Effects of soil preservation for biodiversity monitoring using environmental DNA

Alessia Guerrieri[1], Aurélie Bonin[1], Tamara Münkemüller [2], Ludovic Gielly[2], Wilfried Thuiller[2], and Gentile Francesco Ficetola[1]

[1]University of Milan
[2]Universite Grenoble Alpes

May 6, 2020

Abstract

Environmental DNA metabarcoding is becoming a key tool for biodiversity monitoring over large geographical or taxonomic scales and for elusive taxa like soil organisms. Increasing sample sizes and interest in remote or extreme areas often require the preservation of soil samples and thus deviations from optimal standardized protocols. However, we still ignore the impact of different methods of soil sample preservation on the results of metabarcoding studies and there is no guidelines for best practices so far. Here, we assessed the impact of four methods of soil sample preservation commonly used in metabarcoding studies (preservation at room temperature for 6h, preservation at 4°C for three days, desiccation immediately after sampling and preservation for 21 days, and desiccation after 6h at room temperature and preservation for 21 days). For each preservation method, we benchmarked resulting estimates of taxon diversity and community composition of three different taxonomic groups (bacteria, fungi and eukaryotes) in three different habitats (forest, river bank and grassland) against results obtained under optimal conditions (i.e. extraction of eDNA right after sampling). Overall, the different preservation methods only marginally impaired results and only under certain conditions. When rare taxa were considered, we detected small but significant changes in MOTU richness of bacteria, fungi and eukaryotes across treatments, while the exclusion of rare taxa led to robust results across preservation methods. The differences in community structure among habitats were evident for all treatments, and the communities retrieved using the different preservation conditions were extremely similar. We propose guidelines on the selection of the optimal soil sample preservation conditions for metabarcoding studies, depending on the practical constraints, costs and ultimate research goals.

INTRODUCTION

Environmental DNA (eDNA) can be defined as the mixture of complex, often degraded, DNA that organisms leave behind in their environment (i.e. soil, water, sediments, etc.). By studying short, taxonomically-informative DNA fragments obtained from eDNA samples, it is possible to identify the associated taxa and therefore to survey biodiversity. Coined as "eDNA metabarcoding", this approach has revolutionized several branches of ecology and environmental sciences during the last decade, by providing relatively quick, non-invasive, and standardized assessments of present or past biodiversity of animals, plants and microorganisms (Taberlet, Bonin, Zinger, & Coissac, 2018). Metabarcoding is particularly valuable for monitoring biodiversity over large geographical or taxonomic scales (De Vargas et al., 2015; Delgado-Baquerizo et al., 2018; Zinger et al., 2019b). Furthermore, it gives access to biodiversity components that are elusive to conventional survey methods. For instance, it allows the rapid assessment of microbial soil biodiversity, which is extremely complex, time-consuming and imperfect when using direct observations, culturing techniques or microscopy (Giovannoni, Britschgi, Moyer, & Field, 1990; Ward, Weller, & Bateson, 1990).

Metabarcoding relies on a succession of several steps: 1) sampling; 2) preservation of the collected material until lab processing; 3) DNA extraction; 4) PCR amplification of a selected genomic region; 5) high-throughput

sequencing of amplicons; and 6) analysis of sequences using bioinformatics and statistical tools (Zinger, Bonin, et al., 2019a). Each step is critical to obtain robust taxonomic inventories and diversity estimates, and an increasing number of studies have assessed how methodological choices across the different steps could influence the conclusions of a study (Calderón-Sanou, Munkemuller, Boyer, Zinger, & Thuiller, 2020; Cantera et al., 2019; Chen & Ficetola, 2020; Nichols et al., 2018; Taberlet et al., 2018). So far, despite this growing body of literature, little attention has been accorded to the effect of different preservation conditions of the collected environmental material before lab processing (i.e. step 2). We thus know neither under which conditions the collected material should be stored, nor how long it can be stored to avoid biases in taxonomic inventories.

While more is known for water samples (see e.g. Kumar, Eble, & Gaither, 2020; Majaneva et al., 2018), in the case of soil biodiversity research, methodological analyses on the effects of sample preservation are largely dismissed probably because the majority of metabarcoding studies have so far been performed in temperate areas where access to lab facilities is often easy (Hoffmann, Schubert, & Calvignac-Spencer, 2016; Huerlimann et al., 2020). In such cases, sample preservation is sometimes not necessary at all, or at least not over long periods of time. However, one great promise of metabarcoding is its potential for providing biodiversity data for remote areas, where biodiversity monitoring is essential but difficult. When sampling in remote or inaccessible areas (e.g. tropical and arctic areas; mountain chains), samples are rarely collected nearby lab facilities and an immediate*in situ* DNA extraction is generally not possible due to logistic constraints (but see Zinger, Taberlet, et al., 2019b for a notable exception). More generally, with the ever-increasing number of samples analyzed during a typical metabarcoding study, sample preservation is more and more indispensable, and the time lag between sample collection and subsequent molecular processing makes it particularly relevant to understand the impact of sample preservation, and to identify preservation strategies that do not bias the conclusions of studies.

In an optimal metabarcoding study, communities recovered from preserved samples should ideally be identical to those retrieved if samples had been processed immediately after sampling. However, inappropriate preservation conditions can cause both DNA degradation and the proliferation of certain taxonomic groups, with respect to others, before DNA extraction (Cardona et al., 2012; Orchard, Standish, Nicol, Dickie, & Ryan, 2017). This can in turn affect taxa detection and also the relative contributions of different taxonomic groups to the overall biodiversity. A recent review suggested that the majority of eDNA metabarcoding studies do not provide accurate information about sample treatment before processing (Dickie et al., 2018). Almost half of the studies do not report how samples were stored and conserved, and 30% of them store samples at 0-4degC, and thus at a temperature where many bacteria and fungi continue to be active and potentially affecting the whole sample. About 15% of the studies stored samples in a range of 5-35degC, which is considered as a poor practice, and only 10 % stored them below 0degC (Dickie et al., 2018).

So far, the consequences of preservation practices and the resulting deviations from immediate processing and analyses have rarely been studied quantitatively. Yet, Lauber, Zhou, Gordon, Knight, & Fierer (2010) tested the effect of storing samples from soil, human gut and skin at different temperatures and did not detect any significant effect on bacterial communities, while Orchard et al. (2017) found that storage time and temperature can affect colonization by arbuscular mycorrhizal fungi, with subsequent impacts on the reconstruction of communities. Differences between these studies may be due to their different protocols. However, they also focused on different taxonomic groups, which may react differently to storage period and temperature. Other studies use desiccation for conserving plant and animal tissues for subsequent genomic studies (e.g. Chase & Hills, 1991), which has proven efficient and convenient. Although not widely used for metabarcoding samples, desiccation is another attractive option, and has a potential for being largely implemented in soil sample preservation. A clear understanding of the effect of different preservation methods, especially across various groups of taxa, is thus pivotal for a robust application of eDNA metabarcoding to biodiversity monitoring in general, and that of remote areas in particular.

Here, using eDNA metabarcoding of different taxonomic groups in soil systems, we tested: (i) how preservation methods influence overall richness estimates and what the role of rarely observed taxa is; (ii) how

preservation methods influence identified community structure and its turn-over between different habitats; and (iii) what the best practices are under limited laboratory access. More specifically, we first selected three soil preservation methods (room temperature, 4degC, desiccation by addition of silica gel) because they are commonly used in the literature (room temperature and 4degC) or because they are easy to implement in the field (desiccation and room temperature). Then, we assessed the impact of these preservation methods applied to different durations in order to mimick logistic constraints, and compared the communities obtained with those observed in ideal conditions, i.e. when eDNA is extracted immediately after sampling (within less than one hour). We examined bacterial, fungal and eukaryotic communities to cover a broad taxonomic range, since different taxa can be differentially affected by sample preservation conditions (Cardona et al., 2012; Orchard et al., 2017).

MATERIALS AND METHODS

Soil preservation and experimental treatments

In April 2019, we collected soil samples from three habitats: a grassland (N 45.194deg E 5.776deg), a broadleaved forest (N 45.196deg E 5.774deg), and a vegetated river bank (N 45.195deg E 5.780deg). To allow DNA extraction immediately after sampling, all sites were within 400 m from the Laboratoire d'Ecologie Alpine (LECA) in Grenoble, France. We established five plots within each habitat and the minimum distance between nearby plots was about 20 m. Within each plot, we collected four soil samples (with a minimum distance of one meter) at a depth of 0–20 cm and then pooled them together, for a total of five pooled samples per habitat (approx. 200 g each pooled sample). Soil litter was not included in the samples. Pooled samples (15 in total) were homogenized; subsequently, from each of them we took five subsamples of 15 g of soil (total: 75 subsamples; Fig. 1).

The five soil subsamples of each pooled sample were subjected to five different treatments: 1) eDNA was extracted immediately after sampling (within 1 h; treatment hereafter referred to as "control"); 2) samples were preserved at room temperature (21-23degC) and eDNA was extracted 6 h after sampling; 3) samples were inserted in sterile 50-mL falcon tubes and preserved at 4degC and eDNA was extracted three days after sampling; 4) samples were inserted in hermetic, sterile boxes with 20 g of silica gel immediately after sampling, then stored at room temperature, and eDNA was extracted 21 days after sampling; 5) samples were inserted in hermetic, sterile boxes with 20 g of silica gel 6h after sampling, then stored at room temperature, and eDNA was extracted 21 days after sampling.

Molecular analyses

For all sample treatments, eDNA extraction was performed in a dedicated room using the NucleoSpin(r) Soil Mini Kit (Macherey-Nagel, Germany), after a preliminary step where 15 g of soil were mixed with 20 ml of phosphate buffer for 15 minutes as described in (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012); and with a final elution in 150 µl. We also included one extraction negative control per treatment.

Environmental DNA of bacteria, fungi and eukaryotes was amplified using primers designed for markers Bact02 (Taberlet et al., 2018), Fung02 (Epp et al., 2012; Taberlet et al., 2018) and Euka02 (Guardiola et al., 2015), respectively. Bact02 and Fung02 relate to fragments of about 220-250 bp, while Euka02 generally relates to fragments <150 bp. The three markers are well suited for metabarcoding analyses, as all have a very low number of mismatches in the priming region across target organisms, and the relatively short length of amplified fragments allows their use with potentially degraded DNA (Taberlet et al., 2018). To allow bioinformatic discrimination of PCR replicates after sequencing, eight-nucleotide long tags were added on the 5' end of both forward and reverse primers, so that each PCR replicate was represented by a unique combination of forward and reverse tags. Tags had at least five nucleotide differences among them (Coissac, 2012). Samples were randomized on a 96-well plate, along with the five extraction controls, eight bioinformatic blanks, six PCR negative controls and two PCR positive controls. PCR positive controls were included to check for potential cross-contaminations and to monitor amplification and sequencing performances. The positive control was a 1:10 dilution of the ZymoBIOMICS Microbial Community DNA Standard II (Zymo Research, USA) constituted of genomic DNA of eight bacterial and two fungal strains (i.e., *Pseudomonas*

aeruginosa, Escherichia coli, Salmonella enterica, Lactobacillus fermentum, Enterococcus faecalis, Staphylococcus aureus, Listeria monocytogenes, Bacillus subtilis, Saccharomyces cerevisiae, Cryptococcus neoformans) at known concentrations.

In order to avoid over-amplification of template DNA and to limit chimera formation, we determined the optimal number of amplification cycles and DNA extract dilution using qPCR. The qPCR assay was conducted on 48 randomly selected samples, using 1 µl of 1:1000 diluted SYBR® Green I nucleic acid gel stain (Invitrogen, USA), with a real-time PCR thermal cycler set to standard mode. qPCR was performed for both 1:10 diluted and undiluted template eDNA.

For Bact02 and Fung02, PCR reactions were performed on 1:10 diluted template DNA, using 32 and 44 cycles respectively. For Euka02, we performed 34 cycles on undiluted DNA. All PCR reactions consisted of 10 µl of AmpliTaq Gold 360 Master Mix 2X (Applied Biosystems, Foster City, CA, USA), 2 µl of primers mix at initial concentration of 5 µM of each primer, 0.16 µl of Bovine Serum Albumin (corresponding to 3.2 µg; Roche Diagnostic, Basel, Switzerland) and 2 µl of DNA extract, for a final volume of 20 µl. The PCR profiles had an initial step of 10 min at 95°C, followed by several cycles of a 30 s denaturation at 94°C, a 30 s annealing at 53°C (Bact02), 56°C (Fung02) or 45°C (Euka02), and a 90 s elongation for Bact02 and Fung02, or a 60 s elongation for Euka02 at 72°C, followed by a final elongation at 72°C for 7 minutes. The amplification was performed in 384-well plates, with four replicates for each sample. After amplification, PCR products of the same marker were pooled together in equal volumes and a 5-µl aliquot of the pooled amplicons was visualized by high-resolution capillary electrophoresis (QIAxcel Advanced System, QIAGEN, GERMANY) to verify the expected fragments length and to monitor primer dimers. Pooled amplicons were purified using the MinElute PCR Purification Kit (QIAGEN, GERMANY) following the manufacturer's protocol. Six subsamples of the pool of amplicons were purified separately for each marker, and then combined again before being sent for library preparation and sequencing to Fasteris (SA, Geneva, Switzerland). One library was prepared per marker using the MetaFast protocol (Taberlet et al., 2018) and then sequenced using the MiSeq (Fung02 and Bact02) or HiSeq 2500 (Euka02) platforms (Illumina, San Diego, CA, USA) with a paired-end approach (2 × 250 bp for Fung02 and Bact02, and 2 × 150 bp for Euka02).

Bioinformatic treatment

The bioinformatic treatment of sequence data was performed using the OBITools software suite (Boyer et al., 2016). First, forward and reverse reads were assembled using the *illuminapairedend* program, keeping only sequences with an alignment score higher than 40. Aligned sequences were assigned to the corresponding PCR replicate using the program*ngsfilter* , by allowing two and zero mismatches on primers and tags, respectively. After sequence dereplication using *obiuniq* , bad-quality sequences (i.e. containing "N"), sequences whose length fell outside the expected size interval (below 45 bp for Bact02, below 68 bp Fung02 and below 36 bp for Euka02) and singletons were filtered out. The *obiclean* program was run to detect potential PCR or sequencing errors with the -r option set at 0.5: in a PCR reaction, sequences are tagged as "heads" when they are at least twice as abundant as other related sequences differing by one base. Only the sequences tagged as "heads" in at least one PCR were kept.

Taxonomic assignment was conducted using the *ecotag* program based on a reference database constructed from EMBL (version 136) by running the *ecoPCR* program (Ficetola et al., 2010). More specifically,*ecoPCR* carried out an *in silico* PCR with the primer pair used for the experiment and allowing three mismatches per primer. The obtained reference databases were further curated by keeping only sequences assigned at the species, genus and family levels.

Further data filtering was performed in R version 3.6.1 (R Core Team, 2018) to remove spurious sequences that can bias ecological conclusions drawn from DNA metabarcoding data (Calderón-Sanou et al., 2020). More specifically, we discarded from our dataset MOTUs with a best identity <85% (Fung02, Bact02) or <80% (Euka02), observed less than five times overall or in more than one extraction or PCR negative control (Zinger, Bonin, et al., 2019a). Furthermore, we removed all MOTUs that were detected in less than two PCR replicates of the same sample, as they often represent false positives (Ficetola et al., 2015).

Statistical analyses

For all taxonomic groups, we used Generalized Linear Mixed Models (GLMMs) to test if the different treatment lead to differences in the observed MOTU richness. In GLMMs, the number of MOTUs per sample was calculated and used as a dependent effect, the five treatments were used as predictors, and sample identity was used as a random factor. The model was performed with the generalized poisson distribution error using the R package *glmmTMB* (Brooks et al., 2017), in order to take into account overdispersion (Consul & Famoye, 1992). If GLMM detected significant differences among treatments, we used treatment contrasts to test if each treatment led to communities significantly different from those unraveled by the "control" condition. Treatment contrasts are standard non-orthogonal contrasts, in which each category (treatment) is compared to a user-defined reference category, and are appropriate to compare multiple treatments against one single control category (in this case, immediate extraction; (Field, Miles, & Field, 2015). The uncorrected number of MOTUs tends to overestimate the actual taxonomic richness (Calderon-Sanou et al., 2020). Therefore, we repeated this analysis twice: considering all the observed MOTUs, and excluding the rare MOTUs (i.e. MOTUs with frequency $< 1\%$ in each sample).

Subsequently, we used multivariate analyses to assess the variation of bacteria, fungi and eukaryotic communities across habitats and treatments. Before running multivariate analyses, we calculated the proportion of reads of each MOTU in each sample. Relative abundance values were then transformed using the Box-Cox transformation, which simultaneously solves the double-zero problem and improves the multivariate normality of data (Legendre & Borcard, 2018).

First, we used Nonmetric MultiDimensional Scaling (NMDS) to describe differences in communities among the three habitats, and check whether different treatments yield different interpretations of ecological relationships among samples. NMDS uses an optimization process to find a configuration of points (samples) in a space with a small number of dimensions, and is suitable for metabarcoding analyses that aim to reconstruct variation in community composition as well as possible, without preserving any particular distance measure among objects (Borcard, Gillet, & Legendre, 2011; Chen & Ficetola, 2020; Paliy & Shankar, 2016). Given its robustness and flexibility, NMDS is often used as the first step to characterize the similarity of communities in metabarcoding studies (Chen & Ficetola, 2020; Paliy & Shankar, 2016). NMDS was run on the Euclidean distance computed on Box–Cox-chord-transformed data (Legendre & Borcard, 2018), by building 1,000 ordinations.

Second, we used *ProcMod* , a Procrustes-based analysis (Coissac & Gonindard-Melodelima, 2019), to measure the multivariate correlations between the communities obtained using the different treatments. *ProcMod* can be used to measure the shared variation between matrices, and is particularly appropriate to test relationships between datasets obtained through DNA metabarcoding and metagenomics (Coissac & Gonindard-Melodelima, 2019). Procrustes analyses tend to overfit the data, therefore we used a modified version of Procrustes correlation that is robust to highly-dimensional data and allows a correct estimation of the shared variation between data sets (Coissac & Gonindard-Melodelima, 2019). The Procrustes-based correlation tests were performed using the *corls* function in the R package*ProcMod,* using 1,000 randomizations to test the mean covariance between random matrices (Coissac & Gonindard-Melodelima, 2019).

Third, we used redundancy analysis (RDA) to measure the amount of variation among communities that is explained by differences in habitat and treatments (Legendre & Legendre, 2012; Ter Braak, 1986). With habitat typology and treatment as constraining matrices, we used treatment contrasts to test if each treatment led to communities significantly different from those unraveled by the control treatment. Thus, significant treatment contrasts indicate that results between control and experimental treatments differ in an important way, while non-significant results mean that deviation from ideal conditions is not specifically pronounced. Significance of RDA and treatment contrasts was tested through 10,000 permutations using the *vegan* package in R (Borcard et al., 2011; Oksanen et al., 2019).

For bacteria only, RDA detected significant differences between the control treatment and some of the treatments. We thus ran a similarity percentage analysis with the *simper* R function (Clarke, 1993) from *vegan* to

identify the taxa contributing to the overall pairwise treatment difference (Geyer et al., 2014). Significance was tested using 50,000 permutations. Given the large number of tests performed, the significance of tests was corrected using the False Discovery Rate (FDR) method with the *fdrtool* package (Strimmer, 2008). FDR has greater power than traditional approaches (e.g. Bonferroni correction) when performing multiple comparisons (Benjamini & Hochberg, 1995). All statistical analyses were performed in the R environment.

RESULTS

A total of 6.3, 7.9 and 25.7 million reads were obtained from the Bact02, Fung02 and Euka02 libraries, respectively. After read assembly, quality filtering, spurious sequence and contaminant removal, 481,411; 2,511,721 and 13,232,441 good-quality sequences remained, consisted of 660 (Bact02), 1,075 (Fung02) and 3,611 (Euka02) unique sequences (i.e. MOTUs).

Differences in MOTU richness among treatments

Generalized Linear Mixed Models allowed identifying shifts in the richness of observed MOTUs likely due to unproportional growth of different taxonomic groups or to differential DNA degradation under different preservation conditions of each sample.

When we considered all the detected MOTUs, GLMM detected significant differences in MOTUs richness among treatments for all the markers considered (Bact02: $\chi 4 = 38.9$, $P < 0.001$; Fung02: $\chi 4 = 18.2$, $P = 0.001$; Euka02: $\chi 4 = 21.7$,$P < 0.001$; Fig. 2). Compared to the control, contrasts showed small but significant changes in MOTUs richness under treatment 3 (Bact02: z = 2.54, $P = 0.010$; Fung02: z = -2.17, $P = 0.029$; Euka02: z = 2.65, $P = 0.008$), treatment 4 (Bact02: z = -2.93, $P = 0.003$; Fung02: z = -3.99, $P < 0.001$; Euka02: z = 3.92, $P < 0.001$), and treatment 5 (Bact02: z = -3.74; Fung02: z = -4.02; Euka02: z = 4.18; all $P < 0.001$). Treatment 2 caused a small but significant decrease in MOTUs richness for fungi (z = -2.42; $P = 0.015$), but not for bacteria and eukaryotes ($P = 0.456$, $P = 0.283$, respectively; for all contrasts: Table S1).

Nevertheless, when we repeated analyses by excluding rare MOTUs (i.e. with a frequency <1%), differences in richness were much smaller, and were significant only for bacteria and fungi (Bact02:$\chi 4 = 9.69$, $P = 0.045$; Fung02: $\chi 4 = 14.1$, $P = 0.006$; Euka02: $\chi 4 = 2.22$, $P = 0.693$; Fig. 2). Compared to the control, MOTUs richness decreases for Bact02 under treatment 3 (z = -2.91; $P = 0.003$) and increases for Fung02 under treatments 4 and 5 (z = 2.77; $P = 0.005$; z = 1.75; $P = 0.080$; respectively), while no significant effect was detected for Euka02 under any of the treatments (all $P > 0.170$; for all contrasts: Table S1).

Habitat caused a significant effect in MOTUs richness only for Fung02 either before and after removing rare MOTUs (before: $\chi 1 = 11.8$,$P < 0.001$; after: $\chi 1 = 20.5$, $P< 0.001$).

Ecological similarity of communities among treatments

Nonmetric MultiDimensional Scaling showed a stress value of 0.13 for Bact02, 0.14 for Fung02 and 0.12 for Euka02. For each of the three markers, the NMDS plots obtained for the five sample preservation treatments were extremely similar, and the ecological differences among the three habitats were clearly identified by all the preservation treatments (Fig. 3).

The multivariate correlation between the communities obtained with the five treatments was always very strong (Procrustes-modified correlation: for all comparisons between "control" and treatments r [?] 0.84,$P < 0.0001$; Fig. 4) indicating, for all markers, that most of the variation of retrieved communities was shared across all the treatments. Procrustes correlations were particularly high for Fung02 and Euka02 (all r [?] 0.9), and between "control" and treatments 2 and 3 (all r [?] 0.93; Fig. 4).

Differences between the obtained communities

Redundancy analysis allowed us to measure the amount of variation explained by differences among habitats and by treatments. Overall, 33%, 24%, and 33% of variability were explained by differences in habitat for bacteria, fungi, and eukaryotes, respectively. The community differences among habitats were strongly

significant for the three taxonomic groups (permutation test: all P [?] 0.001). Differences among treatments were much weaker, and explained 9%, 2% and 2% of variation only for bacteria, fungi and eukaryotes, respectively. Nevertheless, differences were significant for bacteria (permutation test: $P < 0.0001$), but not for fungi and eukaryotes (both $P = 1$).

For bacteria, contrasts did not detect significant differences between "control" and treatment 2 or 3. Differences between "control" and treatment 4 and 5 were significant but explained a limited amount of variation (for both treatments, [?]3% of variation explained; $P< 0.0001$; Table 1). We thus used similarity percentage analysis to identify the MOTUs significantly contributing to these differences. Only one MOTU showed a significant contribution ($P = 0.03$ after FDR correction) to the differences between "control" and treatment 4; this MOTU showed a very limited frequency under treatment 4 (Fig. S1). After FDR correction, no MOTU showed a significant contribution to the differences between "control" and treatment 5. The MOTU highlighted by the similarity percentage analysis for treatment 4 also showed a very limited frequency under treatment 5, but these differences were not significant at the 5% threshold (corrected significance after FDR correction = 0.078).

DISCUSSION

Monitoring soil biodiversity with eDNA metabarcoding over large geographical and taxonomic scales and sometimes in remote places has become an important practice in ecological research. Understanding how preservation conditions affect estimates of taxonomic richness and community composition is essential to ensure sound ecological conclusions. Our study shows that soil metabarcoding results are surprisingly robust to preservation conditions, as we observed limited differences in community structure and diversity estimates when samples were preserved using different strategies. However, some taxonomic groups and diversity components are more sensitive than others to certain preservation conditions. This allowed us developing guidelines for preservation depending on the aims of monitoring programs and on focal taxa.

The aim of this study was comparing realistic approaches to soil preservation against an ideal situation. Immediate extraction was our reference approach, as it avoids both DNA degradation (i.e. potential underrepresentation of certain taxa) and continued growth of certain taxonomic groups (i.e. potential overrepresentation of other taxa). Unfortunately, immediate extraction is only possible if sampling occurs nearby facilities, or when a mobile eDNA laboratory is available (e.g. Zinger, Taberlet, et al. (2019b), and logistical constraints often hampers its application in remote areas. We selected preservation conditions among the most achievable, cost-effective and frequent practices to sampling soil for eDNA studies (Dickie et al., 2018). For more details about the design of preservation conditions, see Appendix A.

Influence of preservation methods on richness estimates

Preservation methods generated some small but significant differences in MOTUs richness compared to what is observed in the "control", with some contrasting effects across taxa. When considering all the MOTUs, none of the preservation conditions yielded estimates of alpha-diversity identical to the "control". For instance, just six hours at room temperature caused a significant decrease of MOTUs richness in fungi. It has been shown that estimates of alpha-diversity using metabarcoding are extremely sensitive to methodological choices (Calderón-Sanou et al., 2020). Our study underlines that even preservation for a very short time can further affect the detection of rare MOTUs and highlights the sensitivity of fungi to preservation at room temperature (Delavaux, Bever, Karppinen, & Bainard, 2020). MOTUs richness of all the taxa was also affected by preservation at 4degC, which caused a slight increase of MOTUs richness for bacteria and eukaryotes, and a slight decrease for fungi. The effect of temperature and time storage in fungal and bacterial growth has already been proved (see e.g. Orchard et al., 2017; Pettersson & Baath, 2003). Despite this, in addition to temperature, we can expect that other parameters such as initial soil moisture and pH influences bacterial growth (Baath & Arnebrant, 1994; Drenovsky, Vo, Graham, & Scow, 2004; Fernandez-Calvino & Baath, 2010; Kaiser et al., 2016) with a combined effect. Finally, drought affects the richness of microbial communities in soil ecosystems with differential effects across taxa depending on their ecology (Evans, Wallenstein, & Burke, 2014; Meisner, Jacquiod, Snoek, Ten Hooven, & van der Putten, 2018; Ochoa-

Hueso et al., 2018), and three weeks of preservation with silica gel generally reduced the observed MOTUs richness in bacteria and fungi, while it increased the richness of eukaryotes.

However, our study also shows that specific caution is mostly necessary when rare MOTUs are of interest. The exclusion of rare MOTUs strongly reduced differences between optimal conditions and different preservation. Remaining effects were much weaker for bacteria and fungi, and disappeared for eukaryotes (Fig. 2). This suggests that the effect of preservation approach on taxonomic richness mostly occur on rare species, as already suggested by Meisner et al. (2018) for microbial communities. Several authors have shown that eDNA metabarcoding does not represent the best tool for the detection of rare MOTUs, as some rare MOTUs remain undetected, while many sequences detected at rare frequency are artifactual (Brown et al., 2015). Estimates of α-diversity should therefore always taken with caution, and indices that underweight rare MOTUs (e.g. Shannon or Simpson diversity) can provide more robust estimates (Brown et al., 2015; Calderón-Sanou et al., 2020; Balint et al., 2016).

Differences in community structure

If the study interest is in community structure and not in richness estimates then preservation choices become even less important. In fact, the similarity of communities obtained through the different preservation conditions is surprisingly high (see Procrustes correlation coefficients; Fig. 4); the amount of variation explained by preservation conditions was much lower than the observed differences among habitats (see redundancy analysis), and multivariate analyses consistently allowed to detect community differences among habitats (Fig. 3). In other words, metabarcoding is able to identify the ecological differences among sites, independently of the preservation approach. Even though metabarcoding analyses are sensitive to methodological choices, estimates of relationships between diversity and the environment are often robust (Calderon-Sanou et al., 2020; Ji et al., 2013), and this is a very good news if we want to apply these approaches to broad-scale monitoring programs, aiming at assessing the effects of environmental changes.

Bacteria were the only taxon for which we detected significant differences between the "control" and the preservation conditions, with [?]3% of variability explained by differences between the "control" and the desiccation treatments. The observed differences most likely refer to some taxa that are affected by the dry conditions and could lead to a overrepresentation of some taxa that are more resistant under these conditions. Differences between the desiccation treatments and the control were small, and only one out of 660 MOTUs showed a significant variation in abundance with the control. This MOTU (belonging to the Bacteroidetes phylum of bacteria, see Supporting Information) was generally abundant in the control and preservation conditions 2 and 3 (average frequency of reads around 10%) while it drastically decreased under preservation conditions 4 and 5 (Fig. S1).

Nevertheless, the significant differences observed for some taxa and preservation conditions stress the importance of selecting the preservation method before starting a monitoring program and using it consistently through the whole monitoring, to avoid confusion between the effects of methods and of environmental changes.

Conclusions: guidelines for optimizing preservation conditions

Standardized protocols are essential for repeatable and reliable biodiversity monitoring, and our results allow to propose guidelines to improve and standardize the preservation of soil samples for eDNA metabarcoding analyses (Fig. 5):

1) If sampling occurs close to lab facilities, or a mobile lab is available, extracting DNA as soon as possible is the best approach. Storing samples a few hours at room temperature does not have major impact on the outcome of analyses, especially if the focus is not on rare MOTUs;

2) If lab facilities are available after a short-time transportation, storing samples in the fridge (0-4degC) for a few days is a safe approach as it does not have a significant impact on community composition, and only moderately affects MOTUs richness. However, this approach can be problematic if the aim is to estimate

MOTUs richness, and particularly the occurrence of rare MOTUs. The feasibility of this strategy also depends on the number and volume of samples, and to the possibility of maintaining the cold chain;

3) If monitoring in remote areas, sample desiccation (e.g. using silica gel) and long-term preservation at room temperature is a reasonable approach, and it is particularly convenient when working with a large volume of samples. This approach preserves ecological signal, but can affect the detection of some taxa, particularly among the rarest ones. Therefore, this approach is suboptimal for monitoring programs aiming at detecting rare MOTUs.

An effective application of eDNA metabarcoding to biodiversity monitoring is complex, and protocols of sample preservation are key methodological choices that has to be taken into account when designing a metabarcoding-based monitoring. When working in difficult and remote environments researchers are faced to the trade-offs between a faithful representation of biodiversity, and multiple logistic constraints in the field. Accurate a-priori planning is often the basis of successful monitoring programs and our guidelines can help researchers and practitioners to identify the best approach to sample preservation, depending on the studied taxa and research goals.

ACKNOWLEDGMENTS

We thank P. Taberlet for constructive discussions. This study is funded by the European Research Council under the European Community's Horizon 2020 Programme, Grant Agreement no. 772284 (IceCommunities). WT, TM and LG also thank the rom the French Agence Nationale de la Recherche (ANR) through the GlobNets (ANR-16-CE02-0009) project and from 'Investissement d'Avenir' grant managed by the ANR (Montane: OSUG@2020: ANR-10-LAB-56).

REFERENCES

[dataset] Guerrieri, A., Bonin, A., Munkemuller, T., Gielly, L., Thuiller, W., Ficetola, G.F.; 2020; Raw and filtered supporting data for the study of soil preservation using environmental DNA; XXX

Baath, E., & Arnebrant, K. (1994). Growth rate and response of bacterial communities to pH in limed and ash treated forest soils.*Soil Biology and Biochemistry* , *26* (8), 995–1001. doi: 10.1016/0038-0717(94)90114-7

Balint, M., Bahram, M., Murat Eren, A., Faust, K., Fuhrman, J. A., Bj", B., ... Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews* , *017* , 686–700. doi: 10.1093/femsre/fuw017

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.*Journal of the Royal Statistical Society: Series B (Methodological)* , *57* (1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Borcard, D., Gillet, F., & Legendre, P. (2011). Numerical Ecology with R. In *Numerical Ecology with R* . doi: 10.1007/978-1-4419-7976-6

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. doi: 10.1111/1755-0998.12428

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal* , *9* (2), 378–400. doi: 10.32614/rj-2017-066

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., ... Jumpponen, A. (2015). Scraping the bottom of the barrel: Are rare high throughput sequences artifacts? *Fungal Ecology* , *13* , 221–225. doi: 10.1016/j.funeco.2014.08.006

Calderon-Sanou, I., Munkemuller, T., Boyer, F., Zinger, L., & Thuiller, W. (2020). From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices? *Journal of*

Biogeography , *47* (1), 193–206. doi: 10.1111/jbi.13681

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., ... Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers.*Scientific Reports* , *9* (1), 1–11. doi: 10.1038/s41598-019-39399-5

Cardona, S., Eck, A., Cassellas, M., Gallart, M., Alastrue, C., Dore, J., ... Manichanh, C. (2012). Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiology* ,*12* . doi: 10.1186/1471-2180-12-158

Chase, M. W., & Hills, H. H. (1991). Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *TAXON* ,*40* (2), 215–220. doi: 10.2307/1222975

Chen, W., & Ficetola, G. F. (2020). Numerical methods for sedimentary-ancient-DNA-based study on past biodiversity and ecosystem functioning. *Environmental DNA* , edn3.79. doi: 10.1002/edn3.79

CLARKE, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral Ecology* , *18* (1), 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x

Coissac, E. (2012). OligoTag: A program for designing sets of tags for next-generation sequencing of multiplexed samples. *Methods in Molecular Biology* , *888* , 13–31. doi: 10.1007/978-1-61779-870-2_2

Coissac, E., & Gonindard-Melodelima, C. (2019). Assessing the shared variation among high-dimensional data matrices: a modified version of the Procrustean correlation coefficient. *BioRxiv Bioinformatics* , 1–16. doi: 10.1101/842070

Consul, P. C., & Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics - Theory and Methods* ,*21* (1), 89–109. doi: 10.1080/03610929208830766

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., ... Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* , *348* (6237). doi: 10.1126/science.1261605

Delavaux, C. S., Bever, J. D., Karppinen, E. M., & Bainard, L. D. (2020). Keeping it cool: Soil sample cold pack storage and DNA shipment up to 1 month does not impact metabarcoding results. *Ecology and Evolution* , (November 2019), 1–13. doi: 10.1002/ece3.6219

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-Gonzalez, A., Eldridge, D. J., Bardgett, R. D., ... Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science* , *359* (6373), 320–325. doi: 10.1126/science.aap9516

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., ... Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources* , *18* (5), 940–952. doi: 10.1111/1755-0998.12907

Drenovsky, R. E., Vo, D., Graham, K. J., & Scow, K. M. (2004). Soil water content and organic carbon availability are major determinants of soil microbial community composition. *Microbial Ecology* ,*48* (3), 424–430. doi: 10.1007/s00248-003-1063-2

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., ... Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. *Molecular Ecology* , *21* (8), 1821–1833. doi: 10.1111/j.1365-294X.2012.05537.x

Evans, S. E., Wallenstein, M. D., & Burke, I. C. (2014). Is bacterial moisture niche a good predictor of shifts in community composition under long-term drought. *Ecology* , *95* (1), 110–122. doi: 10.1890/13-0500.1

Fernandez-Calvino, D., & Baath, E. (2010). Growth response of the bacterial community to pH in soils differing in pH. *FEMS Microbiology Ecology* , *73* (1), 149–156. doi: 10.1111/j.1574-6941.2010.00873.x

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessiere, J., ... Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* , *11* (1), 434. doi: 10.1186/1471-2164-11-434

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., ... Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* . doi: 10.1111/1755-0998.12338

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics Using r* . London: SAGE Publications.

Geyer, K. M., Altrichter, A. E., Takacs-Vesbach, C. D., Van Horn, D. J., Gooseff, M. N., & Barrett, J. E. (2014). Bacterial community composition of divergent soil habitats in a polar desert. *FEMS Microbiology Ecology* , *89* (2), 490–494. doi: 10.1111/1574-6941.12306

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton.*Nature* , *345* (6270), 60–63. doi: 10.1038/345060a0

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S., & Turon, X. (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *PLOS ONE* ,*10* (10), e0139633. doi: 10.1371/journal.pone.0139633

Guo, Y., Yang, G. Q., Chen, Y., Li, D., & Guo, Z. (2018). A comparison of different methods for preserving plant molecular materials and the effect of degraded DNA on ddRAD sequencing. *Plant Diversity* ,*40* (3), 106–116. doi: 10.1016/j.pld.2018.04.001

Hoffmann, C., Schubert, G., & Calvignac-Spencer, S. (2016, February 1). Aquatic biodiversity assessment for the lazy. *Molecular Ecology* , Vol. 25, pp. 846–848. doi: 10.1111/mec.13535

Huerlimann, R., Cooper, M. K., Edmunds, R. C., Villacorta-Rath, C., Le Port, A., Robson, H. L. A., ... Jerry, D. R. (2020). Enhancing tropical conservation and ecology research with aquatic environmental DNA methods: an introduction for non-environmental DNA specialists.*Animal Conservation* . doi: 10.1111/acv.12583

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* ,*16* (10), 1245–1257. doi: 10.1111/ele.12162

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schoning, I., ... Daniel, R. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Scientific Reports* , *6* (August), 1–12. doi: 10.1038/srep33696

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA.*Molecular Ecology Resources* , *20* (1), 29–39. doi: 10.1111/1755-0998.13107

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters* , *307* (1), 80–86. doi: 10.1111/j.1574-6968.2010.01965.x

Legendre, P., & Borcard, D. (2018). Box-Cox-chord transformations for community composition data prior to beta diversity analysis.*Ecography* , *41* (11), 1820–1824. doi: 10.1111/ecog.03498

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* . Amsterdam: Elsevier.

Majaneva, M., Diserud, O. H., Eagle, S. H. C., Bostrom, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports* , *8* (1), 4682. doi: 10.1038/s41598-018-23052-8

Meisner, A., Jacquiod, S., Snoek, B. L., Ten Hooven, F. C., & van der Putten, W. H. (2018). Drought legacy effects on the composition of soil fungal and prokaryote communities. *Frontiers in Microbiology* ,*9* (MAR), 294. doi: 10.3389/fmicb.2018.00294

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., ... Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources* , *18* (5), 927–939. doi: 10.1111/1755-0998.12895

Ochoa-Hueso, R., Collins, S. L., Delgado-Baquerizo, M., Hamonts, K., Pockman, W. T., Sinsabaugh, R. L., ... Power, S. A. (2018). Drought consistently alters the composition of soil fungal and bacterial communities in grasslands from two continents. *Global Change Biology* , *24* (7), 2818–2827. doi: 10.1111/gcb.14113

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., ... Maintainer, H. W. (2019). *vegan: Community Ecology Package Version 2.5-6* . www.r-project.org

Orchard, S., Standish, R. J., Nicol, D., Dickie, I. A., & Ryan, M. H. (2017). Sample storage conditions alter colonisation structures of arbuscular mycorrhizal fungi and, particularly, fine root endophyte.*Plant and Soil* , *412* (1–2), 35–42. doi: 10.1007/s11104-016-2867-4

Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology* , Vol. 25, pp. 1032–1057. doi: 10.1111/mec.13536

Pettersson, M., & Baath, E. (2003). Temperature-dependent changes in the soil bacterial community in limed and unlimed soil. *FEMS Microbiology Ecology* , *45* (1), 13–21. doi: 10.1016/S0168-6496(03)00106-5

R Core Team (2018). *R: A Language and Environment for Statistical Computing* . Vienna, Austria: R Foundation for Statistical Computing.

Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics (Oxford, England)* , *24* (12), 1461–1462. doi: 10.1093/bioinformatics/btn209

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Environmental DNA: For Biodiversity Research and Monitoring. In*Environmental DNA: For Biodiversity Research and Monitoring* . doi: 10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* , *21* (8), 2045–2050. doi: 10.1111/j.1365-294X.2012.05470.x

Ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis.*Ecology* , *67* (5), 1167–1179. doi: 10.2307/1938672

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community.*Nature* , *345* (6270), 63–65. doi: 10.1038/345063a0

Zinger, L., Bonin, A., Alsos, I. G., Balint, M., Bik, H., Boyer, F., ... Taberlet, P. (2019a). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions.*Molecular Ecology* , Vol. 28, pp. 1857–1862. doi: 10.1111/mec.15060

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., ... Chave, J. (2019b). Body size determines soil community assembly in a tropical forest. *Molecular Ecology* , *28* (3), 528–543. doi: 10.1111/mec.14919

DATA ACCESSIBILITY

Raw sequencing data as well as filtered data will be deposited in DRYAD (Dryad Digital Repository) upon acceptance.

AUTHOR CONTRIBUTIONS

TM, GFF, LG, AB, WT and AG designed the experiment. LG, TM and AG conducted the field work. AG conducted all molecular analyses and performed the bioinformatic treatment of sequences with the help of AB. AG, AB and GFF ran statistical analyses. AG, GFF and AB drafted the manuscript. All the authors contributed substantially to the revision process, and accepted the final version.

ORCID

Alessia Guerrieri:*https://orcid.org/0000-0002-1519-3517*

Aurelie Bonin:*https://orcid.org/0000-0001-7800-8609*

Tamara Munkemuller:*https://orcid.org/0000-0001-9743-1322*

Ludovic Gielly:*https://orcid.org/0000-0001-5164-6512*

Wilfried Thuiller:*https://orcid.org/0000-0002-5388-5274*

Gentile Francesco Ficetola:*https://orcid.org/0000-0003-3414-5155*

TABLES

Table 1. Treatment contrasts assessing differences between the control (immediate extraction) and four approaches to soil conservation before eDNA extraction. Each conservation treatment was compared against the control in order to determine the percentage of explained variability.

	Bact02 Explained variability (%)	Bact02 *P*	Fung02 Explained var
Treatment 2: room temperature, extraction after 6h;	0.58	0.956	0.53
Treatment 3: 4°C, extraction after 3 days;	0.81	0.563	0.58
Treatment 4: silica gel immediately inserted, extraction after 21 days;	3.14	<0.001	1.03
Treatment 5: silica gel inserted after 6h, extraction after 21 days;	3.16	<0.001	0.73

FIGURES CAPTIONS

Fig. 1. Experimental sampling design.

Fig. 2. MOTUs richness across the different treatments (control: immediate extraction; T2: extraction after 6h at room temperature; T3: extraction after three days at 4°C; T4: immediate preservation in tubes with silica gel, extraction after 21 days; T5: preservation in tubes with silica gel after 6h at room temperature, extraction after 21 days) before (left) and after (right) removing MOTUs with frequency $< 1\%$ in one sample.

Fig. 3. Plots of non-metric dimensional scaling showing dissimilarities of communities among the three habitats: broadleaved forest (black); grassland (green); vegetated riverbank (blue). Each plot shows the results of metabarcoding analysis based on soil samples subjected to five different treatments.

Fig. 4. Procrustes correlation between communities obtained from metabarcoding analyses based on soil samples across environmental conditions subjected to five sample treatments (control: immediate extraction; RT+6h: extraction after 6h at room temperature; 4°C: extraction after three days at 4°C; silicagel: immediate preservation in tubes with silica gel, extraction after 21 days; silicagel+6h: preservation in tubes with silica gel after 6h at room temperature, extraction after 21 days. All correlation coefficients are highly significant (all $P < 0.0001$).

Fig. 5. Guidelines for improving monitoring strategies with eDNA from soil.

APPENDIX A

Design of soil preservation conditions

We did not test full-factorial combinations of different preservation periods and conditions, and we did not consider freezing, which is unrealistic when dealing with large numbers and / or volumes of samples, as is the case for more and more metabarcoding studies. Furthermore, freezing is generally impossible when sampling remote areas, where maintaining a cold chain cannot be ensured given the logistical challenges and, in the best cases, it is replaced by preservation in a cool box (i.e. 4°C or more).

In the design of treatments, we considered approaches allowing preservation at different temperatures and for different periods. For preservation condition 2, we accounted for a certain delay (six hours) between sampling and extraction, that could correspond to local transportation from the sampling area to the nearest base station. Sometimes, even in the case of an *in situ* extraction, samples remain at ambient temperature for hours prior to extraction, especially when monitoring a large area or a tricky ground, but this can have an impact on the final results (Delavaux et al., 2020). Preservation at 4°C is among the most frequent approaches to soil preservation (Dickie et al., 2018). This can be attained through portable refrigerators and requires the cold chain not to be interrupted at any point during transportation, which is only possible when lab facilities are accessible in a relatively short time (Hoffmann et al., 2016; Huerlimann et al., 2020). However, one of the most attractive characteristics of eDNA metabarcoding is its capacity to provide biodiversity data for understudied areas (e.g. tropical and arctic areas; mountain chains). Preservation conditions 4 and 5 of this study refer to the situations where samples are located far from lab facilities, in areas for which traditional biomonitoring is particularly challenging. When sampling these areas, the time lag between sample collection and subsequent molecular analysis can be particularly relevant. The use of silica gel to preserve soil samples is then a good solution (Chase & Hills, 1991; Guo, Yang, Chen, Li, & Guo, 2018): silica gel allows removing up to 26% of its weight in water in a relatively short time (overnight; Taberlet, pers. communication), is cost-effective, easy to transport, and is not an issue for aircraft transportation (no flammable or dangerous preservatives). Silica gel can be added *in situ* during sampling (protocol 4) or within a short time lag, allowing for local transportation from the sampling area to the nearest facilities (protocol 5). Afterwards, samples can be stored at room temperature, protected from sunlight to avoid photo-degradation of DNA.