Global Ecology and Biogeography

A
Mac

What drives study-dependent differences in distance-decay relationships of microbial communities?

Journal:	*Global Ecology and Biogeography*
Manuscript ID	GEB-2020-0363
Manuscript Type:	Meta-Analysis
Keywords:	Bacteria, Archaea, Eukarya, Mantel test, Macroecology, Biogeography, Dispersal limitation, Community dissimilarity

Title: What drives study-dependent differences in distance-decay relationships of microbial communities?

Running title: Meta-Analysis of Microbial Distance-Decay Relationships

Keywords: Bacteria, Archaea, Eukarya, Mantel test, macroecology, biogeography, dispersal limitation, community dissimilarity

Abstract

Aim: Ecological communities that exist closer together in space are generally more compositionally similar than those far apart, as defined by the distance-decay of similarity relationship. However, recent research has revealed substantial variability in the distance-decay relationships of microbial communities between studies of different taxonomic groups, ecosystems, spatial scales, as well as between those using different molecular methodologies (e.g. high-throughput sequencing versus molecular fingerprinting). Here, we test how these factors influence the effect size of microbial distance-decay relationships, to draw generalisations about how microbial β-diversity scales with space.

Location: Global.

Time period: Studies published between 2005-2019 (inclusive).

Major taxa studied: Bacteria, Archaea, and microbial Eukarya.

Methods: We conducted a meta-analysis of microbial distance-decay relationships, using the Mantel correlation coefficient as a measure of effect size. We assembled 452 data points, varying in environmental/ecological context or methodological approaches, and used linear models to test the effects of each variable.

Results: Both ecological and methodological factors had significant impacts on the strength of microbial distance-decay relationships. Specifically, larger spatial extents increased the strength of these relationships, whilst differences also emerged between environments and habitats, with soils showing significantly weaker distance-decay relationships than other habitats. Methodological factors such as sequencing depth were positively related to the strength of distance-decay relationships, and choice of dissimilarity metric was also

important, with phylogenetic metrics generally giving weaker distance-decay relationships than binary or abundance-based indices.

Main conclusions: We conclude that widely studied microbial biogeographic patterns, such as the distance-decay relationship, vary by ecological context but are primarily distorted by methodological choices. Consequently, we suggest that by linking methodological approaches appropriately to the ecological context of a study, we can progress towards generalisable biogeographic relationships in microbial ecology.

Introduction

The distance-decay of community similarity is one of the most widely studied relationships in macroecology (Nekola & White, 1999; Soininen *et al.*, 2007). This relationship quantifies the decrease in compositional similarity (β-diversity) between communities with increasing geographic distance separating them, and demonstrates that nearby communities are more similar to each other than distantly-separated communities. Distance-decay relationships can arise through several different, but often interacting ecological and evolutionary processes, and consequently ecologists have extensively debated the underlying mechanisms that generate such patterns (Nekola & White, 1999; Soininen *et al.*, 2007; Hanson *et al.*, 2012). Spatial structuring of the environment can lead to distance-decay relationships, as communities close together in space are likely to experience more similar environmental conditions, and thus contain more similar communities than those situated in very different environmental conditions. Dispersal limitation can also lead to distance-decay relationships by limiting the connectivity between communities, meaning that communities closer together in space will share more species through localised dispersal than those further apart.

Distance-decay relationships are well documented in a multitude of plant and animal communities. Yet, they are of particular interest to microbial ecologists because microorganisms were typically assumed to have ubiquitous distributions for several reasons. Firstly, their small size facilitates passive dispersal over large geographic distances by vectors such as wind, bio-aerosolization, ocean currents or migrating animals (Bisson *et al.*, 2007; Favet *et al.*, 2013; Joung *et al.*, 2017; Vašutová *et al.*, 2019), thus potentially overcoming dispersal limitation as a contributing factor to microbial community composition. Additionally, microorganisms often maintain high population densities in the environment leading to dispersal by “mass effects”, whereby high dispersal rates from areas of increased

population density maintain populations in less optimal environments (Shmida & Wilson, 1985), helping them to overcome the constraints of spatially-structured environmental gradients. Finally, some microorganisms are able to enter dormant states, whether as vegetative cells or as cysts or spores (Locey *et al.*, 2019), allowing them to survive and disperse through suboptimal environments, simultaneously enhancing their dispersive abilities, and reducing the influence of spatially-structured environmental gradients (Low-Décarie *et al.*, 2016). Combined, these traits theoretically lower microbial β-diversity by increasing the amount of shared species between distant communities, in turn leading to weaker distance-decay relationships compared to macroorganisms. However, empirical tests of microbial distance-decay relationships have yielded mixed results. Many studies have detected little or no evidence of distance-decay relationships in microbial communities (Hazard *et al.*, 2013; Kivlin *et al.*, 2014), whilst others report relationships of varying steepness, across a range of spatial extents, study systems, and taxa (Dumbrell *et al.*, 2010; Martiny *et al.*, 2011; Clark *et al.*, 2017). Thus, despite hundreds of empirical studies, the generality of spatial patterns in microbial communities remains unclear, and we are no closer to understanding whether variability in the spatial scaling relationships of microbial β-diversity originates from ecological or methodological sources.

Variation in microbial distance-decay relationships could be due to different environmental or ecological contexts in studies. Here, we consider environmental context as the variability in the physico-chemical environment (e.g. temperature, pH, topology), and ecological context as the total suite of species present and their interactions. The study systems commonly of interest to microbial ecologists vary in terms of connectivity, which may facilitate or hinder dispersal between communities, thus leading to flatter or steeper distance-decay relationships, respectively. In well connected systems where dispersal is more feasible, such as oceanic waters, distance-decay relationships should be weaker than systems in which dispersal is limited, such as host-associated systems or soil systems, where distance-decay

relationships are weaker in deeper soil horizons (Li *et al.*, 2020). Moreover, study systems differ in the spatially structured environmental gradients they support. Topsoil for example, supports strong environmental gradients over distances of a few meters (e.g. Dumbrell *et al.*, 2010), leading to steep distance-decay relationships due to spatially structured environmental gradients. Additionally, different study taxa are likely to yield variable distance-decay relationships because they may differ in traits that are linked to dispersal efficacy. For example, small cell sizes lead to more efficient long distance dispersal (Wilkinson, 2001; Wilkinson *et al.*, 2012; Norros *et al.*, 2014), thus organisms with larger cell sizes, such as microbial Eukarya, should be more strongly dispersal limited than those with small cell sizes, such as Bacteria (although this may not be true for all taxa e.g. see Kivlin, 2020). Finally, it is known that spatial extent can influence our perception of ecological relationships, which may contribute to variable distance-decay relationships (Steinbauer *et al.*, 2012). Studies incorporating larger spatial extents may find stronger distance-decay relationships as they are more likely to incorporate spatial scales at which the taxa are dispersal limited and/or at which environmental conditions become spatially structured (Martiny *et al.*, 2011).

Whilst the context in which a study was undertaken may contribute to variability in microbial distance-decay relationships, so too could different methodologies. Technological advances have yielded new insight into the structure and functioning of development of environmental microbial communities (Clark *et al.*, 2018). However, rapid turnover in molecular methodologies means that our perception of microbial β-diversity patterns is based on methods that vary substantially in both coverage (ability to detect a greater proportion of the community in a given sample) and resolution (ability to resolve closely related taxa) (Muyzer, 1999; Glenn, 2011). Early methods such as clone library sequencing and community fingerprinting methods (e.g. denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism (TRFLP), or phospholipid fatty acid (PLFA)

analysis) are limited in their ability to detect rare taxa (Bartram *et al.*, 2011), undoubtedly missing taxa that are extremely rare (Low-Décarie *et al.*, 2016). In turn, this could reduce the detected β-diversity, inflating estimated community similarity and weakening distance-decay relationships (Hanson *et al.*, 2012). In contrast, high-throughput sequencing (HTS) platforms (also frequently referred to as next-generation sequencing (NGS)) can deliver sequencing depths upwards of tens or even hundreds of thousands of sequences per sample (Caporaso *et al.*, 2012), thus improving both community coverage (the detected proportion of a given community), and allowing more samples to be examined in a single study (sample coverage). Consequently, variation between molecular methods in their ability to resolve closely related taxa, and to detect rare taxa can be an additional source of variability in microbial beta diversity, which by extension can either weaken or strengthen microbial distance-decay relationships.

In addition to the molecular methods, the choice of analytical methods, such as similarity metric, can influence distance-decay relationships. The similarity of communities can vary according to the identity and abundance of the species present, their phylogenetic relationships, and by external factors such as varying sample sizes. Thus, similarity metrics that vary by one or more of these characteristics would likely result in contrasting distance-decay relationships (Chao *et al.*, 2005; Barwell *et al.*, 2015). For example, phylogenetic indices would be expected to yield weaker distance-decay relationships than other metrics, because communities that have no species in common can still be highly phylogenetically similar if the species share many branches of a phylogenetic tree, thus reducing the decay of similarity over geographic distance (Bryant *et al.*, 2008). On the other hand, quantitative indices compare not only the composition of species present, but also their abundance in each community, reflecting finer-scale changes in community structure, and thus should result in stronger distance-decay relationships by providing an additional axis (species abundances) by which communities can differ.

Here, to disentangle the effects of both contextual (e.g. spatial extent, taxon, or ecosystem) and methodological (e.g. means of identifying/differentiating taxa, or similarity metric) variables on microbial distance-decay relationships, we undertook a meta-analysis to test the following specific hypotheses:

- H_1 Bacteria and Archaea will show weaker (lower effect size) distance-decay relationships than micro-eukaryotic taxa due to their smaller size and higher population densities in most environments.
- H_2 Ecosystems that contain steep physicochemical gradients will have stronger distance-decay relationships due to spatially-structured niche partitioning of communities.
- H_3 The spatial extent of a study will be positively related to the strength of any resulting distance-decay relationships, as larger extent studies incorporate greater environmental heterogeneity and lower dispersal rates between communities.
- H_4 High-throughput sequencing methods will yield stronger distance-decay relationships due to: a) their ability to resolve closely related taxa, b) their greater community coverage (e.g. number of sequences per sample, or number of individuals counted per sample), and/or c) their greater sample coverage.
- H_5 Phylogenetic similarity metrics (e.g. Unifrac, beta nearest taxon index) will result in weaker distance-decay relationships than other metrics as communities can be phylogenetically similar, yet different at fine taxonomic resolutions, whilst quantitative metrics (e.g. Bray-Curtis, Hellinger, Euclidean) will yield the strongest as they reflect changes in both species composition and abundance.

Methods

Meta-Analysis

In order to test our hypotheses, we first gathered available data on microbial distance-decay relationships via a systematic literature search. To do this, five search terms were selected to detect relevant studies (Table 1). All literature searches were conducted using the Web of Science search portal on 18/04/2020, and all results published between 1900-2019 (inclusive) were retained. To further filter the dataset to studies suitable for testing our hypotheses, search results were downloaded and manually screened using the "metagear" (Lajeunesse, 2016) package in R (version 3.4.1; R Core Team, 2019). Here, suitable studies were those that tested the relationship between community similarity and geographic distance in microbial communities, and not studies of "macroorganisms", or studies of strain-level genetic distance (e.g. using multi-locus sequence typing). Furthermore, studies that did not test distance-decay relationships using Mantel correlation, or that used only partial Mantel tests, were also discarded. We did not identify any potentially suitable studies that were published prior to 1967, the year the Mantel test was described (Mantel, 1967), and the earliest suitable study was published in 2005.

Table 1. Details of Web of Science search terms, and the number of results for each search.

Search	Search Term	Number of results
1	TS = (biogeograph*) AND TS = (bacteria* OR archaea* OR microb* OR microorganism*)	2907
2	TS = (macroecolog*) AND TS = (bacteria* OR archaea* OR microb* OR microorganism*)	136
3	TS = ("everything is everywhere") AND TS = (bacteria* OR archaea* OR microb* OR microorganism*)	66
4	TS = ("geographic distance") AND TS = (bacteria* OR archaea* OR microb* OR microorganism*)	220
5	TS = ("distance decay") AND TS = (bacteria* OR archaea* OR microb* OR microorganism*)	186

From these studies, we extracted Mantel correlation coefficients (*r*) as an effect-size measure for each distance-decay relationship. The Mantel test is a permutation-based method used to test for correlation between two distance matrices, or in the context of this study, community (dis)similarity and geographic distance. The Mantel test statistic is an ideal measure of effect size for use in meta-analytical frameworks for several reasons. Firstly, the Mantel correlation test is the most frequently used method for testing distance-decay relationships in microbial ecology (Franklin & Mills, 2007; Ramette, 2007). Secondly, as the Mantel coefficient is a standardised correlation coefficient (i.e. is bound by -1 and 1), it provides an easily interpretable and comparable measure of effect size (Harrison, 2012).

We ensured all Mantel correlation coefficients reflected correlations between geographic distance and community dissimilarity, rather than similarity, by multiplying correlation coefficients by -1 where necessary (so that positive values indicate a typical distance-decay relationship). Partial Mantel statistics (which test for correlation between two matrices whilst controlling for a third) were excluded as they are influenced by other variables included in the test, and are therefore not easily comparable between studies. All Mantel correlation coefficients were transformed to *z*-scores using Fisher's *z* transformation, as recommended by Koricheva *et al.*, (2013). All subsequent statistical analyses were conducted on the transformed *z*-scores, whilst original Mantel correlation coefficients were used to make figures, for ease of interpretation.

In order to test our hypotheses, several variables relating to the context and methodology of each distance-decay relationship were recorded. Details of these variables are described in Box 1.

Box 1. Details of the explanatory variables extracted from each study.

Resolution

Each distance-decay relationship was categorised into either high-resolution (high-throughput or Sanger sequencing), low resolution (molecular e.g. ARISA, TRFLP, DGGE, PhyloChip, PLFA), or low resolution (morphological), based on the method's ability to distinguish between closely related organisms.

Community Coverage

This refers to the sequencing depth in sequencing-based studies, or number of individuals counted in morphology-based studies, per sample. For sequencing studies, we recorded the number of sequences after rarefaction, or if this was not given, the average number of sequences per sample. As there is no comparable measure of coverage for fingerprinting studies, we excluded them from analyses of community coverage.

Sample Coverage

Sample coverage refers to the sample size (e.g. number of communities/samples) of each distance-decay relationship.

Dissimilarity Index

The dissimilarity index used to calculate each distance-decay relationship. Recorded dissimilarity indices were then categorised as quantitative (Bray-Curtis, Horn-Morisita, Euclidean, Hellinger, Theta), qualitative (Jaccard, Raup-Crick, Sørensen, Simpson, βsim), or phylogenetic (weighted or unweighted Unifrac, Rao, β-mean nearest taxon distance, β-mean pairwise distance).

Study Taxon

Each distance-decay relationship was binned into the following broad taxonomic categories based on the taxonomy of the focal organisms (Archaea, Bacteria, Fungi, or other microbial Eukarya), or combination of these categories if a relationship was based on multiple taxa (for example due to using sequencing primers that detect both Archaea and Bacteria). Fungi grouped separately from other micro-Eukaryotes due to their distinct reproductive strategy (e.g. spore-production) and the fact the they are frequently targeted using distinct molecular approaches (e.g. via taxon-specific primer sets), in contrast to most other studies of micro-Eukarya.

Spatial Extent

This is the maximum distance separating communities in km. If this was not stated in text or provided in supplementary material (e.g. in a geographic distance matrix), it was calculated from given geographic coordinates, estimated from a plot of the distance-decay relationship, or estimated from scaled maps.

Environment

We broadly categorised distance-decay relationships based on the type of environment (agriculture, air, aquifer, coastal wetlands/intertidal, desert, dune, forest, glacier, grassland, lake, marine, coastal marshes, mine, river, snow, urban) within which they were sampled. Whilst these categories are not mutually exclusive, we categorised each study based on which environment best represented the environmental context in which each study was undertaken.

Habitat

The type of environmental material that the sampled communities occupied. We categorised distance-decay relationships as: air, host-associated, sediment, snow, soil, water.

Statistical Analyses

In order to determine whether distance-decay relationships varied between categorical variables (as in hypotheses 1, 2, 4, and 5), we used ANOVA tests. In tests where significant differences between groups were found, Tukey's Honest Significant Difference (HSD) tests were used to determine which groups were different. Linear models were used to test relationships between effect sizes and continuous variables such as spatial extent and community coverage. The variables spatial extent and community coverage were initially log transformed to aid model fitting, as they spanned several orders of magnitude. To compare the overall influence of ecological vs methodological factors on microbial distance-decay relationships, we compared two full models (including all relevant variables) using AIC scores, on a subset of the data for which all variables were successfully recorded. We then sought to identify a smaller number of variables that adequately predicted the effect size of microbial distance-decay relationships by using a drop-term likelihood ratio procedure.

Results

Our Web of Science searches resulted in 2,982 unique search results. Manual screening of the abstracts yielded 951 studies that were deemed to be potentially suitable for use in this analysis. A total of 452 Mantel correlation coefficients were successfully obtained from 187 studies represented in 61 journals (Fig. S1). Reported Mantel correlation coefficients ranged from -0.33 to 0.95, with a mean of 0.27 (std. error = 0.011).

Influence of Context on the Distance-Decay Relationship

In order to determine whether contextual factors can influence the strength of distance-decay relationships, the influence of ecological factors including study taxa, study system, and spatial scale were tested. Within the dataset, the most commonly studied taxa were Bacteria (n = 238), followed by Fungi (n = 93), other microbial Eukaryotes (n = 67), and Archaea (n = 26). We found no clear differences in the effect sizes of distance-decay relationships between these taxa ($F_{5,\ 441}$ = 0.97, P = 0.43), although distance-decay relationships incorporating bacterial and fungal communities showed the smallest effect sizes, albeit only from six studies (Fig. 1).

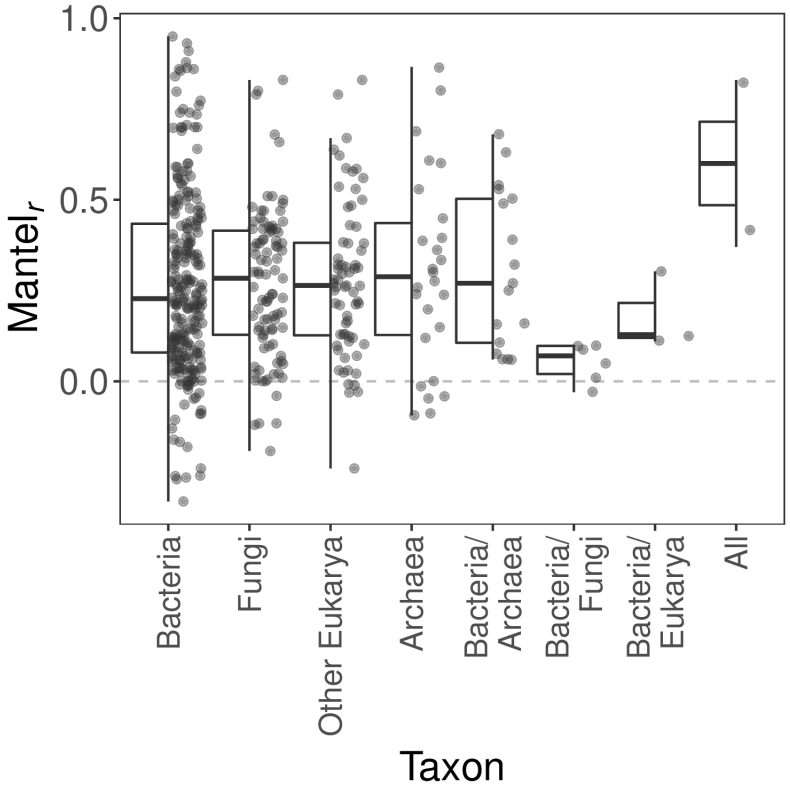


Figure 1. Effect sizes (Mantel$_r$) of distance-decay relationships based on different study taxa. A larger effect size indicates stronger positive correlation between community dissimilarity and geographic distance. The “All” category consists of studies that incorporated all microbial taxonomic groups, whilst combined categories (e.g. Bacteria/Archaea) incorporate communities from multiple taxonomic groups (e.g. bacterial and archaeal communities).

The distance-decay relationships in our dataset originated from 16 different environments. Of these, five were represented by three, or fewer, distance-decay relationships, and so were excluded from further analyses (marsh; n = 3, snow; n = 3, dune, mine, aquifer; n = 1). The most frequently studied environments were grasslands (n = 96), marine (n = 88), and lakes and forests (n = 76 for both). We found clear differences in the effect sizes of distance-decay relationships between environments (Fig. 2A; $F_{10, 432}$ = 3.187, P < 0.001). Specifically, and perhaps counter-intuitively, grassland-based studies tended to have weaker distance-decay relationships than those from aquatic environments such as lakes, rivers, or the marine environment (|coef| > 0.17, P < 0.05 for all comparisons). Urban environments, which included built environments such as sewers and indoor air, also produced low effect sizes, although with only four data points, this difference was not statistically clear (P > 0.43 for all comparisons). A more detailed analysis of the interaction between environment type and habitat revealed that, whilst environments ($F_{9, 420}$ = 3.29, P < 0.001) and habitat ($F_{3, 420}$ = 6.65, P < 0.001) differ from each other, their interaction was not statistically significant ($F_{4, 420}$ = 1.93, P = 0.10). In fact, within environments, only marine host-associated and marine water-based distance-decay relationships were clearly different from each other (Fig. 2B), with host-associated communities showing significantly larger effect sizes (coef = 0.35, P < 0.001).

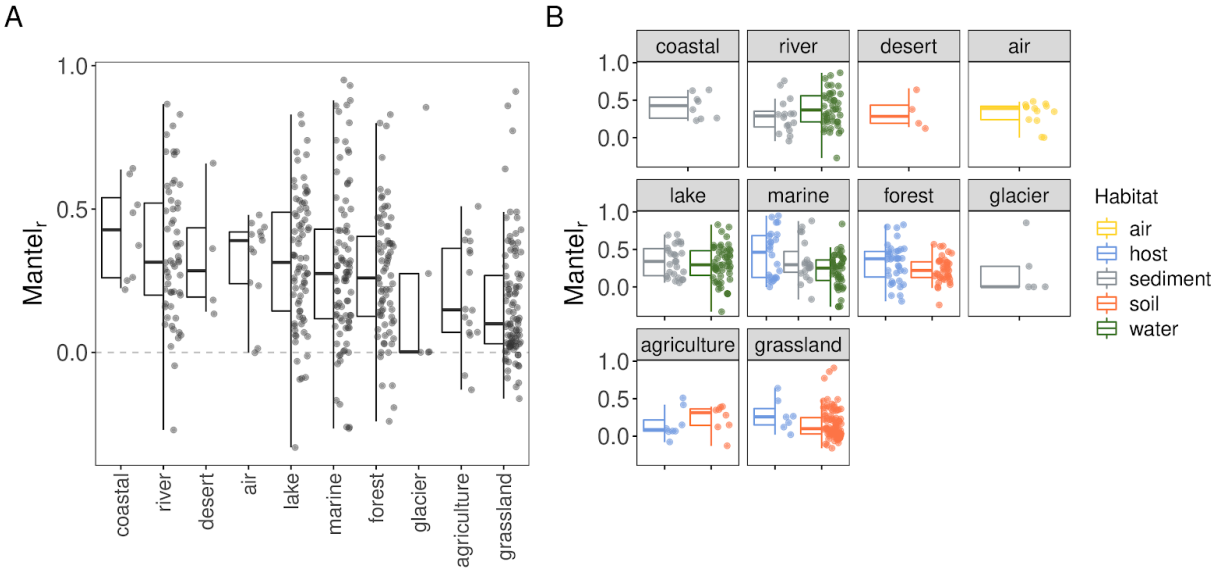


Figure 2. Variation in Mantel correlation coefficients (effect size) of distance-decay relationships between different environments (A) and habitat types (B). Environment categories are arranged from highest to lowest mean effect size.

Finally, we found a positive relationship between the (log) spatial extent and the effect size of microbial distance-decay relationships (Fig. 3; coef = 0.03, t = 4.66, R^2 = 0.05, P < 0.001), such that studies incorporating large spatial scales tend to have stronger distance-decay relationships. As larger spatial scale studies might also incorporate greater sampling coverage, we also tested for collinearity between the spatial scale of a study and the sampling coverage, but found no correlation between these variables (ρ = 0.06, P = 0.19).

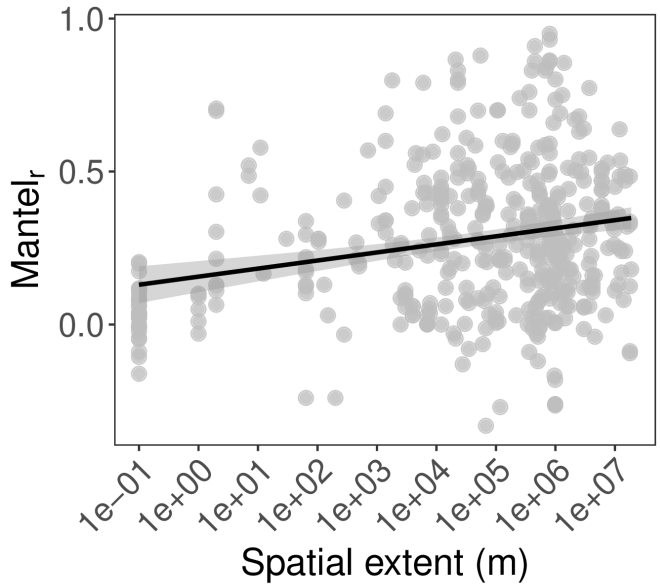


Figure 3. The relationship between spatial extent and the Mantel correlation coefficient of microbial distance-decay relationships. The best fit line represents the fit of a linear regression between the log of spatial extent and Mantel correlation coefficient, and the grey shaded region shows 95% confidence intervals.

Influence of Methodological Factors on the Distance-Decay Relationship

To determine whether the microbial distance-decay relationship may be influenced by methodological factors, we tested for relationships between the method of community characterisation, sampling depth, or choice of community similarity index and the effect size of microbial distance-decay relationships. We grouped community characterisation methods according to their ability to distinguish between closely related taxa. There were no clear differences in the distance-decay effect sizes between methods of differing resolutions ($F_{2, 449}$ = 0.562, P = 0.57), nor were there clear differences between different molecular methods (Fig. S2, $F_{7, 437}$ = 1.97, P = 0.06), considering only those methods that had >4 distance-decay relationships (excluding Ion Torrent; n = 4, phylo-chip; n = 2, and Pac-Bio; n = 1).

Whilst we observed no differences in distance-decay relationships between different resolution methods, we observed a positive relationship between (log_{10}) community coverage and the strength of microbial distance-decay relationships (Fig. 4A; n = 337, coef = 0.04, t = 2.39, P < 0.01). However, this relationship was weak (R^2 = 0.01), and when two distance-decay relationships with extremely high community coverage were removed, the slope was indistinguishable from 0 (coef = 0.03, t = 1.78, R^2 = 0.01, P = 0.08).

The logistics of multiplexing samples on high-throughput sequencing runs means that there is often a trade-off between the community coverage and sampling coverage of a study. However, we found no evidence of negative correlation between these two factors (Pearson's ρ = -0.03, P = 0.54). Neither did we detect any clear relationship between the number of samples (log_{10} sample coverage) and the distance-decay effect size (Fig. 4B; n = 451, coef = -0.04, t = -1.03, P = 0.30).

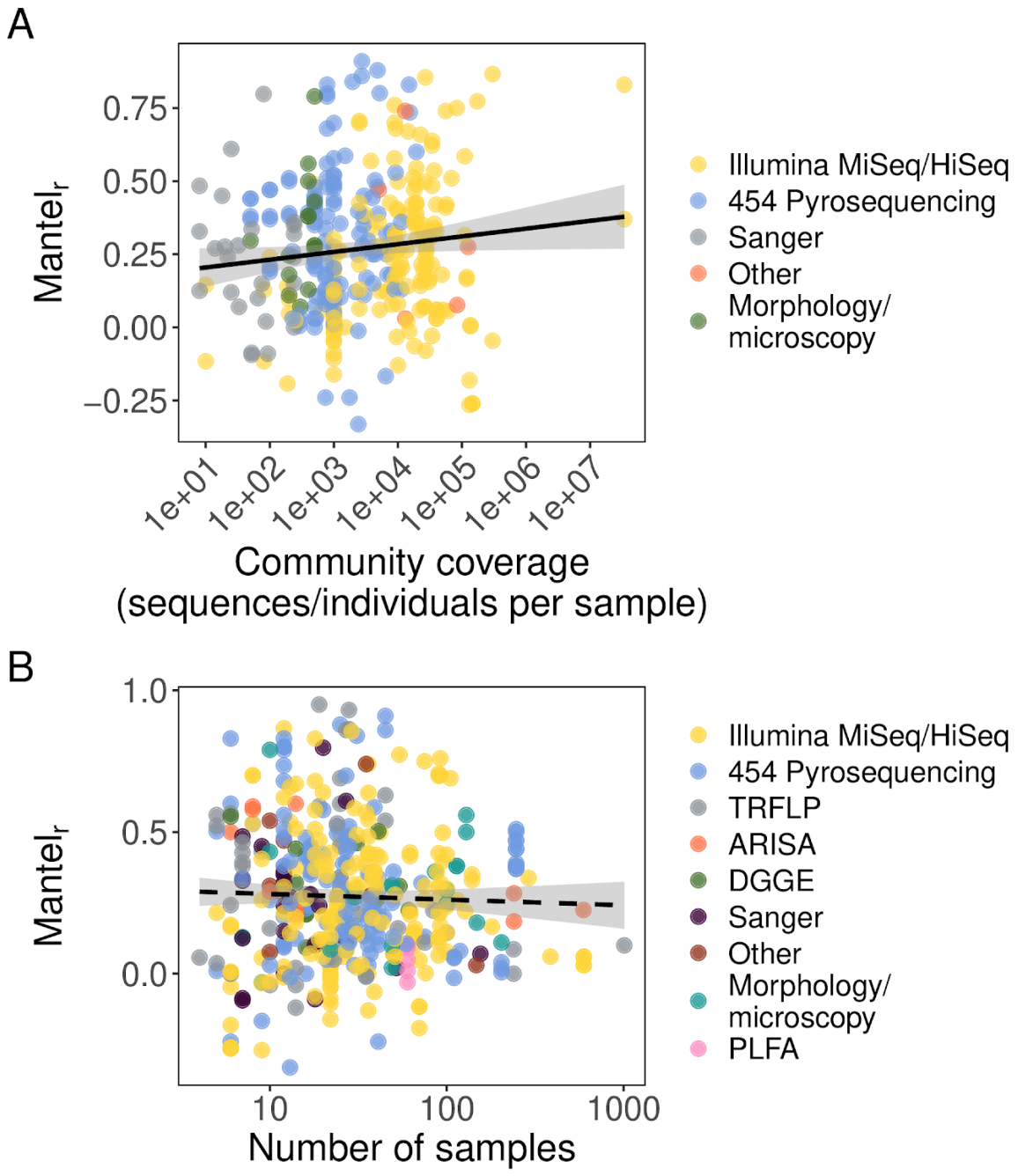
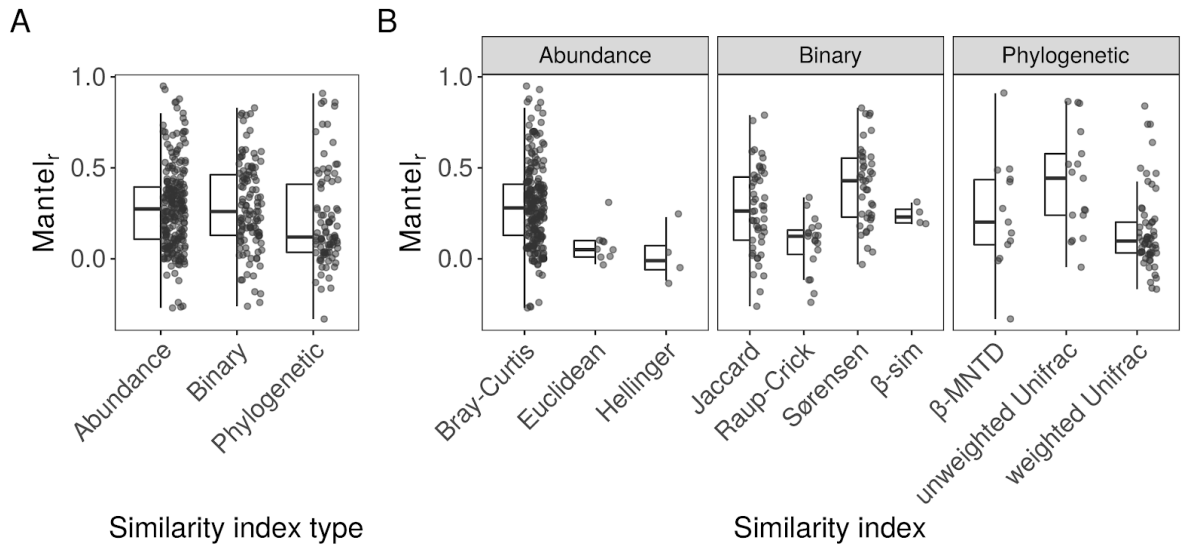


Figure 4. The relationship between the effect size of microbial distance-decay relationships (Mantel correlation coefficient) and A) community coverage, quantified as the number of sequences or individuals counted per sample, and B) sample coverage, quantified as the number of individual samples used to construct distance-decay relationships. Points are individual Mantel correlation coefficients, coloured by the molecular technique used in

characterising the microbial community. Solid lines indicate statistically significant relationships ($P < 0.05$), whilst dashed lines indicate non-significant relationships ($P > 0.05$), and grey ribbons represent 95% confidence intervals. Abbreviated molecular methods in the legend are defined as follows (TRFLP = Terminal Restriction Fragment Length Polymorphism; ARISA = Automated Ribosomal Intergenic Spacer Analysis; DGGE = Denaturing Gradient Gel Electrophoresis; PLFA = Phospholipid Fatty Acid analysis).

Choice of similarity index also had a clear impact on the effect size of microbial distance-decay relationships. As well as recording the specific similarity index used, we categorised these indices into types (binary, abundance, or phylogenetic) to look for broad differences in distance-decay relationships between them. We analysed the nested interaction between similarity index and index type, and found no clear differences between different index types (Fig. 5A; $F_{2, 424} = 1.48$, P = 0.23). However, the interaction between index type and similarity index was significant ($F_{7, 424} = 7.20$, P 0.001). Post-hoc analysis revealed differences between similarity indices within and between index types (Fig. 5B). Distance-decay effect sizes based on the Raup-Crick index were weaker than those based on either Sørensen (coef = -0.38, P < 0.01) or unweighted Unifrac indices (coef = -0.44, P < 0.01), whilst those based on weighted Unifrac were weaker than both Sørensen (coef = -0.29, P < 0.001) and unweighted Unifrac (coef = -0.35 P < 0.05).



309 Figure 5. Variation in effect sizes of microbial distance-decay relationships calculated with
310 different similarity index types (A), or individual indices (B). Only indices with four or more
311 effect sizes were plotted for clarity.

312 *Comparison of Contextual and Methodological Variables*

313 In order to determine whether eco-environmental context or methodological factors better
314 explain the effect size of microbial distance decay relationship, we specified two models,
315 with variables from these two categories, using a subset of the original data for which values
316 were obtained for all variables (n = 323). The two models each had four variables, and used
317 similar degrees of freedom (context model df = 26, methodological model df = 27). The
318 methodological model outperformed the contextual model in terms of both AIC (Akaike
319 Information Criterion) and R^2 measures of model performance (Table 2). Notably, neither
320 model explained a high proportion of the variance, although both AIC and likelihood ratio
321 tests supported both models over a null (intercept only) model.

Table 2. Comparison of models specified using either contextual, or methodological variables. Akaike Information Criterion (AIC) and adjusted R^2 quantify the likelihood and fit of a model relative to the number of predictor variables, respectively.

Model	AIC	Adj-R^2	Likelihood ratio comparison to null (intercept only) model			
			ΔAIC	Sum of squares	F (df)	*P* value
Contextual	146.89	0.11	-13.69	5.34	2.61	< 0.001
Methodological	134.11	0.14	-26.46	6.47	3.17 (25)	< 0.001

Discussion

Previous research into the spatial ecology of microbial communities has not yielded a consistent distance-decay relationship. Our meta-analysis of 452 microbial distance-decay relationships suggests that the reasons for this lack of consistency are two-fold. Firstly, the differing contexts within which studies are conducted contribute variability to reported distance-decay relationships. In particular, we found that differing study systems and spatial extents were associated with variation in microbial distance-decay relationships. Secondly, methodological differences between studies, including dissimilarity index, data resolution, and sample coverage, all significantly affected observed distance-decay relationships. A central tenet of macroecology is the search for universal patterns and relationships; our results suggest generalisable relationships may only emerge when methodological approaches are appropriately coupled to ecological context.

Our comparison of distance-decay relationships between different study systems revealed that agricultural and especially grassland-based studies had weaker relationships than studies of other environments. Within these environments, soils were by far the most

frequently studied habitat, and we initially expected that, as soils maintain strong physicochemical gradients over small vertical and horizontal spatial scales (e.g. Dumbrell *et al.*, 2010), that these distance-decay relationships would be stronger than in other environments or habitats. It is possible that the environmental gradients present in soils do not change linearly over geographic distance, for example if similar environmental conditions are patchily distributed. Alternatively, soil microorganisms may be able to disperse more effectively than previously thought, perhaps via association with other soil organisms (e.g. bacterial migration along fungal hyphae; Warmink *et al.*, 2011), migratory species such as birds (Bisson *et al.*, 2007), wind blown soil particles (Favet *et al.*, 2013), or via bioaerosols (Joung *et al.*, 2017). The depth profile over which soil samples integrate may also play a role in obscuring distance-decay relationships, as surface soils show stronger distance-decay relationships than deeper ones, likely due to the greater intensity of dispersing propagules entering the surface (Li *et al.*, 2020). Furthermore, soils harbour extensive microbial "seed banks" of dormant organisms and/or relic DNA that could weaken the distance-decay relationship (Lennon & Jones, 2011; Carini *et al.*, 2016; Lennon *et al.*, 2018). Dormant cells and relic DNA are not subject to environmental selection yet, are routinely detected in molecular community assays, and thus may diminish the perceived effects of spatially-structured environmental selection on microbial communities (Locey *et al.*, 2019). Thus, in habitats such as soils, distinguishing dormant from active cells could result in stronger distance-decay relationships than those recorded previously, though the extent to which this phenomenon plays a role in other environments is less clear .

Originally, we expected that studies of aquatic microbial communities may show the weakest distance-decay relationships as riverine or oceanic hydrology may provide an effective dispersal mechanism, homogenising microbial communities over larger spatial and environmental gradients over larger spatial scales. In contrast, we found that aquatic communities actually showed stronger distance-decay relationships. Soininen *et al.* (2007)

recorded similar distance-decay rates between terrestrial, marine and aquatic ecosystems, showing that context-dependent distance-decay relationships may be a feature of microbial communities. Host-associated communities showed relatively strong, but variable distance-decay relationships. We suggest that this is caused jointly by the ecology of the host species, in combination with the degree of host-specificity with the associated microbial community. For example, if the host is not dispersal limited, and associates with a large variety of microorganisms, then the distance-decay relationship may be relatively weaker than those of either dispersal limited hosts, or highly specific associated microbiomes.

Scale-dependent relationships have been reported previously (Bissett *et al.*, 2010; Martiny *et al.*, 2011; Soininen *et al.*, 2011), albeit with contrasting results. Our results are comparable to those of Soininen *et al.* (2011), who reported that distance-decay relationships of various microbial communities were generally steeper as greater spatial scales were incorporated. The scale-dependence of this relationship may be explained by greater environmental heterogeneity in large-scale studies, thus communities are subjected to different environmental filters, resulting in more dissimilar communities. In combination with this, communities separated by very large geographic distances should have minimal dispersal between them, assuming connectivity is negatively related to geographic distance. Alternatively, this result may be a statistical artefact, caused by studies with large spatial extents incorporating many zero similarity community comparisons (i.e. communities with no species in common), therefore biasing measured distance-decay relationships (Millar *et al.*, 2011; Steinbauer *et al.*, 2012).

Distance-decay relationships are frequently interpreted as evidence for neutral community assembly processes such as dispersal limitation, in the microbial literature. Across microbial taxa, cell size is a trait thought to influence dispersal efficacy (Wilkinson, 2001; Wilkinson *et al.*, 2012; Zinger *et al.*, 2019), and so larger microorganisms such as micro-Eukarya should

show stronger distance-decay relationships than smaller microorganisms such as Bacteria or Archaea. However, we found no evidence for this, suggesting that phylogenetically structured traits such as cell size may be less important compared to other contextual and methodological factors, or that the broad domain-level classification used here does not sufficiently capture different microbial cell sizes. As discussed previously, distance-decay relationships can arise from spatially autocorrelated environmental gradients as well as dispersal limitation (Nekola & White, 1999). Therefore, the lack of differences in biogeographic patterns observed at the domain level may be the result of a trade-off between dispersal-related processes and environmental filtering. For instance, bacterial distance-decay relationships may be less strongly influenced by dispersal than environmental filtering, and vice versa for Eukarya. Consequently, these influences may balance out at broad taxonomic levels, resulting in similar biogeographic patterns at the domain level.

In comparison to contextual factors, methodological factors were found to have a greater influence on microbial distance-decay relationships. The development of molecular methods, including high-throughput sequencing platforms, has vastly improved our ability to characterise microbial communities (Roesch *et al.*, 2007; Caporaso *et al.*, 2012). However, these methods differ in their resolution, community coverage, and ability to multiplex large numbers of samples, all of which we hypothesised could strengthen or weaken distance-decay relationships by altering our estimation of microbial β-diversity. In contrast, we observed only a weak relationship between distance-decay effect sizes and community coverage, and no clear relationships with different resolution methods, or with the number of samples, suggesting that molecular methodology may not play as large a role in determining microbial biogeographic patterns as previously thought.

The ability to resolve closely related taxa has previously been found to be an important determinant of our ability to detect biogeographical patterns, as such patterns may only emerge when taxa are defined at sufficiently high resolution (Hanson *et al.*, 2012). Yet, other studies show that bioinformatically altering taxonomic resolution frequently has little effect on microbial biogeographic patterns. For example, increasing the similarity threshold at which operational taxonomic units are defined is thought to be equivalent to increasing the taxonomic resolution (Callahan *et al.*, 2017). Yet, empirical biogeographic relationships often appear robust to such manipulation, in a variety of taxa and ecosystems (Clark *et al.*, 2017; Glassman & Martiny, 2018; Meyer *et al.*, 2018), supporting our finding that resolution may not be important. Perhaps most molecular methodologies operate above resolutions at which biogeographic patterns begin to change, or more worryingly, perhaps we are still studying microbial biogeography at too low a resolution.

Aside from resolution, another important variable related to molecular methodology is community coverage. One of the few universal patterns that appears to hold true for most microbial communities is the “long-tailed” species abundance-distributions (Dumbrell *et al.*, 2010; Shoemaker *et al.*, 2017; Maček *et al.*, 2019), which is caused by the majority of microorganisms in a community being rare. The rarer taxa in microbial communities also tend to be the least widespread (Clark *et al.*, 2017; Lindh *et al.*, 2017; Meyer *et al.*, 2018; Shade & Stopnisek, 2019) and thus, only detecting the more abundant, widespread organisms would overestimate compositional similarity across communities, and consequently, weaken distance-decay relationships due to the lower rate of turnover (Meyer *et al.*, 2018). Perhaps of more concern is that even with existing sequencing platforms, our surveys of environmental microbial communities still miss taxa that are vanishingly rare in the environment, such as extremophiles that persist in non-extreme habitats (Low-Décarie *et al.*, 2016). The ability of common species to reflect ecological patterns of the wider

community is debated (Galand *et al.*, 2009; Heino & Soininen, 2010; van Dorst *et al.*, 2014) and is linked to a wider debate on the ecological importance of rare species that is far beyond the scope of this work (e.g. Gaston, 2012). However, rare microorganisms are well known to be of critical importance in the context of environmental perturbations (Shade *et al.*, 2014; Low-Décarie *et al.*, 2016) and in providing ecosystem processes (e.g. sulfate-reduction in peat soils, Hausmann *et al.*, 2016; and anaerobic ammonia-oxidation in river sediments Lansdown *et al.*, 2016) and as a result, ignoring them may further distance biogeographic patterns from ecosystem-level processes.

Against expectation, we observed no clear differences in distance-decay relationships using different similarity metric types, and differences between specific metrics were minimal. Distance-decay relationships based on the weighted Unifrac distance were weaker than those based on other metrics. Phylogenetic metrics, such as Unifrac, cluster communities at a lower resolution, as two communities can be closely genetically related, yet distinct at fine taxonomic resolutions (e.g. species or strain-level). For example, Bryant *et al.* (2008) found that Unifrac similarity was approximately three times higher than the compositional similarity of the same set of bacterial communities. Further, phylogenetic metrics may be inappropriate in less phylogenetically diverse environments (e.g. extreme systems) where phylogenetic diversity can be largely constrained to one taxon (e.g. the Halobacteria in hypersaline environments), leaving few “phylogenetic degrees of freedom” left to separate communities (Fukuyama, 2019). However, this does not account for the observed difference between weighted and unweighted versions of the Unifrac index, the former of which accounts for species’ relative abundance data, whilst the latter is binary (presence/absence based). A criticism of the weighted Unifrac index is that too much weight is placed on abundant taxa (Chen *et al.*, 2012). As abundant species are generally more widespread, placing too much weight on abundant taxa would have the effect of making communities appear artificially similar, exacerbating the effects of using a phylogenetic metric. As we observed no

difference between binary and abundance-based compositional indices, the differences observed with weighted Unifrac appear to be the result of combining phylogenetic and weighted indices. We therefore suggest that weighted phylogenetic metrics may underestimate microbial biogeographic patterns, unless appropriate weight is given to rare and abundant taxa (Chen *et al.*, 2012).

Our analysis of 452 microbial distance-decay relationships also revealed the overwhelming preference of microbial ecologists to use classic dissimilarity indices such as the Bray-Curtis (n = 218), Jaccard (n = 49), Sørensen (n = 42) indices. These choices no doubt reflect a wider trend in ecology as a whole, however, it is pertinent to draw attention to more recently developed metrics that may be more appropriate given the properties of microbial datasets and the hypotheses being tested. Biotic interactions are drivers of microbial β-diversity (Hanson *et al.*, 2012), yet classic dissimilarity metrics do not account for co-occurrence information in communities. To this end, a new family of metrics described by Schmidt *et al.*, (2017) include information on the average interactions of the taxa present, thus providing a novel approach to integrating co-occurrence data into distance-decay relationships. One problematic characteristic of high-throughput sequence datasets is the non-biological variance of sample sizes, which can result in statistical artefacts that confound biogeographic relationships (Baselga, 2007). Here, modifications made to some classic indices by Chao *et al.* (2005) reduce the sensitivity of these indices to variable sample sizes by accounting for unobserved species, thus reducing the need for post-sequencing normalisation of sample sizes (McMurdie & Holmes, 2014). Furthermore, “fuzzy logic”-based similarity indices are able to reduce the impact of false-absences or -presences on estimates of β-diversity, which is beneficial for microbial ecology studies where rarefaction may induce false-absences, and taxonomic assignment errors or contamination may lead to false-presences. Finally, many similarity metrics have been shown to merge compositional turnover (replacement of species) and nestedness (whereby communities are subsets of one

another), thereby blurring the contribution of distinct ecological processes to total community (dis)similarity. To combat this, modified versions of classic indices such as Jaccard, Sorensen, and Bray-Curtis have been developed, allowing the partitioning of community similarity metrics into their turnover and nestedness components (Baselga, 2010; Podani & Schmera, 2011). We echo the call of Green and Bohannan (2006) for microbial ecologists to exercise more care in their choice of dissimilarity metrics, especially as many of these new metrics are implemented in popular and freely accessible software, such as R (e.g. Baselga and Orme, 2012).

Overall, our analyses revealed that methodological factors explain more variation in microbial distance-decay relationships than ecological context, but that both sets of factors alter our perception of this biogeographic pattern. Given the importance of methodological factors in determining the strength of microbial biogeographic patterns, it is intuitive to recommend standardising approaches across studies in order to minimise the statistical signals associated with methodological variance. However, our results show variance due to differing ecological contexts would still hinder drawing generalisable relationships across studies. Instead, we suggest that tailoring methodological choices towards specific ecological contexts may enhance generalisable relationships in microbial ecology. For instance, in searching for consistent relationships between ocean waters and terrestrial soils, it would be unrealistic to sample both at the same spatial grain and extent, as the heterogeneity in the physicochemical environment, and dispersal processes of their microbial communities, are fundamentally different. Similarly, we should not necessarily expect the relationships between soils and river sediments to be comparable, as microorganisms in soils can feasibly disperse in any direction, whereas in rivers or streams dispersal would be largely constrained by flow direction. Consequently, tailoring methodological approaches, such as the sampling design and/or (geographical) distance measure, to better reflect the environmental heterogeneity and dispersal dynamics between

contrasting ecological contexts may enable us to negotiate the hierarchy of interacting factors that obscure macroecological patterns in microbial communities.

Conclusions

Our meta-analysis of >450 microbial distance-decay relationships revealed that factors related to the eco-environmental context within which a study was conducted, as well as the methodology of the study, jointly influence quantification of this classic biogeographic pattern. Against expectation, factors related to molecular methodology had relatively little effect on distance-decay relationships, whilst choice of dissimilarity metric was more important, highlighting that even after using robust, modern molecular methods, analytical choices have the power to obscure or enhance biogeographic patterns. Whilst we were able to detect clear relationships between microbial distance-decay relationships and various contextual and methodological variables, combining these variables explained only a modest amount of variation in our dataset. This lack of explanatory power highlights the fact that microbial biogeographic patterns may depend on a great number of contextual variables beyond those analysed here, and that understanding the environmental, or methodological, factors that drive this context-dependence may enable us to unify the seemingly disparate patterns observed by microbial ecologists over the past few decades.

Data Availability Statement

Full raw data analysed in this manuscript are provided in Table S1. Full raw data and code used in this manuscript will be uploaded to the Dryad data repository upon acceptance of this article.

References

Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G. & Neufeld, J.D. (2011) Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads. *Applied and Environmental Microbiology*, **77**, 3846–3852.

Barwell, L.J., Isaac, N.J.B. & Kunin, W.E. (2015) Measuring β-diversity with species abundance data. *The Journal of Animal Ecology*, **84**, 1112–1122.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808–812.

Bissett, A., Richardson, A.E., Baker, G., Wakelin, S. & Thrall, P.H. (2010) Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Molecular Ecology*, **19**, 4315–4327.

Bisson, I.-A., Marra, P.P., Burtt, E.H., Sikaroodi, M. & Gillevet, P.M. (2007) A Molecular Comparison of Plumage and Soil Bacteria Across Biogeographic, Ecological, and Taxonomic Scales. *Microbial Ecology*, **54**, 65–81.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008) Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences*, **105**, 11505–11511.

Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **11**, 2639–2643.

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

Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S. & Fierer, N. (2016) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology*, **2**, 1–6.

Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.-J. (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, **8**, 148–159.

Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., Collman, R.G., Bushman, F.D. & Li, H. (2012) Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, **28**, 2106–2113.

Clark, D.R., Ferguson, R.M.W., Harris, D.N., Nicholass, K.J.M., Prentice, H.J., Randall, K.C., Randell, L., Warren, S.L. & Dumbrell, A.J. (2018) Streams of data from drops of water: 21st century molecular microbial ecology. *Wiley Interdisciplinary Reviews: Water*, **5**, e1280.

Clark, D.R., Mathieu, M., Mourot, L., Dufossé, L., Underwood, G.J.C., Dumbrell, A.J. & McGenity, T.J. (2017) Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization? *Global Ecology and Biogeography*, **26**, 1435–1446.

van Dorst, J., Bissett, A., Palmer, A.S., Brown, M., Snape, I., Stark, J.S., Raymond, B., McKinlay, J., Ji, M., Winsley, T. & Ferrari, B.C. (2014) Community fingerprinting in a sequencing world. *FEMS microbiology ecology*, **89**, 316–330.

Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C. & Fitter, A.H. (2010) Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME Journal*, **4**, 337–345.

Favet, J., Lapanje, A., Giongo, A., Kennedy, S., Aung, Y.-Y., Cattaneo, A., Davis-Richardson, A.G., Brown, C.T., Kort, R., Brumsack, H.-J., Schnetger, B., Chappell, A., Kroijenga, J., Beck, A., Schwibbert, K., Mohamed, A.H., Kirchner, T., de Quadros, P.D., Triplett, E.W., Broughton, W.J. & Gorbushina, A.A. (2013) Microbial hitchhikers on intercontinental dust: catching a lift in Chad. *The ISME Journal*, **7**, 850–867.

Franklin, R.B. & Mills, A.L. (2007) *Statistical Analysis Of Spatial Structure In Microbial Communities*. *The Spatial Distribution of Microbes in the Environment* (ed. by R.B. Franklin) and A.L. Mills), pp. 31–60. Springer Netherlands, Dordrecht.

Fukuyama, J. (2019) Emphasis on the deep or shallow parts of the tree provides a new characterization of phylogenetic distances. *Genome Biology*, **20**, 131.

Galand, P.E., Casamayor, E.O., Kirchman, D.L. & Lovejoy, C. (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences*, **106**, 22427–22432.

Gaston, K.J. (2012) The importance of being rare. *Nature*, **487**, 46–47.

Glassman, S.I. & Martiny, J.B.H. (2018) Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. *mSphere*, **3**.

Glenn, T.C. (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.

Green, J. & Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution*, **21**, 501–507.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, **10**, 497–506.

Harrison, F. (2012) Getting started with meta-analysis. *Journal of Applied Ecology*, 1–10.

Hausmann, B., Knorr, K.-H., Schreck, K., Tringe, S.G., Glavina del Rio, T., Loy, A. & Pester, M. (2016) Consortia of low-abundance bacteria drive sulfate reduction-dependent degradation of fermentation products in peat soil microcosms. *The ISME Journal*, **10**, 2365–2375.

Hazard, C., Gosling, P., Gast, C.J. van der, Mitchell, D.T., Doohan, F.M. & Bending, G.D. (2013) The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME Journal*, **7**, 498–508.

Heino, J. & Soininen, J. (2010) *Are common species sufficient in describing turnover in aquatic metacommunities along environmental and spatial gradients*.

Joung, Y.S., Ge, Z. & Buie, C.R. (2017) Bioaerosol generation by raindrops on soil. *Nature Communications*, **8**, 1–10.

Kivlin, S.N. Global mycorrhizal fungal range sizes vary within and among mycorrhizal guilds but are not correlated with dispersal traits. *Journal of Biogeography*, **n/a**.

Kivlin, S.N., Winston, G.C., Goulden, M.L. & Treseder, K.K. (2014) Environmental filtering affects soil fungal community composition more than dispersal limitation at regional scales. *Fungal Ecology*, **12**, 14–25.

Koricheva, J., Gurevitch, J. & Mengersen, K. (2013) *Handbook of Meta-analysis in Ecology and Evolution*, Princeton University Press.

Lajeunesse, M.J. (2016) Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for r. *Methods in Ecology and Evolution*, **7**, 323–330.

Lansdown, K., McKew, B.A., Whitby, C., Heppell, C.M., Dumbrell, A.J., Binley, A., Olde, L. & Trimmer, M. (2016) Importance and controls of anaerobic ammonium oxidation influenced by riverbed geology. *Nature Geoscience*, **9**, 357–360.

Lennon, J.T. & Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews. Microbiology*, **9**, 119–130.

Lennon, J.T., Muscarella, M.E., Placella, S.A. & Lehmkuhl, B.K. (2018) How, When, and Where Relic DNA Affects Microbial Diversity. *mBio*, **9**.

Li, P., Li, W., Dumbrell, A.J., Liu, M., Li, G., Wu, M., Jiang, C. & Li, Z. (2020) Spatial Variation in Soil Fungal Communities across Paddy Fields in Subtropical China. *mSystems*, **5**.

Lindh, M.V., Sjöstedt, J., Ekstam, B., Casini, M., Lundin, D., Hugerth, L.W., Hu, Y.O.O., Andersson, A.F., Andersson, A., Legrand, C. & Pinhassi, J. (2017) Metapopulation

theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. *Environmental Microbiology*, **19**, 1222–1236.

Locey, K.J., Muscarella, M.E., Larsen, M.L., Bray, S.R., Jones, S.E. & Lennon, J.T. (2019) Dormancy dampens the microbial distance-decay relationship. *bioRxiv*, 717546.

Low-Décarie, E., Fussmann, G.F., Dumbrell, A.J. & Bell, G. (2016) Communities that thrive in extreme conditions captured from a freshwater lake. *Biology Letters*, **12**, 20160562.

Maček, I., Clark, D.R., Šibanc, N., Moser, G., Vodnik, D., Müller, C. & Dumbrell, A.J. (2019) Impacts of long-term elevated atmospheric CO2 concentrations on communities of arbuscular mycorrhizal fungi. *Molecular Ecology*, **28**, 3445–3458.

Mantel, N. (1967) The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, **27**, 209–220.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-Devine, M.C. (2011) Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*, **108**, 7850–7854.

McMurdie, P.J. & Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, **10**, e1003531.

Meyer, K.M., Memiaghe, H., Korte, L., Kenfack, D., Alonso, A. & Bohannan, B.J.M. (2018) Why do microbes exhibit weak biogeographic patterns? *The ISME Journal*, **12**, 1404–1413.

Millar, R.B., Anderson, M.J. & Tolimieri, N. (2011) Much ado about nothings: using zero similarity points in distance-decay curves. *Ecology*, **92**, 1717–1722.

Muyzer, G. (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, **2**, 317–322.

Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.

Norros, V., Rannik, Ü., Hussein, T., Petäjä, T., Vesala, T. & Ovaskainen, O. (2014) Do small spores disperse further than large spores? *Ecology*, **95**, 1612–1621.

Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence – absence data. *Oikos*, **120**, 1625–1638.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ramette, A. (2007) Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, **62**, 142–160.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G. & Triplett, E.W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, **1**, 283–290.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N. & Gilbert, J.A. (2014) Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio*, **5**.

Shade, A. & Stopnisek, N. (2019) Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology*, **49**, 50–58.

Shmida, A. & Wilson, M.V. (1985) Biological Determinants of Species Diversity. *Journal of Biogeography*, **12**, 1–20.

Shoemaker, W.R., Locey, K.J. & Lennon, J.T. (2017) A macroecological theory of microbial biodiversity. *Nature Ecology & Evolution*, **1**, 1–6.

Soininen, J., Korhonen, J.J., Karhu, J. & Vetterli, A. (2011) Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. *Limnology and Oceanography*, **56**, 508–520.

Soininen, J., McDonald, R. & Hillebrand, H. (2007) The distance decay of similarity in ecological communities. *Ecography*, **30**, 3–12.

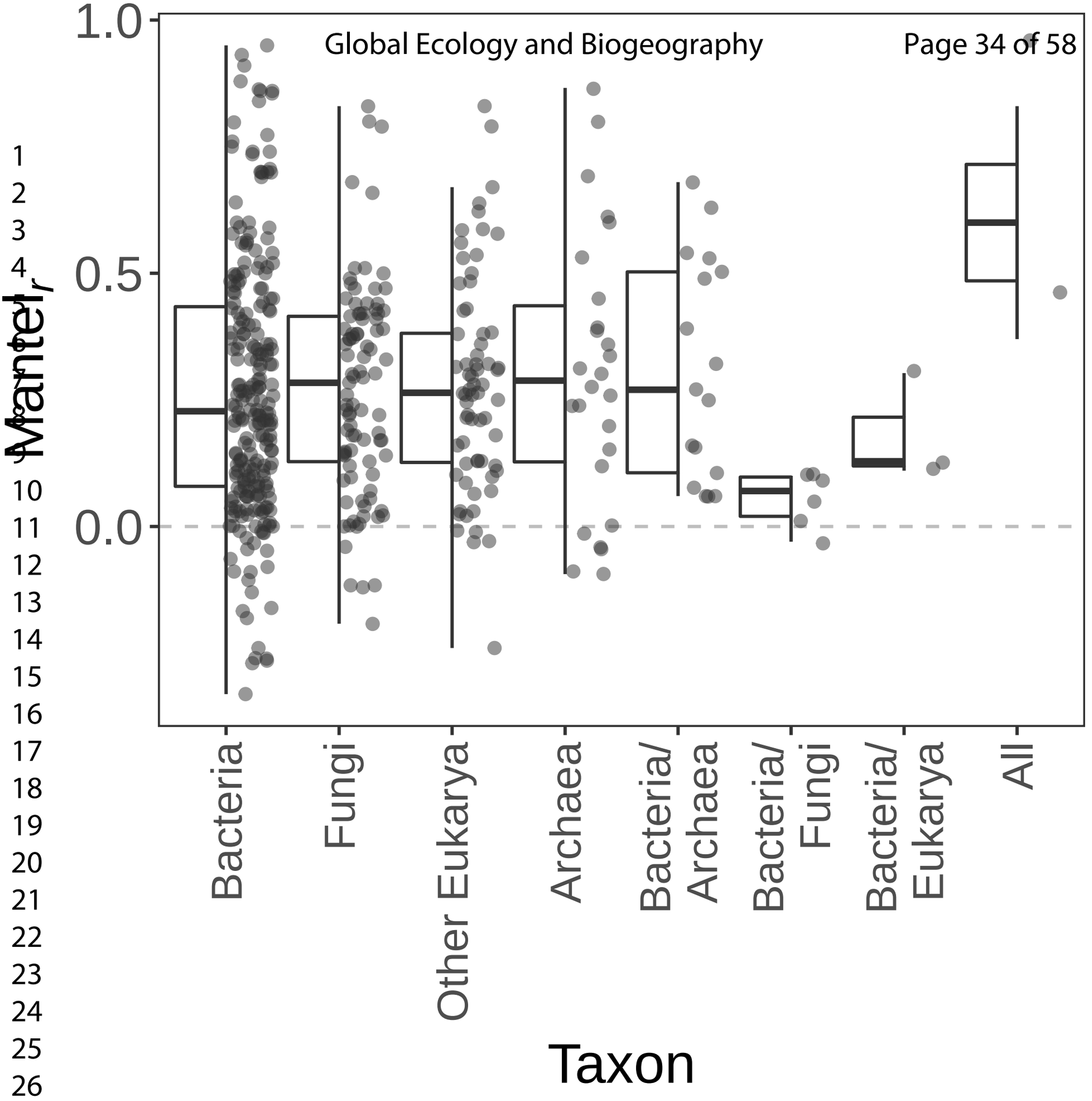
Steinbauer, M.J., Dolos, K., Reineking, B. & Beierkuhnlein, C. (2012) Current measures for distance decay in similarity of species composition are influenced by study extent and grain size. *Global Ecology and Biogeography*, **21**, 1203–1212.

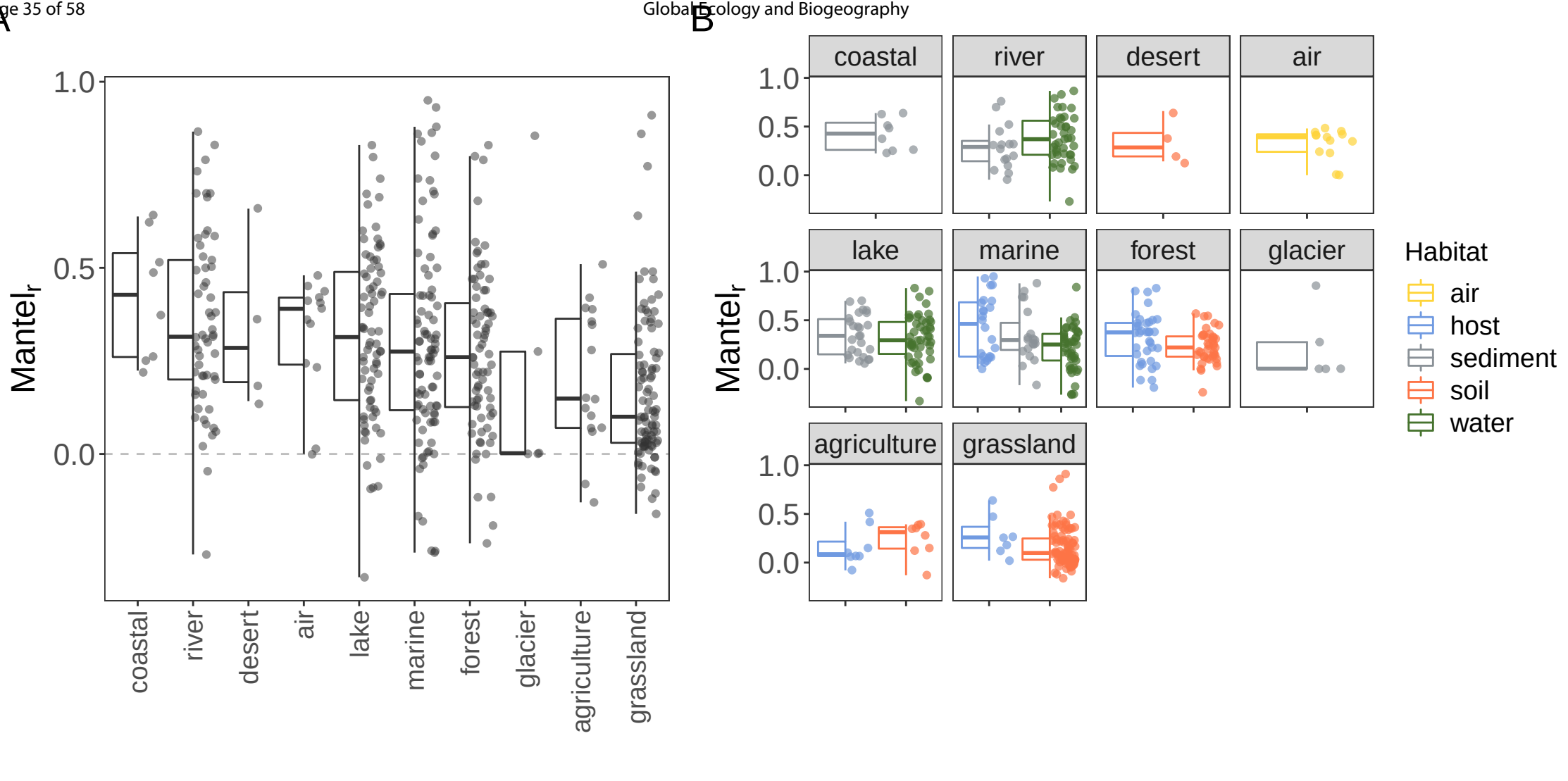
Vašutová, M., Mleczko, P., López-García, A., Maček, I., Boros, G., Ševčík, J., Fujii, S., Hackenberger, D., Tuf, I.H., Hornung, E., Páll-Gergely, B. & Kjøller, R. (2019) Taxi drivers: the role of animals in transporting mycorrhizal fungi. *Mycorrhiza*, **29**, 413–434.

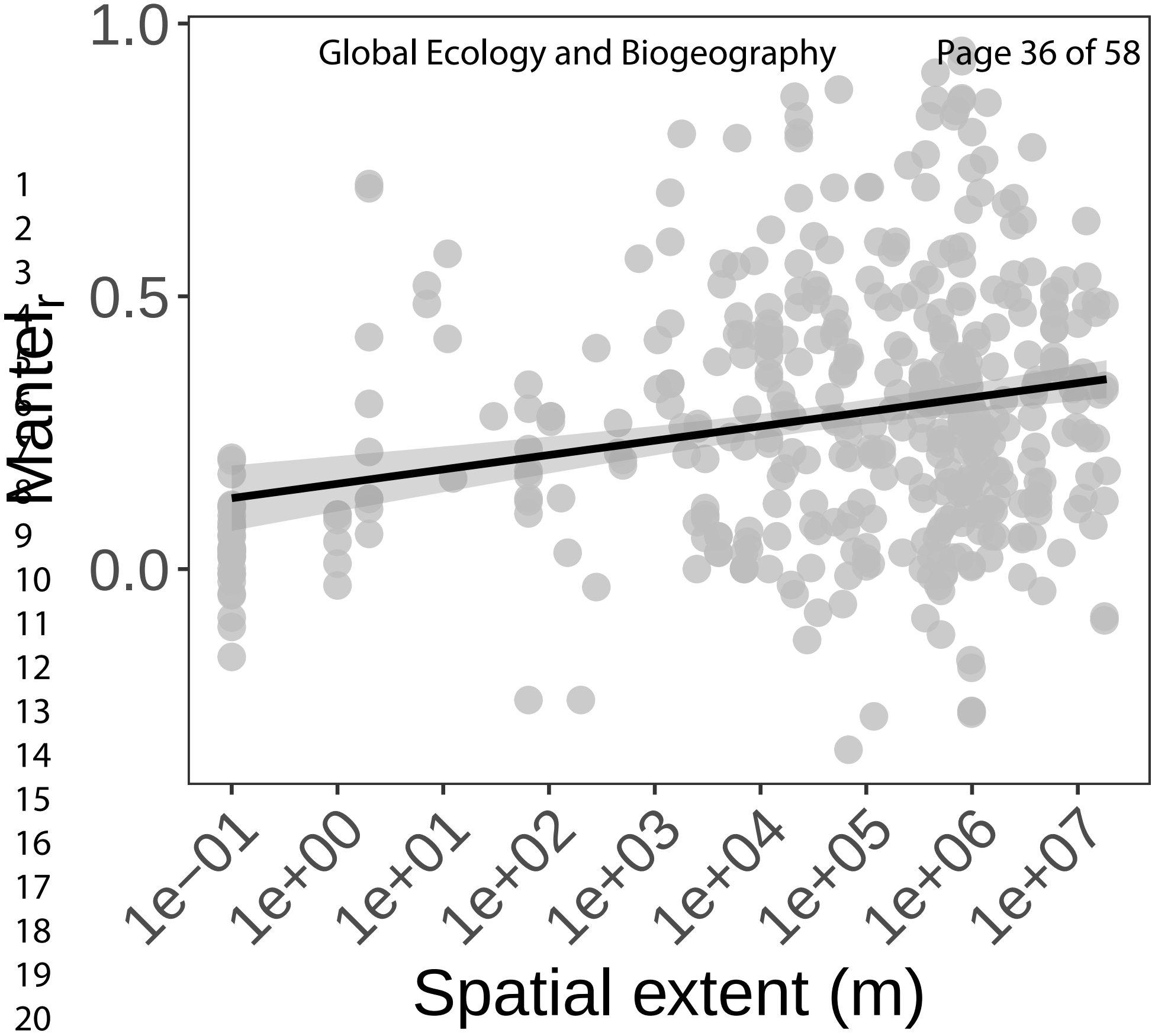
Warmink, J.A., Nazir, R., Corten, B. & van Elsas, J.D. (2011) Hitchhikers on the fungal highway: The helper effect for bacterial migration via fungal hyphae. *Soil Biology and Biochemistry*, **43**, 760–765.

Wilkinson, D.M. (2001) What is the upper size limit for cosmopolitan distribution in free-living microorganisms? *Journal of Biogeography*, **28**, 285–291.

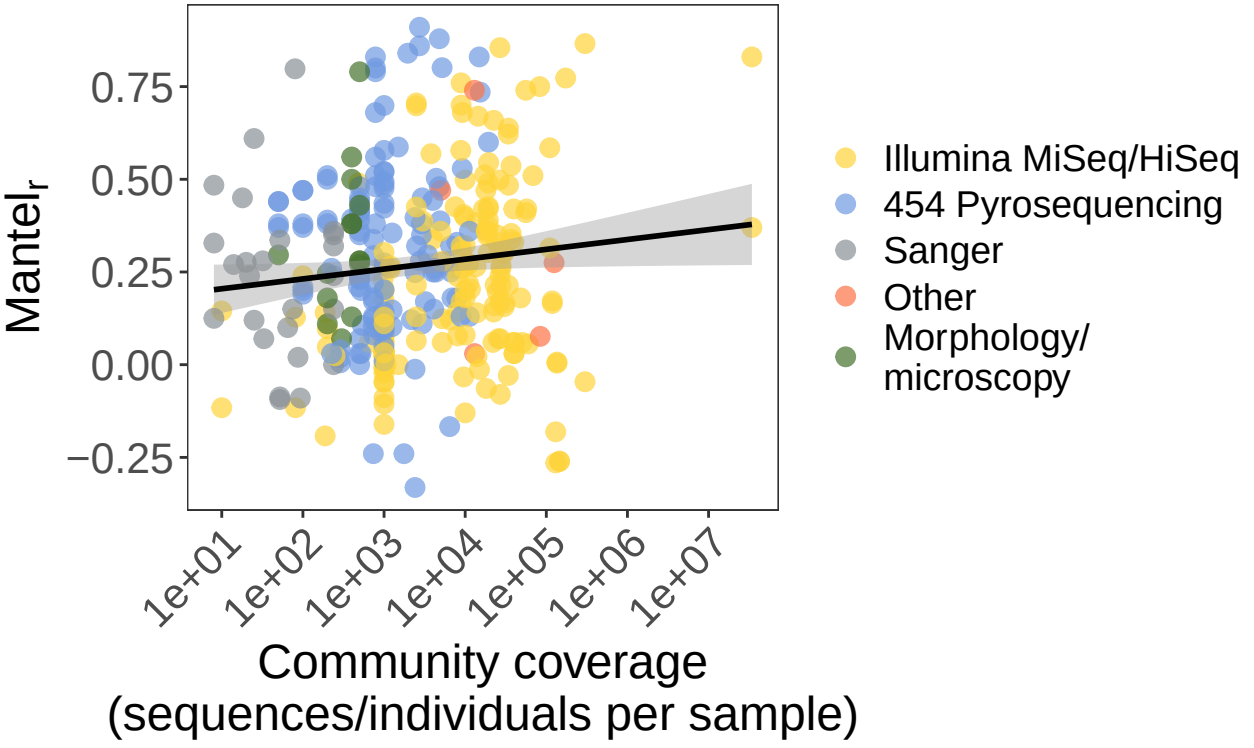
Wilkinson, D.M., Koumoutsaris, S., Mitchell, E.A.D. & Bey, I. (2012) Modelling the effect of size on the aerial dispersal of microorganisms. *Journal of Biogeography*, **39**, 89–97.


1.0
0.5
0.0
Mantel r
Bacteria
Fungi
Other Eukarya
Archaea
Bacteria/ Archaea
Bacteria/ Fungi
Bacteria/ Eukarya
All
Taxon

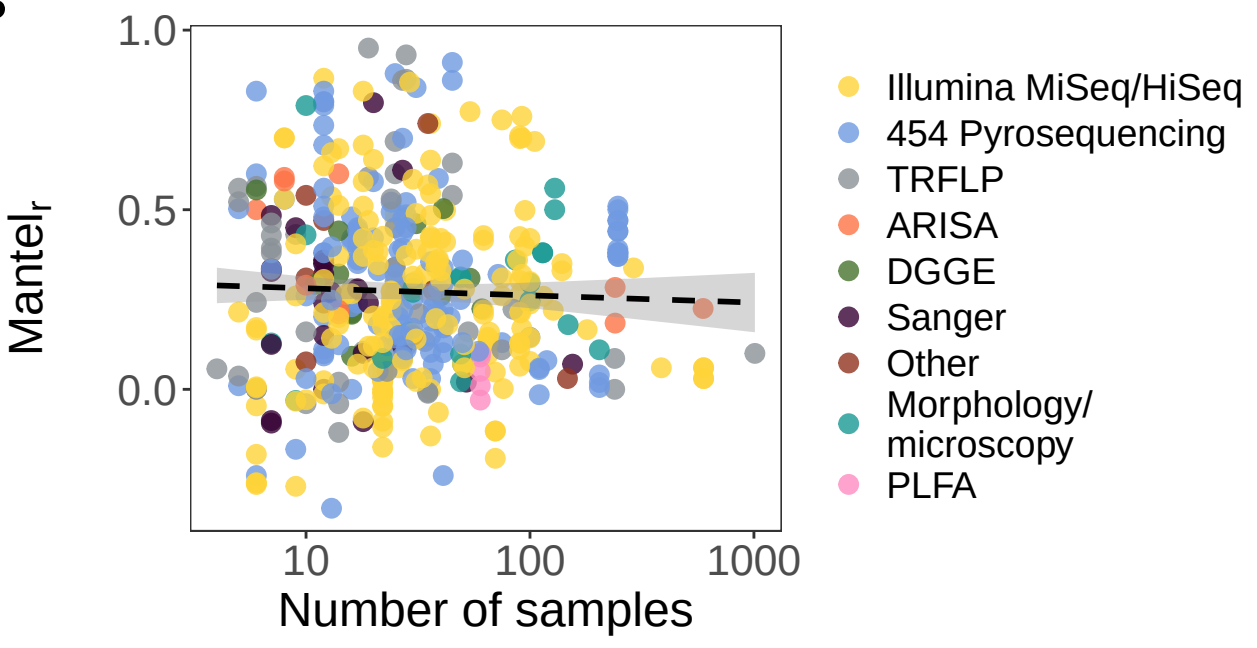

A
B
Mantel$_r$
1.0
0.5
0.0
coastal
river
desert
air
lake
marine
forest
glacier
agriculture
grassland
Habitat
air
host
sediment
soil
water

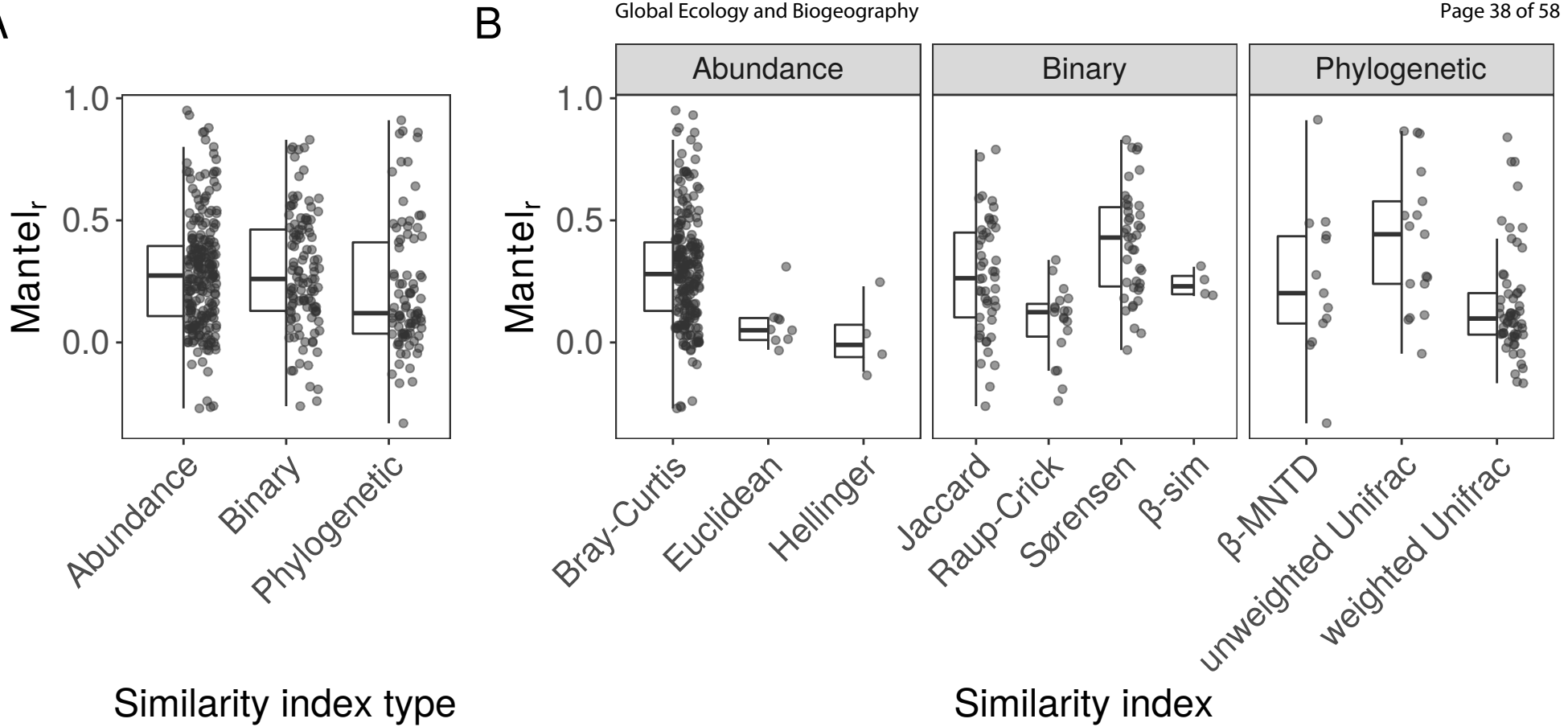

1.0
0.5
0.0
Mantel r
1e−01
1e+00
1e+01
1e+02
1e+03
1e+04
1e+05
1e+06
1e+07
Spatial extent (m)

A


Mantel$_r$
0.75
0.50
0.25
0.00
−0.25
1e+01
1e+02
1e+03
1e+04
1e+05
1e+06
1e+07
Community coverage
(sequences/individuals per sample)
Illumina MiSeq/HiSeq
454 Pyrosequencing
Sanger
Other
Morphology/
microscopy

B


Mantel$_r$
1.0
0.5
0.0
10
100
1000
Number of samples
Illumina MiSeq/HiSeq
454 Pyrosequencing
TRFLP
ARISA
DGGE
Sanger
Other
Morphology/
microscopy
PLFA


A
B
Abundance
Binary
Phylogenetic
Mantel$_r$
1.0
0.5
0.0
Abundance
Binary
Phylogenetic
Similarity index type
Bray-Curtis
Euclidean
Hellinger
Jaccard
Raup-Crick
Sørensen
β-sim
β-MNTD
unweighted Unifrac
weighted Unifrac
Similarity index

Supplementary Information

Title: What drives study-dependent differences in distance-decay relationships of microbial communities?

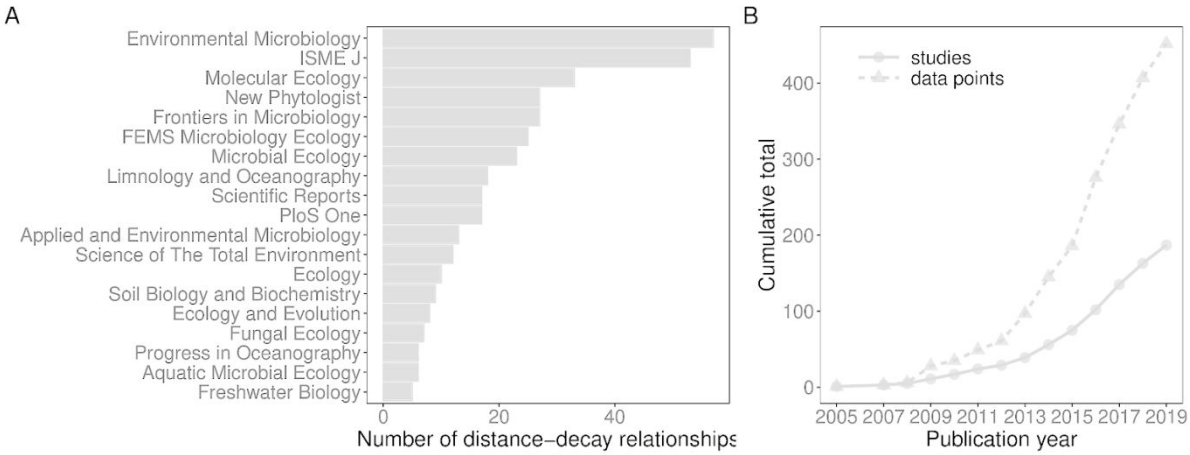


Figure S1. (A) The number of distance-decay relationships obtained from each journal. Only journals with five or more distance-decay relationships are shown for clarity. (B) The cumulative total of suitable studies (circles) and distance-decay relationships (triangles) according to publication year.

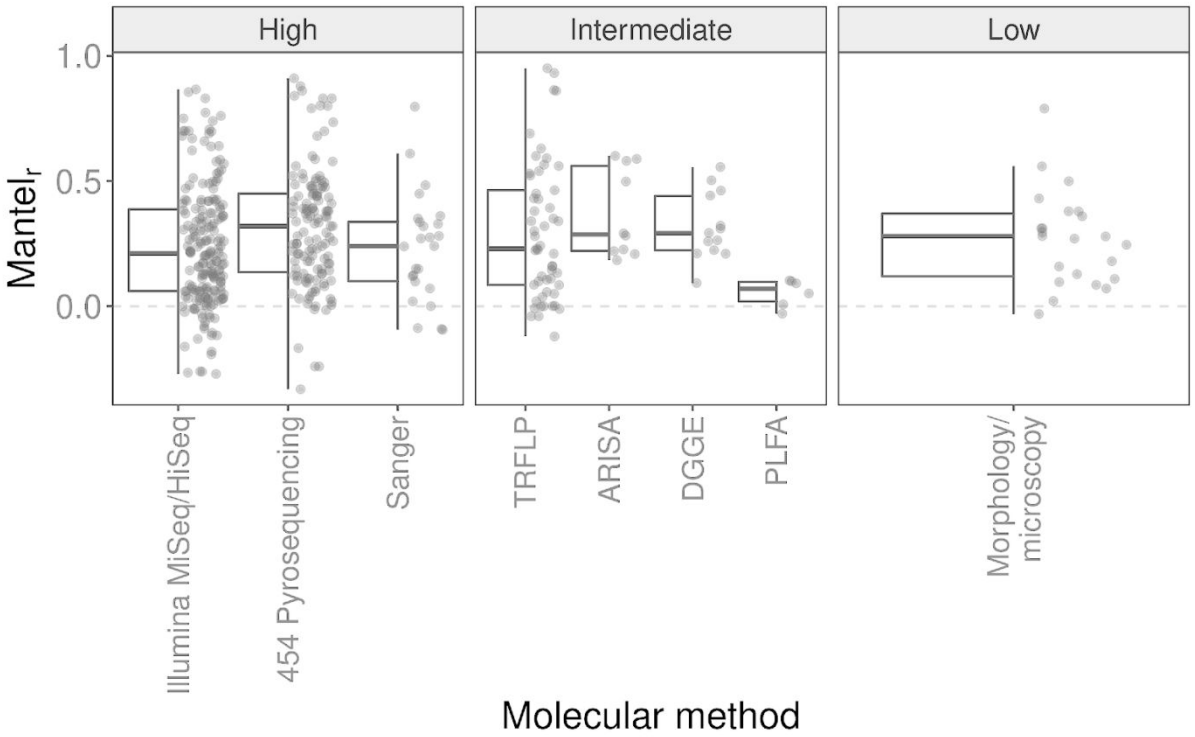


Figure S2. Mantel correlation coefficients of distance-decay relationships where the microbial community was characterised using molecular or morphological methods of varying resolution. High Mantel correlation coefficients indicate a stronger distance-decay relationship. Abbreviated molecular techniques are defined as follows: (TRFLP = Terminal Restriction Fragment Length Polymorphism; ARISA = Automated Ribosomal Intergenic Spacer Analysis; DGGE = Denaturing Gradient Gel Electrophoresis; PLFA = PhosphoLipid Fatty Acid analysis).

Table S1. Metadata extracted from each study confirmed to be suitable for inclusion in this study. Data will be deposited to the Dryad data repository upon acceptance of this manuscript.

title	year	journal	method	markerGen	nSamples	seqDepth	simIndex	otuDefinitio
Plant divers	2017	Soil Biology	illumina		290	4000	bray	0.97
The differer	2017	Catena	pyrosequer	16S	45	2735	u_unifrac	0.97
The differer	2017	Catena	pyrosequer	16S	45	2735	betaMNTD	0.97
Bacterial D	2017	Microbial E	pyrosequer	16S	75	11248	bray	0.97
Similar com	2017	Limnology a	pyrosequer	16S	21	7909	bray	0.97
Similar com	2017	Limnology a	pyrosequer	16S	21	7909	bray	0.97
Latitudinal v	2017	Marine Eco	TRFLP	16S	45	NA	bray	band
Latitudinal v	2017	Marine Eco	TRFLP	16S	45	NA	bray	band
Latitudinal v	2017	Marine Eco	illumina	16S	18	9226	bray	97
Temporal a	2017	PloS One	pyrosequer	16S	6	19358	bray	97
Temporal a	2017	PloS One	ARISA	ITS	6	NA	bray	band
Regional va	2016	Scientific R	illumina	16S	34	10000	bray	97
Microbial eu	2016	Environmen	DGGE	18S	40	NA	bray	band
Microbial eu	2016	Environmen	illumina	18S	14	14437	bray	NA
Decoupling	2016	Science	illumina	metagenom	139	24644	bray	0.99
Decoupling	2016	Science	illumina	metagenom	139	24644	bray	0.99
Geographic	2016	Frontiers in	pyrosequer	16S	27	400	bray	0.97
Forest area	2016	Ecology	pyrosequer	28S	36	500	bray	0.97
Forest area	2016	Ecology	illumina	ITS	36	1500	bray	0.97
Diversity, B	2016	Frontiers in	pyrosequer	16S	9	6442	w_unifrac	0.97
The local e	2016	Environmen	sanger	ITS	52	87	Horn-moris	0.97
Biogeograp	2016	FEMS Micr	TRFLP	16S	18	NA	bray	band
Biogeograp	2016	FEMS Micr	TRFLP	16S	34	NA	bray	band
Biogeograp	2016	FEMS Micr	TRFLP	16S	53	NA	bray	band
Biogeograp	2016	FEMS Micr	TRFLP	16S	28	NA	bray	band
Taxon inter	2016	Molecular E	Ion Torrent	16S	37	124582	bray	0.97
Biogeograp	2016	Frontiers in	illumina	16S	115	7300	bray	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	70	200	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	70	10	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	70	81	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	70	188	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	65	200	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	65	10	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	65	81	Raup-Crick	0.97
Ectomycorr	2016	Molecular E	illumina	ITS	65	188	Raup-Crick	0.97
Interactions	2016	ISME	illumina	16S	386	50323	bray	0.97
Patterns an	2016	ISME	pyrosequer	18S	28	4783	bray	1
Patterns an	2016	ISME	morphology	NA	86	NA	bray	NA
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmen	illumina	16S	22	1000	w_unifrac	0.97

Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Spatial scal	2016	Environmei	illumina	16S	22	1000	w_unifrac	0.97
Fungal com	2016	Scientific R	pyrosequer	ITS	30	3433	bray	0.97
Microbial eu	2016	Molecular E	illumina	18S	13	36744	sorensen	1
Bacterial cc	2016	Environmei	pyrosequer	16S	12	2900	bray	0.97
Bacterial cc	2016	Environmei	pyrosequer	16S	12	2900	bray	0.97
Biogeograp	2016	Microbial E	illumina	16S	91	27719	theta	0.97
Biogeograp	2016	Microbial E	illumina	16S	91	27719	jaccard	0.97
Salinity sha	2016	Scientific R	illumina	16S	9	60000	bray	0.97
Contrasting	2016	New Phytol	illumina	16S	26	660	bray	0.97
Contrasting	2016	New Phytol	illumina	ITS	26	970	bray	0.98
Diversity ar	2016	PloS One	pyrosequer	16S	27	6883	jaccard	0.97
Archaeal ar	2016	PloS One	TRFLP	16S	239	NA	Raup-Crick	band
Archaeal ar	2016	PloS One	TRFLP	16S	239	NA	Raup-Crick	band
Decoupled	2016	Environmei	illumina	16S	75	83008	bray	0.97
Decoupled	2016	Environmei	TRFLP	dsrA	75	NA	bray	band
The roles o	2016	Hydrobiolo	morpholog	NA	204	200	bray	NA
Scale-depe	2015	FEMS Micr	illumina	16S	54	173260	bray	0.97
Environmei	2015	Journal of I	morpholog	NA	29	NA	bray	NA
Spatiotemp	2015	Applied Mic	DGGE	16S	16	NA	jaccard	band
Spatiotemp	2015	Applied Mic	DGGE	16S	16	NA	jaccard	band
Spatiotemp	2015	Applied Mic	DGGE	18S	16	NA	jaccard	band
Spatiotemp	2015	Applied Mic	DGGE	18S	16	NA	jaccard	band
Depth shap	2015	Environmei	pyrosequer	18S	39	1500	sorensen	0.97
Bacterial bi	2015	Environmei	illumina	16S	95	10000	w_unifrac	0.97
Bacterial bi	2015	Environmei	illumina	16S	95	10000	jaccard	0.97
Soil bacteri	2015	Environmei	Pac-Bio	16S	12	5000	w_unifrac	0.95
Quantifying	2015	ISME	pyrosequer	26S	36	1257	jaccard	0.98
Quantifying	2015	ISME	pyrosequer	26S	36	1257	jaccard	0.98
Quantifying	2015	ISME	pyrosequer	26S	36	1257	jaccard	0.98
The biogeo	2015	ISME	illumina	16S	42	26322	bray	0.97
Biogeograp	2015	Microbial E	illumina	16S	12	300000	u_unifrac	0.98
Biogeograp	2015	Microbial E	illumina	16S	6	300000	u_unifrac	0.98
Quantifying	2015	Environmei	DGGE	16S	31	NA	jaccard	band
Bacterial ar	2015	Microbial E	pyrosequer	16S	12	5192	bray	0.97
Bacterial ar	2015	Microbial E	pyrosequer	16S	12	15320	bray	0.97
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
Seasonal c	2014	Soil Biolog	plfa	NA	60	NA	euclidean	band
A continent	2015	New Phytol	pyrosequer	ITS	247	50	sorensen	0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	50	bray	0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	50	Morisita-ho	0.95

A continent	2015	New Phytol	pyrosequer	ITS	247	50	jaccard		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	100	sorensen		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	100	bray		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	100	Morisita-ho		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	100	jaccard		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	200	sorensen		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	200	bray		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	200	Morisita-ho		0.95
A continent	2015	New Phytol	pyrosequer	ITS	247	200	jaccard		0.95
Catchment-	2015	ISME	pyrosequer	16S	23	2179	bray	NA	
Biogeograp	2015	Aquatic Mic	pyrosequer	16S	37	500	bray		0.97
Biogeograp	2015	Aquatic Mic	pyrosequer	16S	37	500	bray		0.85
Aquatic bac	2015	Internationa	sanger	16S	20	80	sorensen		0.97
Testing the	2015	Aquatic Mic	morphology	NA	50	300	bray	NA	
Bacterial ar	2015	Aquatic Mic	pyrosequer	16S	8	9200	bray		0.97
Plant divers	2015	Ecology Le	illumina	16S	25	100	bray		0.97
Plant divers	2015	Ecology Le	illumina	16S	25	18000	bray		0.97
Plant divers	2015	Ecology Le	illumina	ITS	25	485	bray		0.97
Environmer	2014	Fungal Eco	pyrosequer	18S	21	1000	euclidean		0.97
Environmer	2014	Fungal Eco	pyrosequer	18S	5	1000	euclidean		0.97
A phylogen	2014	Molecular E	sanger	16S	18	65	u_unifrac		0.97
Biogeograp	2014	Proceeding	illumina	16S	596	40000	bray		0.97
Biogeograp	2014	Proceeding	illumina	16S	596	40000	jaccard		0.97
Biogeograp	2014	Proceeding	illumina	18S	596	40000	bray		0.97
Biogeograp	2014	Proceeding	illumina	18S	596	40000	jaccard		0.97
Distance-D	2014	PloS One	TRFLP	16S	25	NA	sorensen	band	
Distance-D	2014	PloS One	TRFLP	mcrA	25	NA	sorensen	band	
Distance-D	2014	PloS One	TRFLP	16S	25	NA	sorensen	band	
Distance-D	2014	PloS One	TRFLP	16S	25	NA	bray	band	
Distance-D	2014	PloS One	TRFLP	mcrA	25	NA	bray	band	
Distance-D	2014	PloS One	TRFLP	16S	25	NA	bray	band	
Spore dispe	2014	New Phytol	pyrosequer	ITS	16	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	16	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	16	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	16	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	16	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Spore dispe	2014	New Phytol	pyrosequer	ITS	17	500	bray		0.97
Soil fungal	2014	Molecular E	pyrosequer	ITS	204	289	jaccard	NA	
Soil fungal	2014	Molecular E	pyrosequer	ITS	204	289	betaMNTD	NA	
Soil fungal	2014	Molecular E	pyrosequer	ITS	204	289	betaMPD	NA	
Spatial Sca	2014	Microbial E	pyrosequer	16S	30	520	bray		0.97
Spatial Sca	2014	Microbial E	pyrosequer	16S	26	520	bray		0.97
SSU rDNA	2014	PloS One	TRFLP	18S	35	NA	bray	band	
SSU rDNA	2014	PloS One	TRFLP	18S	35	NA	bray	band	
Pyrosequer	2014	Journal of A	pyrosequer	16S	6	1759	bray		0.97
Neotropical	2014	Environmer	pyrosequer	16S	5	4400	jaccard		0.97
Diversity ar	2014	Applied an	pyrosequer	16S	49	4000	bray		0.97
The spatial	2014	Ecology	TRFLP	16S	16	NA	bray	band	
Drivers sha	2014	Molecular E	pyrosequer	16S	30	4346	bray		0.97
Differentiati	2014	Environmer	sanger	mcrA	27	25	w_unifrac		1

Differentiati	2014	Environmer	sanger	mcrA	27		25	bray		1
Biogeograp	2014	Applied an	pyrosequer	16S	25		4800	bray		0.97
Host rules:	2013	FEMS Micr	TRFLP	16S	28	NA		bray	band	
Host rules:	2013	FEMS Micr	TRFLP	16S	28	NA		bray	band	
Host rules:	2013	FEMS Micr	TRFLP	16S	27	NA		bray	band	
Host rules:	2013	FEMS Micr	TRFLP	16S	27	NA		bray	band	
Host rules:	2013	FEMS Micr	TRFLP	16S	19	NA		bray	band	
Host rules:	2013	FEMS Micr	TRFLP	16S	19	NA		bray	band	
Environmer	2013	Ecosphere	TRFLP	16S	32	NA		bray	band	
Dispersal ir	2013	ISME	pyrosequer	ITS	44		100	beta_sim		0.97
Dispersal ir	2013	ISME	pyrosequer	ITS	36		100	beta_sim		0.97
The biogeo	2013	ISME	DGGE	ITS	61	NA		NA	band	
Phylogenet	2013	ISME	pyrosequer	16S	12		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	12		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	27		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	27		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	20		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	28		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	24		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	25		1000	u_unifrac		0.97
Phylogenet	2013	ISME	pyrosequer	16S	12		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	12		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	27		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	27		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	20		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	28		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	24		1000	betaMNTD		0.97
Phylogenet	2013	ISME	pyrosequer	16S	25		1000	betaMNTD		0.97
Geographic	2013	FEMS Micr	TRFLP	18S	24	NA		bray	band	
Geographic	2013	FEMS Micr	pyrosequer	18S	6		14890	bray		0.95
Biogeograp	2013	ISME	pyrosequer	16S	39		2800	bray		0.97
Contempor	2013	ISME	pyrosequer	16S	59		540	w_unifrac		0.97
Microbial bi	2013	Aquatic Mic	DGGE	16S	14	NA		bray	band	
Microbial bi	2013	Aquatic Mic	DGGE	18S	14	NA		bray	band	
Distance D	2013	PloS One	morphology	NA	114		400	bray	NA	
Distance D	2013	PloS One	morphology	NA	129		400	bray	NA	
Distance D	2013	PloS One	morphology	NA	114		400	simpson	NA	
Distance D	2013	PloS One	morphology	NA	129		400	simpson	NA	
Geographic	2012	Environmer	pyrosequer	16S	17		4000	u_unifrac		0.97
Dispersal li	2012	Ecology an	TRFLP	16S	12	NA		bray	band	
Dispersal li	2012	Ecology an	TRFLP	16S	12	NA		bray	band	
Dispersal li	2012	Ecology an	sanger	16S	12		239	u_unifrac		0.97
Dispersal li	2012	Ecology an	sanger	16S	12		239	bray		0.9
Dispersal li	2012	Ecology an	sanger	16S	12		239	bray		0.93
Dispersal li	2012	Ecology an	sanger	16S	12		239	bray		0.95
Dispersal li	2012	Ecology an	sanger	16S	12		239	bray		0.97
Dispersal li	2012	Ecology an	sanger	16S	12		239	bray		0.99
Bacterial as	2012	Biogeoscie	pyrosequer	16S	16		6687	bray		0.97
Biogeograp	2011	Molecular E	pyrosequer	16S	31		1959	w_unifrac		0.97
Ecology an	2012	Frontiers in	TRFLP	16S	84	NA		bray	band	
Bacterial cc	2011	Freshwater	DGGE	16S	41	NA		jaccard	band	
The bacteri	2011	Environmer	TRFLP	16S	1010	NA		bray	band	
Disentangli	2011	Limnology a	morphology	NA	100		200	sorensen	NA	
Disentangli	2011	Limnology a	morphology	NA	100		50	sorensen	NA	
Disentangli	2011	Limnology a	TRFLP	16S	100	NA		sorensen	band	
Metacomm	2011	Ecology	TRFLP	16S	14	NA		hellinger	band	

Metacomm	2011	Ecology	TRFLP	ITS	14	NA		hellinger	band	
Metacomm	2011	Ecology	TRFLP	16S	14	NA		hellinger	band	
Metacomm	2011	Ecology	TRFLP	ITS	14	NA		hellinger	band	
Possible int	2011	ISME	TRFLP	16S	6	NA		bray	band	
Possible int	2011	ISME	TRFLP	16S	6	NA		bray	band	
Possible int	2011	ISME	TRFLP	16S	6	NA		bray	band	
Evidence o	2010	Ecology	sanger	ITS	155		33	bray		0.97
The ecolog	2010	Environmer	pyrosequer	16S	119		750	w_unifrac		0.97
Community	2010	Freshwater	sanger	amoA	17		20	w_unifrac		0.98
Life history	2010	Molecular E	phylochip	16S	10	NA		bray	NA	
Life history	2010	Molecular E	phylochip	16S	10	NA		bray	NA	
Biogeograp	2010	Journal of E	morphology	NA	7		400	sorensen	NA	
Microbial B	2009	Applied and	sanger	16S	7		52	jaccard		0.99
Microbial B	2009	Applied and	sanger	16S	7		52	jaccard		0.97
Microbial B	2009	Applied and	sanger	16S	7		52	jaccard		0.95
Microbial B	2009	Applied and	sanger	16S	7		8	jaccard		0.99
Microbial B	2009	Applied and	sanger	16S	7		8	jaccard		0.97
Microbial B	2009	Applied and	sanger	16S	7		8	jaccard		0.95
Relationshi	2009	Journal of F	morphology	NA	9	NA		sorensen	NA	
Biogeograp	2009	Environmer	ARISA	16S	593	NA		sorensen	band	
Bar-Coded	2009	Applied and	pyrosequer	16S	39		484	bray	NA	
Contrasting	2009	Limnology a	TRFLP	16S	7	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	9	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	4	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	7	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	6	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	5	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	14	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	5	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	7	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	7	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	6	NA		sorensen	band	
Contrasting	2009	Limnology a	TRFLP	16S	5	NA		sorensen	band	
Bacterial co	2009	Freshwater	DGGE	16S	6	NA		jaccard	band	
Relationshi	2008	Microbial E	TRFLP	16S	10	NA		jaccard	band	
Relationshi	2008	Microbial E	TRFLP	16S	10	NA		jaccard	band	
Water mass	2008	Limnology a	DGGE	18S	54	NA		euclidean	band	
Phylogenet	2007	Applied and	sanger	16S	18		93	bray		0.99
Environmer	2007	Ecology	TRFLP	16S	23	NA		bray	band	
Does ecosy	2005	Ecology	DGGE	16S	11	NA		jaccard	band	
Large varia	2016	ISME	pyrosequer	18S	25		6625	bray		0.97
Microhabita	2015	FEMS Micr	illumina	16S	33		1000	w_unifrac		0.97
Methanoge	2012	Biogeocher	sanger	16S	30		75	sorensen		0.97
Environmer	2011	Microbial E	sanger	nifH	13		14	rao		0.99
[FeFe]-hydı	2010	ISME	sanger	hydA	9		18	rao	NA	
Phyllospher	2016	Microbial E	pyrosequer	ITS	12		784	sorensen		0.9
Phyllospher	2016	Microbial E	pyrosequer	ITS	12		784	sorensen		0.95
Phyllospher	2016	Microbial E	pyrosequer	ITS	12		784	sorensen		0.97
Phyllospher	2016	Microbial E	pyrosequer	ITS	12		784	sorensen		0.99
Phyllospher	2016	Microbial E	pyrosequer	16S	12		784	sorensen		0.9
Phyllospher	2016	Microbial E	pyrosequer	16S	12		784	sorensen		0.95
Phyllospher	2016	Microbial E	pyrosequer	16S	12		784	sorensen		0.97
Phyllospher	2016	Microbial E	pyrosequer	16S	12		784	sorensen		0.99
Stochastic (	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic (	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic (	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97

Stochastic	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Stochastic	2016	ISME	pyrosequer	ITS	14		738	Raup-Crick		0.97
Stochastic	2016	ISME	pyrosequer	ITS	41		738	Raup-Crick		0.97
Eutrophicat	2014	Freshwater	morphology	NA	10		500	jaccard	NA	
Eutrophicat	2014	Freshwater	morphology	NA	10		500	jaccard	NA	
High diaton	2019	Marine Bio	morphology	NA	30		500	bray	NA	
High diaton	2019	Marine Bio	morphology	NA	30		500	bray	NA	
High diaton	2019	Marine Bio	morphology	NA	30		500	bray	NA	
Microbial di	2019	Freshwater	illumina	16S	36		55890	w_unifrac		0.97
Microbial di	2019	Freshwater	illumina	16S	42		55890	w_unifrac		0.97
The local e	2019	Catena	illumina	ITS	24		27000	bray		0.97
The local e	2019	Catena	illumina	ITS	24		27000	u_unifrac		0.97
Depth and	2019	Science of	illumina	16S	20		22000	bray		0.97
Depth and	2019	Science of	illumina	18S	20		6600	bray		0.99
Diversity Di	2019	Frontiers in	illumina	18S	21		13595	bray		0.97
Diversity Di	2019	Frontiers in	illumina	18S	15		13595	bray		0.97
Ammonia C	2019	Applied and	sanger	amoA	19		22	w_unifrac		0.95
Ammonia C	2019	Applied and	sanger	amoA	17		32	w_unifrac		0.95
Stochastic	2019	Microbiome	illumina	18S	30		110394	bray		0.97
Stochastic	2019	Microbiome	illumina	18S	30		110394	bray		0.97
Integrated	2019	Frontiers in	illumina	18S	22		7532	bray		0.97
Integrated	2019	Frontiers in	morphology	NA	22	NA		bray	NA	
Not by Sali	2019	Soil Scienc	illumina	ITS	31		30000	bray		0.97
Microbiota	2019	International	Ion Torrent	16S	10		84144	bray		0.94
Large-scale	2019	Microbiolog	illumina	16S	35	NA		bray		0.97
Biogeograp	2019	Science of	illumina	16S	24		11020	bray		0.97
Biogeograp	2019	Science of	illumina	16S	24		11020	w_unifrac		0.97
Biogeograp	2019	Science of	illumina	16S	24		11020	u_unifrac		0.97
Functional	2019	Frontiers in	illumina	18S	180		30890	bray		0.97
On-Site An	2019	Applied and	Ion Torrent	16S	147		13051	bray		0.98
Upland Soil	2019	Science of	illumina	pmoA	30		30381	bray		0.82
Community	2019	Water	illumina	16S	100		16854	bray		0.97
Community	2019	Water	illumina	18S	100		28993	bray		0.97
Phosphorus	2019	FEMS Micr	illumina	16S	9		9563	bray		0.97
Phosphorus	2019	FEMS Micr	illumina	16S	9		9563	bray		0.97
Microbial E	2019	Microbial E	illumina	16S	12		15000	bray		0.97
Microbial E	2019	Microbial E	illumina	ITS	12		9000	bray		0.97
Microbial E	2019	Microbial E	illumina	ITS	12		250	bray		0.97
Historical F	2019	Frontiers in	pyrosequer	16S	10		226	jaccard		0.97
Historical F	2019	Frontiers in	pyrosequer	16S	10		226	bray		0.97
Molecular c	2019	Acta Ocear	pyrosequer	18S	37		100	bray		0.97
Distinct bio	2018	Science of	illumina	16S	50		25571	bray		0.97
How bacter	2018	Molecular E	illumina	16S	105	NA		bray		0.97
Influence o	2018	Scientific R	pyrosequer	ITS	7	NA		bray		0.97
THE EFFE	2018	Journal of F	pyrosequer	rbcL	72		2515	bray	NA	
Multiple prc	2018	Scientific R	illumina	16S	92		8918	bray		0.97
Multiple prc	2018	Scientific R	illumina	16S	92		8918	bray		0.97
Multiple prc	2018	Scientific R	illumina	16S	92		8918	jaccard		0.97
Multiple prc	2018	Scientific R	illumina	16S	92		8918	jaccard		0.97
Association	2018	Ecography	morphology	NA	49	NA		bray	NA	
Association	2018	Ecography	morphology	NA	49	NA		jaccard	NA	
Association	2018	Ecography	morphology	NA	49	NA		bray	NA	
Association	2018	Ecography	morphology	NA	49	NA		jaccard	NA	

Co-occurre	2018	Microbes a	illumina	16S	8	NA		bray		0.97
Co-occurre	2018	Microbes a	illumina	16S	9	NA		bray		0.97
Soil bacteri	2018	Applied Soi	illumina	16S	100		8047	bray		0.97
Soil bacteri	2018	Applied Soi	illumina	16S	100		8047	w_unifrac		0.97
Soil bacteri	2018	Applied Soi	illumina	16S	100		8047	betaMNTD		0.97
Biogeograp	2018	Science of	illumina	16S	51		5209	bray		0.97
Biogeograp	2018	Science of	illumina	16S	51		5209	bray		0.97
Biogeograp	2018	Science of	illumina	16S	51		5209	bray		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		135317	bray		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		145541	bray		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		118482	bray		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		130664	bray		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		135317	jaccard		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		145541	jaccard		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		118482	jaccard		0.97
Vertical anc	2018	Scientific R	illumina	16S	6		130664	jaccard		0.97
Why do mic	2018	ISME	illumina	16S	35		3790	canberra		0.97
Plant growt	2018	Land Degra	illumina	16S	21		18844	bray		0.97
Ecological a	2018	Oecologia	morphology	NA	148		200	sorensen	NA	
Biogeograp	2018	Applied anc	pyrosequer	16S	50		8887	bray		0.987
Biogeograp	2018	Applied anc	pyrosequer	16S	50		11273	bray		0.987
Environmer	2018	Scientific R	pyrosequer	ITS	30		3229	sorensen		0.97
Ammonia-C	2018	Frontiers in	pyrosequer	amoA	26		4900	w_unifrac		0.85
Ammonia-C	2018	Frontiers in	pyrosequer	amoA	26		4100	w_unifrac		0.85
Contrasting	2018	Journal of M	illumina	16S	8	NA		bray		0.97
Contrasting	2018	Journal of M	illumina	16S	8	NA		bray		0.97
Facultative	2018	New Phytol	illumina	ITS	43		48363	bray		0.97
The diversit	2018	Environmer	illumina	18S	36		34239	bray		0.97
The diversit	2018	Environmer	illumina	18S	36		34239	bray		0.97
Distribution	2018	PeerJ	illumina	16S	14		10000	w_unifrac		0.97
Soil organic	2018	Functional	illumina	16S	36		19460	bray		0.97
Impact of E	2018	Internationa	illumina	16S	20	NA		w_unifrac		0.97
Highlighting	2018	Molecular E	illumina	ITS	36		19317	bray		0.97
Highlighting	2018	Molecular E	illumina	ITS	36		19317	bray		0.97
Highlighting	2018	Molecular E	illumina	ITS	28		19317	bray		0.97
Highlighting	2018	Molecular E	illumina	ITS	40		19317	bray		0.97
Patterns an	2017	Frontiers in	illumina	18S	9		9513	bray		0.97
Linking bac	2017	Progress in	ARISA	ISR	10	NA		jaccard	NA	
Linking bac	2017	Progress in	ARISA	ISR	8	NA		jaccard	NA	
Linking bac	2017	Progress in	ARISA	ISR	14	NA		jaccard	NA	
Linking bac	2017	Progress in	ARISA	ISR	8	NA		jaccard	NA	
Linking bac	2017	Progress in	ARISA	ISR	14	NA		jaccard	NA	
Linking bac	2017	Progress in	ARISA	ISR	14	NA		jaccard	NA	
Is microbial	2017	Environmer	illumina	ITS	18		57346	bray		0.97
Is microbial	2017	Environmer	illumina	ITS	18		68490	bray		0.97
Is microbial	2017	Environmer	illumina	16S	17		27106	bray		0.97
Is microbial	2017	Environmer	illumina	16S	18		27106	bray		0.97
Deep nirS a	2017	Environmer	Ion Torrent	nirS	35		13000	w_unifrac		0.88
Elevation, s	2017	Fungal Eco	pyrosequer	ITS	27	NA		bray		0.97
Elevation, s	2017	Fungal Eco	pyrosequer	ITS	27	NA		bray		0.97
Distinct sea	2017	Fungal Eco	pyrosequer	18S	27		880	bray		0.99
Biogeograp	2017	Global Eco	illumina	16S	75		27554	beta_sim		0.97
Biogeograp	2017	Global Eco	illumina	16S	75		26578	beta_sim		0.99
Rhizospher	2017	Journal of E	illumina	ITS	19	NA		w_unifrac		0.97
Rhizospher	2017	Journal of E	illumina	ITS	19	NA		w_unifrac		0.97
Distinct mic	2017	Molecular E	illumina	ITS	31		14000	bray		0.97

Distinct mic	2017	Molecular E	illumina	16S		31	19300	bray		0.97
Distinct mic	2017	Molecular E	illumina	ITS		31	14000	bray		0.97
Distinct mic	2017	Molecular E	illumina	16S		31	19300	bray		0.97
Environmer	2017	Molecular E	illumina	ITS		40	39721	bray		0.97
Environmer	2017	Molecular E	illumina	ITS		40	39721	jaccard		0.97
The Patterr	2017	Frontiers in	illumina	ITS		62	NA	bray		0.97
The Patterr	2017	Frontiers in	illumina	16S		62	NA	bray		0.97
The Patterr	2017	Frontiers in	illumina	ITS		62	NA	beta_bray		0.97
The Patterr	2017	Frontiers in	illumina	16S		62	NA	beta_bray		0.97
Patterns an	2017	Frontiers in	illumina	18S		12	33996	bray		0.97
Patterns an	2017	Frontiers in	illumina	18S		12	33996	bray		0.97
Patterns an	2017	Frontiers in	illumina	18S		10	33996	bray		0.97
City-scale c	2017	Microbiome	illumina	18S		76	1058	jaccard		0.97
Climate cor	2017	FEMS Micr	illumina	16S		88	5000	bray		0.97
Microbial di	2017	FEMS Micr	illumina	16S		36	10000	w_unifrac		0.97
Microbial di	2017	FEMS Micr	illumina	16S		14	10000	w_unifrac		0.97
Microbial di	2017	FEMS Micr	illumina	16S		27	10000	w_unifrac		0.97
Biogeograp	2017	Molecular E	illumina	16S		38	36920	bray		0.97
Biogeograp	2017	Molecular E	illumina	16S		38	36920	bray		0.97
Relative rol	2017	Fungal Eco	ARISA	ITS		240	NA	bray	NA	
Relative rol	2017	Fungal Eco	ARISA	ITS		240	NA	bray	NA	
Transition b	2017	Environmer	illumina	18S		24	61 774	sorensen		1
Distance de	2017	Environmer	illumina	ITS		127	NA	bray		0.95
Ecological S	2017	Frontiers in	pyrosequer	ITS		36	4184	bray		0.97
Biogeograp	2017	FEMS Micr	illumina	16S		29	26800	u_unifrac		0.97
Land scale	2017	Environmer	illumina	16S		14	NA	bray		0.97
Land scale	2017	Environmer	illumina	arsM		14	17434	bray		0.97
Geographic	2017	Genes and	pyrosequer	16S		28	2951	jaccard		0.97
Geographic	2017	Genes and	pyrosequer	16S		28	2951	w_unifrac		0.97
High taxonc	2017	Nature Eco	illumina	16S		22	NA	bray		0.99
Distinct Bio	2017	mSystems	pyrosequer	16S		110	NA	bray		0.97
Distinct Bio	2017	mSystems	pyrosequer	16S		110	NA	bray		0.97
Distinct Bio	2017	mSystems	pyrosequer	18S		110	NA	bray		0.97
Fungal com	2017	Soil Biology	illumina	ITS		13	22466	bray		0.97
Fungal com	2017	Soil Biology	illumina	ITS		13	22466	bray		0.97
Floral organ	2019	Molecular E	illumina	16S		16	1200	bray		0.97
Floral organ	2019	Molecular E	illumina	16S	NA		1200	bray		0.97
Environmer	2019	Science of	illumina	16S		20	11612	bray		0.97
Environmer	2019	Science of	illumina	ITS		20	3018	bray		0.97
Abundant a	2018	Frontiers in	illumina	16S		66	22938	bray		0.97
Ecological p	2018	Water Res	illumina	16S		5	23429	bray	NA	
Benthic Alg	2018	Frontiers in	illumina	23S		18	8843	bray		0.97
Phylum-Lev	2018	Geomicrob	illumina	16S		38	24805	bray		0.97
Community	2017	FEMS Micr	pyrosequer	16S		13	2411	betaMNTD		0.97
Community	2017	FEMS Micr	pyrosequer	16S		13	2411	betaMNTD		0.97
Community	2017	FEMS Micr	pyrosequer	16S		13	2411	bray		0.97
Community	2017	FEMS Micr	pyrosequer	16S		13	2411	bray		0.97
Soil Proper	2019	Frontiers in	illumina	16S		39	18182	bray		0.97
Soil Proper	2019	Frontiers in	illumina	16S		39	18182	beta_bray		0.97
Soil Proper	2019	Frontiers in	illumina	16S		39	18182	nes_bray		0.97
Intensive al	2019	Environmer	illumina	metagenon		18	34192561	Mash	NA	
Intensive al	2019	Environmer	illumina	metagenon		18	34192561	Mash	NA	
Highly struc	2018	ISME	illumina	16S		90	2500	sorensen		0.97
Highly struc	2018	ISME	illumina	UPA		90	2500	sorensen		0.97
Highly struc	2018	ISME	illumina	tufA		90	1000	sorensen		0.98
Highly struc	2018	ISME	illumina	16S		90	2500	bray		0.97

Highly struc	2018	ISME	illumina	UPA	90	2500	bray	0.97
Highly struc	2018	ISME	illumina	tufA	90	1000	bray	0.98
Highly struc	2018	ISME	illumina	16S	90	2500	w_unifrac	0.97
Highly struc	2018	ISME	illumina	UPA	90	2500	w_unifrac	0.97
Highly struc	2018	ISME	illumina	tufA	90	1000	w_unifrac	0.98

taxa	habitat	environmer	spatialExte	mantelR	pValue
fungi	soil	grassland	4000	0.338	0.001
bacteria	soil	grassland	451	0.86	0.001
bacteria	soil	grassland	451	0.91	0.001
bacteria	soil	dune	1700	0.13	0.35
bacteria	water	lake	600	0.18	0.001
bacteria	water	lake	600	0.33	0.001
Bac_arch	host	marine	2500	0.63	0.001
Bac_arch	host	marine	2500	0.54	0.001
Bac_arch	host	marine	2500	0.68	0.001
bacteria	host	marine	130	0.6	0.01
bacteria	host	marine	130	0.5	0.01
Bac_arch	sediment	marine	200	0.322	0.001
eukarya	water	lake	2100	0.26	0.002
eukarya	water	lake	2100	0.67	0.003
bacteria	water	marine	9000	0.35	0.004
bacteria	water	marine	9000	0.33	0.001
archaea	soil	agriculture	3400	0.3931	0.001
fungi	host	forest	7	0.03	0.23
fungi	host	forest	7	0	0.56
bacteria	sediment	marine	964	-0.167	0.726
fungi	host	grassland	1450	0.02	0.32
bacteria	water	marine	4200	0.106	0.161
bacteria	water	marine	4200	0.133	0.028
bacteria	host	marine	4200	0.16	0.001
bacteria	host	marine	4200	0.12	0.015
bacteria	sediment	glacier	1664	0.275	0.001
bacteria	soil	forest	NA	0.259	0.001
fungi	host	forest	NA	0.048	0.301
fungi	host	forest	NA	-0.116	0.834
fungi	host	forest	NA	-0.116	0.832
fungi	host	forest	NA	-0.192	0.947
fungi	soil	forest	NA	0.097	0.177
fungi	soil	forest	NA	0.145	0.155
fungi	soil	forest	NA	0.128	0.151
fungi	soil	forest	NA	0.14	0.116
bacteria	water	lake	1700	0.06	0.1
eukarya	water	marine	163	0.48	0.01
eukarya	water	marine	163	0.36	0.01
bacteria	soil	grassland	0.0001	-0.009	0.912
bacteria	soil	grassland	0.0001	0.199	0.011
bacteria	soil	grassland	0.0001	0.001	0.995
bacteria	soil	grassland	0.0001	0.023	0.779
bacteria	soil	grassland	0.0001	-0.009	0.904
bacteria	soil	grassland	0.0001	-0.048	0.568
bacteria	soil	grassland	0.0001	0.204	0.008
bacteria	soil	grassland	0.0001	0.034	0.644
bacteria	soil	grassland	0.0001	0.079	0.407
bacteria	soil	grassland	0.0001	0.112	0.198
bacteria	soil	grassland	0.0001	0.036	0.649
bacteria	soil	grassland	0.0001	0.116	0.159
bacteria	soil	grassland	0.0001	-0.089	0.31
bacteria	soil	grassland	0.0001	0.032	0.703
bacteria	soil	grassland	0.0001	0.027	0.858
bacteria	soil	grassland	0.0001	0.034	0.666
bacteria	soil	grassland	0.0001	0.059	0.498

bacteria	soil	grassland	0.0001	0.098	0.261	
bacteria	soil	grassland	0.0001	0.063	0.428	
bacteria	soil	grassland	0.0001	0.021	0.819	
bacteria	soil	grassland	0.0001	-0.106	0.166	
bacteria	soil	grassland	0.0001	-0.045	0.602	
bacteria	soil	grassland	0.0001	0.086	0.334	
bacteria	soil	grassland	0.0001	0.174	0.016	
bacteria	soil	grassland	0.0001	-0.023	0.79	
bacteria	soil	grassland	0.0001	0.038	0.598	
bacteria	soil	grassland	0.0001	0.039	0.649	
bacteria	soil	grassland	0.0001	0.076	0.4	
bacteria	soil	grassland	0.0001	0.118	0.223	
bacteria	soil	grassland	0.0001	-0.161	0.072	
fungi	sediment	marine	997.47	0.2959	0.001	
eukarya	water	lake	12270	0.536	0.00002	
bacteria	soil	grassland	923	0.35	0.004	
bacteria	soil	grassland	923	0.38	0.002	
bacteria	host	forest	0.45	0.213	0.001	
bacteria	host	forest	0.45	0.268	0.001	
bacteria	sediment	lake	467	0.056	NA	
bacteria	soil	grassland	1530	0.06	NA	
fungi	soil	grassland	1530	0.23	0.002	
bacteria	sediment	marine	18700	0.18	0.017	
bacteria	host	marine	2.5	0.086	0.01	
archaea	host	marine	2.5	0	NA	
bacteria	sediment	marsh	1300	0.75	0.001	
bacteria	sediment	marsh	1300	0.11	0.001	
eukarya	water	marine	1500	0.11	NA	
bacteria	soil	grassland	3700	0.773	0.001	
eukarya	water	river	1150	0.16	0.005	
bacteria	water	river	115	0.092	0.315	
bacteria	water	river	115	0.209	0.022	
eukarya	water	river	115	0.212	0.02	
eukarya	water	river	115	0.263	0.004	
eukarya	sediment	marine	670	0.587	0.001	
bacteria	water	marine	225	0.498	0.001	
bacteria	water	marine	225	0.398	0.001	
bacteria	soil	forest	500	0.47	0.01	
fungi	host	agriculture	885.49	########	0.1114	
fungi	host	agriculture	885.49	########	0.0502	
fungi	soil	agriculture	885.49	########	0.0006	
bacteria	water	lake	2700	0.498	0.01	
archaea	water	river	21	0.866	0.001	
archaea	sediment	river	21	-0.046	NA	
bacteria	water	river	380	0.461	0.001	
archaea	sediment	marine	1000	0.801	0.001	
bacteria	sediment	marine	1000	0.735	0.001	
bac_fungi	soil	grassland	0.001	0.1	0.05	
bac_fungi	soil	grassland	0.001	0.05	NA	
bac_fungi	soil	grassland	0.001	-0.03	NA	
bac_fungi	soil	grassland	0.001	0.09	0.05	
bac_fungi	soil	grassland	0.001	0.01	NA	
bac_fungi	soil	grassland	0.001	0.1	0.05	
fungi	host	forest	6000	0.44	0.001	
fungi	host	forest	6000	0.38	0.001	
fungi	host	forest	6000	0.37	0.001	

fungi	host	forest	6000	0.44	0.001
fungi	host	forest	6000	0.47	0.001
fungi	host	forest	6000	0.38	0.001
fungi	host	forest	6000	0.37	0.001
fungi	host	forest	6000	0.47	0.001
fungi	host	forest	6000	0.5	0.001
fungi	host	forest	6000	0.39	0.001
fungi	host	forest	6000	0.38	0.001
fungi	host	forest	6000	0.51	0.001
bacteria	water	river	90	0.1214	0.048
bacteria	water	lake	1200	0.23	0.001
bacteria	water	lake	1200	0.07	0.019
bacteria	water	lake	1.8	0.7979	0.04117
eukarya	water	river	500	0.07	0.446
Bac_arch	water	marine	7500	0.529	0.1
archaea	soil	grassland	14800	0.24	0.021
bacteria	soil	grassland	14800	0.47	0.001
fungi	soil	grassland	14800	0.49	0.001
fungi	soil	grassland	360	0.05	0.87
fungi	air	air	110	0.01	0.36
bacteria	water	lake	72	0.1	0.31
Bac_arch	soil	grassland	4	0.06	0.001
Bac_arch	soil	grassland	4	0.06	0.001
eukarya	soil	grassland	4	0.03	0.02
eukarya	soil	grassland	4	0.03	0.025
archaea	sediment	lake	1.4	0.45	0.03
archaea	sediment	lake	1.4	0.6	0.001
bacteria	sediment	lake	1.4	0.34	0.006
archaea	sediment	lake	1.4	0.3	0.005
archaea	sediment	lake	1.4	0.69	0.001
bacteria	sediment	lake	1.4	0.34	0.006
fungi	air	air	12	0	0.51
fungi	air	air	12	0.23	0.04
fungi	air	air	12	0.41	0.04
fungi	air	air	12	0.48	0.01
fungi	air	air	12	0.42	0.04
fungi	air	air	12	0.24	0.08
fungi	air	air	12	0.39	0.01
fungi	air	air	12	0.35	0.03
fungi	air	air	12	0.41	0.03
fungi	air	air	12	0.36	0.02
fungi	air	air	12	0.44	0
fungi	air	air	12	0.45	0.02
fungi	soil	grassland	100	0.019	0.22
fungi	soil	grassland	100	0.005	0.577
fungi	soil	grassland	100	0.04	0.12
bacteria	soil	forest	0.15	0.03	0.05
bacteria	soil	forest	1300	0.2	NA
eukarya	water	marine	540	-0.011	NA
eukarya	water	marine	540	-0.008	NA
bacteria	water	urban	0.2	-0.24	0.175
Bac_arch	water	river	290	0.50273	0.05
Bac_arch	sediment	coastal	1350	0.25	0.001
bacteria	water	marine	0.03	0.28	0.01
bacteria	water	marine	500	0.2497	0.01
archaea	sediment	lake	32	0.12	0.14

archaea	sediment	lake	32	0.61		0.001
bacteria	sediment	marine	55	0.879		0.001
bacteria	host	marine	800	0.863	NA	
bacteria	host	marine	800	0.931	NA	
bacteria	host	marine	800	0.085	NA	
bacteria	host	marine	800	0.86	NA	
bacteria	host	marine	800	0.95	NA	
bacteria	host	marine	800	0.591	NA	
bacteria	soil	desert	60	0.21		0.006
fungi	air	urban	0.5	0.2		0.001
fungi	air	urban	0.5	0.19		0.003
fungi	sediment	coastal	4000	0.224		0.003
bacteria	sediment	lake	3	0.093	NA	
bacteria	sediment	lake	3	0.112	NA	
bacteria	sediment	lake	50	0.699		0.001
bacteria	water	lake	50	0.476		0.01
bacteria	sediment	lake	0.011	0.578		0.05
bacteria	sediment	lake	0.007	0.52		0.001
bacteria	sediment	river	33	0.521		0.01
bacteria	sediment	lake	850	0.267	NA	
bacteria	sediment	lake	3	0.099	NA	
bacteria	sediment	lake	3	0.202	NA	
bacteria	sediment	lake	50	0.435		0.001
bacteria	water	lake	50	0.275		0.01
bacteria	sediment	lake	0.011	0.422		0.001
bacteria	sediment	lake	0.007	0.486		0.001
bacteria	water	river	33	0.494		0.001
bacteria	sediment	lake	850	0.143	NA	
eukarya	water	lake	400	0.53		0.01
eukarya	water	lake	400	0.83		0.05
bacteria	water	lake	20	0.28		0.001
Bac_arch	NA	mine	1600	0.106		0.072
bacteria	water	marine	523	0.441		0.003
eukarya	water	marine	523	0.321		0.013
eukarya	water	river	800	0.38		0.001
eukarya	water	river	800	0.56		0.001
eukarya	water	river	800	0.38		0.001
eukarya	water	river	800	0.5		0.001
bacteria	sediment	lake	1670	0.443		0.03
bacteria	soil	forest	343	0.35		0.022
bacteria	soil	forest	343	0.32		0.027
bacteria	soil	forest	343	0.24		0.064
bacteria	soil	forest	343	0		0.479
bacteria	soil	forest	343	0.15		0.186
bacteria	soil	forest	343	0.36		0.043
bacteria	soil	forest	343	0.35		0.036
bacteria	soil	forest	343	0.32		0.031
bacteria	water	marine	7700	0.25		0.04
bacteria	water	marine	700	0.84		0.0001
bacteria	NA	aquifer	1000	0.223		0.001
bacteria	water	lake	2150	0.503		0.0001
bacteria	soil	grassland	1200	0.1		0.001
eukarya	water	lake	800	0.246		0.001
eukarya	water	lake	800	0.296		0.001
bacteria	water	lake	800	0.145		0.001
bacteria	soil	grassland	508	0.23		0.007

fungi	soil	grassland	508	-0.12	NA	
bacteria	soil	grassland	508	0.02		0.41
fungi	soil	grassland	508	-0.04	NA	
bacteria	sediment	glacier	7	0.001		0.007
bacteria	sediment	glacier	7	0		0.392
bacteria	sediment	glacier	7	0.002		0.169
fungi	host	forest	7.8	0.07		0.001
bacteria	host	forest	14000	0.08		0.5
archaea	water	lake	333.52	0.276		0.017
bacteria	soil	forest	350	0.54		0.007
bacteria	soil	forest	350	0.31		0.02
eukarya	water	marine	342	0.129		0.614
archaea	water	lake	17845.5	0.336		0.116
archaea	water	lake	17845.5	-0.094		0.473
archaea	water	lake	17845.5	-0.087		0.486
bacteria	water	lake	17845.5	0.125		0.166
bacteria	water	lake	17845.5	0.484		0.0016
bacteria	water	lake	17845.5	0.328		0.002
eukarya	water	lake	480	-0.031		0.437
bacteria	soil	grassland	NA	0.225		0.001
bacteria	water	river	2	0.21		0.0001
bacteria	water	lake	6.9	0.393		0.01
bacteria	water	lake	6.6	0.431		0.05
bacteria	water	lake	3	0.057	NA	
bacteria	water	lake	3.9	0.38		0.001
bacteria	water	lake	8.7	0.565		0.01
bacteria	water	lake	7.8	0.037	NA	
bacteria	water	lake	7.1	0.225		0.05
bacteria	water	lake	4.5	0.56		0.01
bacteria	water	lake	9.1	0.428		0.01
bacteria	water	lake	6.2	0.463		0.05
bacteria	water	lake	5.2	0.242	NA	
bacteria	water	lake	4.3	0.522		0.001
bacteria	water	lake	5.9	0.555		0.032
archaea	water	marine	4600	-0.04		0.7939
bacteria	water	marine	4600	0.16		0.3512
eukarya	water	marine	2000	0.30932		0.019
bacteria	water	lake	362	-0.09	NA	
bacteria	sediment	river	7.5	0.05	NA	
bacteria	water	lake	7.5	0.292		0.025
eukarya	water	marine	16000	0.32		0.001
bacteria	host	forest	77.5	0.032		0.28
archaea	soil	marsh	775	0.15		0.67
Bac_arch	sediment	river	51.56	0.27		0.01
bacteria	sediment	river	53.3	0.45		0.01
fungi	host	forest	23	0.83		0.001
fungi	host	forest	23	0.8		0.01
fungi	host	forest	23	0.79		0.001
fungi	host	forest	23	0.68		0.001
bacteria	host	forest	23	0.38		0.01
bacteria	host	forest	23	0.48		0.01
bacteria	host	forest	23	0.51		0.01
bacteria	host	forest	23	0.56		0.001
eukarya	soil	forest	0.064	0.338		0.05
eukarya	soil	forest	0.064	-0.24	NA	
fungi	soil	forest	0.064	0.171		0.05

eukarya	soil	forest	0.064	0.22		0.05
eukarya	soil	forest	0.064	0.13	NA	
fungi	soil	forest	0.064	0.18		0.05
eukarya	soil	forest	0.064	0.102	NA	
eukarya	soil	forest	0.064	0.124	NA	
fungi	soil	forest	0.064	0.294		0.05
eukarya	water	river	6	0.43		0.01
eukarya	water	river	6	0.79		0.001
eukarya	water	marine	0.104	0.27		0.001
eukarya	water	marine	0.104	0.28		0.001
eukarya	water	marine	0.104	0.28		0.001
bacteria	water	lake	250	0.74		0.001
bacteria	water	lake	200	0.41		0.001
fungi	soil	forest	12900	0.17		0.001
fungi	soil	forest	12900	0.24		0.001
Bac_arch	water	river	14.3	0.06		0.31
eukarya	water	river	14.3	0.12		0.07
eukarya	water	marine	66.48	0.383		0.001
eukarya	sediment	coastal	66.48	0.264		0.007
archaea	water	river	1100	0.24		0.02
bacteria	water	river	1100	0.28		0.04
eukarya	water	river	45	0.315		0.001
eukarya	water	river	45	0.585		0.001
eukarya	water	marine	50	0.426		0.001
eukarya	water	marine	50	0.086		0.242
fungi	soil	grassland	406.98	0.301		0.001
Bac_arch	sediment	lake	64.35844	0.0764		0.2525
bacteria	soil	grassland	737	0.383		0.001
bacteria	soil	grassland	800	0.274		0.001
bacteria	soil	grassland	800	0.219		0.001
bacteria	soil	grassland	800	0.271		0.001
eukarya	soil	grassland	0.0124	0.1661		0.0001
bacteria	water	marine	7000	0.03	NA	
Bac_arch	soil	grassland	2000	0.1567		0.046
bacteria	water	river	16.75	0.42		0.01
eukarya	water	river	16.75	0.3		0.01
bacteria	soil	grassland	0.28	0.405		0.056
bacteria	soil	grassland	0.28	-0.033		0.504
bacteria	snow	snow	410	-0.013		0.438
fungi	snow	snow	410	0.306		0.033
eukarya	snow	snow	410	0.024		0.281
bacteria	sediment	marine	220	0.26	NA	
bacteria	sediment	marine	220	0.03	NA	
eukarya	sediment	marine	500	0.21		0.001
bacteria	host	agriculture	NA	0.07	NA	
bacteria	water	river	1200	0.69		0.01
fungi	water	river	940	0.3343		0.001
eukarya	sediment	river	3816.9	0.32		0.001
bacteria	sediment	river	15.63357	0.32		0.04
bacteria	sediment	river	364.114	0.7		0.002
bacteria	sediment	river	15.63357	0.17		0.22
bacteria	sediment	river	364.114	0.76		0.001
eukarya	sediment	river	667.6188	0.098		0.074
eukarya	sediment	river	667.6188	0.021		0.336
eukarya	sediment	river	667.6188	0.31		0.001
eukarya	sediment	river	667.6188	0.313		0.001

bacteria	water	river	105	0.7	0.01
bacteria	water	river	118	-0.27	NA
bacteria	soil	grassland	524.5567	0.2376	0.001
bacteria	soil	grassland	524.5567	0.1432	0.001
bacteria	soil	grassland	524.5567	0.07761	0.015
bacteria	soil	agriculture	3700	0.279	0.001
bacteria	soil	agriculture	3700	0.124	0.055
bacteria	host	agriculture	3700	0.06	0.201
bacteria	water	marine	989	0.007	0.978
bacteria	water	marine	989	-0.26	0.35
bacteria	water	marine	989	0.165	0.557
bacteria	water	marine	989	-0.265	0.34
bacteria	water	marine	989	0.004	0.987
bacteria	water	marine	989	-0.261	0.348
bacteria	water	marine	989	0.172	0.539
bacteria	water	marine	989	-0.181	0.519
bacteria	soil	forest	#########	0.569	0.001
bacteria	soil	agriculture	4129.6	0.347	0.05
eukarya	sediment	grassland	1673.871	0.18	0.001
bacteria	sediment	marine	11250.7	0.13	0.0001
archaea	sediment	marine	11250.7	0.36	0.0001
fungi	sediment	lake	148	0.17	0.014
archaea	soil	agriculture	740.6049	0.388	0.001
bacteria	soil	agriculture	740.6049	0.15	0.007
archaea	water	river	108	0.53	NA
bacteria	water	river	108	0.7	0.01
fungi	host	grassland	1804.736	0.18	0.001
eukarya	sediment	coastal	12000	0.638	0.01
eukarya	sediment	coastal	12000	0.484	0.01
bacteria	water	marine	251.9414	0.18	NA
bacteria	soil	forest	3700	0.545	0.001
bacteria	host	grassland	3000	0.64	0.01
fungi	water	marine	544.152	0.421	0.001
fungi	water	marine	544.152	0.426	
fungi	water	marine	544.152	0.373	
fungi	sediment	marine	544.152	0.302	0.002
eukarya	water	marine	1300	0.26	0.13
bacteria	water	river	175	0.29	NA
bacteria	water	river	175	0.58	0.05
bacteria	water	river	140	0.22	0.05
bacteria	water	river	190	0.59	0.005
bacteria	water	river	190	0.6	0.0005
bacteria	water	river	140	0.21	NA
fungi	host	agriculture	35	0.42	0.007
fungi	host	agriculture	35	0.51	0.001
bacteria	host	agriculture	35	0.07	0.27
bacteria	host	agriculture	35	-0.08	0.61
bacteria	sediment	marine	NA	0.74	0.001
fungi	host	forest	NA	0.15	0.013
fungi	host	forest	NA	0.22	0.002
fungi	sediment	marine	590	0.091	0.085
archaea	sediment	marine	9714.929	0.26	0.001
archaea	sediment	marine	9714.929	0.31	0.001
fungi	host	grassland	3000	0.47	0.01
fungi	host	grassland	3000	0.12	0.18
fungi	soil	grassland	771.278	0.022	0.63

Bac_arch	soil	grassland	771.278	0.49	0.001	
fungi	soil	grassland	771.278	0.14	0.008	
Bac_arch	soil	grassland	771.278	0.39	0.001	
fungi	soil	forest	1.070612	0.33	0.01	
fungi	soil	forest	1.070612	0.42	0.01	
fungi	soil	grassland	1100	0.428	0.0001	
bacteria	soil	grassland	1100	0.221	0.0001	
fungi	soil	grassland	1100	0.415	0.0001	
bacteria	soil	grassland	1100	0.209	0.0001	
eukarya	water	marine	19.95401	0.214	0.085	
eukarya	sediment	coastal	12.54235	0.622	0.001	
eukarya	water	marine	19.39535	-0.029	0.85	
fungi	air	urban	30	0.002	0.4	
bacteria	soil	desert	888	0.36	0.001	
bacteria	soil	agriculture	27.54095	-0.13	NA	
bacteria	sediment	river	27.23884	0.2	NA	
bacteria	water	river	31.57764	0.08	NA	
bacteria	soil	grassland	360	0.3381	0.01	
bacteria	soil	grassland	360	0.3539	0.01	
fungi	host	forest	17.06215	0.2837	0.001	
fungi	host	forest	16.95822	0.1843	0.001	
eukarya	sediment	marine	11626.09	0.25	0.001	
fungi	host	forest	110	0.22	0.001	
fungi	soil	forest	1.840173	0.26	0.003	
bacteria	sediment	glacier	1396.279	0.855	0.001	
bacteria	sediment	coastal	1624.666	0.371	0.001	
bacteria	sediment	coastal	1624.666	0.512	0.001	
bacteria	soil	forest	10000	0.45	0.0001	
bacteria	soil	forest	10000	0.11	0.204	
bacteria	sediment	forest	0.13	0.13	0.09	
bacteria	soil	forest	3000	0.059	0.012	
archaea	soil	forest	3000	-0.015	0.657	
fungi	soil	forest	3000	0.055	0.051	
fungi	soil	desert	925	0.142	0.181	
fungi	soil	desert	925	0.659	0.001	
bacteria	host	grassland	2.55474	0.266	0.038	
bacteria	host	grassland	2.55474	0.258	0.032	
bacteria	soil	grassland	70	0.204	0.017	
fungi	soil	grassland	70	0.387	0.002	
Bac_arch	sediment	river	610	0.159	0.001	
bacteria	water	lake	NA	0.215	NA	
eukarya	sediment	lake	514	0.578	0.001	
archaea	sediment	lake	3656	0.197	0.001	
bacteria	water	lake	67.85	-0.3311	NA	
bacteria	water	lake	67.85	-0.0122	NA	
bacteria	water	lake	67.85	0.2482	0.0285	
bacteria	water	lake	67.85	0.3969	0.0003	
bacteria	soil	grassland	60	0.3574	0.0001	
bacteria	soil	grassland	60	0.3626	0.0001	
bacteria	soil	grassland	60	-0.0642	NA	
all	water	river	674	0.83	0.001	
all	water	river	674	0.37	0.1	
bacteria	host	marine	0.00199	0.7059	0.0001	
eukarya	host	marine	0.00199	0.2147	0.1072	
bac_euk	host	marine	0.00199	0.3029	0.0543	
bacteria	host	marine	0.00199	0.698	0.0002	

eukarya	host	marine	0.00199	0.1289	0.2247
bac_euk	host	marine	0.00199	0.1102	0.2611
bacteria	host	marine	0.00199	0.4252	0.0128
eukarya	host	marine	0.00199	0.0644	0.3428
bac_euk	host	marine	0.00199	0.1286	0.2345