DR. IAN A. DICKIE (Orcid ID : 0000-0002-2740-2128)

PROF. STÉPHANE BOYER (Orcid ID : 0000-0002-0750-4864)

DR. SERGIO E MORALES (Orcid ID : 0000-0002-5862-0726)

Towards robust and repeatable sampling methods in eDNA based studies

Ian A. Dickie[1,2],

Stephane Boyer[3,4]

Hannah Buckley[5]

Richard P. Duncan[6]

Paul Gardner[2]

Ian D. Hogg[7,8]

Robert J. Holdaway[9]

Gavin Lear[10]

Andreas Makiola[1]

Sergio E. Morales[11]

Jeff R. Powell[12]

Louise Weaver[13]

[1]Bio-Protection Research Centre, Lincoln University, Lincoln, New Zealand

[2]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

[3]Institut de Recherche sur la Biologie de l'Insecte – UMR 7261 CNRS / Université de Tours, Parc Grandmont, 37200 Tours, France

[4]Applied Molecular Solutions Research Group, Environmental and Animal Sciences, Unitec Institute of Technology, Private Bag 92025, Victoria Street West, Auckland 1142, New Zealand

[5]School of Science, Auckland University of Technology, Auckland 1010, New Zealand.

[6]Institute for Applied Ecology, University of Canberra, ACT 2617, Australia

[7]School of Science, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

[8]Polar Knowledge Canada, CHARS Campus, Cambridge Bay, NU X0B 0C0, Canada

[9]Landcare Research, PO Box 69040, Lincoln 7640, New Zealand

[10]School of Biological Sciences, The University of Auckland, Auckland 1010, New Zealand.

[11]Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin 9054, New Zealand

[12]Hawkesbury Institute for the Environment, Western Sydney University, Penrith NSW 2751, Australia

[13]Institute of Environmental Science and Research Ltd., Christchurch, New Zealand

Corresponding author:

Ian Dickie

Ian.dickie@canterbury.ac.nz

School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand

+64 3 3692268

Running title: Robust sampling in molecular ecology

Summary

DNA based techniques are increasingly used for measuring the biodiversity (species presence, identity, abundance and community composition) of terrestrial and aquatic ecosystems. While there are numerous reviews of molecular methods and bioinformatic steps, there has been little consideration of the methods used to collect samples upon which these later steps are based. This represents a critical knowledge gap, as methodologically sound field sampling is the foundation for subsequent analyses. We reviewed field sampling methods used for metabarcoding studies of both terrestrial and freshwater ecosystem biodiversity over a nearly three-year period (n = 75). We found that 95% (n = 71) of these studies used subjective sampling methods, inappropriate field methods, and/or failed to provide critical methodological information. It would be possible for researchers to replicate only 5% of the metabarcoding studies in our sample, a poorer level of reproducibility than for ecological studies in general. Our findings suggest greater attention to field sampling methods and reporting is necessary in eDNA-based studies of biodiversity to ensure robust outcomes and future reproducibility. Methods must be fully and accurately reported, and protocols developed that minimise subjectivity. Standardisation of sampling protocols would be one way to help to improve reproducibility, and have additional benefits in allowing compilation and comparison of data from across studies.

Introduction

Methods for determining the composition of ecological communities from environmental DNA (eDNA; see Table 1 for definitions) samples have become increasingly routine as a means to describe biodiversity, for conservation monitoring, and to test ecological hypotheses (reviewed by Taberlet et al., 2012; Holdaway et al., 2017). Specifically, soil, tissue and water samples are collected from the environment, DNA is extracted from those samples and amplified using PCR with taxon-specific primers, the resulting PCR products are sequenced using platforms such as Roche 454 pyrosequencing, Illumina MiSeq, Ion Torrent, and PacBio, and taxonomic entities are identified from unique sequence reads. This approach, known as "metabarcoding", is most commonly used to study communities of Bacteria, Archaea, Fungi and plankton (Taberlet et al., 2012). However, in addition to studying microorganisms in environmental samples, metabarcoding can be used to study the composition of a wide range of Eukaryote taxa (Andersen et al., 2012; Coissac et al., 2012), including pools of larger organisms where the DNA can be extracted and analysed without separating individuals (Dickie and St John, 2016; Holdaway et al., 2017). Other closely-related techniques include analysis of the combined genomic material from all organisms in a sample (rather than specific gene regions), known as environmental metagenomics (Holdaway et al., 2017).

A critical step in all DNA studies is the collection of samples from the environment. Sample collection involves a series of decisions that have important implications for how the data should be analysed and interpreted (Table 2); valid inferences are critically dependent on valid sampling techniques (Crawley, 2015). Further, sample collection at a specific site and specific time can be done only once. This contrasts with later steps (e.g., DNA amplification, sequencing and bioinformatic analyses), which can be repeated or re-run from archived samples.

Previous reviews of metabarcoding methods provide guidance on statistical replication in sampling (Prosser, 2010; Lennon, 2011), methods for processing samples (Lear et al., 2018), data reporting (Field et al., 2009; Chervitz et al., 2011) and bioinformatic analyses (Hiraoka et al., 2016). However, despite the importance of sample collection in underpinning these later steps, we are not aware of any attempt to critically review methods of sample collection for eDNA studies. There are general reviews about how to sample ecological communities (Otypková and Chytrý, 2006; Smith et al., 2017) but we suggest eDNA studies have several unique features that distinguish them from other ecological surveys. First, eDNA studies often focus on microrganisms or species otherwise difficult to observe. As a result, there is little inherent sense of the spatial arrangement and complexity of biodiversity being sampled to guide sampling design. This contrasts with, for example, plant community sampling, where the size of a plot can be adjusted based on the size of plants in the community (e.g., larger plots in forest, smaller in grassland). Second, in eDNA sampling there is a significant time-lag between sample collection and being able to see the results, due to the need for laboratory processing and analysis. This contrasts with traditional sampling, where an initial sense of the data is obtained almost immediately, allowing some opportunity for methods to be revised before substantial investment is made. Third, eDNA sampling is often destructive in that a sample is removed (e.g., soil cores, leaf punches), preventing re-measurement of exactly the same location (unlike, for example, forest surveys of trees). This has implications for studies intending to measure change over time because we often cannot do repeat measurements of the same exact sampling point. Fourth, many eDNA sampling efforts are focussed on species-rich communities with a high degree of spatial and temporal heterogeneity. Fifth, even if *a priori* hypotheses are stated, eDNA sampling is often undertaken with either a primary or secondary goal of characterizing patterns and describing communities. This means that the researcher is frequently interested in obtaining data with a sampling design sufficiently robust to address unanticipated questions that may arise. Finally, sample contamination, both between samples and from exogenous sources, is potentially much more problematic in eDNA sampling than in sampling based on visible organisms.

Here, we set out to (1) review existing methods being used for eDNA sample collection from terrestrial and freshwater systems, (2) suggest criteria that could be used to evaluate sampling methods, (3) provide guidance on how methods could be improved, and (4) identify research questions that need to be resolved in order to improve existing sampling methods. We focus on metabarcoding and metagenomic studies where characterising biodiversity was an objective of the study, and we restrict our review to terrestrial and freshwater systems, largely reflecting the expertise of the authors. Nonetheless, we believe many of the points we raise are generic to studies using other eDNA methods, and to studies in systems we do not consider (e.g., atmospheric, gut content, marine).

Methods

We conducted a review of sampling protocols currently used in metabarcoding studies of terrestrial and freshwater biodiversity to address questions at a scale larger than an individual plot (e.g. for which plots are considered representative of some larger region). To identify methods used in recent studies, we searched the Web of Science Core Collection on 3 December 2015 using the search strategy Topic = ((Soil OR Water OR River OR Lake) AND ("environmental DNA" OR metabarcoding OR metagenomics) AND (Community OR Biodiversity) NOT (Marine OR Ocean)), restricting results to research papers published in 2013, 2014 or 2015. This resulted in an initial list of 275 papers, from which we excluded ancient DNA studies, reviews, laboratory based experiments, entirely industrial, and within-organism microbiome studies. Where more than one paper was published from a single sampling event, these were treated as a single "project". While not intended to be fully comprehensive, the resulting list of 75 independent projects (Table S1) provides a broad spectrum of studies for examining how biodiversity has been sampled using environmental DNA.

We identified the methods used in each study, determined if those methods were likely to introduce bias or subjectivity, and whether the methods were described in sufficient detail that they could be replicated by another researcher. Where methods were cited as being given in another paper or source, we included that source in evaluating the study in question.

Results & Discussion

Sampling Universe

At the broadest scale, all sampling occurs within a "sampling universe". Defining the sampling universe requires specifying the area that samples are intended to be representative of (including, for example, geopolitical constraints, pragmatic limitations, and ecosystem types) and the criteria for excluding portions of that area from potential sampling (including safety and practicality constraints). Specifying the sampling universe is an essential prerequisite for replicating a study, as replication requires knowing what areas should be sampled. Further, results cannot be extrapolated to larger areas unless the sampled universe is defined (Denny and Benedetti-Cecchi, 2012; Smith et al., 2017). More generally, results cannot be interpreted in the absence of a clear definition of what areas they apply to.

There are a few exemplar studies in our review where a clear sampling universe is defined. Two of the studies defined a political boundary and plots were randomly selected from a grid imposed within that boundary (Yang et al., 2014; Terrat et al., 2015). One study defined their universe as being the portion of an island that fell within 200 m of a particular track (Drummond et al., 2015), along with clear criteria for excluding plots. The study by Drummond and colleagues shows that having maps or other pre-existing knowledge of an area is not necessary to define a sampling universe that could be readily repeated by future researchers. Two other studies incompletely defined their sampling universes, for example, defining the sampling areas as those affected by a particular flood, which

gives a clear sampling universe, and "adjacent" areas without further detail (Baldwin et al., 2013), which we considered borderline in terms of reproducibility.

Overall, the majority of papers (92%) did not define a sampling universe, often describing sampling locations in detail, but not how these locations were chosen to be representative of any larger area. Describing sample locations provides specific information about the study, but does not allow other researchers to compare results with comparable areas in other regions. Further, samples can only be taken as indicative of a larger area if the potential area sampled is defined.

We exclude 14 manipulative studies from this calculation. All 14 manipulative studies also did not define a sampling universe, but this is less critical where the goal is to measure response to an imposed treatment because it is assumed that the manipulation associated with the treatment defines the sampling universe.

Representative, Haphazard, Regular or Random Samples

Once the sampling universe is defined, the location of samples or plots *within* that universe must be determined, either using objective or subjective methods. Examples of objective methods included "at the deepest point of the lakes" (Barberan and Casamayor, 2014), "randomly generated within 200 m distance along a contour off a specified track", or the use of grid-based locations (Toju et al., 2013; Yang et al., 2014; Terrat et al., 2015). The defining character of objective protocols is that the precise location of a plot is specified as an exact and absolute location, based on either true random, grid-based, or more complicated sampling designs (e.g., Robertson et al., 2013) or, in some cases, clearly defined criteria (e.g., "deepest point").

With subjective sampling, in contrast, the location of the plot is only loosely specified by the experimental design, with the exact location selected by the researcher, commonly in a way to be both "representative" of a site and to avoid what the researcher views as unusual or disturbed sites. The choice to use subjective criteria to locate plots in representative locations has important effects. Avoiding atypical sites or features within sites will reduce variability in community or ecosystem metrics, likely increasing the probability of detecting statistically significant effects among locations. However, it also results in data that are representative of only a subset of the sampling universe that they purport to describe. Taking pasture sampling as an example, results based on measuring only the portion of a pasture that is not near watering troughs, gates, or disturbed soils, cannot be taken as representative of the total area of pasture. Where sites are selected to be "representative" there is an even larger potential bias based on the individual researcher's view on what is a "representative" ecosystem. This does not imply that samples must be taken everywhere, but rather that the sampling universe must be clearly and objectively defined.

The final alternative is that plots are located haphazardly within broad categories (e.g., any forest site in a particular country). Haphazard sampling is an extreme case of subjective sampling, with implicit rather than stated criteria for sampling. In many cases, haphazard sampling includes a strong element of convenience, such as sampling along existing trails or roads (Anderson, 2001). Haphazard sampling makes replication by others nearly impossible, as a researcher attempting to repeat a study is unlikely to have the exact same implicit criteria.

Excluding manipulative studies, we found only 10% of studies in our survey used a clearly defined objective sampling protocol. A few additional studies (5%) claimed plot locations were "random", but without evidence of how the randomisation was performed. As noted by Crawley (2015) claims of randomisation are common but rarely performed properly. A further 5% of studies stated methods that

were clearly subjective, and > 80% of studies did not clearly indicate how plot locations were determined.

Subjective, and particularly haphazard samples, are not scientifically sound (fit-for-purpose) for drawing conclusions about broader ecosystem questions (Anderson, 2001). Indeed, true randomisation is considered by Crawley (2015) to be one of only two essential concepts in ecological design (with replication being the second). Further, failure to report how samples were located makes replication impossible, as no future researcher could be certain they were using the same criteria for selecting sample locations. Where ecosystems can be mapped before sampling, Smith et al. (2017) provide useful guidance on objective sampling methods using open-source software. Nonetheless, pre-existing maps or knowledge of an area are not absolutely required for pragmatic yet robust sampling, as random distances along randomly oriented transects can be generated in the field and used as one-dimensional grids (Allen, 1993).

Size of Area Sampled

Most studies employ a sampling scheme in which one or more subsamples are taken from within a defined area at a particular location (a plot). Plot area affects measures of biodiversity as several subsamples from a large plot area will tend to capture higher diversity than the same number of subsamples from a smaller plot area, at least in any spatially structured community. In our review, the area of study plots ranged from single points to areas as large as several hectares (e.g., Mendes et al., 2015; Vargas-Gastelum et al., 2015). The only consistencies in plot area were either single sampling points (12 studies) or 20 x 20 m plots (4 studies), with no other plot area used in more than one study. Plot area was not reported in 29 of the 75 studies (39%).

As measurement of biodiversity is directly affected by plot size, reporting plot area is necessary for reproducibility, which 39% of studies failed to do. While the need to report plot area is universal, an appropriate area for plots may depend on the research question being addressed. Despite recognizing this, much of the variation in plot area in our reviewed studies appeared to be arbitrary; authors stated the plot area, but not why that area was chosen. High variability in plot area among studies makes comparing results more difficult. In particular, estimates of species richness cannot be compared unless sampled from the same area. Given the existing prevalence of single point sampling and 20 x 20 m area sampling across existing studies, these provide two useful plot areas at which consistency across studies could be achieved.

Determining the Location of Subsamples within Plots

The location of individual subsamples within plots also varies substantially among studies. Subsamples may be explicitly located based on a regular pattern, truly random, haphazard, or subjectively distributed across the plot. The choice of method requires careful consideration. Subsampling in a regular pattern could, in theory, be spatially synchronised with environmental variation, resulting in a non-representative sample. This is particularly a risk in planted ecosystems (plantations, agricultural fields, orchards; figure 1). True random placement of subsamples within plots avoids this risk, but is less efficient at sampling the entire area of a plot. Randomisation can also be more challenging to employ in practice. Further, it is often unclear if the word "random" is being used in a true sense, as opposed to "haphazard" and potentially subjective (Crawley, 2015). Haphazard and subjective methods provide no clear advantages except for convenience. One risk with any allowance of subjectivity in subsample placement is that portions of the plot may be more or less-likely to be sampled, without that bias being explicit in the protocol. For example, where a plot comprises an intensively managed pasture, subjectively placed subsamples chosen to be 'representative' may under-sample dung or urine patches, whereas regular or random sampling would

capture these locations in proportion to the area they occupy (Figure 1). This is different from a protocol that explicitly excludes these locations or that stratify sampling to better measure specific areas, in that it is subject to observer bias and hence reduces reproducibility of the protocol.

Across studies with more than one subsample per plot, we found 20% of studies used regular or random subsampling, one study stated subsamples were "haphazard", and 8% of studies claimed subsamples were random without giving details of randomisation. Subjective methods were common (17%), often with stated constraints (e.g., minimum distances between samples, distance from features such as trees). Most studies, however, failed to report how subsamples were chosen (54.5% of studies).

Determining How Many Samples to Take

Replication is one of the most critical aspects of any sampling (Crawley 2015), as noted in the aptly titled piece "replicate or lie" (Prosser, 2010). In our review of studies, replication was absent or inadequate in many studies, with 20 studies (27%) having fewer than 6 total replicates. At the other extreme, some studies had more than 350 replicates (Tedersoo et al., 2014; Terrat et al., 2015).

While it is easy to state that replication is needed, determining the optimal number of samples for community analysis depends greatly on the question being asked and will therefore vary across studies. Given the cost of sampling and analysis it is essential not to under-sample, and desirable not to over-sample. Power analysis can help find an optimal sampling level. Power analyses for continuous variables (e.g., species richness) are straightforward, requiring only an assumption of variance and the minimum effect size. There are also examples of using power analysis for detection

of individual species (Olson et al., 2013), patterns of individual species occurrences (Dickie and FitzJohn, 2007), and measures of community similarity (Irvine et al., 2011; Kelly et al., 2015).

Despite the availability of tools allowing replication to be optimised, of the 75 studies we reviewed, none gave a stated rationale for how the number of replicates was determined. On the other hand, 100% of studies reported their sample size, the only one of our assessed variables for which this was true.

Subsampling

While the plot defines the unit of replication, in many cases multiple subsamples are taken within plots. For example, multiple soil cores may be taken from a plot and then either pooled before measurement or measured independently to derive an 'average' value that characterises the replicate. One advantage of subsampling is that it allows for characterisation of a given area which can be re-sampled in the future, whereas single point sample cannot be resampled. The ability to re-sample allows for measurement of change through time. Across studies we found a range of 2 to 100 subsamples taken per plot (Drummond et al., 2015; Pansu et al., 2015). Although the number of subsamples was generally easily determined, it was hard to find out how that number of subsamples was selected. While not extensively studied, one study using T-RFLP found that eight subsamples were sufficient to distinguish bacterial communities among different land-uses in Australia (Osborne et al., 2011).

Subsamples can either be kept independent or pooled before DNA analysis. Pooling can substantially reduce the cost of subsequent analyses and can be appropriate where the research aim is to characterise large-scale patterns. The effects of pooling vary by taxa and depending on the measurement of interest. Osborne et al., (2011) found that pooling detected less variability than not pooling, but caused no change in the observed differences between sites. However, another study found that pooling substantially reduced the ability to detect rare species, particularly for fungi as compared to bacteria (Manter et al., 2010). One further advantage of not pooling subsamples is that it permits the estimation of spatial variance within plots. This may be a key objective in its own right, or may help partition variance (e.g., within plot versus temporal, if re-sampling is intended).

Pooling is very common, either at the level of plot or, in a few cases, within subset categories (i.e., soil depth (Tveit et al., 2013); water depth or oxygen status (Peura et al., 2015); or distance from trees within plots (De Beeck et al., 2015). It is also possible to pool subsets of subsamples. For example, Keshri et al. (2015) pooled nine subsamples into three. Finally, it is possible to keep subsamples independent through DNA extraction and PCR and subsequently pool (Wilkins et al., 2015); two studies suggest that pooling before or after PCR has little effect on the perceived community (Manter et al., 2010; Osborne et al., 2011).

The decision whether or not to pool will depend on a careful evaluation of costs, including trade-offs between increased replication and increased precision per replicate, the spatial heterogeneity of the organism(s) being studied, and potential effects of pooling on the community metrics being analysed. Across studies, we only found three examples where multiple samples were taken within a statistical replicate and not pooled. In at least two of these, not pooling samples allowed the authors to measure both within and between plot variability in community composition (Drummond et al., 2015; Navarrete et al., 2015).

Substrate

Future reproducibility of a study relies on clarity on what actual substrate was sampled. Particularly in the case of soil sampling, the definition of soil can either include or exclude leaf litter and other organic layers. The choice of which layers to include or exclude could have dramatic effects on the perceived community, given vertical stratification of soil biota into different soil horizons (Dickie et al., 2002). Across the 51 studies in our dataset that sampled soil, 82% did not specify whether litter was included in samples.

A further decision is whether to specify to a specified depth, or as in a few cases (4% of studies), sample by soil horizon, where a soil horizon is defined by physical features. Sampling by horizons can help reduce variability in samples and may be essential to test specific hypotheses. However, determination of soil horizons can require significant technical expertise and may vary across researchers. Further, specific horizons may not be present at all sites, or may be confusing (e.g., buried organic horizons, exposed C-horizons).

Sample quality assurance

Soil and other ecological sampling can be a fantastically filthy job, often conducted in adverse weather conditions and at remote locations. This presents a very real challenge to obtaining clean DNA samples, as any foreign material could compromise the results. Major sources of field contamination include: pre-existing DNA on sampling equipment, DNA from the researcher and their personal microbiome, carry-over between samples, and, for some samples, unintentional movement of DNA from the surface into a sample. Determining if a particular taxon is a contaminant can be

challenging as, for example, some common human-associated microbes such as the fungus *Malassezia* are also common in environmental samples (Amend, 2014).

Field negative controls

Most molecular ecologists routinely include negative controls in PCR reactions, recognising the power of PCR to detect very low levels of DNA. Typically, negative controls account only for contamination in the laboratory. Arguably, accounting for contamination in the field may be more important, albeit also more challenging. Field equipment can, and probably should, be sampled through swabbing in order to gain some insight into the potential for contamination. Sample storage media and containers can also be tested. No study in our review reported whether or how field negative controls were included. This stands in contrast to the almost universal use of negative controls in laboratory stages of analysis, an environment where maintaining sample integrity is relatively straightforward.

Sample contamination

Field sampling equipment can also contaminate samples. Some studies avoid contamination through single-use, pre-sterilised equipment, particularly for water sampling. Where sampling equipment is re-used, soaking in a solution of sodium hypochlorite, such as commercial bleach, is an effective method of decontamination (Prince and Andrus, 1992; Kemp and Smith, 2005) provided the length of exposure and solution concentration are sufficient. Household bleach is a variable concentration solution of sodium hypochlorite (3 - 8%). For stringent decontamination, a solution as strong as 2 to 3% sodium hypochlorite may be needed (Kemp and Smith, 2005), but general cleaning should be effective with as low as 0.55% [10% v/v dilution of commercial bleach (Prince and Andrus, 1992)]. The effectiveness of bleach depends on the length of time of exposure, requiring at least a few minutes at typical concentrations. Particularly where large numbers of samples are being collected, it

may be most efficient to have multiple sets of sampling equipment, allowing multiple samples to be taken before having to decontaminate.

Some confusion exists over the difference between sterilizing and decontaminating sampling equipment. Alcohol, in particular, has been reported by some authors as the sole cleaning agent for field gear (e.g., Prober et al., 2015). Alcohol sterilizes by killing microbes, but does not remove DNA contamination. Indeed, ethanol is routinely used in DNA precipitation and processing.

Across the 75 studies in our survey, 59 did not specify anything about decontamination of sampling equipment, eight stated sampling devices were "clean" or "sterile" without giving details, three used ethanol, one (a water sampling) repeatedly purged a pump, two used bleach and one specified autoclaved spoons. Given that alcohol (which does not effectively remove DNA) is as widely reported as bleach or autoclaving, it is likely that many of the studies that fail to report decontamination procedures may be using ineffective techniques.

Movement of DNA from the surface into samples is a specific issue for some sampling, including sampling of soils or water at depth or sampling of the interior of deadwood. To avoid this the surface can be removed or samples can be broken or split open, such that the sampling tool does not pass from the outside surface into the interior. This can include the use of customised tools [e.g., the A-Xenic Extractor (Dickie et al., 2012)].

Sample storage and transport

eDNA results may be affected by sample storage, from the moment a sample is collected, through transport to the laboratory, and during storage in the laboratory before DNA extraction. At least one study suggests that storage conditions caused minor changes in community composition patterns relative to the much larger effects of the environment where samples were collected (Lauber et al., 2010). However, the relative importance of sample storage may be greater where more subtle patterns are being tested. Nonetheless, other studies have shown declines in the abundance of some species (Orchard et al., 2017), growth of cold-adapted microbes in storage (e.g., Wood et al., 2015), and fragmentation of DNA (Cardona et al., 2012) during storage. Important decisions include the temperature samples are stored at, whether naturally anaerobic samples are maintained in anaerobic state, and the length of time samples are stored. Recommended times between collection and freezing are as low as 2 hours (von Wintzingerode et al., 1997). In some cases it may be possible to preserve samples chemically, both preventing microbial growth and loss of DNA (Seutin et al., 1991; Frantzen et al., 1998).

Many studies either freeze samples quickly (10% of studies), or keep samples between 0 and 4 °C (34% of studies). A further 11% store samples at room temperature which, although sub-optimal, can be unavoidable in certain situations (e.g., very remote sampling), and may not be problematic for all samples (e.g., where samples are minimally disrupted, such as animal faeces that has already been at room temperature for some time). Overall, we found that 40% of studies failed to specify the temperature at which samples were transported. A further 4% specified -80°C as an immediate storage temperature, with no further details given of how this was achieved. We considered this suspect, given that achieving -80°C in field conditions would be a substantial challenge (noting that liquid nitrogen would reduce sample temperatures to -196°C).

Metadata

The term metadata refers to the information surrounding data, not the actual sample data itself. In other words, metadata is the what, why, where, when and how for a sample. Without the accompanying metadata, reuse of previously collected data is difficult. Many attempts to collect metadata are framed in terms of the minimum amount of information to give context to samples and the analyses undertaken. Flexible frameworks also allow hierarchical terms or pieces of data to be collected using well-defined vocabularies, e.g. ontologies (Soldatova and King, 2005; Reichman et al., 2011).

Reporting methods

Recording and description of methods is a fundamental component of metadata. If methods are not fully documented, peer review cannot be effective in ensuring scientific validity. Equally importantly, the ability of future researchers to replicate a study is dependent on full method knowledge (Lithgow et al., 2017). Perhaps the most striking result of reviewing the literature was how poorly methods were described (Figure 2). In more than 90% of studies, key methods were not described in sufficient detail to allow replication. This was true even after considering methods cited to other sources.

Site metadata

In scientific manuscripts, it is good practice to only report methods that are directly relevant to the results presented in that manuscript. One consequence of this, however, is that it is difficult or impossible to evaluate what metadata may or may not have been collected in a study. This also has the unfortunate side effect that potentially available metadata cannot be easily "discovered". In some

cases, this means data that could be relevant to novel investigations and meta-analyses are overlooked.

This is not an insurmountable challenge to overcome. There are notable examples from vegetation surveys where national databases allow comprehensive metadata to be archived and queried, even if not used in the initial manuscripts published from a study (Wiser et al., 2001). Publishers of molecular studies have a strong history of requiring archiving sequence data in public databases, but metadata archiving remains haphazard. Recent efforts to develop metadata archives for genomic data from individual organisms (Deck et al., 2017) might provide some role model for eDNA biodiversity metadata.

Conclusions and Recommendations

Overall, our review suggests that the majority of current eDNA studies are based on incompletely reported and, in many cases, questionable methods. Across the various studies and aspects of their sampling design, we found

1. Only 5% of studies provided sufficient information to allow sampling to be repeated by an independent researcher.

2. There is very little consistency across studies in methods being used to sample eDNA.

3. For the most part, there is no documented reason or rationale for these differences.

4. Sampling is often based on methods that make samples non-representative of the sampling universe, bias results, and make results unnecessarily difficult to compare across studies.

5. There are potentially major methodological issues in terms of dealing with DNA contamination and sample handling across many studies, including temperature of storage, lack of effective DNA decontamination, and lack of controls for field contamination.

The net effect of poor reporting and subjective methods was that of the 75 studies we examined in detail, we found only 4 (5%) that we were confident could be repeated by future researchers based on reported methods. While, in part, poor reporting may reflect word limits on the length of publications, there is no reason that sampling protocols cannot be fully detailed in electronic supplements.

Poor reporting and subjective methods are a general issue in ecological science, but the problem appears particularly acute in eDNA studies. In a recent review, Haddaway and Verhoeven (2015) found that 14 – 58% of reviewed literature in ecology lacked sufficient details for future repeatability or analysis. Our finding of only 5% reproducibility suggest the situation may be worse in eDNA studies than in the broader ecological literature. Similarly, Smith et al. (2017) found that 43% of general ecological studies failed to report how spatial sampling units were chosen, compared to our finding of 80% in eDNA studies, while 21% of ecological studies used systematic or random methods, compared to only 10% of eDNA studies. This may, in part, reflect a focus on ongoing methodological development in eDNA studies, rather than a focus on actual measurement of biodiversity. Nonetheless, only 5% of the 75 studies we included were primarily about methods, and all the studies presented data that could be interpreted as being potentially representative of ecosystems.

Prosser (2010) noted that the high cost of molecular methods did not obviate the need for statistical replication, but rather made solid experimental design even more important. The same is true of sampling methodology. Further, because eDNA based surveys typically conduct all of their sampling

before any results are obtained, a failure to sample correctly can result in loss of months to years of analysis time.

Based on our review of methods, we suggest that improved sampling protocols are essential in molecular ecology. In considering ways to improve protocols, we suggest that robust sampling protocols should:

1. *Ensure data are fit for the purpose they were collected for, including being statistically robust and able to be analysed [essential].* We believe that it is self-evident that sample collection should generate data that can be used for the purpose(s) for which it was originally collected, including generation of statistically meaningful results.

2. *Allow for replication by another researcher or at a future date, and hence be both completely described and free from subjective decisions that may differ across researchers [essential].* Reproducibility is a fundamental principle of the scientific method, as recognized since the early 11th century (Steinle, 2016; Lithgow et al., 2017).

3. *Be cost effective and as simple as possible [recommended].* Given limited research budgets, sampling should not be more expensive or difficult than is necessary. Having simple protocols also helps longevity (criterion 4). Nonetheless, being wasteful would not invalidate results, so this is not essential.

4. *Have longevity, including minimising particular skills needed, such that the same method is able to be repeated in the future, but also be adaptable to new methods and change in personnel [recommended].* Using methods with longevity will allow compilation of data into larger analyses, including potentially unanticipated measures of change over time. This partially relies on avoiding methods requiring highly-specialised expertise, as loss of specialist personnel can render these methods non-repeatable.

5. *Be based on evidence or a stated rationale [recommended].* Basing a sampling protocol on an evidence base makes it more likely that methods are robust, and also helps convince other researchers to use a protocol consistently.

6. *Ensure samples are robust to unanticipated analyses and outcomes [recommended].* Data may be useful for analyses beyond the scope of what was originally intended. Given the expense and effort of collecting DNA from environmental samples, it is advantageous to have protocols that are robust to these additional analyses, even though not required for the original study to be valid.

One promising way to improve sampling practices would be to develop standard, fully documented protocols. A major advantage of standardised protocols is that they allow for concerted investment of time and research funding into validation and improvement of those protocols. Standard protocols in vegetation ecology have allowed for integration of results across research projects, researchers, and over long time-periods at national scale (Wiser et al., 2001). Standardized protocols also allow for very efficient reporting of methods, requiring only a citation and noting of any deviations. Towards that goal, we make specific recommendations in Table 3 where we believe that general "best practice" can be identified against the criteria above. Nonetheless, some aspects of eDNA sampling require further research before robust recommendations can be made. In particular, the contribution of spatial and temporal variability to observed community metrics needs to be understood, in order to optimize the scale of sampling, number of subsamples, and temporal re-sampling. In suggesting a movement towards standardization, we recognise that some specific research questions can require specific sampling designs. In these cases, applying a standardized protocol may be inappropriate and even counter-productive. However, most of the current variation in methods across studies does not appear to be necessary.

DNA-based biodiversity assessment remains a relatively young field of science (Taberlet et al., 2012; Holdaway et al., 2017). Our purpose in this review is not to be overly critical of the pioneering research in this field, but rather to point out areas where significant improvements can be easily made for the future. Nonetheless, our evaluation of existing eDNA studies of biodiversity does not suggest a robust, reproducible field of science. Peer reviewers and editors have a responsibility to ensure that the methods published in papers are sufficiently well described to permit future researchers to understand and replicate a study, either in the main text, in cited protocols, or in electronic supplements. Methods should be evaluated to ensure they are robust to address the research topic and more general future use of the data. Ultimately, however, it is up to individual researchers to ensure that the design and execution of field sampling receives the same attention to detail and care that is currently focussed on laboratory methods and bioinformatics.

Data availability

The complete data is in table S1.

References

Allen RB (1993) A permanent plot method for monitoring changes in indigenous forests: a field manual. *Manaaki Whenua Landcare Research*. https://nvs.landcareresearch.co.nz/content/permplotmethod.pdf (last accessed 28 February, 2018).

Amend A (2014) From dandruff to deep-sea vents: *Malassezia*-like fungi are ecologically hyper-diverse. *PLoS Pathogens* **10,** e1004277.

Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, Orlando L, Gilbert MTP, Willerslev E (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* **21,** 1966–1979.

Anderson DR (2001) The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin (1973-2006)* **29,** 1294–1297.

Baldwin DS, Colloff MJ, Rees GNA, Chariton AA, Watson GO, Court LNA, Hartley DM, Morgan MJ, King AJ, Wilson, S. J, Hodda M, Hardy CM (2013) Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Molecular Ecology* **22,** 1746–1758.

Barberan A, Casamayor EO (2014) A phylogenetic perspective on species diversity, beta-diversity and biogeography for the microbial world. *Molecular Ecology* **23,** 5868–5876.

Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, Azpiroz F, Roca J, Guarner F, Manichanh C (2012) Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC microbiology* **12,** 158.

Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, Sansone S-A, Stoeckert CJ, Taylor CF, Taylor R (2011) Data standards for Omics data: the basis of data sharing and reuse. *Bioinformatics for Omics Data: Methods and Protocols* 31–69.

Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology* **21,** 1834–1847.

Crawley MJ (2015) Statistics: an introduction using R, 2nd ed. John Wiley & Sons Ltd., Hoboken, NJ, USA. 339 pp.

De Beeck MO, Ruytinx J, Smits MMA, Vangronsveld J, Colpaert JV, Rineau F (2015) Belowground fungal communities in pioneer Scots pine stands growing on heavy metal polluted and non-polluted soils. *Soil Biology & Biochemistry* **86,** 58–66.

Deck J, Gaither MR, Ewing R, Bird CE, Davies N, Meyer C, Riginos C, Toonen RJ, Crandall ED (2017) The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples. *PLoS biology* **15,** e2002925.

Denny M, Benedetti-Cecchi L (2012) Scaling up in ecology: mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics* **43,** 1–22.

Dickie IA, FitzJohn RG (2007) Using terminal-restriction fragment length polymorphism (T-RFLP) to identify mycorrhizal fungi; a methods review. *Mycorrhiza* **17,** 259–270.

Dickie IA, Fukami T, Wilkie JP, Allen RB, Buchanan PK (2012) Do assembly history effects attenuate from species to ecosystem properties? A field test with wood-inhabiting fungi. *Ecology Letters* **15,** 133–141.

Dickie IA, Xu B, Koide RT (2002) Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* **156,** 527–535.

Dickie IA, St John MG (2016) Second-generation molecular understanding of mycorrhizas in soil ecosystems. F. Martin (ed) *Molecular Mycorrhizal Symbiosis*. John Wiley & Sons Ltd., Hoboken, NJ, USA. 473–491.

Drummond AJ, Newcomb RD, Buckley TRA, Xie D, Dopheide A, Potter BCM, Heled, Joseph, Ross HA, Tooman L, Grosser SA, Park D, Demetras NJ, Stevens MIA, Russell JC, Anderson SH,

Carter A, Nelson, Nicola (2015) Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* **4,** 46.

Field D, Sansone S-A, Collis A, Booth T, Dukes P, Gregurick SK, Kennedy K, Kolar P, Kolker E, Maxon M (2009) Omics data sharing. *Science* **326,** 234–236.

Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK, Kohn MH (1998) Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology* **7,** 1423–1428.

Haddaway NR, Verhoeven JTA (2015) Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecology and Evolution* **5,** 4451–4454.

Hiraoka S, Yang C-c, Iwasaki W (2016) Metagenomics and bioinformatics in microbial ecology: current status and beyond. *Microbes and Environments* **31,** 204–212.

Holdaway RJ, Wood JR, Dickie IA, Orwin KH, Bellingham PJ, Richardson SJ, Lyver PO, Timoti P, Buckley TR (2017) Using DNA metabarcoding to assess New Zealand's terrestrial biodiversity. *New Zealand Journal of Ecology* **41,** 251-262.

Irvine KM, Dinger EC, Sarr D (2011) A power analysis for multivariate tests of temporal trend in species composition. *Ecology* **92,** 1879–1886.

Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Collman RG, Bushman FD, Li H (2015) Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* **31,** 2461–2468.

Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic science international* **154,** 53–61.

Keshri J, Yousuf B, Mishra A, Jha, Bhavanath (2015) The abundance of functional genes, cbbL, nifH, amoA and apsA, and bacterial community structure of intertidal soil from Arabian Sea. *Microbiological Research* **175,** 57–66.

Lauber CL, Zhou N, Gordon JI, 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,** 80–86.

Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckley TR, Cruickshank R, Dopheide A, Handley KM, Hermans S, Kamke J, Lee CK, MacDiarmid R, Morales SE, Orlovich DA, Smissen R, Wood J, Holdaway R (2018) Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *New Zealand Journal of Ecology* **42,** 10-50A.

Lennon JT (2011) Replication, lies and lesser-known truths regarding experimental design in environmental microbiology. *Environmental Microbiology* **13,** 1383–1386.

Lithgow GJ, Driscoll M, Phillips P (2017) A long journey to reproducible results. *Nature* **548,** 387–388.

Manter DK, Weir TL, Vivanco JM (2010) Negative effects of sample pooling on PCR-based estimates of soil microbial richness and community structure. *Applied and Environmental Microbiology* **76,** 2086–2090.

Mendes LW, Tsai SM, Navarrete AA, de, Hollander M, van Veen JA, Kuramae EE (2015) Soil-Borne Microbiome: Linking Diversity to Function. *Microbial Ecology* **70,** 255–265.

Navarrete AA, Tsai SM, Mendes LW, Faust, Karoline, de Hollander M, Cassman NA, Raes, Jeroen, van Veen JA, Kuramae EE (2015) Soil microbiome responses to the short-term effects of Amazonian deforestation. *Molecular Ecology* **24,** 2433–2448.

Olson ZH, Briggler JT, Williams RN (2013) An eDNA approach to detect eastern hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildl Res* **39,** 629.

Orchard S, Standish RJ, Nicol D, Dickie IA, Ryan MH (2017) Sample storage conditions alter colonisation structures of arbuscular mycorrhizal fungi and, particularly, fine root endophyte. *Plant and Soil* **412,** 35–42.

Osborne CA, Zwart AB, Broadhurst LM, Young AG, Richardson AE (2011) The influence of

sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiology Ecology* **78,** 70–79.

Otypková Z, Chytrý M (2006) Effects of plot size on the ordination of vegetation samples. *Journal of Vegetation Science* **17,** 465–472.

Pansu J, Winkworth RC, Hennion FA, Gielly L, Taberlet P, Choler P (2015) Long-lasting modification of soil fungal diversity associated with the introduction of rabbits to a remote sub-Antarctic archipelago. *Biology Letters* **11,** 20150408.

Peura S, Sinclair L, Bertilsson S, Eiler, Alexander (2015) Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. *Scientific Reports* **5,** 12102.

Prince AM, Andrus L (1992) PCR: how to kill unwanted DNA. *Biotechniques* **12,** 358–360.

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, DNAY, Hofmockel KS, Kirkman KP, Knops JM, La Pierre KJ, MacDougall AS, McCulley RL, Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N (2015) Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters* **18,** 85–95.

Prosser JI (2010) Replicate or lie. *Environmental Microbiology* **12,** 1806–1810.

Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and opportunities of open data in ecology. *Science* **331,** 703–705.

Robertson BL, Brown JA, McDonald T, Jaksons P (2013) BAS: Balanced acceptance sampling of natural resources. *Biometrics* **69,** 776–784.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69,** 82–90.

Smith ANH, Anderson MJ, Pawley MDM (2017) Could ecologists be more random? Straightforward alternatives to haphazard spatial sampling. *Ecography* **40,** 1251–1255

Soldatova LN, King RD (2005) Are the current ontologies in biology good ontologies. *Nature Biotechnology* **23,** 1095–1098.

Steinle F 2016. Stability and Replication of Experimental Results: A Historical Perspective. In: Atmanspacher H, Maasen S, eds. *Reproducibility: Principles, Problems, Practices, and Prospects*. John Wiley & Sons, Hoboken, NJ, USA 39 – 64.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21,** 1789–1793.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A (2014) Global diversity and geography of soil fungi. *Science* **346,** 1256688.

Terrat S, Dequiedt S, Horrigue W, Lelievre M, Cruaud C, Saby NPA, Jolivet C, Arrouays D, Maron PA, Ranjard L (2015) Improving soil bacterial taxa–area relationships assessment using DNA meta-barcoding. *Heredity* **114,** 468.

Toju H, Yamamoto S, Sato H, Tanabe, S. A, Gilbert GS, Kadowaki K (2013) Community composition of root-associated fungi in a *Quercus*-dominated temperate forest: codominance of mycorrhizal and root-endophytic fungi. *Ecology and Evolution* **3,** 1281–1293.

Tveit A, Schwacke R, Svenning MM, Urich, Tim (2013) Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *ISME Journal* **7,** 299–311.

Vargas-Gastelum L, Romero-Olivares AL, Escalante, E. A, Rocha-Olivares A, Brizuela C, Riquelme, Meritxell (2015) Impact of seasonal changes on fungal diversity of a semi-arid ecosystem revealed by 454 pyrosequencing. *FEMS Microbiology Ecology* **91,** fiv044.

von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in

environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21,** 213–229.

Wilkins LGE, Rogivue A, Fumagalli LA, Wedekind C (2015) Declining diversity of egg-associated bacteria during development of naturally spawned whitefish embryos (*Coregonus* spp.). *Aquatic Sciences* **77,** 481–497.

Wiser SK, Bellingham PJ, Burrows LE (2001) Managing biodiversity information: development of New Zealand's National Vegetation Survey databank. *New Zealand Journal of Ecology* **25,** 1–17.

Wood JR, Dickie IA, Moeller HV, Peltzer DA, Bonner KI, Rattray G, Wilmshurst JM (2015) Novel interactions between non-native mammals and fungi facilitate establishment of invasive pines. *Journal of Ecology* **103,** 121–129.

Yang C, Wang X, Miller JA, de Blecourt, Marleen, Ji Y, Yang C, Harrison RD, Yu, W. D (2014) Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. *Ecological Indicators* **46,** 379–389.

Table 1. Glossary of terms as used in this review.

Term	Definition
Biodiversity	Any measurement of taxonomic identity and/or abundance within the context of a community of organisms.
eDNA	Environmental DNA, meaning DNA extracted from samples of soil, water, air, or other substances in order to detect and identify species present as microscopic or cryptic organisms in the sample, or with tissue or free DNA in the sample.
Metabarcoding	Amplification of a specific gene region from multiple organisms in an eDNA sample, typically for the purpose of quantifying diversity, species turnover, or community composition.
Metagenomics	Amplification of all DNA from multiple organisms in a sample, including for measurement of species or functional gene diversity.
Plot	Plot refers to a defined spatial extent from which samples are taken, although not all samples have an associated plot. Plots may be experimental replicates, but having multiple plots does not guarantee proper statistical replication.
Sample	Sample refers to a single volume of water, soil, or other substrate from which DNA can be extracted. Samples may be pooled from multiple subsamples, as is often the case when associated with a plot, or taken from a single location.
Sampling protocol	The methods used to locate plots, collect samples within plots, and transport plots to the laboratory. We include some discussion of the number of total replicates, which overlaps somewhat with experimental design, but do not focus on experimental design.
Sampling universe	The total area that sample plots are taken to represent. This is equivalent to "population" in statistical sampling.
Subsample	Non-independent samples taken within a plot or body of water. May be pooled into a single sample before or after PCR.

Table 2. Summary table of some of the critical decisions in any sample collection.

Decision	Options	Implications
Definition of sampling universe (What is being sampled?)	What ecosystem or area are samples representative of?	Ability to scale-up results (e.g., from the plot to a larger ecosystem area; (Denny and Benedetti-Cecchi, 2012)) depends on having defined the sampling universe. Critical to interpretation of results and future reproducibility.
Location of plots or samples (Where is it being sampled?)	Within the sampling universe, the locations of individual samples may be random, regular, haphazard, or subjective.	Reproducibility depends on clearly defined criteria for locating plots. Subjective or otherwise biased sampling may produce results that are not representative of the sampling universe.
Exclusion of sampling points (What is not being sampled?)	Avoidance of unusual features vs. inclusion of all; explicit vs. implicit, unstated, or arbitrary decision criteria.	Degree to which the entire ecosystem is sampled, ability for others to replicate.
How many samples to take	Number of replicates, whether replication is based on pilot studies and/or power analysis.	Whether biologically relevant differences can be detected.
Number of subsamples taken within area	One to many.	Cost and time to sample, amount of variance of results. Ability to measure spatial variability [if samples not pooled].
Size and shape of area sampled [if area based]	Scales from cm to km; circular, transect, square, etc.	Variance of results, ability to detect small scale patterns, impact of edge to total area ratio, maximum linear distance within an area.
Location of subsamples within plot [if plot based]	True random, haphazard, or regular; exclusion criteria.	Ability for others to replicate. Ability to measure spatial variability [if not pooled].
Pooling of subsamples	Separate subsamples within plots versus pooling into one sample per plot.	Cost and complexity versus loss of spatial variability information.
Substrates sampled	Leaves, litter, soil, roots, water column, benthos.	Consistency across landscape, variability of results, ability to compare across studies.
Depth versus horizon based	Sampling a given depth, or	Ability to link to soil

sampling [soils only]	sampling a given horizon.	chemistry, cross-site comparison, technical expertise required to identify horizons, ability to sample all locations.
Total volume of substrate sampled	Volume of material sampled in the field.	Variance reduced by larger volume, larger volumes more likely to capture larger organisms.
Definition of sampling universe (What is being sampled?)	What ecosystem or area are samples representative of?	Ability to scale-up results (e.g., from the plot to a larger ecosystem area; (Denny and Benedetti-Cecchi, 2012)) depends on having defined the sampling universe. Critical to interpretation of results and future reproducibility.
Location of plots or samples (Where is it being sampled?)	Within the sampling universe, the locations of individual samples may be random, regular, haphazard, or subjective.	Reproducibility depends on clearly defined criteria for locating plots. Subjective or otherwise biased sampling may produce results that are not representative of the sampling universe.

Table 3. Specific recommendations and examples of good practice for the decision points identified in Table 2.

Decision	Recommendation
Definition of sampling universe (What is being sampled?)	Explicitly define the entire potential sampling area and any exclusion criteria. Definition can combine political and biogeographic constraints.
Location of plots or samples (Where is it being sampled?)	Pre-determine potential locations of plots following objective protocol (e.g., grid-based, true random).
Exclusion of sampling points (What is not being sampled?)	Clearly state what points would not be sampled if they were a pre-determined location.
How many samples to take	Conduct *a priori* power analysis to determine sufficient replication.
Number of subsamples taken within area [if area based]	Equal numbers of samples across all plots. Further research needed to determine optimal sample numbers for different ecosystems.
Size and shape of area sampled [if area based]	Where possible, standardise across studies; either point or 20 x 20 m area being the most common at present.
Location of subsamples points within plot [if area based]	Take subsamples at pre-determined, defined points within plots.
Pooling of subsamples	Consider whether quantification of within-sample heterogeneity would justify not pooling, preferably based on power analysis.
Substrates sampled	State exactly what substrate is being sampled, for example noting whether litter is removed from the top of soil samples.
Depth versus horizon based sampling [soils only]	Specify the mean depth and variance of samples. Horizon based sampling is acceptable, but probably too variable across researchers to be widely recommended.
Total volume of substrate sampled	Ensure sufficient volume for possible re-extraction of DNA and for long-term archival storage.

Sampling protocols / quality assurance	Use sterile single-use sampling tools or 10% bleach for field decontamination with sufficient exposure time. Include a field negative control (swab of field equipment) to verify decontamination procedure. Freeze samples at earliest opportunity.
Record metadata	Use pre-printed field data sheets that specify all necessary metadata. Report metadata along with results. Insure sampling methods are fully reported in publications. Deposit metadata in public archives (e.g. Dryad).