

# MOLECULAR ECOLOGY RESOURCES

## Towards robust and repeatable sampling methods in eDNA based studies

Journal:	<i>Molecular Ecology Resources</i>
Manuscript ID	MER-18-0074.R1
Manuscript Type:	Invited Technical Reviews
Date Submitted by the Author:	10-May-2018
Complete List of Authors:	Dickie, Ian; University of Canterbury, Bio-Protection Research Centre, School of Biological Sciences Boyer, Stéphane; CNRS / Université François-Rabelais de Tours, Insect Biology Research Institute(IRBI) ; Unitec Institute of Technology, Environmental and Animal Sciences Buckley, Hannah; Auckland University of Technology, School of Science Gardner, Paul; University of Canterbury, Bio-Protection Research Centre, School of Biological Sciences Duncan, Richard; University of Canberra, Institute for Applied Ecology Hogg, Ian; University of Waikato, School of Science; Polar Knowledge Canada Holdaway, Robert; Landcare Research New Zealand Lear, Gavin; School of Biological Sciences, University of Auckland Makiola, Andreas; Lincoln University, Bio-Protection Research Centre Morales, Sergio; University of Otago, Department of Microbiology and Immunology Powell, Jeff; Hawkesbury Institute for the Environment, University of Western Sydney, Weaver, Louise; Institute of Environmental Science and Research Ltd
Keywords:	Environmental DNA, Experimental design, Contamination, Metabarcoding, Metadata, Sampling

1 Towards robust and repeatable sampling methods in eDNA based studies

2

3 Ian A. Dickie<sup>1,2</sup>,

4 Stephane Boyer<sup>3,4</sup>

5 Hannah Buckley<sup>5</sup>

6 Richard P. Duncan<sup>6</sup>

7 Paul Gardner<sup>2</sup>

8 Ian D. Hogg<sup>7,8</sup>

9 Robert J. Holdaway<sup>9</sup>

10 Gavin Lear<sup>10</sup>

11 Andreas Makiola<sup>1</sup>

12 Sergio E. Morales<sup>11</sup>

13 Jeff R. Powell<sup>12</sup>

14 Louise Weaver<sup>13</sup>

15

16 <sup>1</sup>Bio-Protection Research Centre, Lincoln University, Lincoln, New Zealand

17 <sup>2</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

18 <sup>3</sup>Institut de Recherche sur la Biologie de l’Insecte – UMR 7261 CNRS / Université de Tours,

19 Parc Grandmont, 37200 Tours, France

20 <sup>4</sup>Applied Molecular Solutions Research Group, Environmental and Animal Sciences, Unitec

21 Institute of Technology, Private Bag 92025, Victoria Street West, Auckland 1142, New

22 Zealand

23 <sup>5</sup>School of Science, Auckland University of Technology, Auckland 1010, New Zealand.

24 <sup>6</sup>Institute for Applied Ecology, University of Canberra, ACT 2617, Australia

25 <sup>7</sup>School of Science, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

26 <sup>8</sup>Polar Knowledge Canada, CHARS Campus, Cambridge Bay, NU X0B 0C0, Canada

27 <sup>9</sup>Landcare Research, PO Box 69040, Lincoln 7640, New Zealand

28 <sup>10</sup>School of Biological Sciences, The University of Auckland, Auckland 1010, New Zealand.

29 <sup>11</sup>Department of Microbiology and Immunology, University of Otago, PO Box 56, Dunedin  
30 9054, New Zealand

31 <sup>12</sup>Hawkesbury Institute for the Environment, Western Sydney University, Penrith NSW 2751,  
32 Australia

33 <sup>13</sup>Institute of Environmental Science and Research Ltd., Christchurch, New Zealand

34

35 Keywords: contamination, environmental DNA, experimental design, metabarcoding,  
36 metadata, sampling

37

38 Corresponding author:

39 Ian Dickie

40 [ian.dickie@canterbury.ac.nz](mailto:ian.dickie@canterbury.ac.nz)

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

42 +64 3 3692268

43

44 Running title: Robust sampling in molecular ecology

45

46 **Summary**

47

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

65

## 66 **Introduction**

67

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

85

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

91 sequencing and bioinformatic analyses), which can be repeated or re-run from archived  
92 samples.

93

94 Previous reviews of metabarcoding methods provide guidance on statistical replication in  
95 sampling (Prosser, 2010; Lennon, 2011), methods for processing samples (Lear et al., 2018),  
96 data reporting (Field et al., 2009; Chervitz et al., 2011) and bioinformatic analyses (Hiraoka  
97 et al., 2016). However, despite the importance of sample collection in underpinning these  
98 later steps, we are not aware of any attempt to critically review methods of sample collection  
99 for eDNA studies. There are general reviews about how to sample ecological communities  
100 (Otypková and Chytrý, 2006; Smith et al., 2017) but we suggest eDNA studies have several  
101 unique features that distinguish them from other ecological surveys. First, eDNA studies  
102 often focus on microorganisms or species otherwise difficult to observe. As a result, there is  
103 little inherent sense of the spatial arrangement and complexity of biodiversity being sampled  
104 to guide sampling design. This contrasts with, for example, plant community sampling,  
105 where the size of a plot can be adjusted based on the size of plants in the community (e.g.,  
106 larger plots in forest, smaller in grassland). Second, in eDNA sampling there is a significant  
107 time-lag between sample collection and being able to see the results, due to the need for  
108 laboratory processing and analysis. This contrasts with traditional sampling, where an initial  
109 sense of the data is obtained almost immediately, allowing some opportunity for methods to  
110 be revised before substantial investment is made. Third, eDNA sampling is often destructive  
111 in that a sample is removed (e.g., soil cores, leaf punches), preventing re-measurement of  
112 exactly the same location (unlike, for example, forest surveys of trees). This has implications  
113 for studies intending to measure change over time because we often cannot do repeat  
114 measurements of the same exact sampling point. Fourth, many eDNA sampling efforts are  
115 focussed on species-rich communities with a high degree of spatial and temporal

116 heterogeneity. Fifth, even if *a priori* hypotheses are stated, eDNA sampling is often  
117 undertaken with either a primary or secondary goal of characterizing patterns and describing  
118 communities. This means that the researcher is frequently interested in obtaining data with a  
119 sampling design sufficiently robust to address unanticipated questions that may arise. Finally,  
120 sample contamination, both between samples and from exogenous sources, is potentially  
121 much more problematic in eDNA sampling than in sampling based on visible organisms.

122

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

132

### 133 **Methods**

134

135 We conducted a review of sampling protocols currently used in metabarcoding studies of  
136 terrestrial and freshwater biodiversity to address questions at a scale larger than an individual  
137 plot (e.g. for which plots are considered representative of some larger region). To identify  
138 methods used in recent studies, we searched the Web of Science Core Collection on 3  
139 December 2015 using the search strategy Topic = ((Soil OR Water OR River OR Lake) AND  
140 ("environmental DNA" OR metabarcoding OR metagenomics) AND (Community OR

141 Biodiversity) NOT (Marine OR Ocean)), restricting results to research papers published in  
142 2013, 2014 or 2015. This resulted in an initial list of 275 papers, from which we excluded  
143 ancient DNA studies, reviews, laboratory based experiments, entirely industrial, and within-  
144 organism microbiome studies. Where more than one paper was published from a single  
145 sampling event, these were treated as a single "project". While not intended to be fully  
146 comprehensive, the resulting list of 75 independent projects (Table S1) provides a broad  
147 spectrum of studies for examining how biodiversity has been sampled using environmental  
148 DNA.

149

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

154

## 155 **Results & Discussion**

156

### 157 **Sampling Universe**

158

159 At the broadest scale, all sampling occurs within a "sampling universe". Defining the  
160 sampling universe requires specifying the area that samples are intended to be representative  
161 of (including, for example, geopolitical constraints, pragmatic limitations, and ecosystem  
162 types) and the criteria for excluding portions of that area from potential sampling (including  
163 safety and practicality constraints). Specifying the sampling universe is an essential  
164 prerequisite for replicating a study, as replication requires knowing what areas should be  
165 sampled. Further, results cannot be extrapolated to larger areas unless the sampled universe is



166 defined (Denny and Benedetti-Cecchi, 2012; Smith et al., 2017). More generally, results  
167 cannot be interpreted in the absence of a clear definition of what areas they apply to.

168

169 There are a few exemplar studies in our review where a clear sampling universe is defined.

170 Two of the studies defined a political boundary and plots were randomly selected from a grid

171 imposed within that boundary (Yang et al., 2014; Terrat et al., 2015). One study defined their

172 universe as being the portion of an island that fell within 200 m of a particular track

173 (Drummond et al., 2015), along with clear criteria for excluding plots. The study by

174 Drummond and colleagues shows that having maps or other pre-existing knowledge of an

175 area is not necessary to define a sampling universe that could be readily repeated by future

176 researchers. Two other studies incompletely defined their sampling universes, for example,

177 defining the sampling areas as those affected by a particular flood, which gives a clear

178 sampling universe, and "adjacent" areas without further detail (Baldwin et al., 2013), which

179 we considered borderline in terms of reproducibility.

180

181 Overall, the majority of papers (92%) did not define a sampling universe, often describing

182 sampling locations in detail, but not how these locations were chosen to be representative of

183 any larger area. Describing sample locations provides specific information about the study,

184 but does not allow other researchers to compare results with comparable areas in other

185 regions. Further, samples can only be taken as indicative of a larger area if the potential area

186 sampled is defined.

187

188 We exclude 14 manipulative studies from this calculation. All 14 manipulative studies also

189 did not define a sampling universe, but this is less critical where the goal is to measure

190 response to an imposed treatment because it is assumed that the manipulation associated with  
191 the treatment defines the sampling universe.

192

### 193 **Representative, Haphazard, Regular or Random Samples**

194

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

204

205 With subjective sampling, in contrast, the location of the plot is only loosely specified by the  
206 experimental design, with the exact location selected by the researcher, commonly in a way  
207 to be both "representative" of a site and to avoid what the researcher views as unusual or  
208 disturbed sites. The choice to use subjective criteria to locate plots in representative locations  
209 has important effects. Avoiding atypical sites or features within sites will reduce variability in  
210 community or ecosystem metrics, likely increasing the probability of detecting statistically  
211 significant effects among locations. However, it also results in data that are representative of  
212 only a subset of the sampling universe that they purport to describe. Taking pasture sampling  
213 as an example, results based on measuring only the portion of a pasture that is not near  
214 watering troughs, gates, or disturbed soils, cannot be taken as representative of the total area

215 of pasture. Where sites are selected to be "representative" there is an even larger potential  
216 bias based on the individual researcher's view on what is a "representative" ecosystem. This  
217 does not imply that samples must be taken everywhere, but rather that the sampling universe  
218 must be clearly and objectively defined.

219

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

227

228 Excluding manipulative studies, we found only 10% of studies in our survey used a clearly  
229 defined objective sampling protocol. A few additional studies (5%) claimed plot locations  
230 were "random", but without evidence of how the randomisation was performed. As noted by  
231 Crawley (2015) claims of randomisation are common but rarely performed properly. A  
232 further 5% of studies stated methods that were clearly subjective, and > 80% of studies did  
233 not clearly indicate how plot locations were determined.

234

235 Subjective, and particularly haphazard samples, are not scientifically sound (fit-for-purpose)  
236 for drawing conclusions about broader ecosystem questions (Anderson, 2001). Indeed, true  
237 randomisation is considered by Crawley (2015) to be one of only two essential concepts in  
238 ecological design (with replication being the second). Further, failure to report how samples  
239 were located makes replication impossible, as no future researcher could be certain they were

240 using the same criteria for selecting sample locations. Where ecosystems can be mapped  
241 before sampling, Smith et al. (2017) provide useful guidance on objective sampling methods  
242 using open-source software. Nonetheless, pre-existing maps or knowledge of an area are not  
243 absolutely required for pragmatic yet robust sampling, as random distances along randomly  
244 oriented transects can be generated in the field and used as one-dimensional grids (Allen,  
245 1993).

246

### 247 **Size of Area Sampled**

248

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

258

259 As measurement of biodiversity is directly affected by plot size, reporting plot area is  
260 necessary for reproducibility, which 39% of studies failed to do. While the need to report plot  
261 area is universal, an appropriate area for plots may depend on the research question being  
262 addressed. Despite recognizing this, much of the variation in plot area in our reviewed studies  
263 appeared to be arbitrary; authors stated the plot area, but not why that area was chosen. High  
264 variability in plot area among studies makes comparing results more difficult. In particular,

265 estimates of species richness cannot be compared unless sampled from the same area. Given  
266 the existing prevalence of single point sampling and 20 x 20 m area sampling across existing  
267 studies, these provide two useful plot areas at which consistency across studies could be  
268 achieved.

269

### 270 **Determining the Location of Subsamples within Plots**

271

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

289 better measure specific areas, in that it is subject to observer bias and hence reduces  
290 reproducibility of the protocol.

291

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

298

### 299 **Determining How Many Samples to Take**

300

301 Replication is one of the most critical aspects of any sampling (Crawley 2015), as noted in  
302 the aptly titled piece "replicate or lie" (Prosser, 2010). In our review of studies, replication  
303 was absent or inadequate in many studies, with 20 studies (27%) having fewer than 6 total  
304 replicates. At the other extreme, some studies had more than 350 replicates (Tedersoo et al.,  
305 2014; Terrat et al., 2015).

306

307 While it is easy to state that replication is needed, determining the optimal number of samples  
308 for community analysis depends greatly on the question being asked and will therefore vary  
309 across studies. Given the cost of sampling and analysis it is essential not to under-sample, and  
310 desirable not to over-sample. Power analysis can help find an optimal sampling level. Power  
311 analyses for continuous variables (e.g., species richness) are straightforward, requiring only  
312 an assumption of variance and the minimum effect size. There are also examples of using  
313 power analysis for detection of individual species (Olson et al., 2013), patterns of individual

314 species occurrences (Dickie and FitzJohn, 2007), and measures of community similarity  
315 (Irvine et al., 2011; Kelly et al., 2015).

316

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

321

### 322 **Subsampling**

323

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

335

336 Subsamples can either be kept independent or pooled before DNA analysis. Pooling can  
337 substantially reduce the cost of subsequent analyses and can be appropriate where the  
338 research aim is to characterise large-scale patterns. The effects of pooling vary by taxa and

339 depending on the measurement of interest. Osborne et al., (2011) found that pooling detected  
340 less variability than not pooling, but caused no change in the observed differences between  
341 sites. However, another study found that pooling substantially reduced the ability to detect  
342 rare species, particularly for fungi as compared to bacteria (Manter et al., 2010). One further  
343 advantage of not pooling subsamples is that it permits the estimation of spatial variance  
344 within plots. This may be a key objective in its own right, or may help partition variance  
345 (e.g., within plot versus temporal, if re-sampling is intended).

346

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

354

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

362

363 **Substrate**



364

365 Future reproducibility of a study relies on clarity on what actual substrate was sampled.

366 Particularly in the case of soil sampling, the definition of soil can either include or exclude

367 leaf litter and other organic layers. The choice of which layers to include or exclude could

368 have dramatic effects on the perceived community, given vertical stratification of soil biota

369 into different soil horizons (Dickie et al., 2002). Across the 51 studies in our dataset that

370 sampled soil, 82% did not specify whether litter was included in samples.

371

372 A further decision is whether to specify to a specified depth, or as in a few cases (4% of

373 studies), sample by soil horizon, where a soil horizon is defined by physical features.

374 Sampling by horizons can help reduce variability in samples and may be essential to test

375 specific hypotheses. However, determination of soil horizons can require significant technical

376 expertise and may vary across researchers. Further, specific horizons may not be present at

377 all sites, or may be confusing (e.g., buried organic horizons, exposed C-horizons).

378

### 379 **Sample quality assurance**

380

381 Soil and other ecological sampling can be a fantastically filthy job, often conducted in

382 adverse weather conditions and at remote locations. This presents a very real challenge to

383 obtaining clean DNA samples, as any foreign material could compromise the results. Major

384 sources of field contamination include: pre-existing DNA on sampling equipment, DNA from

385 the researcher and their personal microbiome, carry-over between samples, and, for some

386 samples, unintentional movement of DNA from the surface into a sample. Determining if a

387 particular taxon is a contaminant can be challenging as, for example, some common human-

388 associated microbes such as the fungus *Malassezia* are also common in environmental  
389 samples (Amend, 2014).

390

#### 391 *Field negative controls*

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

401

#### 402 *Sample contamination*

403 Field sampling equipment can also contaminate samples. Some studies avoid contamination  
404 through single-use, pre-sterilised equipment, particularly for water sampling. Where  
405 sampling equipment is re-used, soaking in a solution of sodium hypochlorite, such as  
406 commercial bleach, is an effective method of decontamination (Prince and Andrus, 1992;  
407 Kemp and Smith, 2005) provided the length of exposure and solution concentration are  
408 sufficient. Household bleach is a variable concentration solution of sodium hypochlorite (3 -  
409 8%). For stringent decontamination, a solution as strong as 2 to 3% sodium hypochlorite may  
410 be needed (Kemp and Smith, 2005), but general cleaning should be effective with as low as  
411 0.55% [10% v/v dilution of commercial bleach (Prince and Andrus, 1992)]. The effectiveness  
412 of bleach depends on the length of time of exposure, requiring at least a few minutes at

413 typical concentrations. Particularly where large numbers of samples are being collected, it  
414 may be most efficient to have multiple sets of sampling equipment, allowing multiple  
415 samples to be taken before having to decontaminate.

416

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

422

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

429

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

435

436 *Sample storage and transport*

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

451

452 Many studies either freeze samples quickly (10% of studies), or keep samples between 0 and  
453 4 °C (34% of studies). A further 11% store samples at room temperature which, although  
454 sub-optimal, can be unavoidable in certain situations (e.g., very remote sampling), and may  
455 not be problematic for all samples (e.g., where samples are minimally disrupted, such as  
456 animal faeces that has already been at room temperature for some time). Overall, we found  
457 that 40% of studies failed to specify the temperature at which samples were transported. A  
458 further 4% specified -80°C as an immediate storage temperature, with no further details given  
459 of how this was achieved. We considered this suspect, given that achieving -80°C in field  
460 conditions would be a substantial challenge (noting that liquid nitrogen would reduce sample  
461 temperatures to -196°C).

462

463 **Metadata**

464

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

472

473 *Reporting methods*

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

481

482 *Site metadata*

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

487 easily "discovered". In some cases, this means data that could be relevant to novel  
488 investigations and meta-analyses are overlooked.

489

490 This is not an insurmountable challenge to overcome. There are notable examples from  
491 vegetation surveys where national databases allow comprehensive metadata to be archived  
492 and queried, even if not used in the initial manuscripts published from a study (Wiser et al.,  
493 2001). Publishers of molecular studies have a strong history of requiring archiving sequence  
494 data in public databases, but metadata archiving remains haphazard. Recent efforts to develop  
495 metadata archives for genomic data from individual organisms (Deck et al., 2017) might  
496 provide some role model for eDNA biodiversity metadata.

497

#### 498 **Conclusions and Recommendations**

499

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

- 503 1. Only 5% of studies provided sufficient information to allow sampling to be repeated by an  
504 independent researcher.
- 505 2. There is very little consistency across studies in methods being used to sample eDNA.
- 506 3. For the most part, there is no documented reason or rationale for these differences.
- 507 4. Sampling is often based on methods that make samples non-representative of the sampling  
508 universe, bias results, and make results unnecessarily difficult to compare across studies.
- 509 5. There are potentially major methodological issues in terms of dealing with DNA  
510 contamination and sample handling across many studies, including temperature of storage,  
511 lack of effective DNA decontamination, and lack of controls for field contamination.

512

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

518

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

531

532 Prosser (2010) noted that the high cost of molecular methods did not obviate the need for  
533 statistical replication, but rather made solid experimental design even more important. The  
534 same is true of sampling methodology. Further, because eDNA based surveys typically  
535 conduct all of their sampling before any results are obtained, a failure to sample correctly can  
536 result in loss of months to years of analysis time.

537

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

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

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

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

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

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



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

567

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

584

585 DNA-based biodiversity assessment remains a relatively young field of science (Taberlet et  
586 al., 2012; Holdaway et al., 2017). Our purpose in this review is not to be overly critical of the

587 pioneering research in this field, but rather to point out areas where significant improvements  
588 can be easily made for the future. Nonetheless, our evaluation of existing eDNA studies of  
589 biodiversity does not suggest a robust, reproducible field of science. Peer reviewers and  
590 editors have a responsibility to ensure that the methods published in papers are sufficiently  
591 well described to permit future researchers to understand and replicate a study, either in the  
592 main text, in cited protocols, or in electronic supplements. Methods should be evaluated to  
593 ensure they are robust to address the research topic and more general future use of the data.  
594 Ultimately, however, it is up to individual researchers to ensure that the design and execution  
595 of field sampling receives the same attention to detail and care that is currently focussed on  
596 laboratory methods and bioinformatics.

597

598

### 599 **Acknowledgements**

600

601 This work was the outcome of a workshop funded as part of the Biological Heritage National  
602 Science Challenge, New Zealand. We thank all participants in the workshop, four reviewers  
603 and Alex Dumbrell for helpful comments. Additional funding from the BioProtection  
604 Research Centre supported IAD and AM, funding from the Australian Research Council  
605 supported JRP, and MBIE funding to ESR supported LW.

606

### 607 **Data availability**

608 The complete data is in table S1.

609

### 610 **References**

611

- 612 Allen RB (1993) A permanent plot method for monitoring changes in indigenous forests: a  
613 field manual. *Manaaki Whenua Landcare Research*.  
614 <https://nvs.landcareresearch.co.nz/content/permpplotmethod.pdf> (last accessed 28  
615 February, 2018).
- 616 Amend A (2014) From dandruff to deep-sea vents: *Malassezia*-like fungi are ecologically  
617 hyper-diverse. *PLoS Pathogens* **10**, e1004277.
- 618 Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjaer KH, Orlando L,  
619 Gilbert MTP, Willerslev E (2012) Meta-barcoding of ‘dirt’ DNA from soil reflects  
620 vertebrate biodiversity. *Molecular Ecology* **21**, 1966–1979.
- 621 Anderson DR (2001) The need to get the basics right in wildlife field studies. *Wildlife Society*  
622 *Bulletin (1973-2006)* **29**, 1294–1297.
- 623 Baldwin DS, Colloff MJ, Rees GNA, Chariton AA, Watson GO, Court LNA, Hartley DM,  
624 Morgan MJ, King AJ, Wilson, S. J, Hodda M, Hardy CM (2013) Impacts of inundation  
625 and drought on eukaryote biodiversity in semi-arid floodplain soils. *Molecular Ecology*  
626 **22**, 1746–1758.
- 627 Barberan A, Casamayor EO (2014) A phylogenetic perspective on species diversity, beta-  
628 diversity and biogeography for the microbial world. *Molecular Ecology* **23**, 5868–5876.
- 629 Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, Azpiroz F, Roca J, Guarner F,  
630 Manichanh C (2012) Storage conditions of intestinal microbiota matter in metagenomic  
631 analysis. *BMC microbiology* **12**, 158.
- 632 Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P, Sansone S-  
633 A, Stoeckert CJ, Taylor CF, Taylor R (2011) Data standards for Omics data: the basis  
634 of data sharing and reuse. *Bioinformatics for Omics Data: Methods and Protocols* 31–  
635 69.
- 636 Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of

- 637 plants and animals. *Molecular Ecology* **21**, 1834–1847.
- 638 Crawley MJ (2015) *Statistics: an introduction using R*, 2nd ed. John Wiley & Sons Ltd.,  
639 Hoboken, NJ, USA. 339 pp.
- 640 De Beeck MO, Ruytinx J, Smits MMA, Vangronsveld J, Colpaert JV, Rineau F (2015)  
641 Belowground fungal communities in pioneer Scots pine stands growing on heavy metal  
642 polluted and non-polluted soils. *Soil Biology & Biochemistry* **86**, 58–66.
- 643 Deck J, Gaither MR, Ewing R, Bird CE, Davies N, Meyer C, Riginos C, Toonen RJ, Crandall  
644 ED (2017) The Genomic Observatories Metadatabase (GeOMe): A new repository for  
645 field and sampling event metadata associated with genetic samples. *PLoS biology* **15**,  
646 e2002925.
- 647 Denny M, Benedetti-Cecchi L (2012) Scaling up in ecology: mechanistic approaches. *Annual*  
648 *Review of Ecology, Evolution, and Systematics* **43**, 1–22.
- 649 Dickie IA, FitzJohn RG (2007) Using terminal-restriction fragment length polymorphism (T-  
650 RFLP) to identify mycorrhizal fungi; a methods review. *Mycorrhiza* **17**, 259–270.
- 651 Dickie IA, Fukami T, Wilkie JP, Allen RB, Buchanan PK (2012) Do assembly history effects  
652 attenuate from species to ecosystem properties? A field test with wood-inhabiting fungi.  
653 *Ecology Letters* **15**, 133–141.
- 654 Dickie IA, Xu B, Koide RT (2002) Vertical niche differentiation of ectomycorrhizal hyphae  
655 in soil as shown by T-RFLP analysis. *New Phytologist* **156**, 527–535.
- 656 Dickie IA, St John MG (2016) Second-generation molecular understanding of mycorrhizas in  
657 soil ecosystems. F. Martin (ed) *Molecular Mycorrhizal Symbiosis*. John Wiley & Sons  
658 Ltd., Hoboken, NJ, USA. 473–491.
- 659 Drummond AJ, Newcomb RD, Buckley TRA, Xie D, Dopheide A, Potter BCM, Heled,  
660 Joseph, Ross HA, Tooman L, Grosser SA, Park D, Demetras NJ, Stevens MIA, Russell  
661 JC, Anderson SH, Carter A, Nelson, Nicola (2015) Evaluating a multigene

- 662 environmental DNA approach for biodiversity assessment. *GigaScience* **4**, 46.
- 663 Field D, Sansone S-A, Collis A, Booth T, Dukes P, Gregurick SK, Kennedy K, Kolar P,  
664 Kolker E, Maxon M (2009) Omics data sharing. *Science* **326**, 234–236.
- 665 Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK, Kohn MH (1998) Empirical evaluation  
666 of preservation methods for faecal DNA. *Molecular Ecology* **7**, 1423–1428.
- 667 Haddaway NR, Verhoeven JTA (2015) Poor methodological detail precludes experimental  
668 repeatability and hampers synthesis in ecology. *Ecology and Evolution* **5**, 4451–4454.
- 669 Hiraoka S, Yang C-c, Iwasaki W (2016) Metagenomics and bioinformatics in microbial  
670 ecology: current status and beyond. *Microbes and Environments* **31**, 204–212.
- 671 Holdaway RJ, Wood JR, Dickie IA, Orwin KH, Bellingham PJ, Richardson SJ, Lyver PO,  
672 Timoti P, Buckley TR (2017) Using DNA metabarcoding to assess New Zealand's  
673 terrestrial biodiversity. *New Zealand Journal of Ecology* **41**, 251–262.
- 674 Irvine KM, Dinger EC, Sarr D (2011) A power analysis for multivariate tests of temporal  
675 trend in species composition. *Ecology* **92**, 1879–1886.
- 676 Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Collman RG, Bushman FD, Li H  
677 (2015) Power and sample-size estimation for microbiome studies using pairwise  
678 distances and PERMANOVA. *Bioinformatics* **31**, 2461–2468.
- 679 Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the  
680 surface of bones and teeth. *Forensic science international* **154**, 53–61.
- 681 Keshri J, Yousuf B, Mishra A, Jha, Bhavanath (2015) The abundance of functional genes,  
682 *cbbL*, *nifH*, *amoA* and *apsA*, and bacterial community structure of intertidal soil from  
683 Arabian Sea. *Microbiological Research* **175**, 57–66.
- 684 Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N (2010) Effect of storage conditions on the  
685 assessment of bacterial community structure in soil and human-associated samples.  
686 *FEMS Microbiology letters* **307**, 80–86.

- 687 Lear G, Dickie I, Banks J, Boyer S, Buckley HL, Buckley TR, Cruickshank R, Dopheide A,  
688 Handley KM, Hermans S, Kamke J, Lee CK, MacDiarmid R, Morales SE, Orlovich  
689 DA, Smissen R, Wood J, Holdaway R (2018) Methods for the extraction, storage,  
690 amplification and sequencing of DNA from environmental samples. *New Zealand*  
691 *Journal of Ecology* **42**, 10-50A.
- 692 Lennon JT (2011) Replication, lies and lesser-known truths regarding experimental design in  
693 environmental microbiology. *Environmental Microbiology* **13**, 1383–1386.
- 694 Lithgow GJ, Driscoll M, Phillips P (2017) A long journey to reproducible results. *Nature*  
695 **548**, 387–388.
- 696 Manter DK, Weir TL, Vivanco JM (2010) Negative effects of sample pooling on PCR-based  
697 estimates of soil microbial richness and community structure. *Applied and*  
698 *Environmental Microbiology* **76**, 2086–2090.
- 699 Mendes LW, Tsai SM, Navarrete AA, de, Hollander M, van Veen JA, Kuramae EE (2015)  
700 Soil-Borne Microbiome: Linking Diversity to Function. *Microbial Ecology* **70**, 255–  
701 265.
- 702 Navarrete AA, Tsai SM, Mendes LW, Faust, Karoline, de Hollander M, Cassman NA, Raes,  
703 Jeroen, van Veen JA, Kuramae EE (2015) Soil microbiome responses to the short-term  
704 effects of Amazonian deforestation. *Molecular Ecology* **24**, 2433–2448.
- 705 Olson ZH, Briggler JT, Williams RN (2013) An eDNA approach to detect eastern  
706 hellbenders (*Cryptobranchus a. alleganiensis*) using samples of water. *Wildl Res* **39**,  
707 629.
- 708 Orchard S, Standish RJ, Nicol D, Dickie IA, Ryan MH (2017) Sample storage conditions  
709 alter colonisation structures of arbuscular mycorrhizal fungi and, particularly, fine root  
710 endophyte. *Plant and Soil* **412**, 35–42.
- 711 Osborne CA, Zwart AB, Broadhurst LM, Young AG, Richardson AE (2011) The influence of

- 712 sampling strategies and spatial variation on the detected soil bacterial communities  
713 under three different land-use types. *FEMS Microbiology Ecology* **78**, 70–79.
- 714 Otypková Z, Chytrý M (2006) Effects of plot size on the ordination of vegetation samples.  
715 *Journal of Vegetation Science* **17**, 465–472.
- 716 Pansu J, Winkworth RC, Hennion FA, Gielly L, Taberlet P, Choler P (2015) Long-lasting  
717 modification of soil fungal diversity associated with the introduction of rabbits to a  
718 remote sub-Antarctic archipelago. *Biology Letters* **11**, 20150408.
- 719 Peura S, Sinclair L, Bertilsson S, Eiler, Alexander (2015) Metagenomic insights into  
720 strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes.  
721 *Scientific Reports* **5**, 12102.
- 722 Prince AM, Andrus L (1992) PCR: how to kill unwanted DNA. *Biotechniques* **12**, 358–360.
- 723 Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW,  
724 DNAY, Hofmockel KS, Kirkman KP, Knops JM, La Pierre KJ, MacDougall AS,  
725 McCulley RL, Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N  
726 (2015) Plant diversity predicts beta but not alpha diversity of soil microbes across  
727 grasslands worldwide. *Ecology Letters* **18**, 85–95.
- 728 Prosser JI (2010) Replicate or lie. *Environmental Microbiology* **12**, 1806–1810.
- 729 Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and opportunities of open data  
730 in ecology. *Science* **331**, 703–705.
- 731 Robertson BL, Brown JA, McDonald T, Jaksons P (2013) BAS: Balanced acceptance  
732 sampling of natural resources. *Biometrics* **69**, 776–784.
- 733 Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for  
734 DNA analyses. *Canadian Journal of Zoology* **69**, 82–90.
- 735 Smith ANH, Anderson MJ, Pawley MDM (2017) Could ecologists be more random?  
736 Straightforward alternatives to haphazard spatial sampling. *Ecography* **40**, 1251–1255

- 737 Soldatova LN, King RD (2005) Are the current ontologies in biology good ontologies.  
738 *Nature Biotechnology* **23**, 1095–1098.
- 739 Steinle F 2016. Stability and Replication of Experimental Results: A Historical Perspective.  
740 In: Atmanspacher H, Maasen S, eds. *Reproducibility: Principles, Problems, Practices,*  
741 *and Prospects*. John Wiley & Sons, Hoboken, NJ, USA 39 – 64.
- 742 Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Towards next-generation  
743 biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 1789–1793.
- 744 Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-  
745 Palacios AM, Thu PQ, Suija A (2014) Global diversity and geography of soil fungi.  
746 *Science* **346**, 1256688.
- 747 Terrat S, Dequiedt S, Horigue W, Lelievre M, Cruaud C, Saby NPA, Jolivet C, Arrouays D,  
748 Maron PA, Ranjard L (2015) Improving soil bacterial taxa–area relationships  
749 assessment using DNA meta-barcoding. *Heredity* **114**, 468.
- 750 Toju H, Yamamoto S, Sato H, Tanabe S, A, Gilbert GS, Kadowaki K (2013) Community  
751 composition of root-associated fungi in a *Quercus*-dominated temperate forest:  
752 codominance of mycorrhizal and root-endophytic fungi. *Ecology and Evolution* **3**,  
753 1281–1293.
- 754 Tveit A, Schwacke R, Svenning MM, Urich, Tim (2013) Organic carbon transformations in  
755 high-Arctic peat soils: key functions and microorganisms. *ISME Journal* **7**, 299–311.
- 756 Vargas-Gastelum L, Romero-Olivares AL, Escalante, E. A, Rocha-Olivares A, Brizuela C,  
757 Riquelme, Meritxell (2015) Impact of seasonal changes on fungal diversity of a semi-  
758 arid ecosystem revealed by 454 pyrosequencing. *FEMS Microbiology Ecology* **91**,  
759 fiv044.
- 760 von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity  
761 in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology*



- 762        *Reviews* **21**, 213–229.
- 763    Wilkins LGE, Rogivue A, Fumagalli LA, Wedekind C (2015) Declining diversity of egg-  
764        associated bacteria during development of naturally spawned whitefish embryos  
765        (*Coregonus* spp.). *Aquatic Sciences* **77**, 481–497.
- 766    Wisner SK, Bellingham PJ, Burrows LE (2001) Managing biodiversity information:  
767        development of New Zealand’s National Vegetation Survey databank. *New Zealand*  
768        *Journal of Ecology* **25**, 1–17.
- 769    Wood JR, Dickie IA, Moeller HV, Peltzer DA, Bonner KI, Rattray G, Wilmshurst JM (2015)  
770        Novel interactions between non-native mammals and fungi facilitate establishment of  
771        invasive pines. *Journal of Ecology* **103**, 121–129.
- 772    Yang C, Wang X, Miller JA, de Blecourt, Marleen, Ji Y, Yang C, Harrison RD, Yu, W. D  
773        (2014) Using metabarcoding to ask if easily collected soil and leaf-litter samples can be  
774        used as a general biodiversity indicator. *Ecological Indicators* **46**, 379–389.

775

776

777

778

779

780

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.

781

782

783

784

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

<b>Decision</b>	<b>Options</b>	<b>Implications</b>
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 [soils only]	Sampling a given depth, or sampling a given horizon.	Ability to link to soil 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.

785

786

787 **Table 3.** Specific recommendations and examples of good practice for the decision points  
 788 identified in Table 1.  
 789

<b>Decision</b>	<b>Recommendation</b>
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 <i>a priori</i> 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).

---

790

For Review Only

791  
792

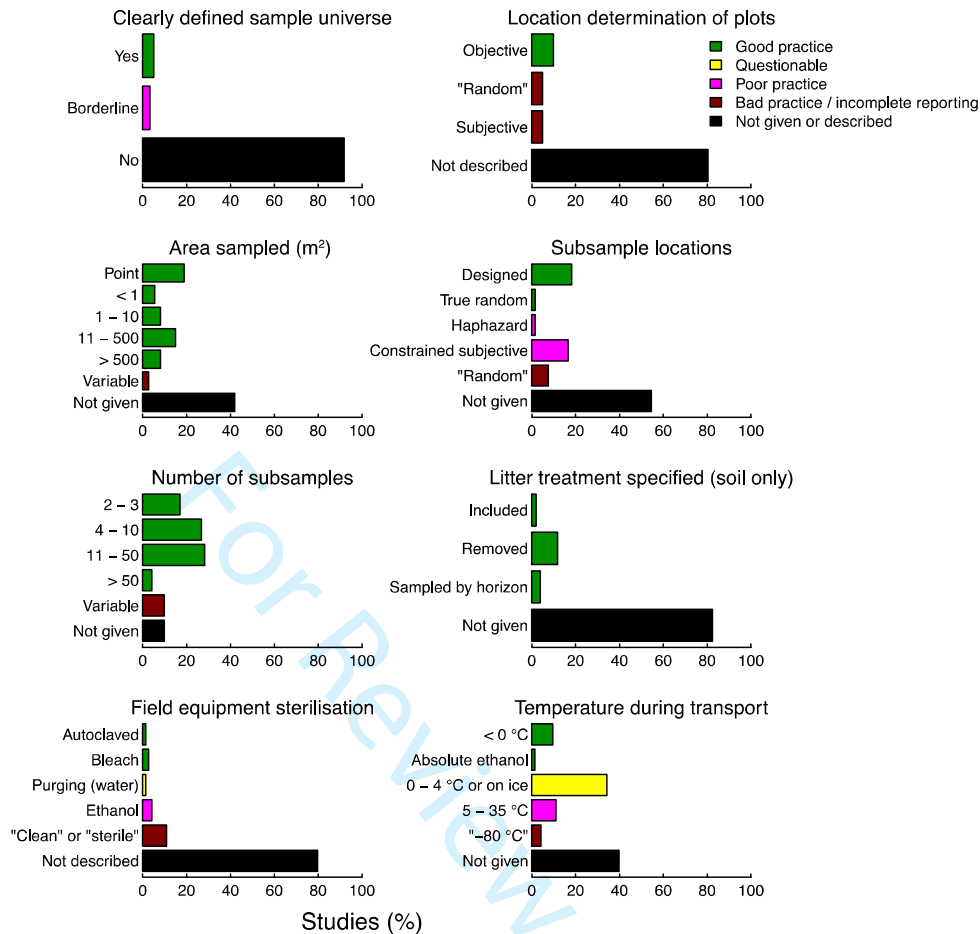


793



794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804

Figure 1. Regular placement of sampling points within plots can be biased if the sampled environment has repeating patterns, such as in this orchard (top). Conversely, regular sampling with precisely defined points is much more likely to sample unpleasant and/or difficult to sample features in proportion to their true prevalence in the environment, such as soil from under dung (bottom), than methods that allow any degree of subjectivity.



805

806

807

808

809

810

Figure 2. Summary of results for some key decision points across studies. Shading indicates whether a choice is best practice through decreasing levels of acceptability (see legend). Black indicates a failure to report methods. Method choices in quotations are reported methods where we suspect the report is not an accurate reflection (e.g., "random" with no method of randomisation, or storing samples at -80 °C immediately in the field).





Figure 1 [top]. Regular placement of sampling points within plots can be biased if the sampled environment has repeating patterns, such as in this orchard (top). Conversely, regular sampling with precisely defined points is much more likely to sample unpleasant and/or difficult to sample features in proportion to their true prevalence in the environment, such as soil from under dung (bottom), than methods that allow any degree of subjectivity.

190x142mm (300 x 300 DPI)



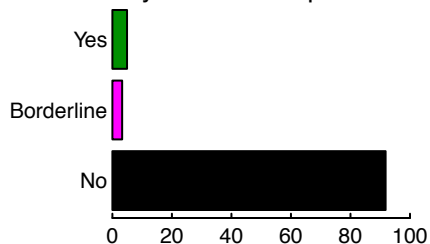


Figure 1 [bottom].

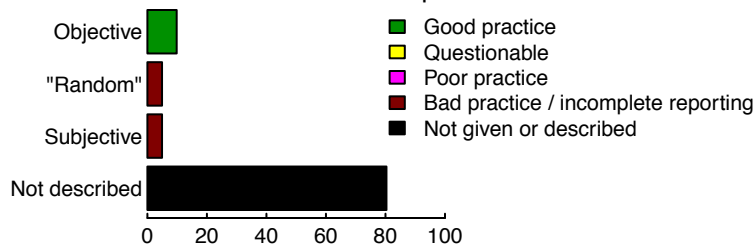
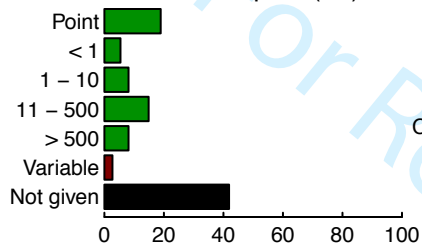
190x127mm (300 x 300 DPI)

www.Only

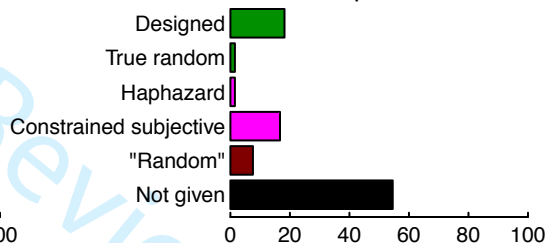
## Clearly defined sample universe



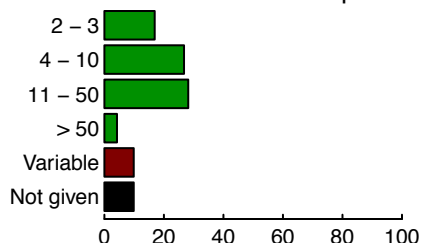
## Location determination of plots

Area sampled (m<sup>2</sup>)

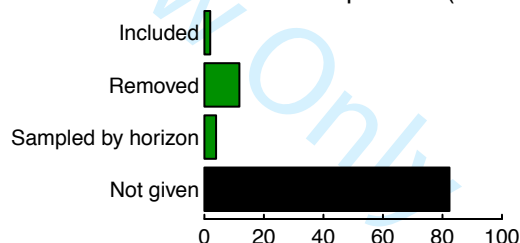
## Subsample locations



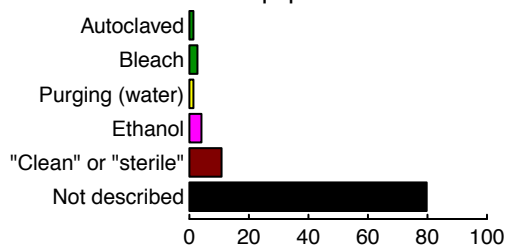
## Number of subsamples



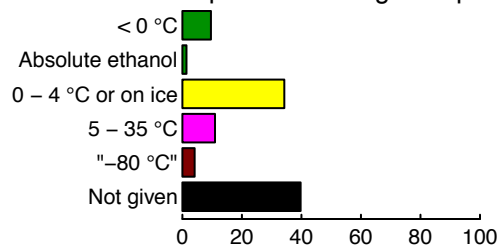
## Litter treatment specified (soil only)



## Field equipment sterilisation



## Temperature during transport



Studies (%)