteen out of 88 PCR replicates (20.45%) gave a positive outcome based on gel electrophoresis (at least one visible band) and were therefore sequenced, along with 42 negative controls. We obtained at least one PCR replicate with positive outcome for each of the three speleothem sections (MB3, MB4, MB5-D) and for all four processed specimens (two specimens were available for the MB5-D section). Thus, 4/7 of the extraction replicates (powder samples) resulted in the detection of at least one plant taxon (see Table 1 for details on replicates scheme). No systematic differences in PCR outcome were observed among extraction replicates performed from different initial amounts of calcite powder

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

Table 1. Replicates scheme: number and type of extraction and PCR replicates (repl.) per speleothem section and specimen. Suffixes E1, E2, and E3 identify different (temporally separated) DNA extractions, while suffixes a and b identify extraction replicates performed using 250 mg and 500 mg of speleothem powder, respectively.

Speleothem section	Specimen	DNA extraction (powder sample)	DNA extraction repl.	PCR repl.	Positive PCR repl.	Positive PCR repl./specimen
MB3	MB3-208	MB3-208_E3	MB3_208_E3_a	4	0	4
			MB3_208_E3_b	4	4	
MB4	MB4	MB4_E1	MB4_E1_a	5	1	1
			MB4_E1_b	5	0	
		MB4_E2	MB4_E2_a	4	0	
			MB4_E2_b	4	0	
MB5-D	MB5-D-1	MB5-D-1_E1	MB5-D-1_E1_a	14	7	9
			MB5-D-1_E1_b	12	2	
		MB5-D-1_E2	MB5-D-1_E2_a	4	0	
			MB5-D-1_E2_b	4	0	
	MB5-D-2	MB5-D-2_E1	MB5-D-2_E1_a	12	2	4
			MB5-D-2_E1_b	12	2	
		MB5-D-2_E2	MB5-D-2_E2_a	2	0	
			MB5-D-2_E2_b	2	0	

(250 mg versus 500 mg). After sequencing and bioinformatics filtering, a total of 1,353,919 merged reads were obtained and 15 operational taxonomic units (OTUs) were identified as present with a non-negligible (>100) number of reads (for details, see Appendix S3 in Supplementary Information).

Negative controls

We obtained 243,340 sequence reads in negative controls (average: 5794 ± 19,066), with three OTUs being recorded with non-negligible quantities, corresponding to the following taxa: Mediterranean cypress (*Cupressus sempervirens* L.), European yew (*Taxus baccata* L.), and alder (*Alnus* sp. Mill.) (Figure 2). Specifically, Mediterranean cypress was recorded in three PCR replicates of extraction blanks (with 43,175; 19,166; and 37,019 number of reads, respectively), all corresponding to the same extraction batch (E3). Alder was recorded in two PCR replicates of one extraction blank (E3; 51,680 and 33,352 number of reads). These two taxa were absent from all PCR negative controls. European yew was recorded only in one PCR blank (55,264 number of reads) and was not recorded in extraction blanks, despite the latter being amplified in multiple PCR replicates.

Overall, we recorded contamination in 1/24 PCR negative controls and in 3/18 PCR replicates of negative control of extractions, all of them deriving from the same extraction batch. Notably, the three taxa recorded from negative controls were absent from filtered speleothem samples (with a 100-reads threshold for negligible presence; see Appendix S3). Excluding these three OTUs from the dataset, the average number of reads per blank decreased to 49, showing no other relevant contaminants.

Samples

We obtained 1,099,305 reads from the speleothem samples (average: 61,072 ± 17,484), corresponding to 12 OTUs detected with non-negligible number of reads (>100). The 12 OTUs were assigned to 11 plant taxa (each one reported at the lowest reliable taxonomic level): (1) Pinaceae, *Picea* sp. (spruce); (2) Pinaceae, *Pinus* sp. (pine); (3) *Abies* sp. (fir); (4) Cupressaceae, *Juniperus communis* (common juniper); (5) Urticaceae, *Parietaria* sp. (pellitory; two OTUs); (6) Poaceae (grasses); (7) Fagaceae, *Fagus sylvatica* (European beech); (8) Plantaginaceae (the plantain family); (9) Oleaceae (the olive and ash family); (10) Betulaceae, subfamily Coryloideae, which includes the following genera: *Corylus* (hazel), *Carpinus* (hornbeam), and *Ostrya* (hop-hornbeam); and (11) Caprifoliaceae, subfamily Valerianoideae, which in the European Alps includes genus *Valeriana*, *Valerianella*, and *Centranthus*. All these taxa were considered absent from negative controls according to the adopted cut-off threshold (<100 reads), except for *Fagus sylvatica*, which was recorded with up to 150 reads in four blank PCR replicates. According to our criteria, the recorded taxa should therefore be considered reliable endogenous presences (i.e., not deriving from external contaminations), with the above-mentioned exception of *Fagus sylvatica* (see further discussion below). For all of the 18 sequenced PCR replicates, at least one plant taxon was detected. A synthetic representation of the recorded taxa composition for each stalagmite section is reported in Figure 2, together with number of extraction and PCR replicates (for details on number of reads/taxa for each PCR replicate, see Table S1 in Supplementary Information).

Four taxa were recorded in the MB3 section (in four positive PCR replicates, all corresponding to the same extraction): (1) Betulaceae, subfamily Coryloideae; (2) Caprifoliaceae, subfamily Valerianoideae; (3) *Picea* sp.; and (4) *Fagus sylvatica*. Each taxon was recorded in only one PCR replicate, except for Caprifoliaceae, subfamily Valerianoideae (two PCR replicates) and always with high number of reads (>40,000), except for *Fagus sylvatica*, which was recorded with 125 reads. *Fagus sylvatica* was recorded with comparable abundance in four negative controls and therefore, based on DNA data, cannot be considered a reliable presence in the MB3 section.

Two taxa were recorded in the MB4 section (one positive PCR replicate): (1) Betulaceae, subfamily Coryloidea (~69,000 reads)

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

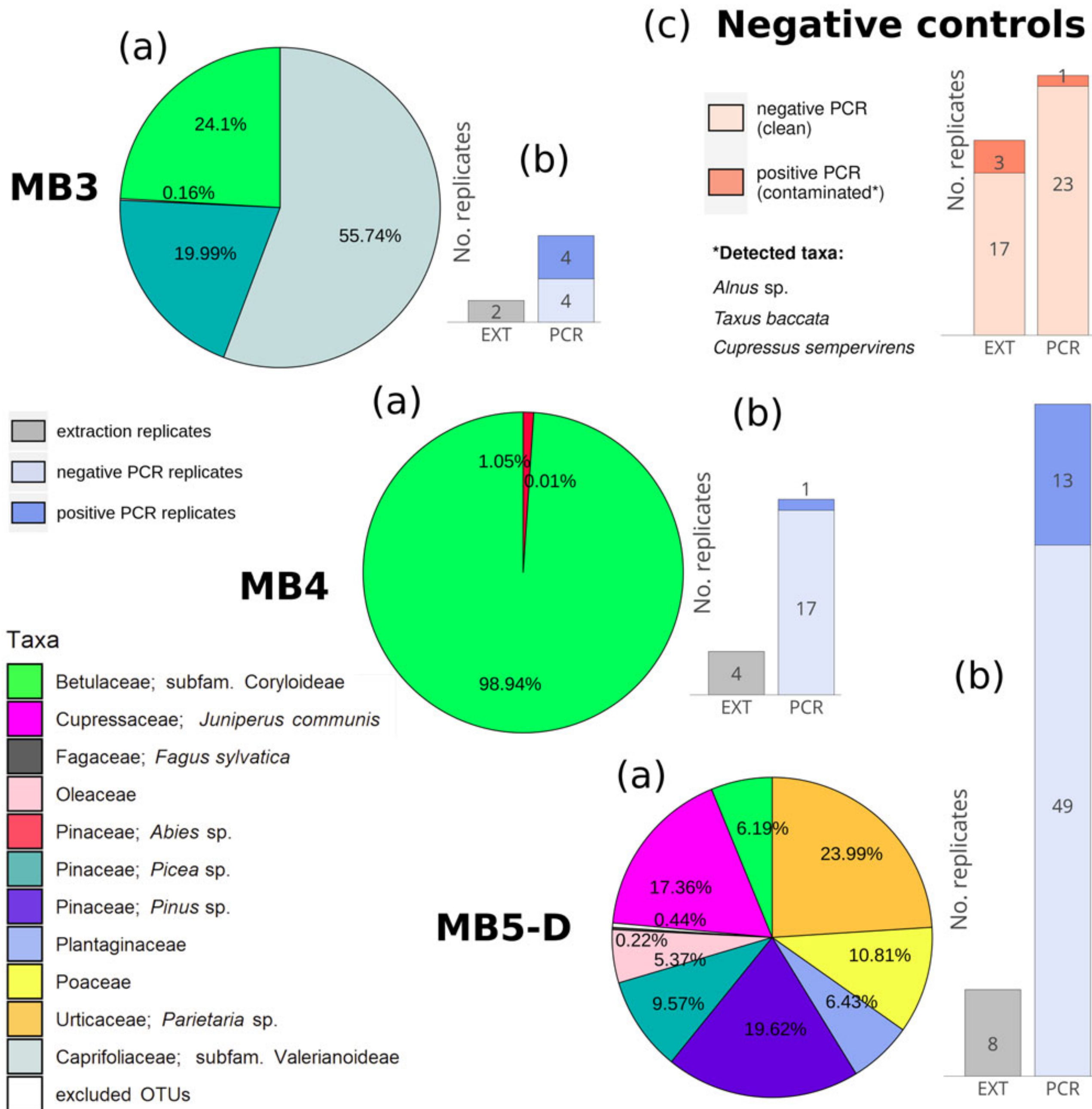

Figure 2. Results of sedaDNA metabarcoding. (**a**) Plant taxa recorded in the three speleothem sections from Milchbach Cave (Switzerland); only OTUs recorded with No. of reads >100 were considered (see Methods); (**b**) number of replicates and (**c**) number of negative controls analyzed (EXT and PCR refer to extraction and PCR blanks, respectively).

and (2) *Abies* sp. (733 reads). *Abies* sp., even if recorded with relatively low abundance, was however almost completely absent from all negative controls (max number of reads = 3).

The MB5-D section resulted in the highest rate of success both in terms of positive PCR replicates (13; representing 2/4 different extractions; i.e., powder samples) and plant diversity, with the following taxa: (1) *Pinus* sp.; (2) *Picea* sp.; (3) *Juniperus communis*; (4) *Parietaria* sp.; (5) Betulaceae, subfamily Coryloideae; (6) Poaceae; (7) Plantaginaceae; (8) Oleaceae; and (9) *Fagus sylvatica*. However, only 1–4 taxa were recorded in each of the PCR replicates. Pellitory (*Parietaria* sp.) was found in six PCR replicates, always with high number of reads (>10,000), except for one case (number of reads = 547). Common juniper (*Juniperus communis*) and pine (*Pinus* sp.) were recorded in five PCR replicates, always with number of reads >10,000, except for pine in two cases (but always >3000 reads). Spruce (*Picea* sp.) was recorded in two PCR replicates, always with number of reads >20,000. Each of the remaining taxa was recorded in only one PCR replicate, in all cases with a high number of reads (>40,000), except for *Fagus sylvatica* (European beech), which was recorded with 1825 reads. The abundance of *Fagus sylvatica* in this speleothem section is more than an order of magnitude higher than in negative controls, but its presence in blanks still prevents us from considering it an endogenous presence with high confidence.

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

DISCUSSION

Indirect authentication of sedaDNA sequences using palynomorphological data

To test the reliability of PCR-based metabarcoding of sedaDNA from speleothems, we focused on speleothems already known to contain high concentrations in biological particles. For two of the three 7600 yr-old Milchbach stalagmite (Luetscher et al., 2011) sections analyzed in this study, namely MB3 and MB5-D, palynomorphological-based data that were available from Festi et al. (2016) allowed us to make a direct comparison between the two methodologies (see Table S2 in Supplementary Information for pollen counts of the recorded plant taxa). This comparison has the potential to reveal whether or not there is a similarity between the speleothems' biological content determined from sedaDNA and that from pollen data. Considering the two sections together, all the taxa identified by target DNA barcode region (a small fragment of the chloroplast gene trnL, ∼100–185 bp long) have also been detected from the palynomorphological analysis, with the only exception of Caprifoliaceae, subfamily Valerianoideae (found only in two PCR replicates of the MB3 sample).

Focusing on MB5-D, the most pollen-rich section (3767 pollen counts versus 256 for MB3), four of the nine most abundant plant taxa highlighted by the palynomorphological analysis (number of pollen counts >100) corresponded to those retrieved by sedaDNA (Figure 3): (1) *Pinus* spp.; (2) *Corylus avellana* (based on the taxonomic level reached by the target DNA barcode region we identified the presence of Betulaceae, subfamily Corlyloidea, which includes *Corylus*); (3) Poaceae; and (4) *Fraxinus excelsior* (DNA-based taxonomic level: family Oleaceae, represented in the native Alpine flora only by the genera *Fraxinus* and *Ligustrum*).

In general, the palynological analysis resulted in detection of a higher number of taxa, compared with sedaDNA: 46 versus 9 in the MB5-D section. Many diverse and not mutually exclusive reasons for explaining this stark difference can be invoked. In morphological analyses, even a single pollen grain can be spotted and attributed to a specific taxon. In contrast, eDNA metabarcoding identifies taxa only if a discrete number of sequences belongs to a specific taxon. Furthermore, molecular investigations are limited by DNA degradation, while pollen walls are known to be one of the most resistant cellular portions (Brooks and Shaw, 1978). Pollen grains may thus be suitable for morphological analyses, even when DNA is not preserved. In addition, standard DNA extraction protocols for sedimentary samples are designed for processing a specific (rather low) amount of material per sample (in our case, 500 mg), which limits the possibility of using multiple primers and performing multiple PCR replicates, which in turn reduces the probability of detecting rare taxa. Ficetola et al. (2015) suggested that at least eight PCR replicates should be performed if detection probability is not high, but their conclusions were based on empirical studies targeting specific groups of organisms and using specific primers.

The problem of rare species detection in eDNA metabarcoding is complex and still debated. For example, Shirazi et al. (2021) found that, for soil samples and using generic primers (ITS1 and ITS2), even if performing a very high number of PCR replicates (24), low-abundance taxa were often unique to one or a few replicates. Indeed, one major flaw when using generic primers (such as for past vegetation reconstruction) is that the results might be skewed towards preferential amplification of certain taxa, while others, particularly rare species (Gonzalez et al., 2012) remain undetected (see Pedersen et al., 2015, and reference therein). In sedaDNA samples, a further complication may derive from the presence of enzymatic inhibitors in sediment extracts, as well as from their removal using reagents included in commercial kits (which can also remove short, damaged DNA fragments). The combination of these two drawbacks can severely affect the retrieval of low-abundance ancient DNA molecules (see Murchie et al., 2021, and reference therein).

Further insights may derive from a quantitative point of view. The palynomorphological analysis highlighted the MB5-D section as much richer than MB3 in terms of total pollen grains (number of pollen counts: 3767 and 256, respectively; MB5-D/MB3 ratio: 14.71). This difference in pollen content is reflected by the DNA-based results: considering the overall results of the different PCR replicates, MB5-D yielded many more sequence reads identified as plant taxa than MB3 (825,593 versus 199,961; MB5-D/MB3 ratio: 4.13). A correspondent quantitative difference emerged in the total number of detected plant taxa: the palynomorphological analysis found 46 taxa in MB5-D and 28 in MB3 (MB5-D/MB3 ratio: 1.64); similarly, our DNA-based results found nine taxa in MB5 and four in MB3 (MB5-D/MB3 ratio: 2.25). Here it is worth noting that the opportunistic experimental design of our study makes a fair comparison difficult for the DNA-based results of these two sections (i.e., different PCR replicates due to different availability of starting material; see Table 1).

In summary, the plant sedaDNA sequences quantitatively and qualitatively seem to reflect endogenous presences rather than potential external origins. None of the plant taxa we found in the speleothem samples was detected in the negative controls with a significant number of reads, despite the large numbers of extraction and PCR blanks included in our experiment. This can be viewed as rule-of-thumb evidence for excluding contamination from modern DNA as the most likely explanation of sedaDNA results.

Issues associated with contaminations from modern DNA and low PCR success rates

We adopted a stringent approach for contamination detection and removal. The rationale is straightforward: a totally contamination-free experimental setting is almost impossible to achieve, even under the most strict conditions (Weyrich et al., 2019), and therefore, if contamination is unavoidable, it has to be carefully monitored. In our experiments, we sequenced a considerably high number of negative controls (42), only four of which were found to be contaminated with significant quantities of plant DNA. The contaminating taxa (cypress, yew, and alder) were not detected in any of the PCR replicates from the speleothems above the adopted 100 reads cut-off threshold. The results for *Fagus sylvatica* lie in a "gray zone": adopting a cautionary approach, its detection in four blanks, even in low quantities (<150 reads), prevents us assigning 100% credibility to its presence in the ancient samples. Beyond this single exception, results from speleothem samples and negative controls are not overlapping, indicating that the plant DNA found in the examined speleothem sections is very unlikely derived from contaminations.

Nevertheless, a more critical evaluation of the plant species recorded in speleothems by sedaDNA highlights two points: (1) all of them can be found commonly in modern vegetational assemblages of Alpine areas, including the surroundings of the

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

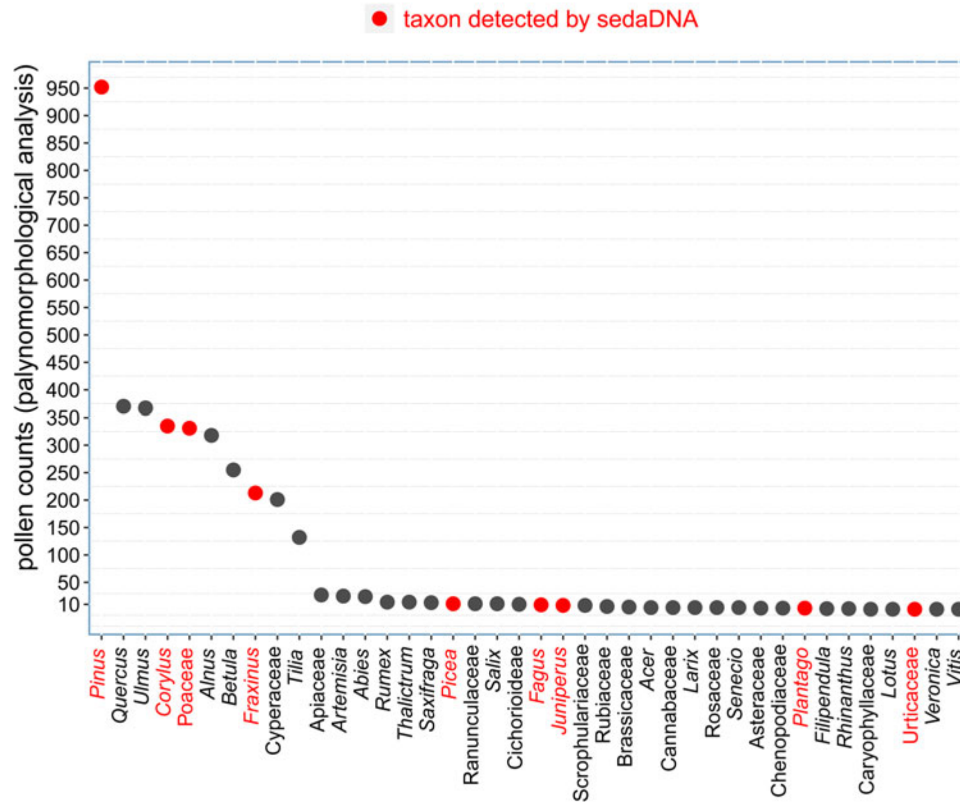

Figure 3. Speleothem section MB5-D: palynomorphological versus sedaDNA analysis. Plant taxa are ordered by pollen abundance (No. of counts); taxa that also were detected by sedaDNA are highlighted in red. Pollen counts are derived from Festi et al. (2016). Note that, due to the different taxonomic resolutions, we adopted the following approximated correspondences: *Corylus* (pollen data) = Betulaceae, subfamily Coryloidea (sedaDNA); *Fraxinus* (pollen) = Oleaceae (sedaDNA). See text for further explanations.

analytical laboratory; and (2) several of them are characterized by high pollen production and/or persistence (e.g. *Pinus*, *Corylus*, *Fraxinus*, Poaceae, etc.).

We recorded PCR success rates in sample replicates (18/88; 20.5%; see Table 1) that are not much higher than those in the negative controls (extraction and PCR blanks: 4/44; 9.1%). These two critical aspects leave our results open to question. It should be noted, however, that low rates of success and high heterogeneity of results among PCR replicates of the same samples (from zero to high number of reads, for a considered taxon) are not unusual in many sedaDNA studies (Capo et al., 2021).

Lastly, when comparing sedaDNA-based data with those from morphological identification, our results again are not unexpected. In this study, the taxa identified by sedaDNA (11) are fewer than those (46) identified by morphological pollen analyses. This is consistent with the outcome of several other sedaDNA studies that were performed using the same barcode gene, trnL (Jørgensen et al., 2012; Parducci et al., 2013, 2015; Pedersen et al., 2013). For example, Parducci et al. (2015) identified a total of 32 pollen taxa in peat samples dating back at ca. 8500–5900 yr BP, and thus spanning the age of our speleothems, whereas sedaDNA identified only nine taxa. This ratio of 28% is comparable to the 24% ratio measured in our study. It should be noted, however, that the performance of sedaDNA compared with pollen data may vary according to the nature and conditions of the sediment/site and the taxonomic group: opposite examples also exist, with more taxa detected by molecular than morphological analysis, particularly for lake sediments and cold environments (e.g., Clarke et al., 2020; see Capo et al., 2021, for a more detailed discussion). In summary, in the context of sedaDNA and with specific regards to "difficult" mineral deposits, low PCR success rates and limited overlap with results deriving from morphological identification cannot be considered a unique peculiarity of our study.

Limitations of PCR-based amplicon sequencing for speleothem sedaDNA and future perspectives

To our knowledge, only three studies have been published so far on aDNA extracted from speleothems. Dhami et al. (2018) used an amplicon sequencing approach, but the authors only focused on the identification of microbial communities associated with biomineralization processes. Zepeda Mendoza et al. (2016) and Stahlschmidt et al. (2019) presented the first attempts to exploit speleothems as past biodiversity archives. Both studies adopted a shotgun metagenomics approach. With only six identified genera (including animals and plants), Stahlschmidt et al. (2019) admitted that their samples were likely characterized by low DNA preservation. Zepeda Mendoza et al. (2016) found a large spectrum of taxa, but concluded that the analyzed speleothems could not be used confidently as paleoarchives due to the high degree of DNA damage (perhaps induced by fungal activity) that negatively affected sequence quality and therefore the taxonomic resolution (for plants, results are limited to the class level). The authors therefore suggested the use of target amplicon metabarcoding to provide more reliable and accurate results. In

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

the present study, we followed this direction, but, despite strictly adopting the best practices recommended for sedaDNA (Capo et al., 2021), some uncertainties remain.

Some drawbacks are typical of the PCR-based metabarcoding methodology, such as (1) the unavoidable trade-off between "universality" (i.e., extensive taxonomic coverage) and taxonomic resolution in the choice of barcode markers (Hollingsworth et al., 2011); (2) preferential amplification of certain taxa over other ones, which leads to taxonomic bias (Elbrecht and Leese, 2015; Zinger et al., 2019); and (3) the extremely high sensitivity (due to PCR amplification) to contamination by DNA from external sources. As demonstrated in previous work (e.g., Weyrich et al., 2019), contaminating DNA molecules are ubiquitous, even in ultra-clean labs and commercial reagents. The risk of detecting contamination from modern DNA sources is particularly severe when processing ancient eDNA samples (Ficetola et al., 2015, 2016), which generally are characterized by low concentrations of damaged DNA and require extreme PCR amplification protocols, such as a high number of cycles: under these conditions, even a few modern intact DNA molecules can be amplified preferentially over a larger amount of ancient damaged molecules (Zinger et al., 2019). To partially overcome these drawbacks, we (1) selected a short region of a barcode marker (trnL), which is known for its "universal" detection power (Taberlet et al., 2007); and (2) extensively monitored contamination, from DNA extraction to PCR and post-metabarcoding analysis. Nevertheless, the outcome was still not optimal.

We think that a significant improvement can be offered by alternative approaches that are not directly based on PCR such as hybridization capture, also called target capture (Shokralla et al., 2016; Wilcox et al., 2018). In hybridization capture, a library derived from eDNA is hybridized with specific probes. After discarding the libraries with no sign of hybridization, the resulting libraries are massively sequenced. Pioneer attempts to use targeted capture for ancient environmental DNA, such as in Murchie et al. (2021), which aimed at reconstructing plant and animal communities from permafrost samples dating to the Pleistocene-Holocene transition, yielded promising results with a generally better performance compared with the standard metabarcoding approach. We strongly encourage this direction, also for speleothem and other ancient sedimentary samples. Target capture is particularly well suited for sedaDNA for the following main reasons: (1) it is PCR-free, thus avoiding limitations such as high sensitivity to contaminations from modern DNA, preferential amplification, and heterogeneity among PCR replicates; (2) it offers higher taxonomic resolution due to the possibility to design probes with high sequence homology to the target taxa; and (3) it preserves the typical damage patterns found at or near the ends of aDNA molecules (such as cytosine deamination; Armbrecht et al., 2021). The latter point is of crucial importance, providing the possibility to authenticate ancient DNA by means of specific bioinformatic tools (see for example MapDamage; Jónsson et al., 2013) and to unambiguously identify contaminations from modern DNA. This is not possible for amplicon sequencing given that, during PCR amplification, primers either bind to or exclude the ends of the template molecules where damage is present, leading to the loss of the damage signal.

The aforementioned drawbacks of PCR-based metabarcoding prompt us to suggest that this approach should be avoided when analyzing DNA retrieved from ancient sediments such as speleothems. It can be argued that our case study, with the retrieval from speleothems of a limited number of plant taxa that commonly also are found in modern environments, may represent a particularly unfortunate example. If some particular plant species, currently absent in Alpine latitudes and altitudes, but which might have occurred in (warmer) past paleoenvironments, had been found in the samples, one probably would have been tempted to consider these results reliable. However, if a method can be considered valid only ex post, that is after checking the results, this method is severely biased. We conclude that speleothems could hold promise as biological archives to be studied with molecular methods, provided different approaches than those based on direct PCR amplification of the extracted DNA are used.

Acknowledgments. The authors thank the anonymous reviewers for their valuable and constructive comments on an earlier version of the manuscript.

Supplementary Material. The Supplementary Material for this article can be found at https://doi.org/10.1017/qua.2022.46.

REFERENCES

Aeschimann, D., Lauber, K., Moser, D.M., Theurillat, J.P., 2004. *Flora Alpina: Atlas des 4500 Plantes Vasculaires des Alpes*. Belin, Paris.

Albanese, D., Fontana, P., De Filippo, C., Cavalieri, D., Donati, C., 2015. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Scientific Reports* **5**, 9743. https://doi.org/10.1038/srep09743.

Armbrecht, L., Hallegraeff, G., Bolch, C.J.S., Woodward, C., Cooper, A., 2021. Hybridisation capture allows DNA damage analysis of ancient marine eukaryotes. *Scientific Reports* **11**, 3220. https://doi.org/10.1038/s41598-021-82578-6.

Bastin, B., 1979. L'analyse pollinique des stalagmites: une nouvelle possibilité d'approche des fluctuations climatiques du Quaternaire. *Annales de la Société Géologique de Belgique* **101**, 13–19.

Bell, K.L., de Vere, N., Keller, A., Richardson, R.T., Gous, A., Burgess, K.S., Brosi, B.J., 2016. Pollen DNA barcoding: current applications and future prospects. *Genome* **59**, 629–640.

Blyth, A.J., Frisia, S., 2008. Molecular evidence for bacterial mediation of calcite formation in cold high-altitude caves. *Geomicrobiology Journal* **25**, 101–111.

Brooks, J., Shaw, G., 1978. Sporopollenin: a review of its chemistry, palaeochemistry and geochemistry. *Grana* **17**, 91–97.

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P.D., Vuillemin, A., Ariztegui, D., *et al.*, 2021. Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: Overview and recommendations. *Quaternary* **4**, 6. https://doi.org/10.3390/quat4010006.

Cheng, H., Edwards, R.L., Shen, C.C., Polyak, V.J., Asmerom, Y., Woodhead, J., Hellstrom, J., *et al.*, 2013. Improvements in ^{230}Th dating, ^{230}Th and ^{234}U half-life values, and U-Th isotopic measurements by multi-collector inductively coupled plasma mass spectrometry. *Earth and Planetary Science Letters* **371**, 82–91.

Clarke, C.L., Alsos, I.G., Edwards, M.E., Paus, A., Gielly, L., Haflidason, H., Mangerud, J., *et al.*, 2020. A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. *Quaternary Science Reviews* **247**, 106564. https://doi.org/10.1016/j.quascirev.2020.106564.

Dhami, N.K., Mukherjee, A., Watkin, E.L., 2018. Microbial diversity and mineralogical-mechanical properties of calcitic cave speleothems in natural and in vitro biomineralization conditions. *Frontiers in Microbiology* **9**, 40. https://doi.org/10.3389/fmicb.2018.00040.

Dredge, J., Fairchild, I.J., Harrison, R.M., Fernandez-Cortes, A., Sanchez-Moral, S., Jurado, V., Gunn, J., *et al.*, 2013. Cave aerosols: distribution and contribution to speleothem geochemistry. *Quaternary Science Reviews* **63**, 23–41.

Edgar, R.C., 2016. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*, 074161. https://doi.org/10.1101/074161.

Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press

with an innovative metabarcoding protocol. *PloS ONE* **10**, e0130324. https://doi.org/10.1371/journal.pone.0130324.

Faegri, K., Kaland, P.E., Krzywinski, K., 1989. *Textbook of Pollen Analysis (IV)*. John Wiley & Sons Ltd., Chichester, UK.

Fairchild, I.J., Baker, A., 2012. *Speleothem Science: from Process to Past Environments*. Wiley-Blackwell, Chichester, UK.

Festi, D., Hoffmann, D.L., Luetscher, M., 2016. Pollen from accurately dated speleothems supports alpine glacier low-stands during the Early Holocene. *Quaternary Research* **86**, 45–53.

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

Ficetola, G.F., Taberlet, P., Coissac, E., 2016. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* **16**, 604–607.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.J., Griggo, C., Gielly, L., Domaizon, I., *et al.*, 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications* **5**, 3211. https://doi.org/10.1038/ncomms4211.

Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., Mira, A., 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PloS ONE* **7**, e29973. https://doi.org/10.1371/journal.pone.0029973.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. *PloS ONE* **6**, e19254. https://doi.org/10.1371/journal.pone.0019254.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684

Jørgensen, T., Haile, J., Möller, P.E.R., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., *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.

Lechleitner, F.A., Dittmar, T., Baldini, J.U., Prufer, K.M., Eglinton, T.I., 2017. Molecular signatures of dissolved organic matter in a tropical karst system. *Organic Geochemistry* **113**, 141–149.

Leontidou, K., Vernesi, C., De Groeve, J., Cristofolini, F., Vokou, D., Cristofori, A., 2018. DNA metabarcoding of airborne pollen: new protocols for improved taxonomic identification of environmental samples. *Aerobiologia* **34**, 63–74.

Leontidou, K., Vokou, D., Sandionigi, A., Bruno, A., Lazarina, M., De Groeve, J., Li, M., *et al.*, 2021. Plant biodiversity assessment through pollen DNA metabarcoding in Natura 2000 habitats (Italian Alps). *Scientific Reports* **11**, 18226. https://doi.org/10.1038/s41598-021-97619-3.

Li, F.-S., Phyo, P., Jacobowitz, J., Hong, M., Weng, J.K., 2019. The molecular structure of plant sporopollenin. *Nature Plants* **5**, 41–46.

Luetscher, M., Hoffmann, D., Frisia, S., Spötl, C., 2011. Holocene glacier history from alpine speleothems, Milchbach Cave, Switzerland. *Earth and Planetary Science Letters* **302**, 95–106.

Luetscher M., Moseley G., Festi D., Hof F., Edwards R.L., Spötl C., 2021. A last interglacial speleothem record from the Sieben Hengste cave system (Switzerland): implications for alpine paleovegetation. *Quaternary Science Reviews,* **262**, 106974. https://doi.org/10.1016/j.quascirev.2021.106974.

Matley, K.A., Sniderman, J.K., Drinnan, A.N., Hellstrom, J.C. 2020. Late-Holocene environmental change on the Nullarbor Plain, southwest Australia, based on speleothem pollen records. *The Holocene* **30**, 672–681.

Murchie, T.J., Kuch, M., Duggan, A.T., Ledger, M.L., Roche, K., Klunk, J., Karpinski, E., *et al.*, 2021. Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quaternary Research* **99**, 305–328.

Parducci, L., Matetovici, I., Fontana, S. L., Bennett, K. D., Suyama, Y., Haile, J., Kjaer K.H., Larsen, N.K., Drouzas A.D., Willerslev, E., 2013. Molecular-and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology* **22**, 3511–3524.

Parducci, L., Väliranta, M., Salonen, J.S., Ronkainen, T., Matetovici, I., Fontana, S.L., Eskola, T., Sarala, P., Suyama, Y., 2015. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20130382. https://doi.org/10.1098/rstb.2013.0382.

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

Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Sarkissian, C.D., Haile, J., Hellstrom, M., Spens J., *et al.*, 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20130383. https://doi.org/10.1098/rstb.2013.0383.

Richardson, R.T., Lin, C.H., Sponsler, D.B., Quijia, J.O., Goodell, K., Johnson, R.M., 2015. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. *Applications in Plant Sciences* **3**, 1400066. https://doi.org/10.3732/apps.1400066.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584. https://doi.org/10.7717/peerj.2584.

Shirazi, S., Meyer, R.S., Shapiro, B., 2021. Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution* **11**, 15766–15779.

Shokralla, S., Gibson, J.F., King, I., Baird, D.J., Janzen, D.H., Hallwachs, W., Hajibabaei, M., 2016. Environmental DNA barcode sequence capture: targeted, PCR-free sequence capture for biodiversity analysis from bulk environmental samples. *BioRxiv* doi.org/10.1101/087437.

Stahlschmidt, M.C., Collin, T.C., Fernandes, D.M., Bar-Oz, G., Belfer-Cohen, A., Gao, Z., Jakeli, N., *et al.*, 2019. Ancient mammalian and plant DNA from Late Quaternary stalagmite layers at Solkota Cave, Georgia. *Scientific Reports* **9**, 6628. https://doi.org/10.1038/s41598-019-43147-0.

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 trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**, e14. https://doi.org/10.1093/nar/gkl938.

Taberlet, P., Bonin, A., Coissac, E., Zinger, L., 2018. *Environmental DNA: for Biodiversity Research and Monitoring*. Oxford University Press, Oxford, UK.

ter Schure, A.T.M., Bajard, M., Loftsgarden, K., Høeg, H.I., Ballo, E., Bakke, J., Støren, E.W.N., *et al.*, 2021. Anthropogenic and environmental drivers of vegetation change in southeastern Norway during the Holocene. *Quaternary Science Reviews* **270**, 107175. https://doi.org/10.1016/j.quascirev.2021.107175.

Varotto, C., Pindo, M., Bertoni, E., Casarotto, C., Camin, F., Girardi, M., Maggi, V., Cristofori, A., 2021. A pilot study of eDNA metabarcoding to estimate plant biodiversity by an alpine glacier core (Adamello glacier, North Italy). *Scientific Reports* **11**, 1208. https://doi.org/10.1038/s41598-020-79738-5.

Weyrich, L.S., Farrer, A.G., Eisenhofer, R., Arriola, L.A., Young, J., Selway, C.A., Handsley-Davis, M., Adler, C.J., Breen, J., Cooper, A., 2019. Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources* **19**, 982–996.

Wilcox, T.M., Zarn, K.E., Piggott, M.P., Young, M.K., McKelvey, K.S., Schwartz, M.K., 2018. Capture enrichment of aquatic environmental DNA: a first proof of concept. *Molecular Ecology Resources* **18**, 1392–1401.

Willerslev, E., Hansen, A. J., Poinar, H.N., 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution* **19**, 141–147.

Zepeda Mendoza, M.L., Lundberg, J., Ivarsson, M., Campos, P., Nylander, J.A., Sallstedt, T., Dalen, L. 2016. Metagenomic analysis from the interior of a speleothem in Tjuv-Ante's Cave, Northern Sweden. *PloS ONE* **11**, e0151577. https://doi.org/10.1371/journal.pone.0151577.

Zinger, L., Bonin, A., Alsos, I.G., Bálint, M., Bik, H., Boyer, F., Chariton, A.A., *et al.*, 2019. DNA metabarcoding—need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28**, 1857–1862.

https://doi.org/10.1017/qua.2022.46 Published online by Cambridge University Press