



Why do microbes exhibit weak biogeographic patterns?

Kyle M. Meyer¹ · Hervé Memiaghe^{2,3} · Lisa Korte⁴ · David Kenfack⁵ · Alfonso Alonso⁴ · Brendan J. M. Bohannan¹

Received: 15 May 2017 / Revised: 25 August 2017 / Accepted: 20 November 2017 / Published online: 16 April 2018
© International Society for Microbial Ecology 2018

Abstract

Analysis of patterns in the distribution of taxa can provide important insights into ecological and evolutionary processes. Microbial biogeographic patterns almost always appear to be weaker than those reported for plant and animal taxa. It is as yet unclear why this is the case. Some argue that microbial diversity scales differently over space because microbial taxa are fundamentally different in their abundance, longevity and dispersal abilities. Others have argued that differences in scaling are an artifact of how we assess microbial biogeography, driven, for example, by differences in taxonomic resolution, spatial scale, sampling effort or community activity/dormancy. We tested these alternative explanations by comparing bacterial biogeographic patterns in soil to those of trees found in a forest in Gabon. Altering taxonomic resolution, excluding inactive individuals, or adjusting for differences in spatial scale were insufficient to change the rate of microbial taxonomic turnover. In contrast, we account for the differences in spatial turnover between these groups by equalizing sampling extent. Our results suggest that spatial scaling differences between microbial and plant diversity are likely not due to fundamental differences in biology, and that sampling extent should be taken into account when comparing the biogeographic patterns of microorganisms and larger organisms.

Introduction

Biogeography describes the distribution of taxa over space and time, and it has led to fundamental insights into the mechanisms maintaining and generating species diversity [1]. Numerous studies have established that microbial communities can exhibit biogeographic patterns, and in many cases these patterns are qualitatively similar to those of macro-organisms [2–4]. Microbial biogeographic

patterns, however, tend to be much weaker than those of macro-organisms. For example, the accumulation of taxa with increasing area and the decay of community similarity with geographic distance (two very well studied biogeographic patterns) tend to be lower for microorganisms than for plants and animals [2–5]. It is as yet unclear why this occurs.

Understanding why microorganisms differ quantitatively from plants and animals in their distribution is important for several reasons. First, biogeographic patterns can provide insight into the fundamental processes that determine biodiversity. Quantitative differences in biogeographic patterns could suggest that these fundamental processes are different for microbes and larger organisms. Second, biogeography forms a foundation for conservation and environmental management, including bioprospecting. Understanding whether or not microbial and plant/animal biogeography are governed by different rules is important for designing effective management and conservation strategies [6–8].

Some have suggested that microbes have weak biogeographic patterns because they are fundamentally different in ways that alter their biogeography; for example, due to high abundance, longevity, or dispersal abilities [9]. Others however, have suggested that these differences are artifacts of how microbial biogeography is studied [10, 11]. These artifacts could include: (1) that the operational taxonomic

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41396-018-0103-3>) contains supplementary material, which is available to authorized users.

✉ Kyle M. Meyer
kmeyer@uoregon.edu

¹ Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA

² Department of Landscape Architecture, University of Oregon, Eugene, OR, USA

³ Institut de Recherche en Écologie Tropicale (IRET), Libreville, Gabon, USA

⁴ Smithsonian Conservation Biology Institute, Washington, DC, USA

⁵ Smithsonian Tropical Research Institute, Washington, DC, USA

units (OTUs) used for characterizing microbes are not an appropriate analog to plant or animal species [4, 12–14], (2) that microbial communities tend to contain high numbers of inactive individuals and most microbial surveys do not distinguish active from inactive individuals [15, 16], (3) that the spatial scales over which biogeographic patterns are assessed differ between microbial and plant/animal studies [3], and (4) that microbial communities tend to be of much higher diversity than plant/animal communities, and thus more prone to severe undersampling, which in turn may result in under-estimating rates of taxonomic turnover [11]. We consider the implications of each of these potential artifacts below.

How taxonomic groups are defined strongly differs between macro-organisms and microorganisms. For microbial taxa, morphological traits are rarely useful for separating lineages, and the physiological measurements necessary to distinguish taxa are possible only for the minority of taxa that can be grown in culture. Thus, researchers commonly delineate taxa using the sequence similarity of marker genes (most commonly ribosomal genes [17]). This sequence similarity is used to create OTUs, defined by an arbitrary sequence similarity cutoff (e.g., 97%). It has been suggested that OTUs defined at 97% sequence similarity tend to contain much higher levels of diversity than typical plant or animal species, and thus may be more comparable to a higher taxonomic level, e.g., a genus or family [4, 18]. It has been demonstrated that the choice of OTU similarity cutoff can impact diversity patterns [13] including biogeographic patterns [4].

Not all microbial taxa are active in a given place and time [15]. Numerous microbial taxa are capable of entering a state of dormancy (i.e., physiological inactivity), and the percentage of microbial cells in this state can be as high as 80–97% in certain environments [15, 16]. This pool of inactive taxa has been likened to a seed bank in that member taxa may emerge into a state of activity/growth in response to various biotic or abiotic cues much like plant seeds in the soil. The typical DNA-based surveys used to assess microbial community membership do not distinguish between active and inactive taxa. Locey [19] argued that if dormancy increases the rate of immigration (by allowing immigrants to avoid initial adverse conditions) and decreases the rate of extinction (by allowing taxa to avoid death from, e.g., starvation or exclusion by a competitively superior individual), then microbial communities containing dormant taxa should exhibit lower temporal turnover since the likelihood of a newcomer being a new species would decrease over time [19]. The same argument could be used for spatial turnover, i.e., that over time the seed bank should tend to accumulate most regional taxa regardless of whether they are suited to the local environment. Thus, including inactive taxa in our

surveys could decouple community turnover from environmental turnover and result in an underestimation of rates of community turnover.

It is well established that biogeographic patterns can change quantitatively with spatial scale. This is true for both microbes [10, 20] and larger organisms [21–25]. It has been suggested that environmental filtering is a more important driver of biogeographic patterns at smaller spatial scales [10, 22, 26] while dispersal limitation and/or diversification are more important drivers of large-scale spatial patterns [2, 27, 28]—although dispersal limitation can also play a role at local scales as well [10, 29]. Microbial and plant/animal biogeographic surveys are often performed at different spatial scales and this could potentially confound our interpretations of how diversity of these groups scales quantitatively. Including, for example, more small-scale spatial comparisons in a survey could make rates of community turnover appear lower when compared to a survey comprised mainly of large-scale comparisons.

Finally, incomplete sampling of communities is a problem that exists throughout ecology [30], but is particularly pronounced for microbial communities, which tend to be especially diverse. Under-sampling tends to be biased against rare community members. Rare members are often more restricted in range and hence could be important in determining biogeographic patterns. Woodcock et al. [11] showed that the rate at which microbial species richness increases with area can be strongly influenced by the intensity of sampling effort. However, it has also been suggested that rare taxa exert relatively minimal effects on microbial biogeographic patterns compared to the effects of species abundances and levels of population aggregation [21, 31]. The impacts of under-sampling on biogeography in environmental surveys has rarely been assessed and, to our knowledge, never in the context of accounting for the differences between microbial and plant/animal biogeographic patterns.

Here we compare the rates of the decay of taxonomic similarity over geographic distance between the soil bacterial community and the tree community in a forest in the Rabi Forest Monitoring Plot, Gabon. The distance decay of community similarity is a fundamental pattern in the biogeography of plant/animal [21, 23–25, 31] and microbial [2–4, 29, 32, 33] taxa. Our design allows us to compare this relationship across spatial scales ranging from centimeters to 100 s of meters. We test the following hypotheses: (1) microbial species definitions will influence the rate at which microbial community similarity changes over space, (2) excluding inactive microbial taxa will result in the steepening of microbial distance–decay patterns, (3) microbial and tree distance decay patterns will become more similar when compared at the same spatial scales, and (4) the effects of under-sampling a community can account for the

differences between microbial and tree distance–decay rates.

Materials and methods

Experimental design

The study was conducted at the Smithsonian Center for Tropical Forest Science's (CTFS) 25 ha plot located in the Rabi oil field in Southwestern Gabon (2° 13' 22" S, 9° 55' 2" E), within the Gamba Complex of Protected Areas [34]. This plot, which is part of the Smithsonian Forest Global Earth Observatory (ForestGEO) network, was established for the purposes of studying forest dynamics and spatial ecology. The Rabi plot is particularly advantageous in that all trees with ≥ 1 cm diameter at breast height (dbh) have been censused [35], which allows for direct comparisons between spatial patterns of trees and microbes in the same landscape.

Microbial sampling took place at the end of the dry season in September 2013. Within the 25 ha plot, we sampled using a spatially explicit nested design (Supplementary Figure 1a) whereby three 100 m \times 100 m quadrats were established, with 10 m \times 10 m, 1 m \times 1 m, 0.1 m \times 0.1 m quadrats nested within each, giving high coverage of a range of spatial scales. Soil cores were taken from the corners of each quadrat giving a total of 39 samples. Soil cores were taken using standard coring methods to a depth of 15 cm, following the removal of the litter layer. For each sampling point three representative soil cores were taken, homogenized, then either subsampled and preserved for molecular analysis (described below) or kept on ice and transported back to the US for soil chemical analysis (described below).

Tree data were obtained from the data set of the first census of the Rabi plot [35]. Data for all tree individuals with dbh ≥ 1 cm for all areas of the 25 ha plot overlapping with the soil bacterial census were extracted (ref. [35], Supplementary Figure 1b). We assessed tree community turnover by comparing the tree species composition of each of the 20 m \times 20 m quadrats included in the study.

Molecular analysis

From each set of homogenized soil cores, 3 ml (~1 g) of soil was added to 9 ml Lifeguard solution (Mobio, California, USA) in the field, then shipped cold and stored at -80 °C in order to stabilize nucleotides for later extraction. Soil DNA and RNA were co-extracted from each sample using MoBio's Powersoil RNA Isolation kit with the DNA Elution Accessory Kit (MoBio, California, USA) following manufacturer's instructions, using 3 ml of the soil:Lifeguard

mixture (~0.25 g soil). Extractions were quantified using Qubit (Life Technologies, USA). RNA was reverse transcribed to cDNA using Superscript III first-strand reverse transcriptase and random hexamer primers (Life Technologies, USA).

The V3 and V4 region of the 16S rRNA gene of the DNA and cDNA were PCR amplified using the primers 319F and 806R. This region is considered a molecular barcode for identifying bacterial taxa in the environment [36]. Sequencing libraries were prepared using a 2-step PCR with a dual-indexing approach [36, 37]. The first round of amplification consisted of 22 cycles with Phusion HiFi polymerase. Products were cleaned using Agencourt AMPure XP (Beckman Coulter, California, USA), then amplified for an additional six cycles. The final library was sent to the Dana-Farber Cancer Institute Molecular Biology Core Facilities for 300 bp paired-end sequencing on the Illumina MiSeq platform.

Soil chemical analysis

Soil chemical parameters were measured from each soil core to estimate the impact of the chemical environment on microbial community composition. All soil chemical analyses were performed by A & L Western Agricultural Lab (Modesto, CA, USA). In total, percent organic matter (loss on ignition [38]), extractable phosphorus (Weak Bray [39] & sodium bicarbonate [40]), nitrate-N, extractable cations (K, Mg, Ca, Na), sulfate-S [41], pH, buffer pH, and cation exchange capacity (CEC) [42], were measured.

Data processing and statistical analysis

Paired end reads were joined then demultiplexed in QIIME [43] before quality filtering. Primers were removed using a custom script. UPARSE was used to quality filter and truncate sequences (416 bp, EE 0.5) [44]. Sequences were retained only if they had an identical duplicate. OTUs were clustered de novo at 97% using USEARCH [45]. OTUs were checked for chimeras using the gold database in USEARCH. To assign taxonomy, we used repset from UPARSE in QIIME using greengenes version 13_5 (RDP classifier algorithm). Finally, we averaged 100 rarefactions at 3790 observations per sample to achieve approximately equal sampling depth and avoid bias associated with a single rarefaction, which excluded four samples.

Statistical analyses were performed in the R platform [46]. Canberra pairwise community distances were calculated for both the bacterial and tree communities using the vegdist function in the package 'vegan' [47]. Canberra was chosen because of its incorporation of abundance data and sensitivity to rare community members [48]. Turnover was estimated for both the bacterial and tree communities by

regressing pairwise similarity against pairwise geographic distance [21]. Mantel tests were used to test for significant associations between geographic and community distance in base R. Distance–decay slopes were compared using the function `diffslope` in the package ‘`simba`’ [49], which employs a randomization approach across samples from each data set and compares difference in slope to the original configuration of samples. The p values computed are the ratio between the number of cases where the differences in slope exceed the difference in slope of the initial configuration and the number of permutations (1000).

The relative impacts of the environment and geographic distance on microbial community dissimilarity were assessed using partial mantel tests on distance matrices as implemented by the `mantel.partial` function in the ‘`vegan`’ package [47] in R. Environmental dissimilarity was calculated using the Gower general dissimilarity coefficient [50] as implemented by the function `daisy` in the ‘`cluster`’ package [51] in R. The influence of individual soil parameters on community dissimilarity was assessed using redundancy analysis as implemented by the `rda` function in ‘`vegan`’ [47] following Hellinger transformation of the community data.

OTU clustering experiment

To test whether species definition impacts biogeographic patterns, OTUs were clustered at 95, 97, 99, and 100% similarity thresholds, each time using the aforementioned bioinformatic pipeline. Clustering at these levels resulted in 1179, 2243, 6611, and 14 864 OTUs, respectively. RNA- and DNA-derived OTU tables were then separated and averaged across 100 rarefactions to achieve approximately equal sampling depth. DNA-derived OTU tables were rarefied to 4709, 3100, 3324, and 2479 observations per sample (the minimum number of observations per sample that would allow us to retain all samples), respectively. RNA-derived OTU tables were rarefied to 3693, 3100, 2375, and 2049 observations per sample, respectively. Linear models of community turnover (described above) were compared against the tree community turnover linear model for each OTU threshold using the random permutation approach described above.

RNA- versus DNA-inferred community comparison

To ask whether distinguishing the active bacterial community members from the inactive members would impact biogeographic patterns, we inferred bacterial community membership using two molecular methods: analysis of community RNA and analysis of community DNA. By inferring community membership via RNA we enrich for

taxa that are active, whereas communities inferred via DNA will tend to include a higher proportion of inactive members. Distance–decay linear regression slopes were compared between the RNA- and DNA-inferred communities clustered at the 97% OTU similarity threshold using the aforementioned permutation approach.

Spatial scale

To assess whether bacterial community distance–decay rates more closely resemble tree community distance–decay rates at the same spatial scale, we subset the bacterial community to only include comparisons at the same spatial scale as trees. We also asked whether bacterial distance–decay patterns differed at different spatial scales by subsampling our data to include only small- to medium-scale comparisons (tens of centimeters to tens of meters) and medium- to large-scale comparisons (tens of meters to hundreds of meters).

Effects of undersampling

We used rarefaction to assess the impact of undersampling on biogeographic patterns for both tree and bacterial communities. We wrote a custom R function (provided in the supplementary code) that repeatedly subsamples (1000 times) a community at a given depth and computes a distance–decay linear regression for each sampling event. For this study we used a 97% OTU cutoff for the DNA-inferred community.

Results

Distance–decay of community similarity

Community similarity (1- Canberra dissimilarity) significantly decreased with geographic distance for both the bacterial (Mantel $r = 0.569$, $p = 0.001$) and tree (Mantel $r = 0.476$, $p = 0.001$) communities (Fig. 1). The soil chemical environment showed slight spatial autocorrelation over the distances covered (Mantel $r = 0.11$, $p < 0.01$), but was relatively uniform (Supplementary Table 1). Variation in the soil chemical environment overall was significantly correlated with bacterial community variation (partial Mantel $r = 0.233$, $p = 0.011$) after having controlled for the effects of distance. Soil pH and sulfur concentration were significantly associated with variation in bacterial community structure (RDA $F_{1,35} = 2.603$, $P < 0.01$ and $F_{1,35} = 2.597$, $P < 0.01$, respectively). The rate at which community similarity decayed over space differed significantly between bacterial and tree communities (difference in slope: 0.02, $p < 0.001$) with the tree community exhibiting a steeper rate

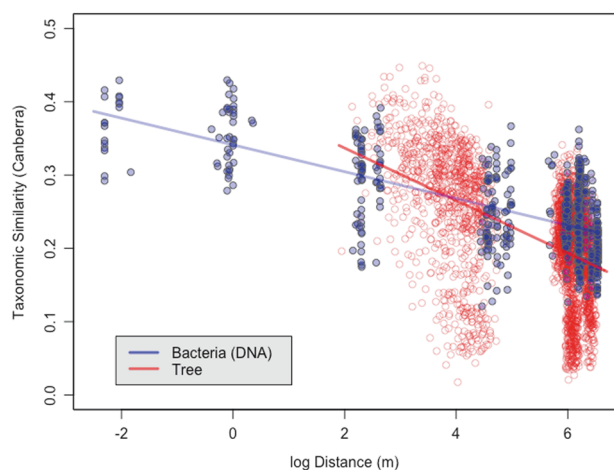


Fig. 1 Distance–decay plot of the bacterial community (inferred from DNA, OTU cutoff = 97%) versus the tree community on the Rabi plot, Gabon. Colors for bacterial samples are transparent

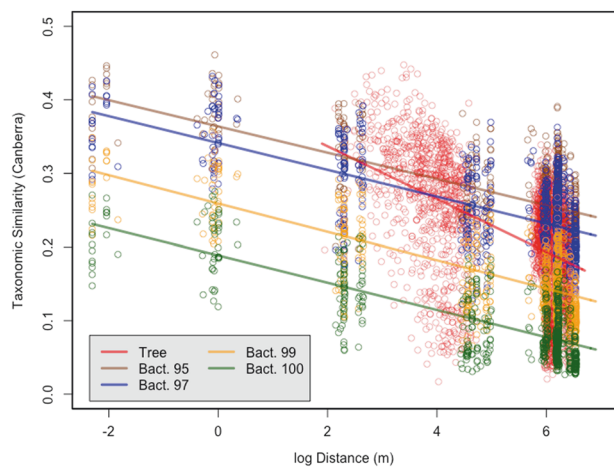


Fig. 2 The impacts of changing OTU threshold on distance–decay patterns of the DNA-derived soil bacterial community at the Rabi plot, Gabon

of turnover (-0.0359 ± 0.001) than the bacterial community (Fig. 1, -0.0183 ± 0.0008).

The impact of OTU clustering

We asked whether altering the sequence similarity cutoff used to define our taxa (analogous to moving from subspecies to species to genera and families) could impact the rate of bacterial community turnover in our data and account for the differences between the tree and bacterial community turnover rates. Neither broadening (i.e., to 95%) nor narrowing (i.e., to 99 and 100%) sequence similarity cutoffs significantly altered the rate of community turnover (Fig. 2), and thus the bacterial community distance–decay rate was lower than that of the trees, regardless of OTU definition. The range of taxonomic similarity values, however, did change with taxonomic definition. Broader cutoffs

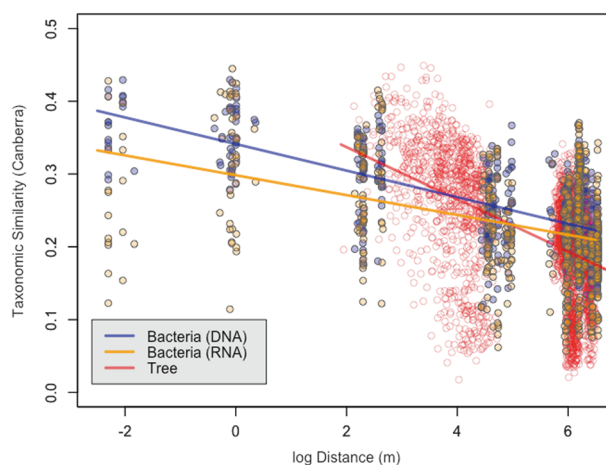


Fig. 3 Distance–decay patterns of DNA- and RNA-inferred bacterial communities at the Rabi plot, Gabon. Colors for bacterial samples are transparent

tended to exhibit higher levels of taxonomic similarity while narrower cutoffs exhibited lower ranges of taxonomic similarity.

Excluding inactive taxa

We tested whether excluding inactive taxa from our survey would render the microbial distance–decay rate more similar to that of the tree community. Excluding inactive taxa, however, did not result in a steeper distance–decay slope in our study (Fig. 3). The RNA-inferred (active) community distance–decay slope (-0.0137 ± 0.001) was significantly flatter than the DNA-inferred (active + inactive) community distance–decay slope (-0.0183 ± 0.0008 , difference in slope = 0.0047, $p = 0.005$) and both community distance–decay rates were lower than the tree community distance–decay rate (-0.0359 ± 0.001). For both communities, geographic distance was a more important predictor of community variation than turnover in the soil chemical environment. Variation in the DNA-inferred community structure was more predictable overall by our meta-data (geographic distance and soil chemical environment) than the RNA-inferred community. In fact, variation in the soil chemical environment was not a significant predictor of variation in the RNA-inferred community.

We also asked whether the OTU clustering threshold of the RNA-inferred community impacted the slope of the distance–decay relationship. Distance–decay slopes across 95, 97, 99, and 100% thresholds were statistically indistinguishable from one another, but decreased in the range of similarity level with higher OTU threshold (Supplementary Figure 2). All slopes were flatter than the tree community distance–decay slope.

The RNA-inferred community was dominated by Proteobacteria, Actinobacteria, and Acidobacteria, comprising

61.9%, 18.9%, and 11.0% of sequences, respectively. The DNA-inferred community was similar, being dominated by Proteobacteria, Acidobacteria, and Actinobacteria, comprising 54.4%, 21.1%, and 13.0% of sequences, respectively. The DNA- and RNA-inferred communities had an average of 486.2 ± 16 and 332.8 ± 17 OTUs per sample, respectively, and shared an average of 238 ± 12 OTUs per sample. The RNA-inferred community was not a complete subset of the DNA-inferred community, containing on average $27.9 \pm 0.01\%$ OTUs not detected in the DNA-inferred community. The DNA-inferred community contained on average $51.6 \pm 0.02\%$ OTUs that were not detected in the RNA.

Spatial scale

We asked first whether comparing microbial and tree communities at the same spatial scale might account for the discrepancy between tree and bacterial distance–decay patterns and second whether there was an alternate spatial scale at which the bacterial distance–decay slope might resemble more closely that of trees. The microbial distance–decay slope across all scales did not significantly differ from the slope derived from the subset of spatial distances shared with trees (difference in slope = 0.0007, $p = 0.323$, Fig. 4). Thus, when compared at the same spatial scales, the microbial distance–decay slope was still significantly shallower than the tree distance–decay slope (difference in slope = 0.018, $p < 0.001$). At the small (centimeters to meters) scale subset, the distance–decay slope was not significantly different from zero, although it tended to be shallower than the distance–decay slope calculated from the entire data set. At the largest subset (hundreds of meters), the slope was not significantly different from the

slope derived from the entire dataset (difference in slope = 0.0022, $p = 0.072$).

Sampling effort

Both tree and bacterial communities in our study showed a positive frequency–abundance relationship (Supplementary Figure 3a, b) whereby abundant taxa tended to be more widespread across the study plot and low abundance taxa tended to be more restricted in distribution. We simulated the effects of under-sampling on the distance–decay relationship by using rarefaction on both the tree and bacterial communities. For both communities we saw the same trend; the more thoroughly sampled a community was, the steeper the distance–decay rate (Fig. 5a). We then asked whether the effects of sampling effort could account for the differences in distance–decay slope between trees and microbes. We found that if we sampled the bacterial community as deeply as we could, the distance–decay slope was within the 95% CI range of the tree community when the tree community was dramatically under-sampled (Fig. 5b).

Discussion

Although numerous studies have reported differences in the biogeographic patterns of microbial taxa and plants/animals, there have been very few studies that have attempted to disentangle the drivers of these differences. We first demonstrated that the microbial community distance–decay rate at our site was lower than that of the tree community. We then asked: (1) whether the microbial species definition (i.e., the OTU sequence similarity threshold) had an influence on the rate at which microbial community similarity changes over space, (2) if excluding inactive microbial taxa (by inferring microbial community structure via RNA sequencing) would result in the steepening of microbial distance–decay patterns, (3) whether microbial and tree distance decay patterns would become more similar when compared at the same spatial scales, and (4) whether the effects of under-sampling a community would account for the differences between microbial and tree distance–decay rates.

Various studies have suggested that broadening taxonomic resolution (for example, by comparing genera or families, rather than species) can decrease the strength of biogeographic patterns [4, 13, 18], although not always [52]. To test this idea, we clustered OTUs at four different sequence similarity thresholds (i.e., 95, 97, 99, and 100% sequence similarity)—analogous to moving from families/genera to species and subspecies—and observed no change to the rate at which community similarity changes over distance for both the RNA- and DNA-inferred communities.

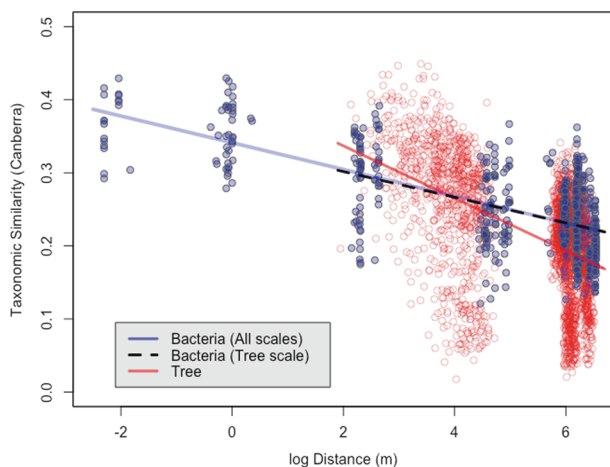


Fig. 4 The distance–decay slope of soil bacterial communities considered at spatial subsets. Colors for bacterial samples are transparent

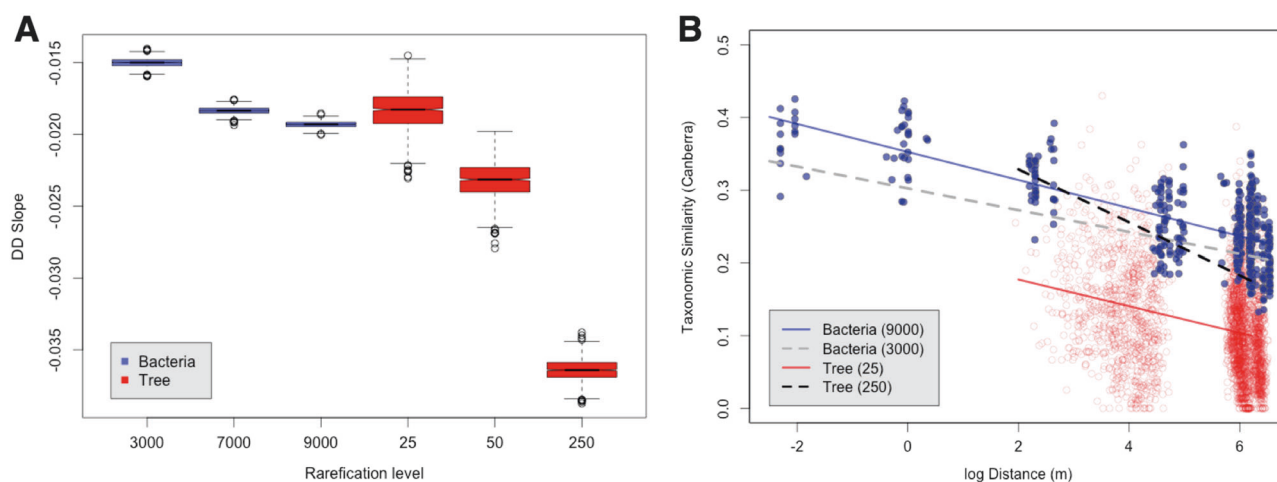


Fig. 5 Sampling effort impacts the distance–decay slope in bacterial and tree communities. **a** The range of distance–decay slopes derived from different levels of sampling intensity for the bacterial and tree communities. Results shown represent 1000 sampling efforts at each

level of rarefaction. **b** Sampling effort can account for the differences in distance–decay rate between bacteria and trees. Colors for bacterial samples are transparent

Our results are in contrast to Horner-Devine et al. [4] who reported that narrowing the sequence similarity cutoff for taxon definition resulted in a steeper bacterial distance–decay slope in a temperate salt marsh ecosystem. There are a number of potential explanations for why we did not observe this in our study. Our findings might be different because the contribution of environmental variation to bacterial community turnover was lower in our study than that reported by Horner-Devine et al. [4]. If the distance decay of community similarity is driven strongly by the distance decay of environmental similarity, and if narrowing taxonomic resolution results in groups with narrower environmental tolerances, then a steeper distance decay pattern should result. Another possibility is that the traits required for survival under any given set of environmental conditions were strongly phylogenetically conserved in the taxa in our study. This would result in less of an impact of changing taxonomic (i.e., phylogenetic) resolution on the breadth of environmental tolerances (and ultimately, the rate of distance–decay). Thus, while it has been suggested by various authors that OTU definition may quantitatively impact the biogeographic patterns of microbial communities, we find no support for this hypothesis in our study.

The soil environment contains especially high proportions of physiologically inactive (i.e., dormant) microbial taxa [15, 16] and most DNA-based microbial surveys include both active and inactive taxa. Biogeographic surveys of plants and animals, in contrast, rarely include dormant individuals (e.g., seeds). Given that dormancy can allow taxa to persist outside of optimal environmental conditions, the inclusion of inactive taxa could decouple microbial community turnover from environmental

turnover. We hypothesized that if landscape level distance–decay relationships were largely driven by environmental turnover, then including inactive taxa in a microbial survey would flatten the distance–decay slope. Thus, by excluding the inactive taxa (and focusing solely on the active taxa) we expected that the microbial distance–decay slope would become steeper and that this would—at least in part—account for the differences in biogeographic patterns between trees and microbes in our study. This, however, was not what we observed. The active (RNA-inferred) community showed a flatter distance–decay relationship than the DNA-inferred (active + inactive) community and variation in the active community showed less of a statistical association with soil chemical variables than the DNA-inferred community. While this observation was at odds with our expectation, our hypothesis relied on the assumption that the environmental factors responsible for microbial activity would be spatially autocorrelated. Alternatively, if climatic variables such as rainfall events—which tend to be relatively uniform over a landscape—are stronger determinants of soil activity, then we would expect the active community to be more uniform over space, which is what we observed. Seasonal rewetting events in California grasslands have been shown to strongly influence the composition of the active fraction of soil microbial communities [53], and our microbial sampling took place at the end of the dry season when seasonal rewetting was underway. Thus, we find no support for the hypothesis that the inclusion of inactive taxa is responsible for the weakening of the distance–decay relationship in microbial communities.

Both plant/animal and microbial communities have been reported to have different drivers of biogeographic patterns

at different spatial scales [10, 18, 22, 23, 28, 54]. Studies of microbial biogeography are often conducted at smaller spatial scales than those of plants and animals (although not always, e.g., [4]), and this could result in differences in the relative strength of the biogeographic patterns observed. We asked first whether comparing microbial and tree communities at the same spatial scale might account for the difference between tree and bacterial distance–decay patterns and second whether there was an alternate spatial scale at which the bacterial distance–decay slope might resemble more closely that of trees. At the same spatial scales as the tree community (tens of meters to hundreds of meters) the bacterial distance–decay slope was statistically indistinguishable from the slope derived from all spatial scales, indicating that the differences between bacterial and tree community distance decay rates in our study are not likely due to a mismatch in scale. While it has previously been reported that distance–decay rates at smaller spatial scales tend to be lower than those calculated from data sets spanning a larger range of spatial scales [10, 20, 31], we did not detect a significant distance–decay relationship at the smaller spatial subsets in our study. Moreover, Martiny et al. [10] have shown that larger spatial scales tend to exhibit steeper distance–decay slopes than slopes derived from the entire data set. Although this was not the case for our largest spatial subsets, our largest subset was still at a smaller spatial scale and spanned less spatial scales than their survey. Thus adjusting for differences in scale does not account for the differences in microbial and tree distance–decay slopes in our study.

Undersampling communities is a problem that exists throughout ecology [30]. This problem is particularly pronounced in microbial ecology where exhaustively sampling any environment can be impractical if not impossible. In most studies of microbial communities, collector's curves are far from saturation, and unique taxa continue to accumulate with increased sampling effort [11, 55]. Undersampling can lead to a weakening of biogeographic patterns if taxa have a positive frequency–abundance relationship [11, 56], whereby abundant community members tend to be more widespread and less abundant taxa tend to be more restricted in distribution. This occurs because undersampling results in decreased detection of low abundance taxa (with restricted distributions) and the community will thus appear to have less taxonomic turnover across space. Both microbial and plant/animal communities have been reported to have positive frequency–abundance relationships [56–60], and this was the case for both the tree and bacterial communities in our study. We tested whether differences in sampling intensity could be driving the discrepancy between distance–decay rates of tree and bacterial communities in our study by simulating sampling effort using rarefaction. We show that when microbial

communities are deeply sampled, their community distance–decay rates become within the range of very under-sampled tree community distance decay rates, suggesting that sampling intensity plays a strong role in driving the discrepancy of biogeographic patterns between these communities. This finding is congruent with results reported by Woodcock et al. [11] where it was shown in synthetic communities that lower sampling effort could flatten the slope of the taxa–area relationship. While this finding has been previously suggested by Woodcock et al. [11], it has until now remained untested on data from the field, especially in the context of accounting for differences between microbial and macro-organismal biogeographic patterns.

There are a number of important caveats when comparing the spatial patterns of macro- and microorganisms. First, groups like bacteria and trees are greatly different in their levels of diversity, and this disparity could further complicate comparing the spatial patterns of these two groups. Future work could target narrower groups of microorganisms, such as individual phyla or classes, to test whether narrower groups display alternate spatial patterns than larger aggregations of groups. Subgroups within the phylum Acidobacteria have been shown to differ in their distance–decay rate [61], as well as from the phylum-level distance–decay rate. It however remains untested whether this is generalizable across other groups. Another important consideration in microbial biogeography is whether to focus on the turnover of microbial taxonomic structure or the distribution of traits [62]. The global distribution of N-cycling traits has been shown to be much more predictable by environmental conditions than the distribution of the taxa encoding those traits [63]. This has also been shown for the distribution of functional groups involved in marine biogeochemical cycles [64]. Future efforts could focus on traits related to dispersal (e.g., spore formation) to further explore how these attributes contribute to spatial patterns. Other important considerations for future work could be to incorporate more soil environmental parameters (e.g., soil moisture), to potentially increase predictive power, or to expand sampling regimes of both macro- and microorganisms to larger spatial scales (e.g., regional or continental).

Whether our findings are generalizable across other environments, taxonomic groups, or spatial scales remains untested, but since frequency–abundance relationships are common [56–60], it seems likely that the influence of sampling effort on biogeographic patterns will be generalizable to other systems. Our results emphasize the importance of deeper sampling if we are to learn about the ecology of endemic microbial taxa. Furthermore, our findings support the idea that microbial taxa not only qualitatively fit the same biogeographic patterns as plants and animals, but they may do so quantitatively as well. Indeed more intensive sampling efforts may reveal that the spatial

scaling of microbial diversity is not so fundamentally different from that of other forms of life.

Data accessibility

DNA and cDNA sequence FASTA files, OTU tables, soil environmental data, and an R script for repeated rarefaction and distance–decay analysis will be available for download from <https://doi.org/10.6084/m9.figshare.5001314.v1>. Tree community data can be accessed upon request from <http://www.ctfs.si.edu/site/Rabi>.

Acknowledgements We thank the Government of Gabon, Centre National de la Recherche Scientifique et Technologique for permission (Permit N° 866/MENESTFPRSCJS/CENAREST/CG/CAB) to conduct this study. We also thank P. Voua Otomo, G. Moussavou, P. Bissiemou, D. Nguema, E. Mounoumoulossi Nounoungou, and W. Mbadinga-Mbadinga for their help in collecting samples. The Smithsonian Conservation Biology Institute, the Gabon-Oregon Transnational Center on Environment and Development, and Shell Gabon provided financial and logistical support. This is contribution # 154 of the Gabon Biodiversity Program.

Author contributions KMM and BJMB designed research. KMM, BJMB, and LK collected samples. KMM and HM performed analysis. KMM wrote the manuscript with inputs from all authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Lomolino MV, Riddle BR, Brown JH. Biogeography 3rd ed. In: Lomolino MV, Riddle BR, & Brown JH, editors. Sunderland, MA: Sinauer Associates Sunderland; Species and Extinction, Chapter 7. 2006.
- Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, et al. Spatial scaling of microbial eukaryote diversity. *Nature*. 2004;432:747–50.
- Hillebrand H, Watermann F, Karez R, Berninger UG. Differences in species richness patterns between unicellular and multicellular organisms. *Oecologia*. 2001;126:114–24.
- Horner-Devine MC, Lage M, Hughes JB, Bohannon BJM. A taxa-area relationship for bacteria. *Nature*. 2004;432:750–3.
- Zhou J, Kang S, Schadt CW, Garten CT. Spatial scaling of functional gene diversity across various microbial taxa. *Proc Natl Acad Sci USA*. 2008;105:7768–73.
- Diamond JM. The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biol Conserv*. 1975;7:129–46.
- Simberloff D, Abele LG. Refuge design and island biogeographic theory: effects of fragmentation. *Am Nat*. 1982;120:41–50.
- Soule ME, Simberloff D. What do genetics and ecology tell us about the design of nature reserves? *Biol Conserv*. 1986;35:19–40.
- Finlay BJ. Global dispersal of free-living microbial eukaryote species. *Science*. 2002;296:1061–3.
- Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-devine MC. Drivers of bacterial β -diversity depend on spatial scale. *Proc Natl Acad Sci USA*. 2011;108:7850–54.
- Woodcock S, Curtis TP, Head IM, Lunn M, Sloan WT. Taxa-area relationships for microbes: the unsampled and the unseen. *Ecol Lett*. 2006;9:805–12.
- Fuhrman JA, Campbell L. Marine ecology: microbial microdiversity. *Nature*. 1998;393:410–11.
- Storch D, Šizling AL. The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? *Folia Geobot*. 2008;43:329–44.
- Tiedje JM. Approaches to the comprehensive evaluation of prokaryote diversity of a habitat. In: D. Allsopp, D.L. Hawksworth and R.R. Colwell, editors, *Microbial diversity and ecosystem function*. CAB International, Wallingford, England; 1995; p. 73–97.
- Lennon JT, Jones SE. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol*. 2011;9:119–30.
- Blagodatskaya E, Kuzyakov Y. Active microorganisms in soil: critical review of estimation criteria and approaches. *Soil Biol Biochem*. 2013;67:192–211.
- Pace NR. A molecular view of microbial diversity and the biosphere. *Science*. 1997;276:734–40.
- Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol*. 2012;10:497–506.
- Locey KJ. Synthesizing traditional biogeography with microbial ecology: the importance of dormancy. *J Biogeogr*. 2010;37:1835–41.
- Franklin RB, Mills AL. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol Ecol*. 2003;44:335–46.
- Nekola JC, White PS. The distance decay of similarity in biogeography and ecology. *J Biogeogr*. 1999;26:867–78.
- Preston FW. Time and space and the variation of species. *Ecology*. 1960;41:612–27.
- Soininen J, McDonald R, Hillebrand H. The distance decay of similarity in ecological communities. *Ecography*. 2007;30:3–12.
- Tuomisto H, Ruokolainen K, Yli-halla M. Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*. 2003;299:241–5.
- Condit R, Pitman N, Leigh EG, Chave J, Terborgh J, Foster RB, et al. Beta-diversity in tropical forest trees. *Science*. 2002;295:666–9.
- Martiny JBH, Bohannon BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol*. 2006;4:102–12.
- Papke RT, Ramsing NB, Bateson MM, Ward DM. Geographical isolation in hot spring cyanobacteria. *Environ Microbiol*. 2003;5:650–9.
- Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*. 2003;301:976–8.
- Bell T. Experimental tests of the bacterial distance-decay relationship. *ISME J*. 2010;4:1357–65.
- Cam E, Nichols JD, Hines JE, Sauer JR, Alpizar-Jara R, Flather CH. Disentangling sampling and ecological explanations underlying species-area relationships. *Ecology*. 2002;83:1118–30.
- Morlon H, Chuyong G, Condit R, Hubbell S, Kenfack D, Thomas D, et al. A general framework for the distance-decay of similarity in ecological communities. *Ecol Lett*. 2008;11:904–17.
- Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA*. 2006;103:626–31.
- Zinger L, Boetius A, Ramette A. Bacterial taxa-area and distance-decay relationships in marine environments. *Mol Ecol*. 2014;23:954–64.

34. Lee ME, Alonso A, Dallmeier F, Campbell P, Pauwels OSG. The gamba complex of protected areas: an illustration of Gabon's Biodiversity. *Bull Biol Soc Wash*. 2006;12:229–42.
35. Memiaghe HR, Lutz JA, Korte L, Alonso A, Kenfack D. Ecological importance of small-diameter trees to the structure, diversity and biomass of a tropical evergreen forest at Rabi, Gabon. *PLoS ONE*. 2016;11:e0154988.
36. Fadrosch DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. 2014;2:1–7.
37. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79:5112–20.
38. Dean WE. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. *J Sediment Res*. 1974;44:242–248.
39. Kamprath EJ, Watson ME. Conventional soil and tissue tests for assessing the phosphorus status of soils. The role of phosphorus in agriculture; In: F.E. Khasawneh, E.C. Sample, E.J. Kamprath editor, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, USA. 1980. p. 433–69.
40. Olsen SR. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. Washington: United States Department of Agriculture Circular; 1954.
41. Fox RL, Olson RA, Rhoades HF. Evaluating the sulfur status of soils by plant and soil tests. *Soil Sci Soc Am J*. 1964;28:243–6.
42. Chapman HD. Cation-exchange capacity. In: Methods of soil analysis. American Society of Agronomy, Soil Science Society of America; C.A. Black. Madison, WI, USA: 1965. p. 891–901.
43. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat Methods*. 2010;7:335–6.
44. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature*. 2013;10:996–1000.
45. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
46. Vienna, Austria, R Foundation for statistical computing. Development Core Team. R: a language and environment for statistical computing; 2010.
47. Oksanen J, Blanchet FG, Roeland K, Legendre P, Minchin P, O'Hara RB, et al. vegan: Community ecology package; 2015, <https://CRAN.R-project.org/package=vegan>.
48. Jost L, Chao A, Chazdon RL. Compositional similarity and Beta diversity. In: Magurran AE, McGill B, eds.. Biological diversity: frontiers in measurement and assessment.. Oxford: Oxford University Press; 2011. p. 66–84.
49. Jurasinski G, Retzer V. simba: A collection of functions for similarity analysis of vegetation data; 2012, <https://CRAN.R-project.org/package=simba>.
50. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics*. 1971;27:857–71.
51. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2016). cluster: Cluster Analysis Basics and Extensions.
52. Green J, Bohannan BJM. Biodiversity scaling relationships: are microorganisms fundamentally different? In: Storch D, Marquet PA, Brown JH, eds.. Scaling biodiversity. Cambridge: Cambridge University Press; 2007. p. 129–49.
53. Barnard RL, Osborne CA, Firestone MK. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *ISME J*. 2013;7:2229–41.
54. Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol Ecol*. 2010;19:4315–27.
55. Rosenzweig ML. Species diversity in space and time. Cambridge: Cambridge University Press; 1995.
56. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol*. 2006;8:732–40.
57. Gaston KJ, Blackburn TM, Lawton JH. Interspecific abundance-range size relationships: an appraisal of mechanisms. *J Anim Ecol*. 1997;66:579–601.
58. Östman Ö, Drakare S, Kritzbeg ES, Langenheder S, Logue JB, Lindström ES. Regional invariance among microbial communities. *Ecol Lett*. 2010;13:118–27.
59. Brown JH. On the relationship between abundance and distribution of species. *Am Nat*. 1984;124:255–79.
60. Thompson K, Hodgson JG, Gaston KJ. Abundance–range size relationships in the herbaceous flora of central England. *J Ecol*. 1998;86:439–48.
61. Navarrete AA, Venturini AM, Meyer KM, Klein AM, Tiedje JM, Bohannan BJ. et al. Differential response of Acidobacteria subgroups to forest-to-pasture conversion and their biogeographic patterns in the western Brazilian Amazon. *Front Microbiol*. 2015;6:1–10.
62. Green JL, Bohannan BJM, Whitaker RJ. Microbial biogeography: from taxonomy to traits. *Science*. 2008;320:1039–43.
63. Nelson MB, Martiny AC, Martiny JBH. Global biogeography of microbial nitrogen-cycling traits in soil. *Proc Natl Acad Sci USA*. 2016;113:8033–40.
64. Louca S, Parfrey LW, Doebli M. Decoupling function and taxonomy in the global ocean microbiome. *Science*. 2016;353:1272–7.