Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Consequences of environmental DNA pooling

Ralph Wainer[1], Hayley Ridgway[1], Eirian Jones[2], Duane Peltzer[3], and Ian Dickie[1]

[1]Bio-Protection Research Centre
[2]Lincoln University
[3]Landcare Research New Zealand

October 20, 2020

Abstract

DNA-based techniques are increasingly used to assess biodiversity both above- and belowground. Most effort has focussed on bioinformatics and sample collection, whereas less is known about the consequences of mixing collected environmental DNA (eDNA), post-extraction and pre-PCR. We applied varying degrees of pooling to stand-alone eDNA samples collected across a non-native plant invasion density gradient, and compared the fungal communities of pooled and unpooled samples. Pooling soil eDNA decreased observable fungal rarefied richness in our samples, led to phylum-specific shifts in proportional abundance, and increased the sensitivity of detection for the invasive plant's overall impact on fungal diversity. We demonstrate that pooling fungal eDNA could change the outcome of similar eDNA studies where the aim is to: 1) identify the rare biosphere within a soil community, 2) estimate species richness and proportional abundance, or 3) assess the impact of an invasive plant on soil fungi. Sample pooling might be appropriate when determining larger-scale overarching responses of soil communities, as pooling increased the sensitivity of measurable effects of an invasive plant on soil fungal diversity.

ABSTRACT

DNA-based techniques are increasingly used to assess biodiversity both above- and belowground. Most effort has focussed on bioinformatics and sample collection, whereas less is known about the consequences of mixing collected environmental DNA (eDNA), post-extraction and pre-PCR. We applied varying degrees of pooling to stand-alone eDNA samples collected across a non-native plant invasion density gradient, and compared the fungal communities of pooled and unpooled samples. Pooling soil eDNA decreased observable fungal rarefied richness in our samples, led to phylum-specific shifts in proportional abundance, and increased the sensitivity of detection for the invasive plant's overall impact on fungal diversity. We demonstrate that pooling fungal eDNA could change the outcome of similar eDNA studies where the aim is to: 1) identify the rare biosphere within a soil community, 2) estimate species richness and proportional abundance, or 3) assess the impact of an invasive plant on soil fungi. Sample pooling might be appropriate when determining larger-scale overarching responses of soil communities, as pooling increased the sensitivity of measurable effects of an invasive plant on soil fungal diversity.

KEYWORDS: diversity, environmental DNA, experimental design, metabarcoding, sampling, soil DNA extraction

INTRODUCTION

High throughput DNA sequencing technology (Caporaso *et al.* 2012) is increasingly used for determining the composition of ecological communities, testing ecological hypotheses (Holdaway *et al.* 2017) and could revolutionize biodiversity and conservation monitoring (Lindahl *et al.* 2013). One technique in particular, DNA metabarcoding, can identify the presence of a multitude of species across a wide taxonomic range (Taberlet

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

et al. 2012), which previously could only be achieved through time-consuming morphological identification of individual organisms (Lawton *et al.* 1998). The growing use of DNA metabarcoding to sequence environmental DNA (eDNA, i.e., DNA extracted from soil, water, air or other substances) has revealed the need for robust sampling protocols and experimental designs, as well as guidelines on how to process obtained samples in the laboratory environment (Zinger *et al.* 2019). Previous reviews of metabarcoding methods have focussed on statistical replication in sampling (Lennon 2011), the processing of collected samples (Lear *et al.* 2018), as well as data reporting and bioinformatics analysis (Hiraoka *et al.* 2016). However, there are very few DNA metabarcoding studies on the effect of mixing extracted eDNA samples together, i.e., "pooling" samples prior to being sequenced. Given how common pooling is when undertaking community studies (Dickie *et al.* 2018) and that both sampling and subsequent pooling techniques underpin inferences from these data (Crawley 2015), the consequences of pooling large numbers of eDNA samples prior to sequencing deserves more attention.

Some studies suggest that pooling samples pre-PCR has little effect on community data, but decreases detected variability compared to not pooling (Manter *et al.* 2010, Osborne *et al.* 2011), making pooling economically advantageous but potentially less informative. It is possible to partially overcome the issue of decreased detected variability by amplifying several diluted subsets of a pooled sample, thereby increasing the likelihood that less abundant species are successfully amplified and detected. However in such cases, it might be better to take multiple stand-alone samples allowing for additional spatial variability analyses (Dickie *et al.* 2018). This is because rarity is spatially-dependent; species may be uncommon either by 1) having low abundance but ubiquitous distribution, or 2) by being abundant at fine scales but heterogeneously distributed (Green *et al.* 2004). Such spatial distributions could be characteristic to different fungal taxa. For instance, it has been observed that wood-inhabiting members of the fungal phylum Ascomycota are more specific to certain tree species compared with the more homogeneously distributed fungal phylum Basidiomycota (Purahong *et al.* 2018).

Some valuable insights on soil eDNA pooling have been provided via a small number of mixed samples pooled pre-extraction (Song *et al.*2015). However to the best of our knowledge, there is yet no comprehensive study which examines how sample pooling post-extraction affects the species richness and proportional abundance measurements of fungal eDNA, nor how identifying the presence of an ecological gradient (e.g., the effect on a dominant invasive plant species on belowground species richness) might be hindered or exaggerated by pooling eDNA samples. Given the cost of field sampling, wet-lab processing and sequencing, it is desirable to neither under- nor oversample when conducting an eDNA-based ecological survey. Methodologically sound eDNA sample preparation is the foundation for subsequent analyses and examining the consequences of sample pooling is one way to assess the reliability and reproducibility of eDNA surveys. Here, we applied four degrees of eDNA pooling to individual soil core extracts collected from six plots along an exotic plant's invasion gradient, followed by Illumina sequencing of indexed PCRs targeting the ribosomal internal transcribed spacer (ITS) region (Schoch *et al.* 2012) within soil fungi.

Our aim was to investigate how varying degrees of eDNA sample pooling affects species richness and proportional abundance of soil organisms, specifically fungi, and to examine whether sample pooling enhances or dampens the overarching effect of a globally-invasive exotic plant (*Cytisus scoparius*) on soil fungal communities. Species richness estimations can be highly sensitive to sampling errors and detection biases (Flynn *et al.* 2015, Dopheide *et al.* 2019) whereas comparisons of community composition tend to be less affected by these errors (Leray and Knowlton 2015, Taberlet *et al.* 2018). Alongside richness, we therefore also determine how pooling affects proportional abundances within our sampled communities, particularly at the level of fungal phylum.

Pooled eDNA samples may have a higher species richness than any individual stand-alone sample that the pooled sample is composed of, given that a pooled sample can cover a broader area. However, we suggest that this increased detection of richness occurs at the cost of a reduced detection of the rare biosphere.

We hypothesised that:

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

- Pooling eDNA samples will increase species richness, but rare species will be under-represented.
- Sample pooling will cause distortions in the proportional abundance of fungal taxa.
- As sample pooling will likely decrease within-plot variation of fungal communities, caused by multiple unaccounted abiotic and biotic factors, the use of pooled samples should therefore be more sensitive to larger scale between-plot comparisons, which in our case is the presence of an exotic plant across an invasion gradient.

To test these hypotheses, we applied varying degrees of pooling to eDNA samples collected across an exotic plant invasion gradient and compared the fungal communities of computationally pooled samples with our physically pooled samples (pooled after DNA extraction).

MATERIALS AND METHODS

Study site and field experiment

The study site was located in the Saint James conservation area in New Zealand's South Island (-42.460273 Lat., 172.830938 Long.; elevation = 800–900 m.a.s.l.; mean annual temperature = 10.3°C; mean annual rainfall = 1158 mm). A description of the site's vegetation is given in Broadbent *et al.* (2017). Permanent 20 × 20 m vegetation plots were laid out at the site by Manaaki Whenua – Landcare Research, following standard field protocols (Hurst and Allen 1993). We selected six plots across a *C. scoparius* density gradient, all within 1 km of each other. Field sampling took place from 14 February 2017 to 17 April 2017. From each plot, 24 individual georeferenced soil cores were taken, totalling in 144 spatially explicit soil samples. Soil samples were dug up vertically to a depth of 150–200 mm, each soil sample weighing 200–250 g. Litter and leaf matter was discarded, and measurements of broom abundance (% cover) were taken at each extraction point.

Each soil sample (stored at 4°C) was broken up manually and a ˜10 g mixed soil sample was obtained by systematically extracting 10 × ˜1 g subsamples (Figure S1). Any roots or stones larger than 5 mm in diameter were excluded from the mixed subsampled soil, which was kept frozen at -18°C until DNA extraction.

Experimental design and wet-lab processing

Both the kit used for soil DNA extraction and the chosen fungal primers were recommended by Lear *et al.* (2018). DNA extraction was performed on the 144 soil cores using DNeasy PowerSoil® HTP 96 Kits (Quiagen), according to the manufacturer's instructions. Five µL subsamples of the 144 stand-alone soil extracts were then mixed in equal proportions to create pooled soil samples, composed of 3, 6, 12 or 24 combined soil extracts (Figure S2). We chose the pooling partitions in such a way that smaller partitions "fitted" into larger ones. To avoid potential pipetting errors, pools of three samples were first created (8 × 3-sample pools per plot), which were then used to produce pools of six samples by combining equal quantities of two 3-sample pools. Pools of six were in turn mixed together to create pools of 12, which were likewise combined to create a pool of all 24 samples extracted from a plot. Following this method, a sum of 15 pooled samples were created per plot (alongside the 24 individual samples per plot) and 90 pooled samples were prepared in all along with 144 individual samples. A total of 234 eDNA extracts (both pooled and individual) underwent the same following PCR amplification steps.

Based on amplification protocols outlined by the Earth Microbiome Project (Gilbert *et al.* 2014), two single-indexed DNA libraries were assembled using the fITS7 general fungal primer (5'- GTG ART CAT CGA ATC TTT G -3') (Ihrmark *et al.* 2012) and the ITS4 reverse primer (5'- TCC TCC GCT TAT TGA TAT GC -3') (White *et al.* 1990). Amplifications were performed in a 25 µL volume containing 0.2 µL FastStart™ DNA polymerase (Merck), 0.5 µL dNTP mixture (10 mM each), 2.5 µL PCR buffer (with 20 mM $MgCl_2$), 2 µL 2.5 µM of each forward and reverse primer, 1.25 µL 10 µM molecular grade Bovine Serum Albumin, 1 µL 10× diluted DNA template and 15.55 µL filtered deionized water. PCR conditions were denaturation step of 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 57°C and 30 s at 72°C, with a final step at 72°C for 7 min (and held at 4°C). All PCRs were carried out in duplicate along with positive and negative controls. Duplicate PCR product were combined prior to normalization. SequalPrep™ Normalization Plates

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

(ThermoFisher Scientific) were used to both clean and normalize. PCR product was eluted in 12 μL elution buffer (reduced from 20 μL to increase final concentration). Illumina MiSeq^{TM}sequencing (run option: 2×250 base PE v2) (Caporaso *et al.* 2012) was carried out by Massey Genome Service, New Zealand.

Bioinformatics and statistical analysis

We merged forward and reverse Illumina reads using a 32-bit version of USEARCH v11.0.667 (Edgar 2010). We removed any sequences with less than 200 bp or which had more than one expected error using VSEARCH 2.10.4 (Rognes *et al.* 2016). Singletons and doubletons were removed (Leray and Knowlton 2017), and remaining sequences clustered to 97% similarity. OTUs were matched using BLAST v2.5.0+ (Altschul *et al.* 1997) against the UNITE public database (accessed July 2019) (Nilsson *et al.* 2018). We removed all recorded OTUs which were not within the kingdom Fungi and all OTUs which had a <200 bp match to any known species. Extraction blanks, and positive and negative controls were checked for contamination and OTUs which were found within our negative controls (0.34% of all OTUs) were deleted. In order to further limit the effects of PCR and sequencing artefacts (Vesty*et al.* 2017), we removed any OTU that occurred less than three times in a sample. We also deleted any sample which had <1000 reads (0.23% of samples).

We used R version 3.5.0 (Team 2013) for creating graphs and conducting analyses. We created *in silico* samples which were pooled computationally to correspond with our physically pooled samples. When examining the composition of our samples according to fungal phylum, we generated randomly rarefied community versions of our dataset via the*rrarefy* function in "vegan" (Oksanen *et al.* 2013) before measuring the proportion belonging to a specific fungal phylum. This process was iterated 100 times and a mean proportion across iterations was calculated for each fungal phylum.

To quantify the effects of pooling on fungal rarefied richness and proportional abundance, we used linear mixed-effect models via the R package "lme4 (v1.121)" (Bates *et al.* 2014), setting sampling plot as a random effect. When fitting our linear mixed-effect models to rarefied richness over *C. scoparius* % coverage, we optimised our model using lme4's *update* function, reducing the model by excluding the interaction when the P value for the interaction term was higher than 0.05.

RESULTS

In total there were 3471 fungal OTUs, the three most dominant fungal phyla and subphyla being Ascomycota (2039 OTUs, 58.7%), Basidiomycota (1048 OTUs, 30.2%) and Mortierellomycotina (119 OTUs, 3.4%) (Figure S3; Table S4).

Effect of pooling on richness

Across all plots, physically pooling eDNA samples prior to PCR amplification increased the rarefied richness of pooled samples compared to individual samples, yet the rarefied richness was highest for each plot when the individual samples were computationally combined to correspond with the pooled samples (Figure 1). The higher the degree of physical pooling (i.e., the more stand-alone samples a pooled sample was initially composed of), the higher the loss in rarefied species richness when compared to individual samples pooled computationally. The loss in rarefied species richness remained prevalent when examining individual fungal phyla (Figure S6).

Effects of pooling on composition

Again across all plots, pooling eDNA samples shifted the proportional abundance of three out of six tested fungal taxa, which consisted of the most dominant fungal phyla and subphyla in the dataset (Figure 2). Any degree of eDNA pooling resulted in a downwards shift in the proportional abundance of Ascomcota and upward shift in the proportional abundance of Basidiomycota and Mortierellomycotina, with no significant effect of the number of samples in a pool. Pooling showed no significant effect on the proportional abundance of Glomeromycotina, Chytridiomycota or Mucoromycotina.

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Fungal phyla which occurred at relatively lower frequencies were less detectable the higher the degree of pooling (Table 1); yet the proportion of Mortierellomycotina OTUs increased the higher the degree of pooling (Figure S8), which could also be observed for each of the 6 plots individually. A similar general pattern was observed when considering propotional rank sequence abundance of all plots simultaneously (Figure S9), or of each plot individually. Whereas stand-alone samples were dominated by Ascomycota, when physically pooled, Basidiomycota and Mortierellomycotina both increased in rank sequence abundance. When considering the OTUs with the highest rank abundance within all individual samples ($n = 143$), 9/10 were Ascomycota, yet only 2/10 of the most abundant OTUs were Ascomycota when analysing all pooled eDNA samples ($n = 90$).

Effects of pooling on perceived impact of C. scoparius

C. scoparius coverage increased rarefied diversity for all fungi and for the three most abundant fungal phyla independently. Different levels of pooling showed dissimilar responses of rarefied fungal diversity to *C. scoparius* coverage (Figure S10), with higher levels of pooling showing a steeper upward trend for all fungi, and for Basidiomycota and Mortierellomycotina, yet not for Ascomycota. When considering individual samples, *C. scoparius* coverage had little effect on the rarefied richness of Mortierellomycotina, yet when considering pooled samples, the rarefied richness of Mortierellomycotina increased with *C. scoparius* coverage.

As *C. scoparius* coverage increased, the rarefied richness of Basidiomycota, the second most abundant fungal phylum, decreased in proportion to all fungal phyla when pooled computationally yet increased when pooled physically (Figure S11). Also, as *C. scoparius*coverage increases, the proportion of Mortierellomycotina relative to all fungal phyla decreased except for when the samples were pooled by 12 or by 24, in which case *C. scoparius* coverage had negligible effect.

DISCUSSION

We demonstrate that pooling soil eDNA pre-PCR will decrease observable fungal rarefied richness and lead to fungal phylum-specific shifts in proportional abundance compared with computationally pooled soil eDNA extracts. Simultaneously, pooling increased the observed sensitivity of overall fungal diversity to increasing abundance of an invasive plant. Pooling fungal eDNA might change the outcome of similar eDNA studies where the aim is to identify the rare biosphere within a soil community, to estimate species richness and proportional abundance, or to assess the impact of an invasive plant on soil fungi.

Effect of pooling on richness

Soil commonly harbours a large diversity of microorganisms in close proximity, in part due to habitat heterogeneity at a fine scale (Kang and Mills 2006). It was therefore expected that combining multiple samples captures more OTU richness compared to individual samples (Song*et al.* 2015). Despite a pooled sample encompassing a larger area, pooled eDNA has also been known to have a lower fungal species richness when compared to individual samples (Branco *et al.* 2013). This decrease in species richness is due to rare taxa being poorly represented in pooled samples (Ohman and Lavaniegos 2002). One discriminatory trait of PCR is its competitive process (Siebert and Larrick 1992), where the level of amplification achieved is positively correlated with the amount of starting template DNA. Pooling is likely to dilute rare DNA templates to the extent that amplification of rare templates may be insufficient for detection.

Effects of pooling on composition

Different fungal phyla are known to have broader or more restricted distributions (Purahong *et al.* 2018, Zebarth *et al.* 2018). The shifts in proportional abundance observed in our data, where pooling decreased the proportional abundance of the community's most dominant fungal phylum (Ascomycota) in favour of less common taxa (Basidiomycota and Mortierellomycotina), may be driven by Ascomycota having more spatially restricted, locally dominant OTUs. On the other hand, OTUs of Basidiomycota and Mortierellomycotina may occur more homogenously distributed in a given area, yet with lower dominance when present. Pooling homogenously distributed OTUs might result in an eDNA mixture where spatially restricted OTUs appear less abundant.

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Effects of pooling on perceived impact of C. scoparius

The significant interaction of *C. scoparius* cover and perceived species richness as a function of pooling suggests that pooling may be a useful technique when studying large-scale effects such as the impacts of an invasive plant on soil communities, as suggested by Ellingsøe & Johnsen (2002). Pooling reduced the variability between samples, thereby providing a more general expression of the overall community structure in a given plot (Ellingsøe and Johnsen 2002, Manter *et al.* 2010).

The differing results based on pooling, with pooled samples showing an increase in the proportion of Basidiomycota (relative to all fungi) as*C. scoparius* increases whereas stand-alone samples showed a decrease, is concerning. This suggests that dissimilar (and in this case, mutually exclusive) ecological effects are observed depending on sample processing. The question remains as to which result more accurately reflects the effects of *C. scoparius* invasion on fungal communities. One possibility is that pooling reflects changes primarily in widespread OTUs, whereas stand-alone samples are driven more by spatially restricted OTUs.

Conclusions and applications

Pooling is a common practice in eDNA studies (Dickie *et al.*2018), both at the level of plot (Osborne *et al.* 2011) or within subset categories such as soil depth (Tveit *et al.* 2013). Although there have been two notable cases where pooling before or after PCR was observed to have little effect on the perceived community (Manter *et al.* 2010, Osborne *et al.* 2011), in both cases this was specific to bacteria whereas fungi may be more susceptible to pooling effects because of their more spatially heterogeneous distribution compared to bacteria (Manter *et al.* 2010). Such a patchy fungal spatial heterogeneity, which has long been observed at fine scales (Horton and Bruns 2001), could be an underlying factor causing locally dominant but spatially rare taxa to become overly diluted in pooled eDNA samples, rendering them untraceable to metabarcoding techniques.

The decision whether or not to pool can be highly context-dependent and relies on weighing up costs of field sampling, DNA extraction, wet-lab processing and the sequencing technology (or other approach) used to identify the samples. The spatial heterogeneity of the studied organisms as well as the trade-offs between increased replication and improved precision per replicate need to be likewise taken into account (Dickie*et al.* 2018). It is uncommon in eDNA studies that multiple stand-alone samples are taken within a statistical replicate (Dickie*et al.* 2018), even so, such an approach should be encouraged if both within and between plot variability in community composition is to be examined (Drummond *et al.* 2015, Navarrete *et al.* 2015).

When considering the possible benefits of pooling, particularly in relation to how *C. scoparius* coverage increases fungal rarefied richness, the intra-plot variance which pooling decreases may not necessarily be 'distracting noise' (Ranjard *et al.* 2003), but potentially valuable ecological information. However, if the goal of a study is to test large scale patterns, unexplained intra-plot variance can be an obvious hurdle. With this in mind, we have two recommendations regarding the use of eDNA pooling:

- If the objective of an eDNA survey is study the rare biosphere or the proportional abundance of fungi in a given environment, then our results support Lear *et al.* (2018)'s recommendation that sample pooling should be avoided in favour of analysing more small subsamples. This is particularly important for some taxa: Ascomycota, Mucoromycota, Glomeromycotina or Chytridiomycota.
- If however the objective of an eDNA survey is to study a large scale overarching effect across several plots such as that of an invasive plant species on fungal communities, then our results support Ellingsøe and Johnsen (2002)'s recommendation to use larger sample sizes, or in our case pooled samples, as these reduce intra-plot variation allowing broader-scale effects to override the abundant fine-scale variation present in fungal communities (Dickie *et al.* 2002, Tedersoo *et al.* 2003, Taylor *et al.* 2014).

ACKNOWLEDGEMENTS

This project was supported by Centre of Research Excellence funding to the Bio-Protection Research Centre from the Tertiary Education Commission and funding from the Biological Heritage National Science Challenge. Thank you to M. St John, C. Morse and K. Orwin at Manaaki Whenua – Landcare Research for

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

access to their permanent sampling plots at Molesworth Station. We are also grateful to Xiao Xiao Lin and the team at Massey Genome Service for their support with sequencing. The authors declare no conflict of interest.

REFERENCES

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25** :3389-3402.

Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, and B. Dai. 2014. Linear mixed-effects models using Eigen and S4. R package version 1.

Branco, S., T. D. Bruns, and I. Singleton. 2013. Fungi at a small scale: spatial zonation of fungal assemblages around single trees. *PLoS One* **8** .

Broadbent, A. A., K. H. Orwin, D. A. Peltzer, I. A. Dickie, N. W. Mason, N. J. Ostle, and C. J. Stevens. 2017. Invasive N-fixer impacts on litter decomposition driven by changes to soil properties not litter quality.*Ecosystems* **20** :1151-1163.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, and M. Bauer. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6** :1621-1624.

Crawley, M. 2015. Statistics. An Introduction using R. 2nd rd. Wiley.

Dickie, I. A., S. Boyer, H. L. Buckley, R. P. Duncan, P. P. Gardner, I. D. Hogg, R. J. Holdaway, G. Lear, A. Makiola, and S. E. Morales. 2018. Towards robust and repeatable sampling methods in eDNA-based studies.*Molecular Ecology Resources* **18** :940-952.

Dickie, I. A., B. Xu, and R. T. Koide. 2002. Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* **156** :527-535.

Dopheide, A., D. Xie, T. R. Buckley, A. J. Drummond, and R. D. Newcomb. 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and Evolution***10** :120-133.

Drummond, A., R. Newcomb, T. Buckley, D. Xie, A. Dopheide, B. Potter, J. Heled, H. Ross, L. Tooman, and S. Grosser. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment.*GigaScience* **4** , 46.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26** :2460-2461.

Ellingsoe, P., and K. Johnsen. 2002. Influence of soil sample sizes on the assessment of bacterial community structure. *Soil Biology and Biochemistry* **34** :1701-1707.

Flynn, J. M., E. A. Brown, F. J. Chain, H. J. MacIsaac, and M. E. Cristescu. 2015. Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecology and Evolution***5** :2252-2266.

Gilbert, J. A., J. K. Jansson, and R. Knight. 2014. The Earth Microbiome project: successes and aspirations. *BMC Biology* **12** :69.

Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie. 2004. Spatial scaling of microbial eukaryote diversity. *Nature* **432** :747.

Hiraoka, S., C.-c. Yang, and W. Iwasaki. 2016. Metagenomics and bioinformatics in microbial ecology: current status and beyond.*Microbes and Environments* :ME16024.

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Holdaway, R. J., J. R. Wood, I. A. Dickie, K. H. Orwin, P. J. Bellingham, S. J. Richardson, P. O. B. Lyver, P. Timoti, and T. R. Buckley. 2017. Using DNA metabarcoding to assess New Zealand's terrestrial biodiversity. *New Zealand Journal of Ecology***41** :251-262.

Horton, T. R., and T. D. Bruns. 2001. The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. *Molecular Ecology* **10** :1855-1871.

Hurst, J., and R. Allen. 1993. A permanent plot method for monitoring indigenous forests: field protocols. Landcare Research New Zealand.

Ihrmark, K., I. Bodeker, K. Cruz-Martinez, H. Friberg, A. Kubartova, J. Schenck, Y. Strid, J. Stenlid, M. Brandstrom-Durling, and K. E. Clemmensen. 2012. New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* **82** :666-677.

Kang, S., and A. L. Mills. 2006. The effect of sample size in studies of soil microbial community structure. J*ournal of Microbiological Methods* **66** :242-250.

Lawton, J. H., D. Bignell, B. Bolton, G. Bloemers, P. Eggleton, P. Hammond, M. Hodda, R. Holt, T. Larsen, and N. Mawdsley. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391** :72.

Lear, G., I. Dickie, J. Banks, S. Boyer, H. L. Buckley, T. R. Buckley, R. Cruickshank, A. Dopheide, K. M. Handley, and S. Hermans. 2018. Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *New Zealand Journal of Ecology***42** :10-50A.

Lennon, J. T. 2011. Replication, lies and lesser-known truths regarding experimental design in environmental microbiology. *Environmental Microbiology* **13** :1383-1386.

Leray, M., and N. Knowlton. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity.*Proceedings of the National Academy of Sciences***112** :2076-2081.

Leray, M., and N. Knowlton. 2017. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding.*PeerJ* **5** :e3006.

Lindahl, B. D., R. H. Nilsson, L. Tedersoo, K. Abarenkov, T. Carlsen, R. Kjoller, U. Koljalg, T. Pennanen, S. Rosendahl, and J. Stenlid. 2013. Fungal community analysis by high-throughput sequencing of amplified markers–a user's guide. *New Phytologist* **199** :288-299.

Manter, D. K., T. L. Weir, and J. M. Vivanco. 2010. Negative effects of sample pooling on PCR-based estimates of soil microbial richness and community structure. *Applied and Environmental Microbiology***76** :2086-2090.

Navarrete, A. A., S. M. Tsai, L. W. Mendes, K. Faust, M. de Hollander, N. A. Cassman, J. Raes, J. A. van Veen, and E. E. Kuramae. 2015. Soil microbiome responses to the short-term effects of Amazonian deforestation. *Molecular Ecology* **24** :2433-2448.

Nilsson, R. H., K.-H. Larsson, A. F S. Taylor, J. Bengtsson-Palme, T. S. Jeppesen, D. Schigel, P. Kennedy, K. Picard, F. O. Glockner, L. Tedersoo, I. Saar, U. Koljalg, and K. Abarenkov. 2018. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research***47** :D259-D264.

Ohman, M. D., and B. E. Lavaniegos. 2002. Comparative zooplankton sampling efficiency of a ring net and bongo net with comments on pooling of subsamples. California Cooperative Oceanic Fisheries Investigations Report:162-173.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. O'hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. Package 'vegan'. Community ecology package, version 2:1-295.

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Osborne, C. A., A. B. Zwart, L. M. Broadhurst, A. G. Young, and A. E. Richardson. 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.

Purahong, W., T. Wubet, D. Kruger, and F. Buscot. 2018. Molecular evidence strongly supports deadwood-inhabiting fungi exhibiting unexpected tree species preferences in temperate forests. *The ISME Journal* **12** :289-295.

Ranjard, L., D. P. Lejon, C. Mougel, L. Schehrer, D. Merdinoglu, and R. Chaussod. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. *Environmental Microbiology***5** :1111-1120.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahe. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ***4** :e2584.

Schoch, C. L., K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen, and F. B. Consortium. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences***109** :6241-6246.

Siebert, P. D., and J. W. Larrick. 1992. Competitive PCR. *Nature***359** :557-558.

Song, Z., D. Schlatter, P. Kennedy, L. L. Kinkel, H. C. Kistler, N. Nguyen, and S. T. Bates. 2015. Effort versus reward: preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS One* **10** .

Taberlet, P., A. Bonin, E. Coissac, and L. Zinger. 2018. Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21** :2045-2050.

Taylor, D. L., T. N. Hollingsworth, J. W. McFarland, N. J. Lennon, C. Nusbaum, and R. W. Ruess. 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning.*Ecological Monographs* **84** :3-20.

Team, R. C. 2013. R: A language and environment for statistical computing.

Tedersoo, L., U. Koljalg, N. Hallenberg, and K. H. Larsson. 2003. Fine scale distribution of ectomycorrhizal fungi and roots across substrate layers including coarse woody debris in a mixed forest. *New Phytologist* **159** :153-165.

Tveit, A., R. Schwacke, M. M. Svenning, and T. Urich. 2013. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *The ISME Journal* **7** :299-311.

Vesty, A., K. Biswas, M. W. Taylor, K. Gear, and R. G. Douglas. 2017. Evaluating the impact of DNA extraction method on the representation of human oral bacterial and fungal communities. *PLoS One***12** .

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications **18** :315-322.

Zebarth, B., C. Goyer, S. Neupane, S. Li, A. Mills, S. Whitney, A. Cambouris, and I. Perron. 2018. Soil microbial communities have distinct spatial patterns in agricultural fields.

Zinger, L., A. Bonin, I. G. Alsos, M. Balint, H. Bik, F. Boyer, A. A. Chariton, S. Creer, E. Coissac, and B. E. Deagle. 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology* **28** :1857-1862.

Data accessibility statement: We confirm that should the manuscript be accepted, the data will be archived in an appropriate public repository (Dryad or Figshare) and the data DOI included at the end of the article.

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Author Contributions: IAD obtained funding. RFW, HJR, EEJ, and IAD designed the experiment. RFW set up the experiment and collected data using permanent sampling plots laid out by IAD and DAP. RFW conducted analyses and wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

Hosted file

`image1.emf` available at https://authorea.com/users/368776/articles/487801-consequences-of-environmental-dna-pooling

Table 1. Taxonomic composition of OTUs across all plots, split into the top 3rd ("dominant"), middle 3rd ("core") and bottom 3rd("rare") proportional rank abundance percentile for each degree of pooling. Note that "0" values are in bold to accentuate the loss of rare fungal taxa the higher the degree of pooling. The same data is visualised in Figure S8 in terms of proportions.

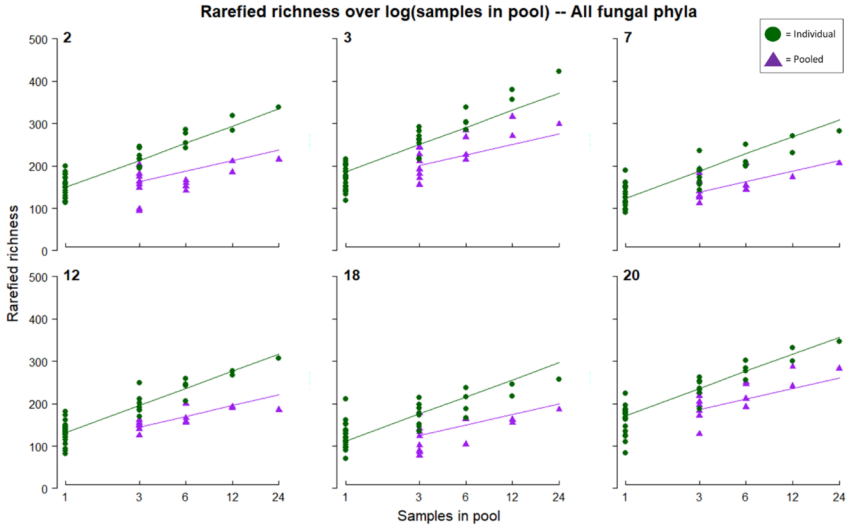

Figure 1. Rarefied richness over number of samples per pool (note log scale axis) for all fungal phyla. Purple triangles denote the physical pooling of samples whereas green circles represent the corresponding individual samples, the rarefied diversity of which has been pooled computationally to correspond with the physically pooled samples. Individual plot numbers are indicated on the top left of each graph and lines follow linear mixed-effect model fit. The P -value estimates in the linear mixed-effect model were all $<$ 0.0001. Accompanying t -values are compiled in Table S5.

Figure 2. *[Next page]* Proportional abundance by fungal phylum over number of samples per pool (note log scale axis). Purple triangles denote the physical pooling of samples whereas green circles represent the corresponding individual samples, the proportional abundance of which has been pooled computationally to correspond with the physically pooled samples. Individual plot numbers are indicated on the top left of each graph and lines follow linear mixed-effect model fit. The P -value estimates in the linear mixed-effect model are presented in the below table (accompanying t -values are compiled in Table S7).

Hosted file

`image3.emf` available at https://authorea.com/users/368776/articles/487801-consequences-of-environmental-dna-pooling

Posted on Authorea 20 Oct 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.160322420.00370902/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.