Perspectives on the Impact of Sampling Design and Intensity on Soil Microbial Diversity Estimates

Syrie M. Hermans¹, Hannah L. Buckley² and Gavin Lear*¹

¹ School of Biological Sciences, The University of Auckland, Auckland, New Zealand, ² School of Science, Auckland University of Technology, Auckland, New Zealand

Soil bacterial communities have long been recognized as important ecosystem components, and have been the focus of many local and regional studies. However, there is a lack of data at large spatial scales, on the biodiversity of soil microorganisms; national or more extensive studies to date have typically consisted of low replication of haphazardly collected samples. This has led to large spatial gaps in soil microbial biodiversity data. Using a pre-existing dataset of bacterial community composition across a 16-km regular sampling grid in France, we show that the number of detected OTUs changes little under different sampling designs (grid, random, or representative), but increases with the number of samples collected. All common OTUs present in the full dataset were detected when analyzing just 4% of the samples, yet the number of rare OTUs increased exponentially with sampling effort. We show that far more intensive sampling, across all global biomes, is required to detect the biodiversity of soil microorganisms. We propose avenues such as citizen science to ensure these large sample datasets can be more realistically achieved. Furthermore, we argue that taking advantage of pre-existing resources and programs, utilizing current technologies efficiently and considering the potential of future technologies will ensure better outcomes from large and extensive sample surveys. Overall, decreasing the spatial gaps in global soil microbial diversity data will increase our understanding on what governs the distribution of soil taxa, and how these distributions, and therefore their ecosystem contributions, will continue to change into the future.

Keywords: soil bacteria, biogeography, biodiversity, national datasets, global datasets

The geographic ranges of biological species, and therefore the biodiversity of ecosystems, are continually changing over ecological and evolutionary timescales. The collation of national and international databases has proven vital to better understand patterns in current species distributions, supporting evidence-based conservation efforts (Jetz et al., 2012), and to predict species range-shifts under, for example, climatic change and future land use scenarios (Thomas et al., 2004; Pompe et al., 2008). However, although climate and land use projections are increasingly highly resolved, often at resolutions of a few kilometers or finer (Chen et al., 2015; Abatzoglou et al., 2018), the spatial grain of resolution for most known species distributions remains far coarser (Jetz et al., 2012). Microorganisms, the most abundant group of organisms on Earth, are key players in global biogeochemical cycles, yet only limited attempts have been made to characterize their distributions across wide geographic ranges using analyses of large datasets. This is especially true for soil microbial communities, where environmental heterogeneity leads to many distinct microbial habitats (Fierer, 2017), and global dissimilarities in soil physico-chemical characteristics
present unique considerations to ensure accurate cataloging of their diversity across landscapes, regions and continents. Substantial efforts are required to reduce the gaps in soil microbial diversity data, which will require studies with adequate sampling depth across all global biomes.

Systematic surveys of microbial life are essential for providing new perspectives on bacterial distributions and the causal processes driving these patterns. Understandably, the significant effort and costs associated with consistently sampled national or global studies means it is common to see research that covers large spatial extents, but with spatially irregular sampling and relatively low replication. Even the most extensive national-scale datasets of soil bacterial biogeography, such as surveys of the British Isles (Griffiths et al., 2011) and Australia (Bissett et al., 2016), use non-uniform sampling designs, and may comprise of sample replication that is biased toward more populated and/or accessible areas. To avoid or account for these biases, random or regular (e.g., grid-based) sampling is considered desirable, but is rarely attempted (Powell et al., 2015; Terrat et al., 2017). Therefore, to inform approaches for expanding global soil microbial datasets, it is useful to understand the effects of these alternative sampling approaches on our estimations of bacterial soil community structure.

Comprehensive global soil microbial biodiversity datasets must be assembled from regional studies; however, the relative comparability and compatibility of regional datasets will determine how useful a given global dataset would be. Thus, here, we explore a dataset that does not suffer from the usual sampling limitations of many regional datasets in the published literature to determine the possible effects of variation in sampling design and replication on detection of soil microbial biodiversity. Using bacterial community data collected across a 16-km regular sampling grid within France as part of the French Soil Quality Monitoring Network (Ranjard et al., 2010; Terrat et al., 2017), we quantified the effects of sampling strategy and intensity for soil bacterial biodiversity estimates (see Supplementary Appendix S1). This analysis shows that the most common OTUs were, in fact, detectable from the analysis of only ∼4% samples collected (Figure 1A, as indicated by the plateau of the curve). This is largely irrespective of whether samples were collected from random locations, in a regular grid format, or proportionally to represent the natural diversity of soil environments (Figures 1B, 2). This pattern held true, even if a geographic subset of the dataset was analyzed (Supplementary Appendix S3 and Supplementary Figure S2). The dominance of a relatively small number of bacterial taxa is similarly reported at the global scale (Delgado-Baquerizo et al., 2018b). Variation in sampling design and intensity that is commonly observed among regional datasets may therefore not be an important consideration for capturing common and dominant bacterial taxa at a global scale.

While intensive sampling of local environments may not be required for the detection of many common taxa, sampling intensity significantly impacts community diversity measures, largely caused by the increased detection of rare OTUs with greater sampling effort (Figure 1A). Taxa may be rare due to their low local abundance, habitat specificity or restricted geographic spread, but can have a disproportionately large influence on ecosystem processes (Jousset et al., 2017; Karimi et al., 2018). Conditionally rare bacterial taxa (Shade et al., 2014) may be more metabolically active, even when present in low abundance (Dimitriu et al., 2010), and their vast genetic resource has been shown to enhance the functionality of more abundant microbes, via the horizontal transfer of beneficial genes (Low-Décarie et al., 2015). The number of rare OTUs detected in the French dataset did not appear to be influenced by sampling design (i.e., grid vs. random sampling) but did increase with increasing sample numbers (Figure 1A). Even with the inclusion of all available samples, a complete plateau in the increase of rare OTUs detected was not reached, although the number of new OTUs detected did decrease with increasing sample numbers (Figure 1A). This suggests that decisions about sampling intensity within national biodiversity monitoring are crucial for generating datasets that will be globally comparable where the distributions of rare taxa are of interest.
Perhaps the most comprehensive and coordinated effort to catalog microbial diversity across a range of environments around the globe is the Earth Microbiome Project (EMP; (Gilbert et al., 2014), highlighting the substantial progress that can be made through cooperative research. However, even within the EMP, multiple regions are grossly under-sampled and poorly represented (Supplementary Appendix S2 and Supplementary Figure S1); the continents of Europe, Asia, South America, Africa and Australia are each represented by ≤10 spatially independent samples of microbial DNA. Similar spatial biases are evident in the study of Delgado-Baquerizo et al. (2018b; see Supplementary Appendix S1 and Supplementary Figure S1 in their article). Adding to this knowledge that the spatial scaling of variation in microbial community structure differs widely across spatial scales (Constancias et al., 2015), substantial efforts must be made to further reduce global gaps in soil bacterial diversity data. Sampling to proportionally represent the relative diversity of different soil environments, or even to over-represent rare environments, or conditions, may be required for valid statistical analysis at global scales, since different environmental gradients dominate community assembly across different biomes and spatial scales. For example, soil pH is often strongly correlated with bacterial diversity (Thomson et al., 2015), to the extent that it can be used to generate global predictions of bacterial diversity (Griffiths et al., 2015). However, there are certain biomes where this is not true, such as grasslands where instead aridity...
drives bacterial diversity (Maestre et al., 2015). Such findings highlight the importance of conducting surveys of microbial life appropriate for data analysis at multiple scales because understanding of what affects bacterial community composition at small scales cannot necessarily be extrapolated to make reliable conclusions at larger scales. Grid-based sampling designs are the most statistically powerful way to achieve this, providing that the resolution of the grid is finer than the scale processes of interest (Hirzel and Guisan, 2002; Mallarino and Wittry, 2004; Nanni et al., 2011).

Increasing the size of national and international soil microbial datasets can be achieved by increased cooperation among research facilities, and perhaps even between researchers and the general public. Taking a leaf out of the macro-organism ecologist’s handbook, pursuing a “citizen science” approach is considered particularly useful for the collection of samples from more remote areas (Bahls, 2014), although consistent and well documented sample treatments must be ensured to allow accurate comparability and reproducibility (Dickie et al., 2018). There are already many examples studies where the public has been engaged to help collect data for ecological surveys of birds, trees and tropical reef species (Mckenzie et al., 2007; Ockendon et al., 2009; Roelfsema et al., 2016). Arguably, collecting and transporting soil samples requires much less time and expertise than identifying and monitoring animals and plants. Since public engagement in macro-organism surveys has been shown to be a successful biodiversity monitoring tool (Devictor et al., 2010), and is increasingly being utilized for soil microbial surveys (e.g., microblitz1), this is an avenue worth exploring to increase global coverage of bacterial community data.

Ensuring better sampling designs and global coverage alone will not be sufficient; ecologists are increasingly interested in understanding the factors affecting the present day distributions of organisms. This requires microbial DNA to be collected in tandem with a suite of relevant physicochemical variables; however, a shortcoming of many of large-scale studies published to date is the limited range of metadata collected, as the high costs associated with exhaustive soil analyses remains a major obstacle. A notable workaround for this problem is where microbial surveys are partnered with soil physicochemical monitoring programs (Dequiedt et al., 2011; Griffiths et al., 2011; Ranjard et al., 2013; Hermans et al., 2017) which include a comprehensive list of soil nutrients, physical characteristics and heavy metal concentrations. The benefits of collecting biodiversity data alongside traditional large scale soil monitoring programs is increasingly being recognized (Orgiazzi et al., 2018). As environmental monitoring agencies become more aware of the utility for microbial data to report on the health and production potential of diverse environments, existing monitoring programs are increasingly likely to be adapted to provide valuable support for microbiological investigations, helping to identify key correlates associated with changes in community composition and taxon presence across diverse spatial and temporal scales.

Microbial ecologists have tended to describe changes in composition and diversity from DNA sequence data, often without naming individual taxa, or even groups of bacteria. Arguably, this approach has inhibited our understanding of the natural history of bacteria (Martiny and Walters, 2018). However, unlike DNA fingerprinting methods, which previously dominated large-scale molecular assessments of microbial community diversity (Gobet et al., 2014), next-generation sequencing (NGS) allows taxa to now be identified from their unique DNA barcodes and grouped at various taxonomic levels. It is essential we go beyond describing general changes in microbial community composition, to looking at individual taxa, or phylogenetic or functional groups of taxa, in more detail, in the same way that traditional ecologists studying plants and animals characterize biodiversity by describing and naming the species present (Fierer, 2017; Martiny and Walters, 2018). Encouragingly, with more paired microbial and metadata being collected, NGS technologies are beginning to be used to assess not only taxonomic data, but also to make predictions of microbial functional community attributes. The expense associated with adequately sequencing complex soil metagenomes using shotgun DNA approaches mean that although microbial functional diversity has been assessed under different biomes and land uses (Fierer et al., 2012; Mendes et al., 2015), coordinated efforts to collect metagenome data from large scale soil datasets remain extremely limited. Nevertheless, scientists can capitalize on the increased availability of soil taxonomic and associated metadata to make informed predictions of the biogeography of microbial taxa and traits. As the spatial extent and grain of soil microbial community surveys increases, the relationship between soil variables such as pH, and concentrations of nutrients or potential pollutants and the distribution and relative abundance of microbial taxa are becoming better understood (Hermans et al., 2017; Karimi et al., 2018). This allows ever stronger predictions to be made regarding the environments where specific organisms or groups of organisms might be found (Delgado–Baquerizo et al., 2018a), even for organisms that are yet to be cultured or are only known from their 16S rRNA sequences.

Rapidly improving molecular methods means we also need to consider how samples collected today can be used with technology that may not yet be available, or financially achievable, for use in large scale biodiversity monitoring methods. Technological changes are very likely to occur for how extracted DNA (or RNA) is analyzed, but improvements and changes may also occur in how raw sample material is processed. For example to extract genetic material. DNA extraction biases have repeatedly been shown to exhibit biases and limitations for different sample and organisms types (Luna et al., 2006; Wagner Mackenzie et al., 2015; Hermans et al., 2018). Future improvements to current DNA extraction techniques, or the development of new methods, could lead to desires to re-analyze previous samples to obtain more accurate representations of the microbial communities that were present. It has previously been shown that bacterial DNA can be extracted from dried soil samples over a century after the soil was stored (Clark and Hirsch, 2008), and that DNA can be maintained for months at −80°C (Gorokhova, 2005). However, more research needs to be
conducted to determine the effect of time and storage conditions on microbial community composition in raw sample material, and the degradation of DNA over years, rather than months. Following current best practice storage methods for the large sample numbers that will be generated by national, and global surveys of microbial diversity is essential. This will provide not only a ‘snapshot in time’ of the current biodiversity of soil bacteria globally, but also allow the application of future biodiversity monitoring methods without repeating the labor intensive, and expensive sampling process.

Significant progress has been made in the last decade to catalog microbial diversity across the globe, yet the lack of systematic approaches for sampling across national and global scales, is leading to unbalanced datasets which are failing to cover all of the planet’s biomes. Greater coordination among researchers, collaboration with soil monitoring agencies and the general public, could facilitate the collection of more spatially extensive and -intensive datasets. Extensive sampling of soils across the globe, to identify the microbial taxa residing within them and their functions, is essential to increase our understanding of natural variation in these communities, the effect that human land use has on microorganisms, and the impact that climatic change may have on future ecosystem function.

**AUTHOR CONTRIBUTIONS**

SH analyzed the data. All authors contributed to the writing of the manuscript.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01820/full#supplementary-material

**REFERENCES**

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and HEGEWISCH, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Sci. Data* 5:170191. doi: 10.1038/sdata.2017.191

Bahls, L. L. (2014). New diatoms from the American West—a tribute to citizen science. *Proc. Acad. Nat. Sci. Phila.* 163, 61–84. doi: 10.1635/053.163.0109

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., et al. (2016). Introducing BASE: the biomes of Australian soil environments soil microbial diversity database. *GigaScience* 5:21. doi: 10.1186/s13742-016-0126-5

Chen, J., Chen, J., Liso, A., Cao, X., Chen, L., Chen, X., et al. (2015). Global land cover mapping at 30m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* 103, 7–27. doi: 10.1016/j.isprsjprs.2014.09.002

Clark, I. M., and Hirsch, P. R. (2008). Survival of bacterial DNA and culturable bacteria in archived soils from the roothamsted broadbalk experiment. *Soil Biol. Biochem.* 40, 1090–1102. doi: 10.1016/j.soilbio.2007.11.021

Dequiedt, S., Saby, N. P. A., Horrigue, W., Villerd, J., Guillonin, J.-P., et al. (2015). Mapping and determinism of soil microbial community distribution across an agricultural landscape. *MicrobiologyOpen* 4, 505–517. doi: 10.1002/mbo3.235

Delgado-Baquerizo, M., Olivero, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., et al. (2018a). A global atlas of the dominant bacteria found in soil. *Science* 359, 320–325. doi: 10.1126/science.aap 9516

Delgado-Baquerizo, M., Reith, F., Dennis, P. G., Hamonts, K., Powell, J. R., Young, A., et al. (2018b). Ecological drivers of soil microbial diversity and soil biological networks in the Southern hemisphere. *Ecology* 99, 583–596. doi: 10.1002/ecy.2137

Dequiedt, S., Saby, N. P. A., Lelièvre, M., Jolivet, C., Thioulouse, J., Toutain, B., et al. (2011). Biogeographical patterns of soil microbial biomass as influenced by soil characteristics and management. *Glob. Ecol. Biogeogr.* 20, 641–652. doi: 10.1111/j.1466-8238.2010.00628.x

Devictor, V., Whittaker, R. J., and Beltrame, C. (2010). Beyond scarcity: citizen participation in habitat suitability modelling. *Ecol. Model.* 221, 331–341. doi: 10.1016/j.ecolmodel.2009.12.006

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

Dimitriu, P. A., Lee, D., and Grayston, S. J. (2010). An evaluation of the functional significance of peat microorganisms using a reciprocal transplant approach. *Soil Biol. Biochem.* 42, 65–71. doi: 10.1016/j.soilbio.2009.10.001

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi: 10.1038/nrmicro.2017.87

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21390–21395. doi: 10.1073/pnas.1215210110

Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The Earth microbiome project: successes and aspirations. *BMC Biol.* 12:69. doi: 10.1186/s12915-014-0069-1

Gobet, A., Boetius, A., and Ramette, A. (2014). Ecological coherence of diversity patterns derived from classical fingerprinting and next generation sequencing techniques. *Environ. Microbiol.* 16, 2672–2681. doi: 10.1111/1462-2920.12308

Gorokhova, E. (2005). Effects of preservation and storage of microcrustaceans in RNA later on RNA and DNA degradation. *Limnol. Oceanogr. Methods* 3, 143–148. doi: 10.4399/lof.2005.3.143

Griffiths, R. I., Thomson, B. C., James, P., Bell, T., Bailey, M., and Whitley, A. S. (2011). The bacterial biogeography of British soils. *Environ. Microbiol.* 13, 1642–1654. doi: 10.1111/j.1462-2920.2011.02480.x

Griffiths, R. I., Thomson, B. C., Plassart, P., Gweon, H. S., Stone, D., Creamer, R. E., et al. (2015). Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. *Appl. Soil Ecol.* 97, 1–134

Hermans, S. M., Buckley, H. L., Case, B. S., Curran-Cournane, F., Taylor, M., and Lear, G. (2017). Bacteria as emerging indicators of soil condition. *Appl. Environ. Microbiol.* 83:e02826-16. doi: 10.1128/AEM.02826-16

Hermans, S. M., Buckley, H. L., and Lear, G. (2018). Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Mol. Ecol. Resour.* 18, 557–569. doi: 10.1111/1755-0998.12762

Hirzel, A., and Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Model.* 157, 331–341. doi: 10.1016/S0304-3800(02)00203-X

Jetz, W., McPherson, J. M., and Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* 27, 151–159. doi: 10.1016/j.tree.2011.09.007

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., et al. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* 11, 853–862. doi: 10.1038/ismej.2016.174

Karimi, B., Terrat, S., Dequiedt, S., Saby, N. P. A., Horrigue, W., Lelièvre, M., et al. (2018). Biogeography of soil bacteria and archaea across France. *Sci. Adv.* 4:eaaat1808. doi: 10.1126/sciadv.aat1808

Low-Décarie, E., Kolber, M., Hommen, P., Lozano, A., Dumbrell, A., Gonzalez, A., et al. (2015). Community rescue in experimental metacommunities.
Mckenzie, A. J., Petty, S. J., Toms, M. P., and Furness, R. W. (2007). Importance of
Martiny, J. B. H., and Walters, K. E. (2018). Towards a natural history of soil
Mallarino, A. P., and Wittry, D. J. (2004). Efficacy of grid and zone soil sampling
Maestre, F. T., Delgado-Baquerizo, M., Jeffries, T. C., Eldridge, D. J., Ochoa, V.,
Gozalo, B., et al. (2015). Increasing aridity reduces soil microbial diversity and
Hermans et al. Global Soil Microbial Diversity
Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N. P. A.,
Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernández-Ugalde, O. (2018).
LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil
Pompe, S., Hanspach, J., Badeck, F., Klots, S., Thuiller, W., and Kuhn, I. (2008).
Climate and land use change impacts on plant distributions in Germany. Biol.
Powell, J. R., Karunaratne, S., Campbell, C. D., Yao, H., Robinson, L., and Singh,
B. K. (2015). Deterministic processes vary during community assembly for
ecologically dissimilar taxa. Nat. Commun. 6:8444. doi: 10.1038/ncomms9444
Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N. P. A.,
Lelievre, M., et al. (2013). Turnover of soil bacterial diversity driven by
wide-scale environmental heterogeneity. Nat. Commun. 4:1434. doi: 10.1038/ncomms12431
Ranjalr, L., Dequiedt, S., Jolivet, C., Saby, N. P. A., Thioulouse, J., Harmand, J., et al.
(2010). Biogeography of soil microbial communities: a review and a description
of the ongoing french national initiative. Agron. Sustain. Dev. 30, 359–365.
doi: 10.1051/agro/2009033
Roelfsema, C., Thurstan, R., Beger, M., Dudgeon, C., Loder, J., Kovacs, E.,
et al. (2016). A citizen science approach: a detailed ecological assessment of
subtropical reefs at point lookout, Australia. PLoS One 11:e0163407. doi: 10.1371/journal.pone.0163407
Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N.,
et al. (2014). Conditionally rare taxa disproportionately contribute to temporal
changes in microbial diversity. mBio 5:e01371-14. doi: 10.1128/mBio.01371-14
Terrat, S., Morrigue, W., Dequiedt, S., Saby, N. P. A., Lelièvre, M., Nowak, V.,
et al. (2017). Mapping and predictive variations of soil bacterial richness across France. PLoS One 12:e0186766. doi: 10.1371/journal.pone.0186766
Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J.,
Collingham, Y. C., et al. (2004). Extinction risk from climate change. Nature
427, 145–148.
Thomson, B. C., Tisserant, E., Plassart, P., Uroz, S., Griffiths, R. I.,
Hannula, S. E., et al. (2015). Soil conditions and land use intensification
effects on soil microbial communities across a range of European
field sites. Soil Biol. Biochem. 88, 403–413. doi: 10.1016/j.soilbio.2015.06.012
Wagner Mackenzie, B., Waite, D. W., and Taylor, M. W. (2015). Evaluating
variation in human gut microbiota profiles due to DNA extraction method
and inter-subject differences. Front. Microbiol. 6:130. doi: 10.3389/fmicb.2015.
00130

Conflict of Interest Statement: The authors declare that the research was
conducted in the absence of any commercial or financial relationships that could
be construed as a potential conflict of interest.

Copyright © 2019 Hermans, Buckley and Lear. This is an open-access article
distributed under the terms of the Creative Commons Attribution License (CC BY).
The use, distribution or reproduction in other forums is permitted, provided the
original author(s) and the copyright owner(s) are credited and that the original
publication in this journal is cited, in accordance with accepted academic practice. No
use, distribution or reproduction is permitted which does not comply with these terms.