Received: 13 July 2020 | Revised: 17 December 2020 | Accepted: 1 January 2021

DOI: 10.1111/geb.13266

META-ANALYSIS

Global Ecology and Biogeography A Journal of Macroecology WILEY

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

Dave R. Clark[1,2] | Graham J.C. Underwood[1] | Terry J. McGenity[1] | Alex J. Dumbrell[1]

[1]School of Life Sciences, University of Essex, Colchester, UK

[2]Institute for Analytics and Data Science, University of Essex, Colchester, UK

Correspondence
Alex J. Dumbrell, School of Life Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.

Funding information
Natural Environment Research Council, Grant/Award Number: PhD studentship 471757

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 and spatial scales and between those using different molecular methodologies (e.g., high-throughput sequencing versus molecular fingerprinting). Here, we test how these factors influence the strength of microbial distance–decay relationships, in order to draw generalizations about how microbial β-diversity scales with space.

Location: Global.

Time period: Studies published between 2005 and 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 the strength of distance–decay relationships. Our final dataset consisted of 452 data points, varying in environmental/ecological context or methodological approaches, and we 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, the strength of these relationships varied between environments and habitats, with soils showing significantly weaker distance–decay relationships than other habitats, whereas increasing spatial extents had no effect. 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 biogeographical 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 generalizable biogeographical relationships in microbial ecology.

KEYWORDS

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

1 | 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 geographical distance separating them and demonstrates that nearby communities are more similar to each other than distant communities. Distance–decay relationships arise through several different, but often interacting, ecological and evolutionary processes; consequently, ecologists have debated extensively the underlying mechanisms that generate such patterns (Hanson et al., 2012; Nekola & White, 1999; Soininen et al., 2007). Spatial structuring of the environment can lead to distance–decay relationships, because communities close together in space are likely to experience more similar environmental conditions, hence contain more similar communities than those situated in 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 localized dispersal than those further apart.

Distance–decay relationships are well documented in a multitude of plant and animal communities (e.g., multiple aquatic taxa, Astorga et al., 2012; tropical amphibians, Basham et al., 2019; multiple taxa, Soininen et al., 2007; urban plants, Sorte et al., 2008). Nonetheless, these relationships are of particular interest to microbial ecologists, because microorganisms were assumed to have ubiquitous distributions for several reasons. First, their small size facilitates passive dispersal over large geographical 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 contributory factor to microbial community composition. Second, 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. Third, some microorganisms are able to enter dormant states, whether as vegetative cells or as cysts or spores (Locey et al., 2020), 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 proportion of shared species between distant communities, in turn leading to weaker distance–decay relationships in comparison to macroorganisms. However, empirical studies have yielded mixed results on the strength of microbial distance–decay relationships, where strength is defined as the degree to which geographical distance and community dissimilarity are correlated. Many studies have detected little or no evidence of distance–decay relationships in microbial communities (Hazard et al., 2013; Kivlin et al., 2014), whereas others have reported relationships of varying strengths, across a range of spatial extents, study systems and taxa (Clark et al., 2017; Dumbrell et al., 2010; Martiny et al., 2011). 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 attributable to different environmental or ecological contexts in studies. Here, we consider environmental context as the variability in the physicochemical 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, thereby leading to weaker or stronger distance–decay relationships, respectively. In well-connected systems where dispersal is more feasible, such as oceanic waters, distance–decay relationships should be weaker than in 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 and heterogeneity they support. Sediments and soils, for example, can support strong environmental gradients over distances of a few metres (Dumbrell et al., 2010) and can be highly heterogeneous at the millimetre scale (Vos et al., 2013), strengthening the correlation between distance and community dissimilarity. Additionally, different study taxa are likely to yield variable distance–decay relationships because they differ in traits that are linked to dispersal efficacy. For example, small cells disperse more efficiently over long distances (Norros et al., 2014; Wilkinson, 2001; Wilkinson et al., 2012), meaning that 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 might 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 might contribute to variable distance–decay relationships (Steinbauer et al., 2012).

Studies incorporating larger spatial extents would be expected to show exponential decay of similarity, because communities are more likely to originate from distinct species pools, with high dispersal limitation. In contrast, studies with smaller spatial extents are generally expected to follow power-law decay, although the spatial scales at which the distance–decay relationship follows either of these forms might also depend on the size of the study organisms (Luan et al., 2020; Martiny et al., 2011; Nekola & McGill, 2014).

Although the context in which a study was undertaken might 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 the development of environmental microbial communities (Clark et al., 2018). However, rapid turnover in molecular methodologies means that our perception of microbial β-diversity patterns integrates 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) (Glenn, 2011; Muyzer, 1999). 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) and often miss them completely (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 of tens or even hundreds of thousands of sequences per sample (Caporaso et al., 2012), thereby both improving community coverage (the detected proportion of a given community) and allowing more samples to be examined in a single study (improving sample coverage). Consequently, variation in the ability of molecular methods to resolve closely related taxa and to detect rare taxa can be an additional source of variability in microbial β-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 varies according to the identity and abundance of the species present, their phylogenetic relationships and external factors, such as varying sample sizes. Thus, similarity metrics that vary by one or more of these characteristics would be likely to result in contrasting distance–decay relationships (Barwell et al., 2015; Chao et al., 2005). 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 exhibit high phylogenetic similarity if the species share many branches of a phylogenetic tree, thereby reducing the decay of similarity over geographical distance (Bryant et al., 2008). In contrast, 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 should therefore 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:

1. Bacteria and Archaea will show weaker (lower correlation between geographical distance and community dissimilarity) distance–decay relationships than micro-eukaryotic taxa owing to their smaller size and higher population densities in most environments.
2. Environments that are able to maintain steep physicochemical gradients, such as sediments and soils, will have stronger (higher correlation between geographical distance and community dissimilarity) distance–decay relationships than those such as seawater or air, where environmental gradients are more diffuse.
3. Spatial extent will be related positively to the strength of the distance–decay relationship because, at large spatial scales, increased dispersal limitation and environmental heterogeneity will decrease the variance in community similarity at a given spatial distance, resulting in stronger distance–decay relationships.
4. High-throughput sequencing methods will yield stronger distance–decay relationships owing 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.
5. Phylogenetic similarity metrics (e.g., Unifrac, beta nearest taxon index) will result in weaker distance–decay relationships than other metrics, because communities can be similar phylogenetically, yet different at fine taxonomic resolutions, and quantitative metrics (e.g., Bray–Curtis, Hellinger and Euclidean) will yield the strongest relationships because they reflect changes in both species composition and abundance.

2 | METHODS

2.1 | 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 April 2020, and all results published between 1900 and 2019 (inclusive) were retained. To filter the dataset to studies suitable for testing our hypotheses, search results were downloaded and screened manually using the "metagear" (Lajeunesse, 2016) package in R (v.3.4.1; R Core Team, 2019). Here, suitable studies were those that tested the relationship between community similarity and geographical distance in microbial communities, and not studies of

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*)	2,907
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 = ("geographical 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

TS indicates the topic search terms entered into each Web of Science search.

"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 before 1967, the year the Mantel test was described (Mantel, 1967), and the earliest suitable study was published in 2005.

From these studies, we extracted Mantel correlation coefficients (r) as an effect-size measure for each distance–decay relationship, which we refer to throughout as distance–decay strength. 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 geographical distance. The Mantel test statistic is an ideal measure of effect size for use in meta-analytical frameworks for several reasons. First, the Mantel correlation test is the most frequently used method for testing distance–decay relationships in microbial ecology (Franklin & Mills, 2007; Ramette, 2007). Second, given that the Mantel coefficient is a standardized correlation coefficient (i.e., it is bound by minus one and plus one), it provides an easily interpretable and comparable measure of effect size (Harrison, 2011).

We ensured that all Mantel correlation coefficients reflected correlations between geographical distance and community dissimilarity, rather than similarity, by multiplying correlation coefficients by minus one where necessary (meaning 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 because 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 Rosenberg et al. (2013). All subsequent statistical analyses were conducted on the transformed z-scores, whereas the 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.

3 | Statistical analyses

In order to determine whether distance–decay relationships varied between categorical variables (as in hypotheses 1, 2, 4 and 5), we used ANOVAs. In tests where significant differences between groups were found, Tukey's honestly significant difference (HSD) tests were used to determine which groups were different from each other. Linear mixed-effect models were used to test separately for relationships between the strength (correlation between geographical distance and community dissimilarity, expressed as the Mantel correlation coefficient) of distance–decay relationships and single continuous variables, such as spatial extent and community coverage, using a random intercept to account for heteroscedasticity owing to some studies contributing multiple relationships in each model. The p-values and R^2 values were calculated for each term in these models using the approach described by Nakagawa and Schielzeth (2013). The variables spatial extent and community coverage were initially $\log_{10}$-transformed to aid model fitting, because they spanned several orders of magnitude. To compare the overall influence of ecological versus methodological factors on microbial distance–decay relationships, we compared two full models (including all relevant variables), using Akaike information criterion (AIC) scores, on a subset of the data for which all variables were recorded successfully. We report the results of all null hypothesis tests in terms of statistical "clarity" rather than "significance", in line with recommendations from Dushoff et al. (2019).

4 | RESULTS

Our Web of Science searches resulted in 2,982 unique search results. Manual screening of the abstracts yielded 951 studies that were deemed potentially to be suitable for use in this analysis. A total of 452 Mantel correlation coefficients were obtained successfully from 187 studies represented in 61 journals (Supporting Information Figure S1). Reported Mantel correlation coefficients ranged from −.33 to .95, with a mean of .27 (SE = 0.011), and a summary of the variables collected is shown in Table 2.

BOX 1 Details of the explanatory variables extracted from each study

Resolution

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

Community coverage

This refers to the depth of sequencing in sequencing-based studies, or the 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. Given that 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 was used to calculate each distance–decay relationship. Recorded dissimilarity indices were then categorized as quantitative (Bray–Curtis, Morisita-Horn, Euclidean, Hellinger or Theta), qualitative (Jaccard, Raup–Crick, Sørensen, Simpson or β-sim) or phylogenetic (weighted or unweighted Unifrac, Rao, β-mean nearest taxon distance or β-mean pairwise distance).

Correlation type

Studies were categorized according to the type of correlation coefficient used in the analysis of the distance–decay relationship (e.g., Spearman's or Pearson's correlation coefficient). The type of correlation was recorded only if the type of correlation coefficient was mentioned explicitly.

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 a combination of these categories if a relationship was based on multiple taxa (for example, owing to the use of sequencing primers that detect both Archaea and Bacteria). Fungi were grouped separately from other micro-Eukaryotes owing to their distinct reproductive strategy (e.g., spore production) and the fact that 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 maximal distance separating communities (in kilometres). If this was not stated in the text or provided in the supplementary material (e.g., in a geographical distance matrix), it was calculated from the geographical coordinates given, estimated from a plot of the distance–decay relationship or estimated from scaled maps.

Environment

We categorized distance–decay relationships broadly, based on the type of environment (agriculture, air, aquifer, coastal wetlands/intertidal, desert, dune, forest, glacier, grassland, lake, marine, coastal marshes, mine, river, snow or urban) within which they were sampled. Although these categories are not mutually exclusive, we categorized each study based on which environment best represented the environmental context in which each study was undertaken. For studies on lakes, we also recorded whether relationships originated from a single lake or across multiple lakes.

Habitat

Habitat was the type of environmental material that the sampled communities occupied. We categorized distance–decay relationships as follows: air, host-associated, sediment, snow, soil or water.

4.1 | 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 strength of distance–decay relationships between these taxa (Table S2, $F_{5,441} = 0.99$, $p = .43$), although distance–decay relationships incorporating bacterial and fungal communities showed the weakest relationships, albeit only from six studies (Figure 1).

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 were therefore excluded from further analyses (marsh, n = 3; snow, n = 3; dune, mine and aquifer, n = 1). The most frequently studied environments were

TABLE 2 Summary of collected data.

Ecological variables		Methodological variables	
Variable	**Summary**	**Variable**	**Summary**
Study taxon[a]	Archaea: *n* = 26	Resolution	High: *n* = 345
	Bacteria: *n* = 238		Intermediate: *n* = 84
	Eukarya: *n* = 67		Low: *n* = 23
	Fungi: *n* = 93		
	Archaea + Bacteria: *n* = 17		
	Bacteria + Eukarya: *n* = 3		
	Bacteria + Fungi: *n* = 6		
	All: *n* = 2		
Spatial extent (km)	Minimum = 0.0001	Community coverage (number of individuals/ sequences)	Minimum = 8
	Mean = 1,543		Mean = 217,357
	Median = 220		Median = 1,257
	Maximum = 18,700		Maximum = 34,192,561
	NA = 15		NA = 115
Environment type	Agriculture: *n* = 16	Dissimilarity index	[b]β-MNTD: *n* = 13
	Air: *n* = 13		[c]β-MPD: *n* = 1
	Aquifer: *n* = 1		β-sim: *n* = 4
	Coastal: *n* = 8		Bray–Curtis: *n* = 218
	Desert: *n* = 4		$Bray–Curtis_{Sim}$: *n* = 3
	Dune: *n* = 1		$Bray–Curtis_{Nes}$: *n* = 1
	Forest: *n* = 76		Canberra: *n* = 1
	Glacier: *n* = 5		Euclidean: *n* = 9
	Grassland: *n* = 96		Hellinger: *n* = 4
	Lake: *n* = 76		Jaccard: *n* = 49
	Marine: *n* = 88		Mash: *n* = 2
	Marsh: *n* = 3		Morisita–Horn: *n* = 4
	Mine: *n* = 1		Rao: *n* = 2
	River: *n* = 57		Raup–Crick: *n* = 19
	Snow: *n* = 3		Simpson: *n* = 2
	Urban: *n* = 4		Sørensen: *n* = 42
			Theta: *n* = 1
			Unweighted Unifrac: *n* = 17
			Weighted Unifrac: *n* = 59
			NA: *n* = 1
Habitat type	Air: *n* = 16	Correlation type	Pearson: *n* = 62
	Host: *n* = 75		Spearman: *n* = 86
	Sediment: *n* = 78		NA: *n* = 304
	Snow: *n* = 3	Sample coverage (number of samples)	Minimum = 4
	Soil: *n* = 141		Mean = 52.88
	Water: *n* = 137		Median = 25
	NA: *n* = 2		Maximum = 1,010
			NA = 1

Note: For categorical variables, the number of individual distance–decay relationships in each category is shown, whereas minima, maxima, median and mean values are shown for continuous variables. Detailed descriptions of each variable are found in Box 1, and raw data can be found in the Supporting Information (Table S1).

NA = not assessed.

[a]The "All" category consists of studies that incorporated all microbial taxonomic groups, whereas combined categories (e.g., Archaea + Bacteria) incorporate communities from multiple taxonomic groups (e.g., archaeal and bacterial communities).

[b]β mean nearest taxon distance.

[c]β mean pairwise distance.

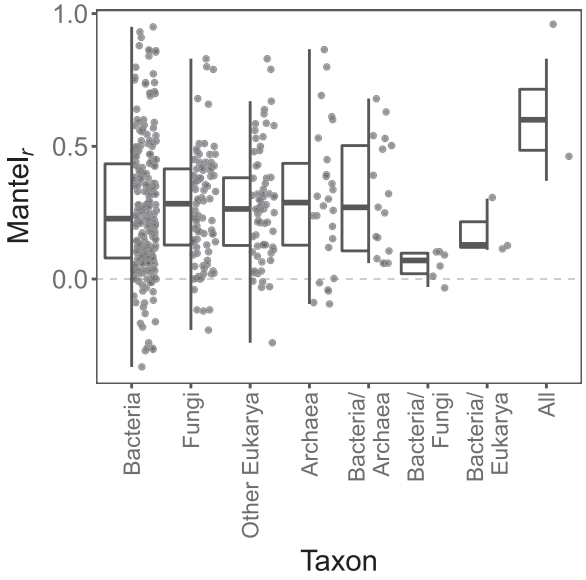

FIGURE 1 The strength (Mantel$_r$) of distance–decay relationships based on different study taxa. A larger Mantel$_r$ value indicates a stronger distance–decay relationship. The "All" category consists of studies that incorporated all microbial taxonomic groups, whereas combined categories (e.g., Bacteria/Archaea) incorporate communities from multiple taxonomic groups (e.g., bacterial and archaeal communities)

grasslands (n = 96), marine (n = 88), and lakes and forests (n = 76 for both). We found clear differences in the strength of distance–decay relationships between environments (Figure 2a, Table S2; $F_{10,432}$ = 3.187, p < .001). Specifically, and perhaps counter-intuitively, grassland-based studies had weaker distance–decay relationships than those from aquatic environments, such as lakes, rivers or the marine environment (magnitude of coefficient |coef| > 0.17, p < .05 for all comparisons). Urban environments, which included built environments, such as sewers and indoor air, also produced weak distance–decay relationships, although with only four data points this difference was not statistically clear (p > .43 for all comparisons). We also found no difference in the strength of distance–decay relationships between studies conducted in single lakes compared with those incorporating multiple lakes ($F_{1,74}$ = 0.11, p = .74), despite the average spatial extent of multiple-lake studies being *c*. 32-fold greater than that of single-lake studies (Supporting Information Figure S2).

A more detailed analysis of the interaction between environment type and habitat revealed that although environments ($F_{9,420}$ = 3.29, p < .001) and habitat ($F_{3,420}$ = 6.65, p < .001) differed from each other, their interaction was not statistically significant ($F_{4,420}$ = 1.93, p = .10). In fact, within environments, only marine host-associated and marine water-based distance–decay relationships were clearly different from each other (Figure 2b), with host-associated communities showing significantly stronger distance–decay relationships (coef = 0.35, p < .001).

The spatial extents of recorded distance–decay relationships ranged from 10 cm to > 18,000 km, and minimal spatial extents varied notably across environments and habitats, with terrestrial- and soil-based studies often conducted over smaller spatial scales (Supporting Information Figure S3). After accounting for differences between studies, we found no evidence of a statistically clear relationship between the spatial extent of a study and the strength of the observed distance–decay relationship (Table S2, coef = 0.02, marginal R^2 = .020, t = 1.58, p = .11). Finally, given that studies at a larger spatial scale might also incorporate greater sampling coverage, we tested for collinearity between the spatial scale of a study and the sampling coverage, but found no correlation between these variables (ρ = .06, p = .19).

4.2 | Influence of methodological factors on the distance–decay relationship

We grouped community characterization methods according to their ability to distinguish between closely related taxa. There were no clear differences in the strength of distance–decay relationships between different resolution methods (Table S2, $F_{2,449}$ = 0.562, p = .57), nor were there clear differences between different molecular methods (Supporting Information Figure S4; $F_{7,437}$ = 1.97, p = .06), considering only those methods that had more than four distance–decay relationships across the entire dataset (excluding Ion Torrent, n = 4; PhyloChip, n = 2; and Pac-Bio, n = 1; Figure 3).

Although we observed no differences in distance–decay relationships between different resolution methods, after accounting for study-dependent differences we found a positive relationship between ($\log_{10}$) community coverage and the strength of microbial distance–decay relationships (Figure 4a, Table S2; n = 337, conditional R^2 = .57, coef = 0.06, t = 2.73, p < .01), although the marginal effect of community coverage was weak (marginal R^2 = .04).

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 = −.03, p = .54), nor did we detect any clear relationship between the number of samples ($\log_{10}$ sample coverage) and the strength of distance–decay relationships, even after accounting for study-specific differences with a mixed effects model (Figure 4b, Table S2; n = 451, coef = −0.06, marginal R^2 = .01, t = −1.40, p = .16).

Choice of similarity index also had a clear impact on the strength of microbial distance–decay relationships. In addition to recording the specific similarity index used, we categorized indices into types (binary, abundance or phylogenetic) to test for broad differences in distance–decay relationships. We analysed the nested interaction between similarity index and index type and found no clear differences between different index types (Figure 5a; $F_{2,424}$ = 1.48, p = .23). However, the interaction between index type and similarity index was significant ($F_{7,424}$ = 7.20, p < .001). Post hoc analysis revealed differences between similarity indices within and between index types (Figure 5b). Distance–decay relationships based on the

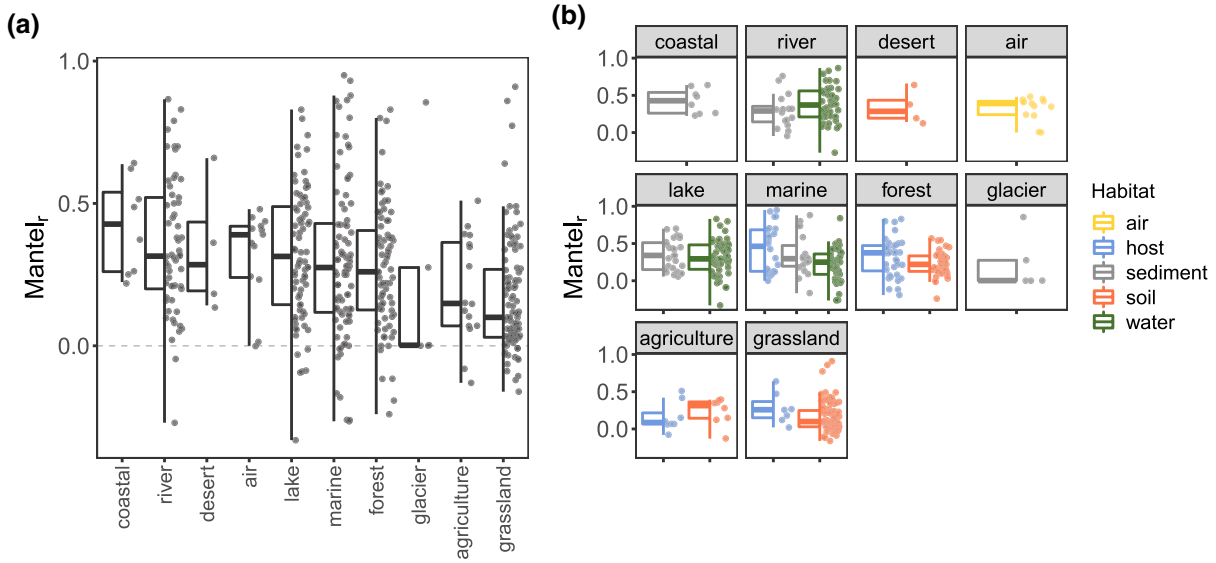

FIGURE 2 Variation in Mantel correlation coefficients of distance–decay relationships (a) between different environments, and (b) between types of habitats. Environment categories are arranged from strongest to weakest mean distance–decay relationship

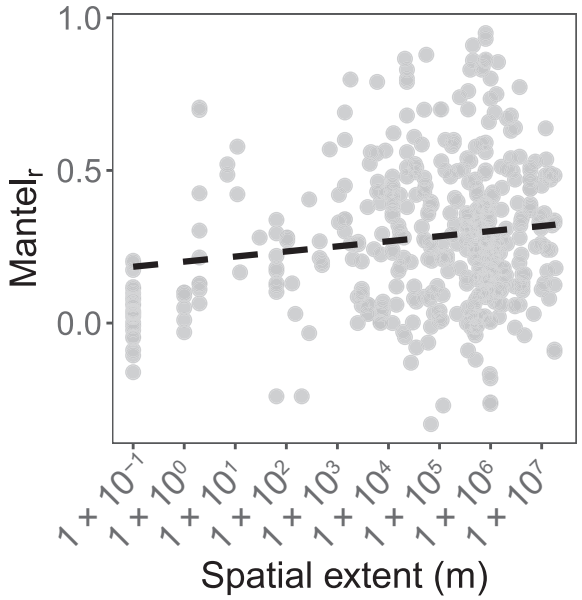

FIGURE 3 The relationship between spatial extent and the Mantel correlation coefficient of microbial distance–decay relationships. The dashed line represents the fit of a mixed-effects model between the $\log_{10}$ of spatial extent and Mantel correlation coefficient, with a study-dependent random intercept

Raup–Crick index were weaker than those based on either Sørensen (coef = −0.38, $p < .01$) or unweighted Unifrac indices (coef = −0.44, $p < .01$), and those based on weighted Unifrac were weaker than both Sørensen (coef = −0.29, $p < .001$) and unweighted Unifrac (coef = −0.35 $p < .05$). Finally, most studies did not state explicitly the type of correlation used to generate each Mantel test ($n = 304$), but of those that did, Spearman's correlation coefficient was more frequently used ($n = 86$) than Pearson's ($n = 62$). We found no clear difference in the strength of microbial distance–decay relationships using these two methods (Table S2, $F_{1,146} = 2.47$, $p = .12$).

4.3 | Comparison of contextual and methodological variables

In order to determine whether eco-environmental context or methodological factors better explain the strength of microbial distance–decay relationships, we specified two models, with variables from these two categories, using a subset of the original data for which values were obtained for all variables ($n = 323$). Each model had four variables and used similar degrees of freedom (context model d.f. = 26; methodological model d.f. = 27). The methodological model outperformed the contextual model in terms of both AIC and R^2 measures of model performance (Table 3). Notably, neither model explained a high proportion of the variance, although both AIC and likelihood ratio tests supported both models over a null (intercept-only) model.

5 | 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 twofold. First, the differing contexts within which studies are conducted contribute variability to reported distance–decay relationships. In particular, we found that differing study systems were associated with variation in microbial distance–decay relationships. Second, methodological

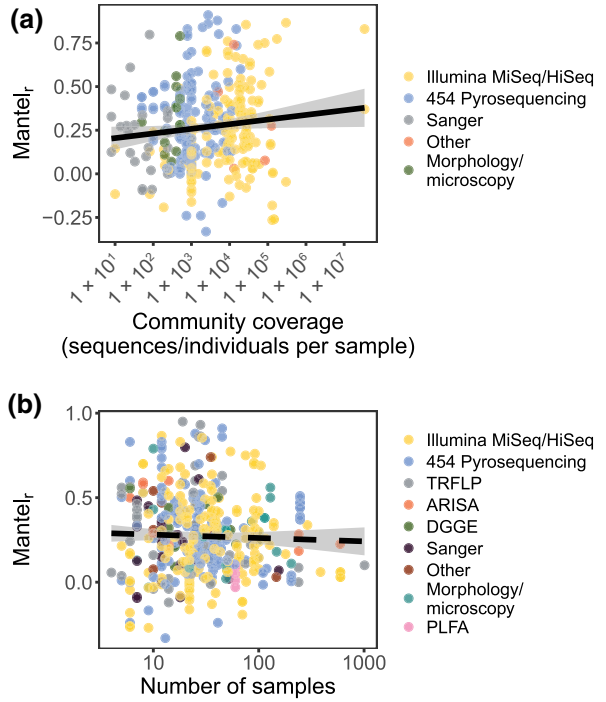

FIGURE 4 The relationship between the strength of microbial distance–decay relationships (Mantel$_r$) and (a) community coverage, quantified as the number of sequences or individuals counted per sample, and (b) sample coverage, quantified as the number of individual samples used to construct distance–decay relationships. Points are individual Mantel correlation coefficients, coloured according to the molecular technique used to characterize the microbial community. Continuous lines indicate statistically significant relationships ($p < .05$), whereas dashed lines indicate non-significant relationships ($p > .05$), and shaded grey ribbons represent 95% confidence intervals. Abbreviated molecular methods in the legend are defined as follows: ARISA = automated ribosomal intergenic spacer analysis; DGGE = denaturing gradient gel electrophoresis; PLFA = phospholipid fatty acid analysis; Sanger = Sanger sequencing of cloned phylogenetically informative genes; TRFLP = terminal restriction fragment length polymorphism

differences between studies, including dissimilarity index, data resolution and sample coverage, all significantly affected the observed distance–decay relationships. A central tenet of macroecology is the search for universal patterns and relationships; our results suggest that generalizable relationships might emerge only when methodological approaches are coupled appropriately to ecological context.

Our comparison of distance–decay relationships between different study systems revealed that agricultural studies, 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, given that 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 geographical distance, for example, if similar environmental conditions are distributed patchily. Alternatively, soil microorganisms might 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 bio-aerosols (Joung et al., 2017). The depth profile over which soil samples integrate might also play a role in obscuring distance–decay relationships, because surface soils show stronger distance–decay relationships than deeper ones, probably owing to the greater intensity of dispersing propagules entering and leaving 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 (Carini et al., 2016; Lennon & Jones, 2011; Lennon et al., 2018). Dormant cells and relic DNA are not subject to environmental selection, yet they are routinely detected in molecular community assays, which is likely to diminish the perceived effects of spatially structured environmental selection on microbial communities (Locey et al., 2020). Thus, in habitats such as soils, distinguishing dormant from active cells could result in stronger distance–decay relationships than those recorded previously, although evidence of the same effect on distance–decay slopes is mixed (Locey et al., 2020; Meyer et al., 2018). The extent to which this phenomenon plays a role in other environments is also unclear.

Originally, we expected the weakest distance–decay relationships to occur in connected aquatic environments, such as rivers and oceans, or within single lakes, because the movement of water might provide an effective dispersal mechanism, homogenizing microbial communities over larger spatial and environmental distances. In contrast, we found that aquatic communities showed stronger distance–decay relationships than terrestrial systems. Soininen et al. (2007) recorded similar distance–decay rates between terrestrial, marine and aquatic ecosystems, showing that context-dependent distance–decay relationships might be a feature of microbial communities. We also found that the strength of distance–decay relationships was not different in studies based on single or multiple lakes, despite the difference in spatial extents of these studies. Lakes act as habitat islands within a terrestrial matrix; therefore, dispersal limitation and environmental heterogeneity should be greater across multiple lakes than within a single lake, resulting in stronger distance–decay relationships in multi-lake studies. One explanation is that catchment-scale environmental parameters, such as geology, might homogenize environmental conditions across multiple lakes, meaning that environmental distances are similar within and between lakes. Alternatively, other biogeographical processes, such as mass effects, might homogenize communities between hydrologically connected lakes (Lindström & Bergström, 2004), especially where lakes are of different sizes (Reche et al., 2005). 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, and the degree of host specificity with the associated microbiome. For example, if the host is not dispersal limited and associates with a large variety of microorganisms, then

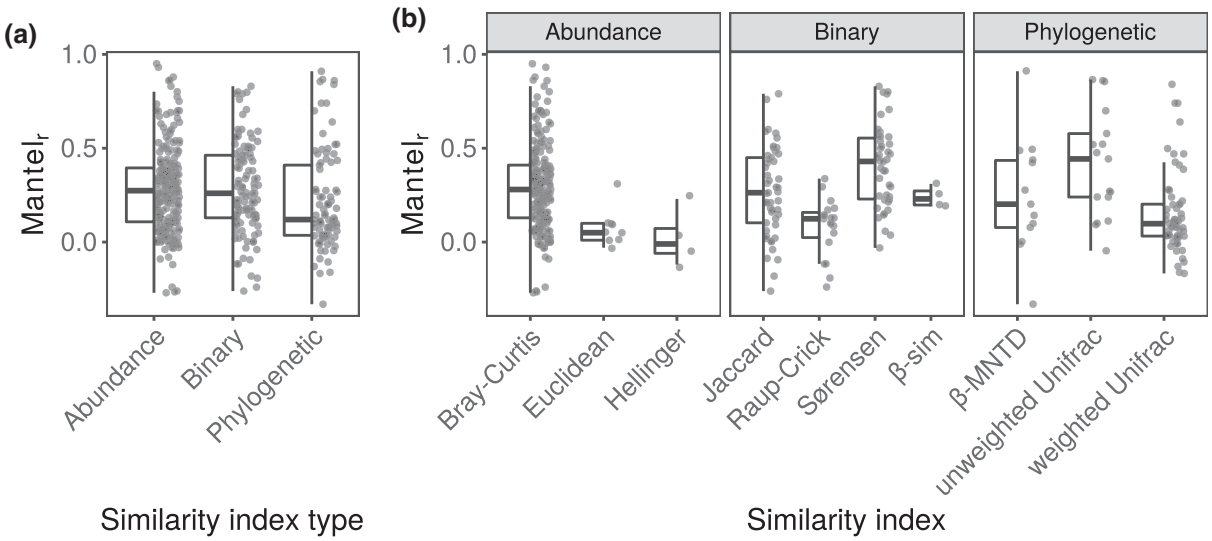

FIGURE 5 Variation in the strength of microbial distance–decay relationships (*Mantel$_r$*) calculated with: (a) different similarity index types, or (b) individual indices. Only indices with four or more distance–decay relationships were plotted, for clarity. β-MNTD, β-Mean Nearest Taxon Distance.

TABLE 3 Comparison of models specified using either contextual or methodological variables

			Likelihood ratio comparison to null (intercept-only) model			
Model	AIC	Adj-R^2	ΔAIC	Sum of squares	F (d.f.)	p-value
Contextual	146.89	.11	−13.69	5.34	2.61	< .001
Methodological	134.11	.14	−26.46	6.47	3.17 (25)	< .001

Note: The Akaike information criterion (AIC) and adjusted R^2 (Adj-R^2) quantify the likelihood and fit of a model relative to the number of predictor variables, respectively.

the distance–decay relationship might be relatively weaker than those of either dispersal-limited hosts or highly specific associated microbiomes.

The scale dependence of various biogeographical relationships is well studied (Bissett et al., 2010; Hillebrand, 2004; Martiny et al., 2011; Soininen et al., 2011), albeit with contrasting results. Soininen et al. (2011) reported that distance–decay relationships of various microbial communities were generally steeper over greater spatial extents, whereas our results suggest that increasing spatial extent does not significantly increase the strength of distance–decay relationships. Given that we analysed distance–decay strength rather than steepness, our results are not necessarily contradictory. A strong distance–decay relationship occurs when, at a given spatial distance, all pairs of communities are equally dissimilar to one another, whereas a steep distance–decay relationship occurs when communities separated by different distances are highly dissimilar to each other. We expected initially that spatial extent might alter the strength of distance–decay relationships because, at greater distances, decreased dispersal and increased environmental heterogeneity should reduce the variance in compositional similarity between pairs of communities (at a given distance). Instead, it could be that the spatial configuration or connectivity of the communities could be more important than spatial extent per se. For example, at a given spatial distance, some pairs of communities could be linked by dispersal and others not, increasing the variation in community similarity at each distance and weakening the distance–decay relationship. In practice, this could occur in lake systems where, at a certain geographical distance, some pairs of communities fall within the same lake and some in different lakes, or when long-distance dispersal vectors link some pairs of communities separated by large distances, but not others, as has been proposed for halophilic microbial communities dispersing on migratory birds, for example (Clark et al., 2017; Kemp et al., 2018). Furthermore, we observed that the minimum spatial extents differed according to the environment in which they were conducted. Studies from terrestrial environments (e.g., grasslands and forests) or those based on soils generally incorporated smaller spatial extents than those based on aquatic systems (with the exception of some host-associated marine studies) or on habitats such as water or air. This could be attributable to the logistics of sampling at small scales. For example, sampling planktonic microbial communities at small (centimetres to metres) scales could be confounded by mixing caused by the sampling process or by tidal movements of water. Additionally, given that many studies analysing microbial distance–decay relationships aimed to discern between environmental and spatial effects on microbial communities, it might be widely assumed that aquatic environments are more

homogeneous and/or that microorganisms are not dispersal limited at these scales compared with more physically stable environments, such as soils or sediments.

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); therefore, 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, might be less important than other contextual and methodological factors or that the broad domain-level classification used here does not capture different microbial cell sizes sufficiently. As discussed previously, distance–decay relationships can arise from spatially autocorrelated environmental gradients and from dispersal limitation (Nekola & White, 1999). Therefore, the lack of differences in biogeographical patterns observed at the domain level might be the result of a trade-off between dispersal-related processes and environmental filtering. For instance, bacterial distance–decay relationships might be less strongly influenced by dispersal than environmental filtering, and vice versa for Eukarya. Consequently, these influences might balance out at broad taxonomic levels, resulting in similar biogeographical 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 characterize microbial communities (Caporaso et al., 2012; Roesch et al., 2007). However, these methods differ in their resolution, community coverage, and ability to multiplex large numbers of samples, all of which we hypothesized could strengthen or weaken distance–decay relationships by altering our estimation of microbial β-diversity. In contrast, we observed only a weak relationship between the strength of distance–decay relationships and community coverage, and no clear effects of different resolution methods or the number of samples, suggesting that molecular methodology might not play as large a role in determining microbial biogeographical 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, because such patterns may emerge only 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 biogeographical 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). Nevertheless, empirical biogeographical 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 might not be important. Perhaps most molecular methodologies operate above resolutions at which biogeographical 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 distribution (Dumbrell et al., 2010; Maček et al., 2019; Shoemaker et al., 2017), 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); therefore, detecting only the more abundant, widespread organisms would overestimate compositional similarity across communities and, consequently, weaken distance–decay relationships owing 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 (van Dorst et al., 2014; Galand et al., 2009; Heino & Soininen, 2010) 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 crucial importance in the context of environmental perturbations (Low-Décarie et al., 2016; Shade et al., 2014) 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 might further disconnect biogeographical patterns from ecosystem-level processes.

Against expectation, we observed no clear differences in distance–decay relationships using different types of similarity metrics, and differences between specific metrics were minimal. Distance–decay relationships based on the weighted Unifrac distance and the Raup–Crick index were weaker than those based on other metrics. The Raup–Crick index is less influenced by concurrent changes in species richness between communities, and as such, is a purer reflection of shifts in β-diversity (Chase et al., 2011). Consequently, by removing the potentially confounding effects of differences in richness, the Raup–Crick index is likely to result in more variable estimates of similarity between communities, which would lead to weaker distance–decay relationships.

Phylogenetic metrics, such as Unifrac, cluster communities at a lower resolution, because two communities can be closely related genetically, 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. Furthermore, phylogenetic metrics might be inappropriate in less phylogenetically diverse environments (e.g., extreme systems), where phylogenetic diversity can be constrained largely to one taxon (e.g., the Haloarchaea

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 relative abundance data of species, whereas 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). Given that abundant species are generally more widespread, placing too much weight on them would have the effect of making communities appear artificially similar, exacerbating the effects of using a phylogenetic metric. Given that 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 suggest, therefore, that weighted phylogenetic metrics might underestimate microbial biogeographical 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) and Sørensen (n = 42) indices. These choices undoubtedly reflect a wider trend in ecology as a whole; however, it is pertinent to draw attention to more recently developed metrics that might 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, thereby providing a new approach to integrating co-occurrence data into distance–decay relationships. Microbiome sequencing data also have several characteristics that can be problematic in the analysis of community (dis)similarities. For example, the non-biological variance of sample sizes in sequence datasets can result in statistical artefacts that confound biogeographical 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, thereby reducing the need for post-sequencing normalization of sample sizes (McMurdie & Holmes, 2014). Furthermore, "fuzzy logic"-based similarity indices are able to reduce the impact of false absences or false presences on estimates of β-diversity, which is beneficial for microbial ecology studies, where rarefaction can induce false absences and taxonomic assignment errors or contamination can lead to false presences. Additionally, most high-throughput sequence datasets are compositional. Compositionality occurs as the arbitrary total number of sequences per sample imposed by the sequencing machine causes species counts (abundances) to be dependent on each other (e.g., if species A increases in abundance, species B and C will appear relatively less abundant, even if their absolute abundance has not changed). Binary indices should be suitable, because occurrences (presence/absences) are not affected by compositionality, unless increases in the abundance of one or more species cause others to drop below the detection limit, in which case fuzzy indices might be appropriate. Alternatively, metrics such as the Aitchison distance perform well when appropriate (centred log-ratio) transformations are applied to counts (Gloor et al., 2017), or recently developed metrics, such as the rank bias overlap index, show promise for analysing similarity between communities based on species abundance ranks (Webber et al., 2010). 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, Sørensen 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 given that many of these new metrics are implemented in popular and freely accessible software, such as R (e.g., Baselga & 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 biogeographical pattern. Given the importance of methodological factors in determining the strength of microbial biogeographical patterns, it is intuitive to recommend standardization of approaches across studies in order to minimize the statistical signals associated with methodological variance. However, our results show that variance attributable to differing ecological contexts would still hinder the drawing of generalizable relationships across studies. Instead, we suggest that tailoring methodological choices towards specific ecological contexts might clarify generalizable 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, because the heterogeneity in the physicochemical environment and the dispersal processes of their microbial communities are fundamentally different. Likewise, we should not necessarily expect the relationships between soils and river sediments to be comparable, because microorganisms in soils can disperse feasibly in any direction, whereas in rivers or streams dispersal would be constrained largely by the direction of flow. Consequently, tailoring methodological approaches, such as the sampling design and/or (geographical) distance measure, to reflect the environmental heterogeneity and dispersal dynamics better between contrasting ecological contexts might enable us to negotiate the hierarchy of interacting factors that obscure macroecological patterns in microbial communities.

5.1 | 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, in addition to the methodology

of the study, jointly influence quantification of this classic biogeographical pattern. Against expectation, factors related to molecular methodology had relatively little effect on distance–decay relationships, whereas the 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 biogeographical patterns. We detected clear relationships between microbial distance–decay relationships and various contextual and methodological variables, yet combining these variables explained only a modest amount of variation in our dataset. This lack of explanatory power indicates that microbial biogeographical patterns depend on a number of contextual variables beyond those analysed here. In future, we suggest that microbial ecologists should place greater emphasis on quantifying habitat connectivity to gain a better understanding of the dispersal processes that lead to spatial patterns, such as the distance–decay relationship. Additionally, we recommend that experimental designs and data-collection strategies should be replicated spatially, taxonomically, temporally or any combination thereof where possible (e.g., Alzarhani et al., 2019; Meyer et al., 2018; Zinger et al., 2019), facilitating a more generalized understanding of the variation in spatial microbial community patterns. The question of whether microbial communities show spatial patterns such as distance–decay relationships should be laid to rest; disentangling the web of ecological and environmental drivers that shape these patterns is the next challenge in microbial biogeography.

ACKNOWLEDGMENTS

This work was funded by a UK Natural Environment Research Council (NERC) quota award DRC studentship (471757).

AUTHOR CONTRIBUTIONS

All authors planned the study and contributed to manuscript preparation. D.R.C. carried out all data collection and data analyses.

DATA AVAILABILITY STATEMENT

Full raw data analysed in this manuscript (provided in Supporting Information Table S1) along with R code necessary to reproduce our statistical analyses are available via the Dryad data repository under the accession https://doi.org/10.5061/dryad.7m0cfxpss

ORCID

Dave R. Clark https://orcid.org/0000-0002-2306-0220
Graham J.C. Underwood https://orcid.org/0000-0001-5605-0697
Terry J. McGenity https://orcid.org/0000-0002-1497-8822
Alex J. Dumbrell https://orcid.org/0000-0001-6282-3043

REFERENCES

Alzarhani, A. K., Clark, D. R., Underwood, G. J. C., Ford, H., Cotton, T. E. A., & Dumbrell, A. J. (2019). Are drivers of root-associated fungal community structure context specific? *The ISME Journal*, *13*, 1330–1344. https://doi.org/10.1038/s41396-019-0350-y

Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Distance decay of similarity in freshwater communities: Do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*, *21*, 365–375. https://doi.org/10.1111/j.1466-8238.2011.00681.x

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. https://doi.org/10.1128/AEM.02772-10

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. (2007). Disentangling distance decay of similarity from richness gradients: Response to Soininen et al. 2007. *Ecography*, *30*(6), 838–841.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*, 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

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.

Basham, E. W., Seidl, C. M., Andriamahohatra, L. R., Oliveira, B. F., & Scheffers, B. R. (2019). Distance–decay differs among vertical strata in a tropical rainforest. *Journal of Animal Ecology*, *88*, 114–124. https://doi.org/10.1111/1365-2656.12902

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. https://doi.org/10.1111/j.1365-294X.2010.04804.x

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. https://doi.org/10.1007/s00248-006-9173-2

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 USA*, *105*, 11505–11511. https://doi.org/10.1073/pnas.0801920105

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. https://doi.org/10.1038/ismej.2017.119

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. https://doi.org/10.1038/ismej.2012.8

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*, 16242.

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. https://doi.org/10.1111/j.1461-0248.2004.00707.x

Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., & Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere*, *2*, 1–11. https://doi.org/10.1890/ES10-00117.1

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. https://doi.org/10.1093/bioinformatics/bts342

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. https://doi.org/10.1002/wat2.1280

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. https://doi.org/10.1111/geb.12670

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. https://doi.org/10.1038/ismej.2009.122

Dushoff, J., Kain, M. P., & Bolker, B. M. (2019). I can see clearly now: Reinterpreting statistical significance. *Methods in Ecology and Evolution*, *10*, 756–759. https://doi.org/10.1111/2041-210X.13159

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., … Gorbushina, A. A. (2013). Microbial hitchhikers on intercontinental dust: Catching a lift in Chad. *The ISME Journal*, *7*, 850–867. https://doi.org/10.1038/ismej.2012.152

Franklin, R. B., & Mills, A. L. (2007). Statistical analysis of spatial structure in microbial communities. *The Spatial Distribution of Microbes in the Environment* (eds. R. B. Franklin & A. L. Mills) (pp. 31–60). Springer Netherlands.

Fukuyama, J. (2019). Emphasis on the deep or shallow parts of the tree provides a new characterization of phylogenetic distances. *Genome Biology*, *20*, 131. https://doi.org/10.1186/s13059-019-1735-y

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 USA*, *106*, 22427–22432. https://doi.org/10.1073/pnas.0908284106

Gaston, K. J. (2012). The importance of being rare. *Nature*, *487*, 46–47. https://doi.org/10.1038/487046a

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*, e00148–e00218. https://doi.org/10.1128/mSphere.00148-18

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, *11*, 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Green, J., & Bohannan, B. J. M. (2006). Spatial scaling of microbial biodiversity. *Trends in Ecology and Evolution*, *21*, 501–507. https://doi.org/10.1016/j.tree.2006.06.012

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. https://doi.org/10.1038/nrmicro2795

Harrison, F. (2011). Getting started with meta-analysis. *Methods in Ecology and Evolution*, *2*, 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. https://doi.org/10.1038/ismej.2016.42

Hazard, C., Gosling, P., van der Gast, C. J., 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. https://doi.org/10.1038/ismej.2012.127

Heino, J., & Soininen, J. (2010). Are common species sufficient in describing turnover in aquatic metacommunities along environmental and spatial gradients. *Limnology and Oceanography*, *55*, 2397–2402. https://doi.org/10.4319/lo.2010.55.6.2397

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, *163*, 192–211. https://doi.org/10.1086/381004

Joung, Y. S., Ge, Z., & Buie, C. R. (2017). Bioaerosol generation by raindrops on soil. *Nature Communications*, *8*, 14668. https://doi.org/10.1038/ncomms14668

Kemp, B. L., Tabish, E. M., Wolford, A. J., Jones, D. L., Butler, J. K., & Baxter, B. K. (2018). The biogeography of Great Salt Lake halophilic Archaea: Testing the hypothesis of avian mechanical carriers. *Diversity*, *10*, 124. https://doi.org/10.3390/d10040124

Kivlin, S. N. (2020). Global mycorrhizal fungal range sizes vary within and among mycorrhizal guilds but are not correlated with dispersal traits. *Journal of Biogeography*, *47*, 1994–2001.

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. https://doi.org/10.1016/j.funeco.2014.04.004

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. https://doi.org/10.1038/ngeo2684

La Sorte, F. A., McKinney, M. L., Pyšek, P., Klotz, S., Rapson, G. L., Celesti-Grapow, L., & Thompson, K. (2008). Distance decay of similarity among European urban floras: The impact of anthropogenic activities on β diversity. *Global Ecology and Biogeography*, *17*, 363–371. https://doi.org/10.1111/j.1466-8238.2007.00369.x

Lennon, J. T., & Jones, S. E. (2011). Microbial seed banks: The ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, *9*, 119–130. https://doi.org/10.1038/nrmicro2504

Lennon, J. T., Muscarella, M. E., Placella, S. A., & Lehmkuhl, B. K. (2018). How, when, and where relic DNA affects microbial diversity. *mBio*, *9*, e00637-18.

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*, e00704-19.

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.

Lindström, E. S., & Bergström, A.-K. (2004). Influence of inlet bacteria on bacterioplankton assemblage composition in lakes of different hydraulic retention time. *Limnology and Oceanography*, *49*, 125–136. https://doi.org/10.4319/lo.2004.49.1.0125

Locey, K. J., Muscarella, M. E., Larsen, M. L., Bray, S. R., Jones, S. E., & Lennon, J. T. (2020). Dormancy dampens the microbial distance–decay relationship. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190243. https://doi.org/10.1098/rstb.2019.0243

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. https://doi.org/10.1098/rsbl.2016.0562

Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., Geisen, S., & Sun, B. (2020). Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nature Communications*, *11*, 6406. https://doi.org/10.1038/s41467-020-20271-4

Maček, I., Clark, D. R., Šibanc, N., Moser, G., Vodnik, D., Müller, C., & Dumbrell, A. J. (2019). Impacts of long-term elevated atmospheric CO_2 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 USA*, *108*, 7850–7854.

McMurdie, P. J., & Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, *10*, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

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. https://doi.org/10.1038/s41396-018-0103-3

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, *2*, 317–322. https://doi.org/10.1016/S1369-5274(99)80055-1

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.

Nekola, J. C., & McGill, B. J. (2014). Scale dependency in the functional form of the distance decay relationship. *Ecography*, *37*, 309–320. https://doi.org/10.1111/j.1600-0587.2013.00407.x

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*, 867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

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. https://doi.org/10.1890/13-0877.1

Podani, J., & Schmera, D. (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos*, *120*, 1625–1638. https://doi.org/10.1111/j.1600-0706.2011.19451.x

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, *62*, 142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Reche, I., Pulido-Villena, E., Morales-Baquero, R., & Casamayor, E. O. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology*, *86*, 1715–1722. https://doi.org/10.1890/04-1587

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. https://doi.org/10.1038/ismej.2007.53

Rosenberg, M. S., Rothstein, H. R., & Gurevitch, J. (2013). Effect sizes: conventional choices and calculations. *Handbook of Meta-Analysis in Ecology and Evolution* (eds. J. Koricheva, J. Gurevitch & K. Mengersen) (pp. 61–71). Princeton University Press, Princeton, NJ.

Schmidt, T. S. B., Rodrigues, J. F. M., & Von Mering, C. (2017). A family of interaction-adjusted indices of community similarity. *The ISME journal*, *11*(3), 791–807.

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*, e01371-14. https://doi.org/10.1128/mBio.01371-14

Shade, A., & Stopnisek, N. (2019). Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology*, *49*, 50–58. https://doi.org/10.1016/j.mib.2019.09.008

Shmida, A., & Wilson, M. V. (1985). Biological determinants of species diversity. *Journal of Biogeography*, *12*, 1–20. https://doi.org/10.2307/2845026

Shoemaker, W. R., Locey, K. J., & Lennon, J. T. (2017). A macroecological theory of microbial biodiversity. *Nature Ecology & Evolution*, *1*, 0107. https://doi.org/10.1038/s41559-017-0107

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. https://doi.org/10.4319/lo.2011.56.2.0508

Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, *30*, 3–12. https://doi.org/10.1111/j.0906-7590.2007.04817.x

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. https://doi.org/10.1111/j.1466-8238.2012.00772.x

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. https://doi.org/10.1111/1574-6941.12308

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. https://doi.org/10.1007/s00572-019-00906-1

Vos, M., Wolf, A. B., Jennings, S. J., & Kowalchuk, G. A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiology Reviews*, *37*, 936–954. https://doi.org/10.1111/1574-6976.12023

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. https://doi.org/10.1016/j.soilbio.2010.12.009

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, *28*, 20. https://doi.org/10.1145/1852102.1852106

Wilkinson, D. M. (2001). What is the upper size limit for cosmopolitan distribution in free-living microorganisms? *Journal of Biogeography*, *28*, 285–291. https://doi.org/10.1046/j.1365-2699.2001.00518.x

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. https://doi.org/10.1111/j.1365-2699.2011.02569.x

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., Barba, M. D., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., & Chave, J. (2019). Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, *28*, 528–543. https://doi.org/10.1111/mec.14919

BIOSKETCH

Dave R. Clark is a microbial ecologist whose research focuses on understanding microbial biodiversity and community structure across a wide range of spatial scales.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Clark DR, Underwood GJC, McGenity TJ, Dumbrell AJ. What drives study-dependent differences in distance–decay relationships of microbial communities?. *Global Ecol Biogeogr*. 2021;00:1–15. https://doi.org/10.1111/geb.13266