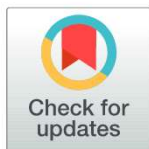


RESEARCH ARTICLE

Alpha-, beta-, and gamma-diversity of bacteria varies across habitats

Kendra E. Walters ^{*}, Jennifer B. H. Martiny

Department of Ecology and Evolutionary Biology, University of California, Irvine, California, United States of America

^{*} kewalter@uci.edu

Abstract

Bacteria are essential parts of ecosystems and are the most diverse organisms on the planet. Yet, we still do not know which habitats support the highest diversity of bacteria across multiple scales. We analyzed alpha-, beta-, and gamma-diversity of bacterial assemblages using 11,680 samples compiled by the Earth Microbiome Project. We found that soils contained the highest bacterial richness within a single sample (alpha-diversity), but sediment assemblages displayed the highest gamma-diversity. Sediment, biofilms/mats, and inland water exhibited the most variation in community composition among geographic locations (beta-diversity). Within soils, agricultural lands, hot deserts, grasslands, and shrublands contained the highest richness, while forests, cold deserts, and tundra biomes consistently harbored fewer bacterial species. Surprisingly, agricultural soils encompassed similar levels of beta-diversity as other soil biomes. These patterns were robust to the alpha- and beta- diversity metrics used and the taxonomic binning approach. Overall, the results support the idea that spatial environmental heterogeneity is an important driver of bacterial diversity.

 OPEN ACCESS

Citation: Walters KE, Martiny JBH (2020) Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. PLoS ONE 15(9): e0233872. <https://doi.org/10.1371/journal.pone.0233872>

Editor: João Carlos Nabout, Universidade Estadual de Goiás - Campus de Ciencias Exatas e Tecnológicas Henrique Santillo, BRAZIL

Received: May 13, 2020

Accepted: September 8, 2020

Published: September 23, 2020

Copyright: © 2020 Walters, Martiny. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: No new data were used in this research. All data used can be found on the Earth Microbiome Project (<https://doi.org/10.1038/nature24621>) or WorldClim (<https://doi.org/10.1002/joc.5086>) websites.

Funding: Funding was provided by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-SC0016410; awarded to JBHM; <https://www.energy.gov/science/ber/biological-and-environmental-research>). This work was supported by the Ridge to Reef NSF Research Traineeship,

Introduction

Bacteria are the most diverse organisms on the planet [1]. Bacterial richness and composition influences ecosystem functioning, whether in host-associated communities, soils, or oceans [2–7]. Nevertheless, we have yet to answer a number of basic questions about bacterial diversity, including “Which habitats contain the highest diversity of bacteria?” More broadly, evaluating geographic patterns in biodiversity across habitats and spatial scales can illuminate the processes influencing and consequences of biodiversity [8–12].

While many studies document spatial patterns of bacterial diversity, most are restricted to a particular geographic region or habitat, such as soil, sediment, or water [13–15]. To understand global trends, however, studies that analyze diversity across habitats and geographic regions are needed. Combining data from independent projects is oftentimes infeasible because community variation can be caused simply by differences in methodology. The Earth Microbiome Project (EMP) comprises 27,751 samples from 97 studies from a wide range of habitats and geographic regions that are processed in the exact same way [16]. Although there

award DGE-1735040 (<https://r2r.bio.uci.edu>) and the National Science Foundation Graduate Research Fellowship Program (to KEW; under Grant No. DGE-1321846). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

are limitations to PCR-based sequencing surveys [17], this dataset is unique for its size in using standardize methods for all samples. Thus, this dataset provides an opportunity for a rigorous comparison of bacterial diversity across many parts of the globe.

A recent overview from the EMP noted, as has previously been observed, that communities of free-living bacteria are more diverse than host-associated bacteria [18–20]. For example, soil and sediment samples have higher alpha-diversity than animal gut or skin microbiomes. We expand on this initial alpha-diversity analysis by additionally evaluating beta- and gamma-diversity across spatial scales while mitigating for unevenly spaced samples (Fig 1). We also take the opportunity to use the large size of the EMP dataset to test whether different diversity metrics and taxa definitions influence our understanding of microbial diversity patterns.

We specifically ask: which habitats support the highest levels of bacterial diversity? We consider three interrelated aspects of biodiversity: alpha-, beta-, and gamma-diversity. We measure alpha-diversity as the observed richness (number of taxa) or evenness (the relative abundances of those taxa) of an average sample within a habitat type. We quantify beta-diversity as the variability in community composition (the identity of taxa observed) among samples within a habitat [21]. Finally, we calculate gamma-diversity as the total observed richness of all samples within in a habitat.

We test several predictions about relative, not absolute, diversity patterns because, even in this large dataset, bacterial diversity remains undersampled. First, we predict that sediment and soil support the highest alpha-diversity within a single sample. These habitats are known to have relatively high bacterial diversity, although their relative rankings have not yet reached a consensus [16,18–20]. Second, we expect that soil, sediment, inland water, and biofilm/mat habitats will exhibit high beta-diversity. These habitats are spatially separated with less dispersal or mixing than air or marine water. Finally, we predict that soils and sediments will exhibit high gamma-diversity as they are expected to have both high alpha- and beta-diversity.

Within the soil habitat, we hypothesize that soils from biomes higher in plant diversity and productivity (e.g., forests and grasslands) support higher alpha-diversity than soils from biomes with low diversity and productivity (e.g., tundra and deserts) [22–24]. Of course, these biomes do not directly influence diversity, but they are defined based on abiotic factors [25], such as temperature or precipitation, that do influence diversity. Further, we expect that agricultural soils will exhibit lower beta-diversity than other biomes as common practices (pesticides, tilling, and fertilizer use) and the low diversity of crop plants influences community composition [26,27]. We also compare the relationship between diversity and biomes to those between diversity, and pH or temperature to assess whether plant diversity or abiotic conditions more strongly influence bacterial diversity. We expect bacterial richness to peak at neutral pH and moderate temperatures [14,28,29], and abiotic factors to be a stronger influence on diversity than plant biomes. Overall, the aim of this study was to compare bacterial diversity trends across habitats. We show that the most diverse habitat depends on the type of diversity (alpha-, beta-, or gamma-diversity).

Materials and methods

Bacterial 16S rDNA (V4 region) sequence data and associated metadata (e.g., sample location, sample type, date of sampling) were downloaded from the Earth Microbiome Project (EMP) on September 1, 2016. Sample processing, sequencing, and core amplicon data analysis were performed by the Earth Microbiome Project (www.earthmicrobiome.org), and all amplicon sequence data and metadata have been made public through the data portal (qiita.microbio.me/emp). Data available from: <https://doi.org/10.1038/nature24621> [16]. We used the EMP closed-reference (Greengenes 13.8) OTU dataset classified at 97% sequence similarity to reduce computational time (instead of the open-reference dataset). The dataset contains

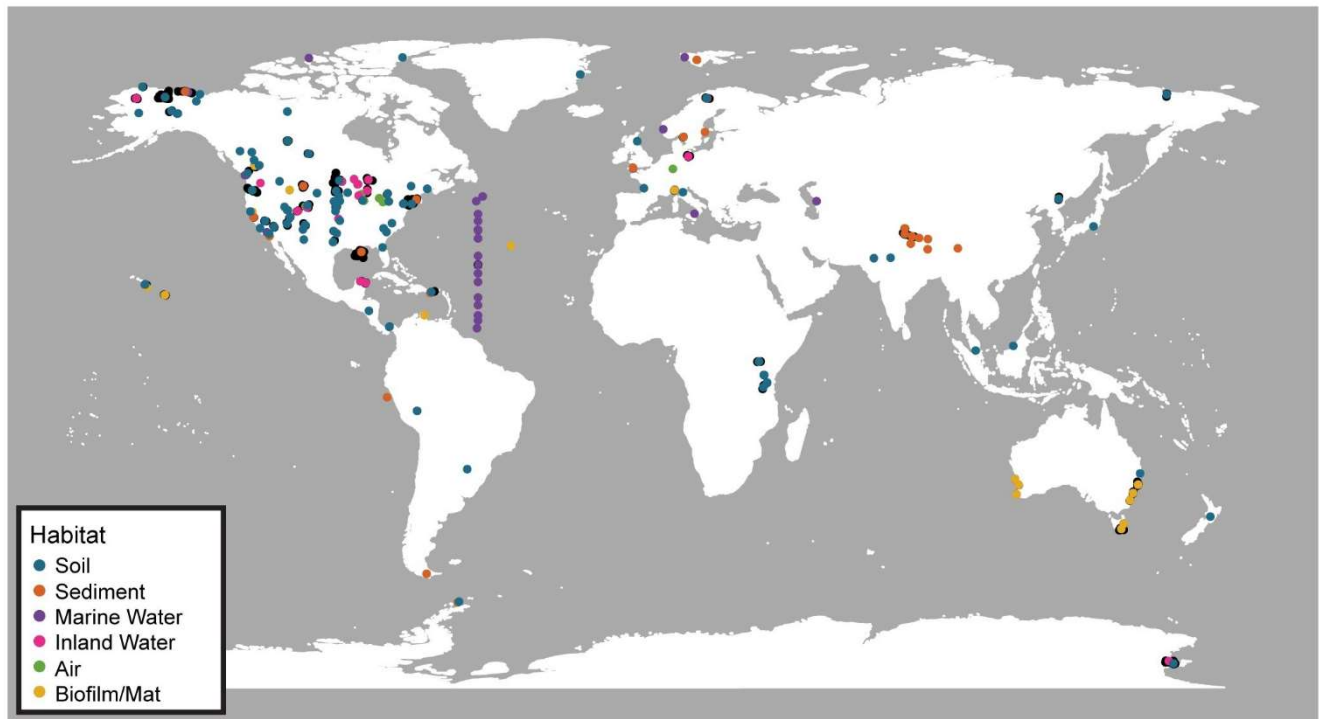


Fig 1. Sample and geocluster locations. Map showing locations of each of the EMP samples used in this study (black dots) and the geoclustered samples for each habitat (colored dots). Geoclusters were created from samples located within 110 km of each other.

<https://doi.org/10.1371/journal.pone.0233872.g001>

27,751 samples, with a median depth of 54,091 sequences per sample. We excluded archaea from the analysis because, relative to bacteria, they make up a small portion of any given community (median = 0.018% of sequences).

Habitat designations

We used the EMP Ontogeny metadata to classify the habitat and, for soil, biome of each sample based on the EMP metadata (Fig 2). When the existing metadata were unclear, we used the latitude and longitude coordinates to assess the environmental context. Samples with insufficient data about their location or habitat were removed from the analysis. Host-associated samples were also removed. Further, we only retained samples that could be classified into one of the following habitats: soil, sediment, marine water, inland water (e.g., rivers and lakes), air, and biofilms/mats. These habitat types were chosen because they represent a wide range of environmental conditions and are well sampled within the EMP dataset. Within the soil habitat, we further classified samples into forest, hot desert, cold desert, grassland, shrubland, tundra, and agricultural soil. We also classified inland water and sediment samples as saline and non-saline. See supplemental materials for descriptions of sample locations (S1 Appendix). After removing samples with less than 15,000 sequences (rarefaction depth in this study), 11,680 free-living (non-host system) samples remained.

Alpha-diversity analysis. To account for differences in sequencing depth, the samples were rarefied to 15,000 sequences with 1,000 resamplings in QIIME [30]. This rarefaction depth provided a high sequence count per sample while minimizing sample loss to 4.74% of samples. All samples with less than 15,000 sequences were removed leaving 11,680 free-living (non-host system) samples. For each resampling, we calculated 24 alpha-diversity metrics on

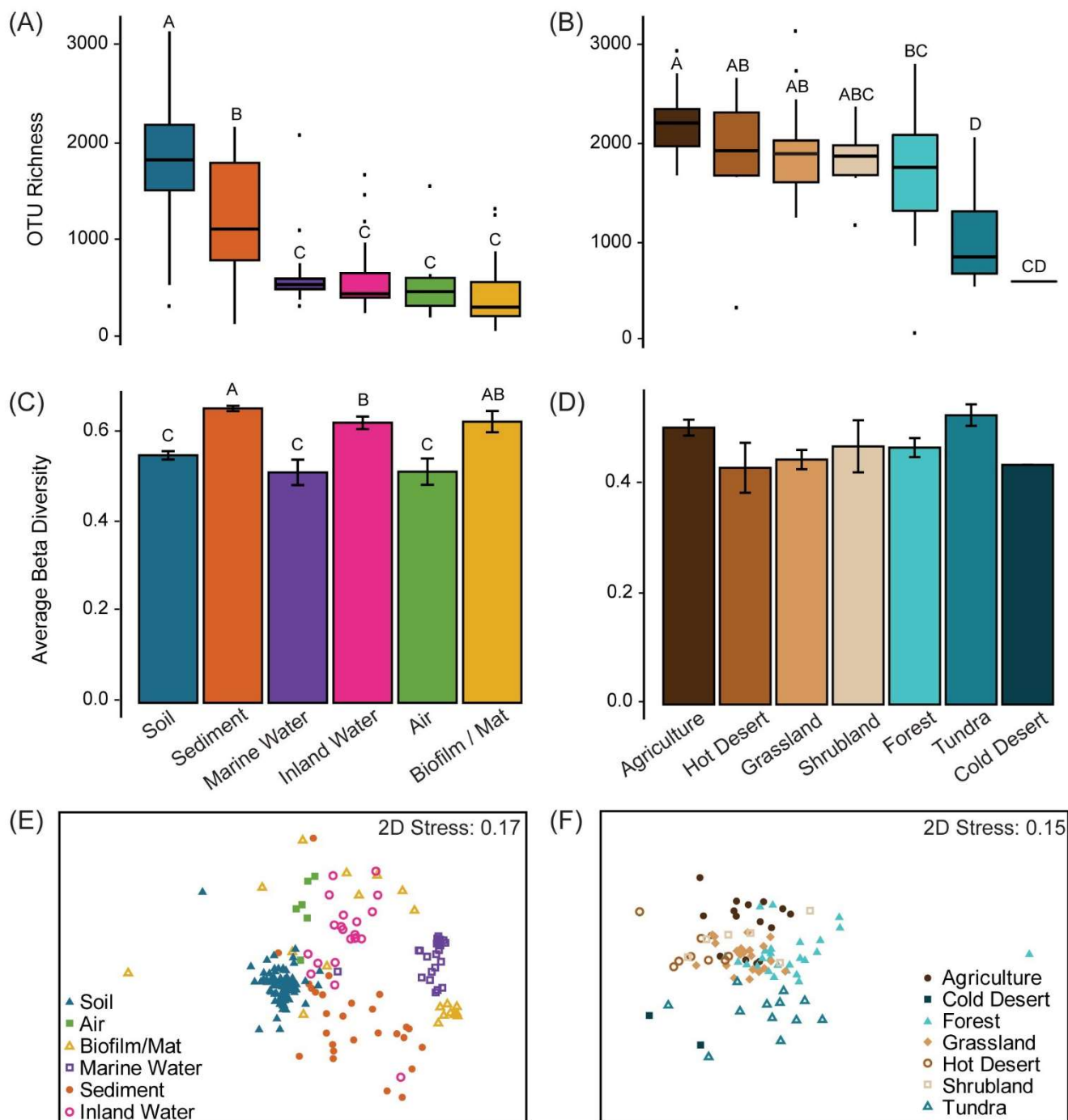


Fig 2. Alpha- and beta-diversity patterns. Alpha- and beta-diversity per habitat for all geoclusters used in study (A, C, and E) and per biome for soil geoclusters (B, D, and F). (A and B) Boxplot of alpha-diversity (OTU richness). (C and D) Mean beta-diversity (distance from centroid) $\pm$ standard error. For all bar and boxplots, letters above indicate significant differences among groups (Tukey test) where groups that share a letter are not significantly different from each other. (E and F) NMDS of geoclusters.

<https://doi.org/10.1371/journal.pone.0233872.g002>

the rarified OTU table in QIIME (Fig 3A). These metrics characterized the community in five general ways: observed richness, estimated richness, evenness/dominance, phylogenetic diversity, and coverage of sampling. We used all 24 metrics throughout the alpha-diversity analysis

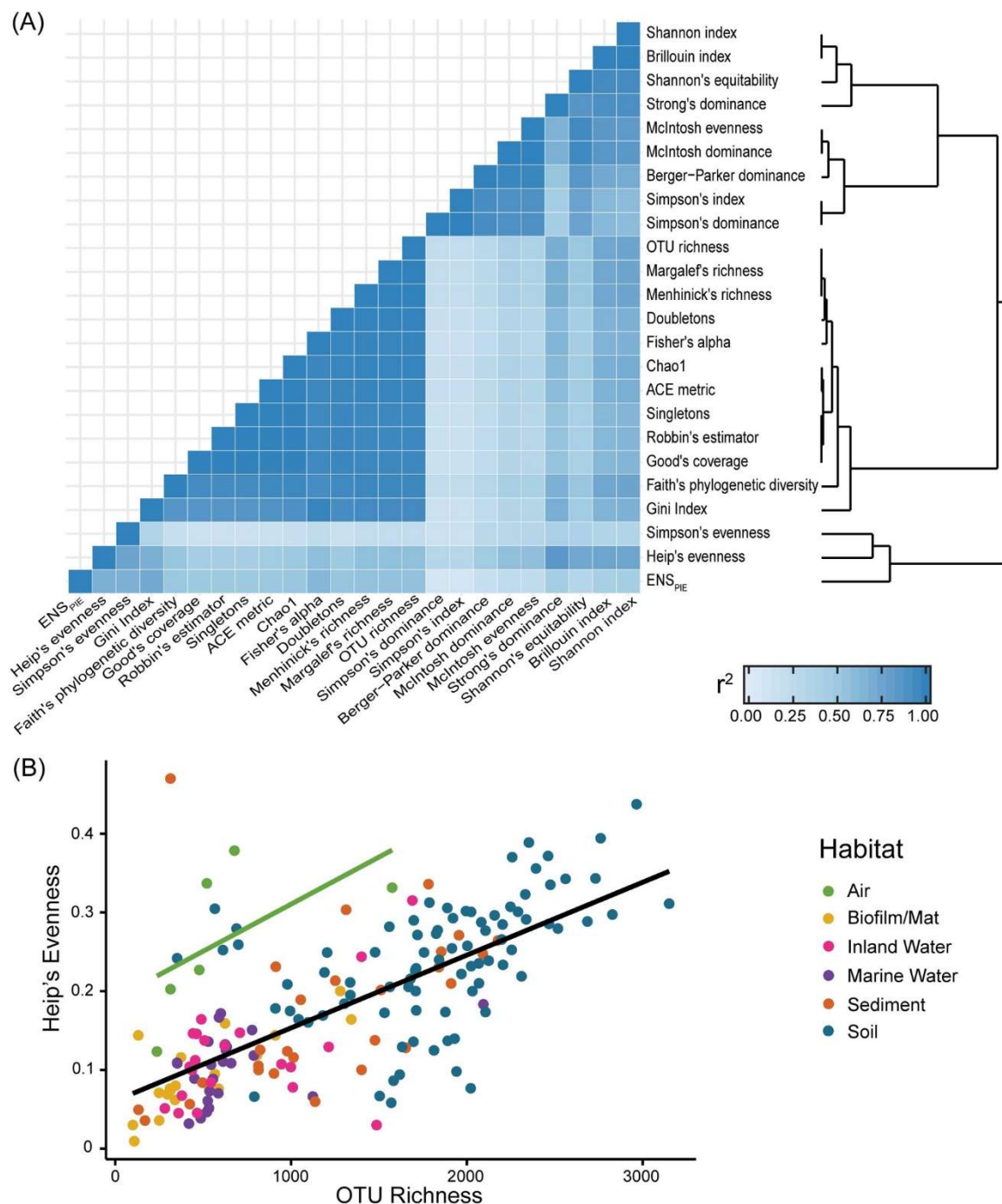


Fig 3. Comparison of diversity metrics. (A) Heatmap shows degree of correlation (r^2 from linear regression with all EMP samples used in analysis). Dendrogram shows relatedness of metrics based on their correlation strength. Note that the metrics are clustered into two groups: one composed of mainly evenness metrics (top cluster on dendrogram) and one composed of mainly richness metrics (bottom cluster on dendrogram). Simpson's evenness, Heip's evenness, and ENS_{pie} fall outside of those two clusters. (B) Dot plot showing relationship between Heip's evenness and OTU richness metrics for geoclusters of each habitat. The green line is a linear regression for air geoclusters, and black line is a linear regression for all geoclusters except for air.

<https://doi.org/10.1371/journal.pone.0233872.g003>

to ensure that our final conclusions were not dependent on the type of metric. We calculated the median value of each metric across the 1,000 replicates.

To minimize the effect of unevenly spaced samples, we averaged the alpha-diversity of the samples within a single geocluster. Many of the samples are highly clumped such that some geographic regions contribute unequally to the habitat's diversity. Geoclusters ($n = 172$) were formed by clustering samples of the same habitat type located within 110 km of each other (distance of 1° latitude at the equator) using `hclust()` and `cutree()` from package 'stats' and `rdist.earth()` from package 'fields' in R [31,32]. While a smaller clustering distance would have yielded a higher geocluster sample size, this conservative distance allowed us to be more confident that our results reflected ecological processes, rather than sampling locations. We calculated the median of each diversity metric for the samples within each geocluster of the same habitat type. The averaged alpha-diversities were then cube root transformed to achieve normality and homoscedasticity. Finally, we tested for significant differences in alpha-diversity among habitats by performing a one-way ANOVA and Tukey's HSD in R.

We tested whether the alpha-diversity results depended on the diversity metric by running a correlation with every pairwise combination of diversity metrics using all 11,680 samples. Likewise, we tested whether our results depended on the resolution of OTU clustering by comparing the 97% similarity OTU table with the single-nucleotide resolution 'sub-OTUs' dataset, Deblur, produced by the EMP. We rarefied the Deblur dataset to 15,000 sequences per samples (1,000 times), calculated Exact Sequence Variance (ESV) richness per sample, and calculated the mean richness across the 1,000 replicates. We then ran a correlation between the OTU richness and ESV richness for every sample present in both datasets ($n = 11,137$).

To further explore what factors might be driving alpha-diversity, we compared bacterial richness with pH and temperature at each sample site. We chose pH and temperature because these were the most widely included in the EMP dataset. For soil samples, however, the temperature data were often missing from the EMP dataset. Thus, for the analysis with just soil samples, we used temperature data from WorldClim. Metadata (pH and temperature) from the EMP were taken at the site at the time of sample collection. Data from WorldClim, a publicly-available dataset, included mean annual temperature averaged from 1970–2000 with spatial resolution of 10 minutes. Data is available from WorldClim Version2: <http://doi.org/10.1002/joc.5086> [33]. We assigned external temperature data to soil samples using latitude and longitude with `extract()` from package 'raster' in R [34]. Temperature and pH were correlated with OTU richness using a second-degree polynomial in R. We tested whether temperature and pH differed among habitats and biomes using an ANOVA in R.

Beta-diversity analysis. To reduce computation time, we used a subset (150 rarefied tables) of the 1000 rarefied OTU tables generated during the alpha-diversity analysis to analyze beta-diversity. The OTU tables were first square root transformed to increase weight given to the rare taxa [35] that make up the majority of microbial communities [36]. For each of the 150 rarefied, square root transformed OTU tables, we calculated the median abundance for each taxon across all the samples within a single geocluster. We then calculated a Bray-Curtis dissimilarity matrix for each of the 150 OTU-by-geocluster tables in QIIME. Finally, we calculated the median of the 150 dissimilarity matrices to yield one median Bray-Curtis dissimilarity matrix.

To visualize compositional differences among habitats, we used NMDS in PRIMER6 [37]. We also tested community composition differences among habitats using PERMANOVA in PERMANOVA+ [38]. Because we averaged OTU abundances for all samples of the same habitat type located within the same geographic area (geocluster), beta-diversity provides an approximation of the amount of community variation from location to location within one habitat (as opposed to variation from sample to sample within one location).

To compare beta-diversity across habitats, we analyzed the variance within each habitat using the function `PERMDISP` in PERMANOVA+. To determine if unequal sampling among

habitats biased these results, we re-calculated the Bray-Curtis values based on a selection of only 20 geoclusters for each habitat from a rarefied OTU-by-geocluster table. We chose 20 geoclusters because that depth included five of the six habitats (excluding air) but avoided the biases expected with sample sizes less than ten [38]. We repeated these subsamplings 100 times and tested for differences in beta-diversity among habitats using `betadisper()`, the PERMDISP test implemented in the R package ‘vegan’ [39]. We compared the relative rankings of these rarefied beta-diversity results to the unrarefied results to determine if rarefaction changed the relationships of variance among habitat groups. Specifically, we considered the rarefied results to match the unrarefied results if 95–100 subsampled tests were significant and showed the same beta-diversity rankings (based on mean distance to centroid) as the unrarefied test.

We tested whether the beta-diversity results depended on the diversity metric by running a correlation with every pairwise combination of nine diversity metrics. For each of the 150 OTU-by-geocluster tables, we calculated nine beta-diversity metrics using `vegdist()` from package ‘vegan’ in R [39]. We then took the mean matrix (of the 150 matrices) for each diversity metric. We performed a Spearman’s mantel test for every pairwise comparison of the nine averaged beta-diversity matrices using `mantel()` from package ‘stats’ [31]. To compare Raup-Crick to the other beta-diversity metrics, we calculated the dissimilarity within, but not among, habitats. Raup-Crick is not an appropriate metric when communities do not share the same species pool [40]. To calculate Raup-Crick, we took the mean of the matrices computed with `raupcrick(..., chase = TRUE)` and `raupcrick(..., chase = FALSE)` from package ‘vegan’ in R [39] to follow the method recommended by Chase et al. [40]. We calculated one Raup-Crick matrix for each habitat for each of the 150 OTU-by-geocluster tables, took the mean for each habitat across the 150 matrices, and then calculated the mean and SE distance from centroid in PRIMER6 [37]. We used the mean and SE to compare the trends in dissimilarity to those generated with the Bray-Curtis metric.

Gamma-diversity analysis. To assess gamma-diversity by habitat, we plotted an OTU accumulation curve for each habitat with `specaccum()` from package ‘vegan’ in R [39] using the 150 OTU-by-geocluster tables. The OTU accumulation curve displays the numbers of geoclusters sampled on the x-axis and observed OTU richness on the y-axis. This plot allowed us to compare cumulative diversity levels across multiple samples distributed across the world. We examined whether habitats likely exhibit different gamma-diversity levels by calculating error bars equal to 1.96 times the standard deviation.

Results

Alpha-diversity

Out of the six habitats compared, soils contained the highest observed richness (i.e., number of observed taxa rarefied at 15,000 sequences) for a single sample, with a median of 1,842 taxa (97% OTUs) per sample given this depth of sequencing (one-way ANOVA: $F = 39.13$, $P < 0.001$, $r^2 = 0.541$; Fig 2A). Sediments were the second most diverse habitat with an average of 1,137 taxa. Marine water, air, inland water, and biofilms/mats had a significantly lower richness (averaging 571, 500, 478, and 342 taxa, respectively) than soils and sediments ($P < 0.001$; Fig 2A) but could not be distinguished from one another by richness.

Because salinity influences bacterial community composition [18], we further tested whether taxon richness varied between non-saline and saline habitats. We found that salinity had no impact on alpha-diversity for sediments (one-way ANOVA: $F = 0.433$, $P = 0.516$; S1 Fig) or inland water ($F = 0.093$, $P = 0.763$; S1 Fig).

Within the soil habitat, we further compared alpha-diversity among seven biomes (agricultural, grassland, shrubland, forest, hot desert, cold desert, and tundra soil). Within soil

samples, richness differed significantly among biomes. Agricultural soils supported the highest richness in a sample, along with hot desert, grassland, and shrubland biomes. Forest soils were less diverse than agricultural soil, and tundra and cold deserts supported the lowest richness (one-way ANOVA: $F = 42.62$, $P < 0.001$, $r^2 = 0.642$; Fig 2B). Notably, the cold desert biome was only represented by two geoclusters (averaging across 117 samples); thus, more data are needed to assess that particular biome's diversity.

The above results were robust to the alpha-diversity metric used. On a sample by sample basis, 24 alpha-diversity indices, including observed richness, were all correlated with each other ($r^2 = 0.09$ – 1.00 , $P < 0.0001$) with a mean r^2 of 0.63 (Fig 3A). The metrics grouped into two main clusters. One cluster encompassed the richness/coverage metrics such as OTU richness, Faith's Phylogenetic Diversity, and Chao1 ($r^2 = 0.86$ – 1.00 , mean $r^2 = 0.97$). The other cluster included the evenness/dominance metrics such as Simpson's and McIntosh dominance index ($r^2 = 0.31$ – 1.00 , mean $r^2 = 0.85$). Further, each metric ranked the habitats from highest to lowest alpha-diversity in the same way, with the exception of air. Air communities were more even than other habitats, given their relative richness level (Fig 3B). Excluding air samples, OTU richness (ANCOVA: $F = 54.68$, $P < 0.0001$, $r^2 = 0.229$), but not habitat ($F = 2.14$, $P = 0.0775$) was a predictor of evenness. When air samples were included, both OTU richness ($F = 57.09$, $P < 0.0001$, $r^2 = 0.173$) and habitat ($F = 7.0989$, $P < 0.0001$, $r^2 = 0.107$) were significant predictors of evenness.

The alpha-diversity patterns were also robust to the taxonomic binning method (Fig 4). We compared the 97% OTU and Exact Sequence Variant (ESV) Deblur datasets provided by the EMP. Not only were OTU richness and ESV richness strongly correlated ($r^2 = 0.933$, $P < 0.0001$), but, on a sample by sample basis, they were also nearly identical (slope = 0.953, $P < 0.0001$). However, the relationship between OTU richness and ESV richness varied among habitats (ANOVA: $F = 786.5$, $P < 0.0001$, $r^2 = 0.331$). In particular, non-saline sediments and inland water demonstrated a higher ESV:OTU richness ratio than other habitats (Fig 4).

Taxon richness displayed a weak hump-shaped relationship with pH and a peak in diversity at a neutral pH (non-linear regression: $P < 0.0001$, $r^2 = 0.047$; S2A Fig). In contrast, taxon richness only weakly correlated with temperature, and this relationship was driven by low-diversity biofilm/mat samples sampled from high temperatures ($P < 0.0001$, $r^2 = 0.036$; S2C Fig). Overall, the bacterial alpha-diversity patterns across all habitats were not obviously related to pH or temperature. Most of the samples were, on average, at a neutral pH, with the soil samples more acidic and the biofilm/mat samples more basic (S2B Fig). Despite this, temperature (ANOVA: $F = 272.3$, $P < 0.0001$, $r^2 = 0.198$) and pH ($F = 245.2$, $P < 0.0001$, $r^2 = 0.215$) differed significantly among habitats (S2D Fig).

Similar to the pattern observed across all habitats, richness within just the soil samples also peaked at a neutral pH (non-linear regression: $P = 0.001$, $r^2 = 0.121$; S3A Fig). In contrast, richness in soils also peaked at a temperature around 10°C ($P < 0.0001$, $r^2 = 0.288$; S3C Fig). Both temperature (ANOVA: $F = 640.2$, $P < 0.0001$, $r^2 = 0.153$) and pH ($F = 199.8$, $P < 0.0001$, $r^2 = 0.594$) differed among soil samples by biome (S3B and S3D Fig). Soils from both hot and cold deserts tended to be basic while agricultural fields, forests, and tundra were acidic. Tundra and cold deserts were the coldest biomes, and shrubland, agriculture, and hot deserts were among the hottest biomes.

Beta-diversity

Sediment, biofilm/mat, and inland water habitats displayed the highest beta-diversity among geographic locations or geoclusters (not within a single sample), whereas soil, air, and marine

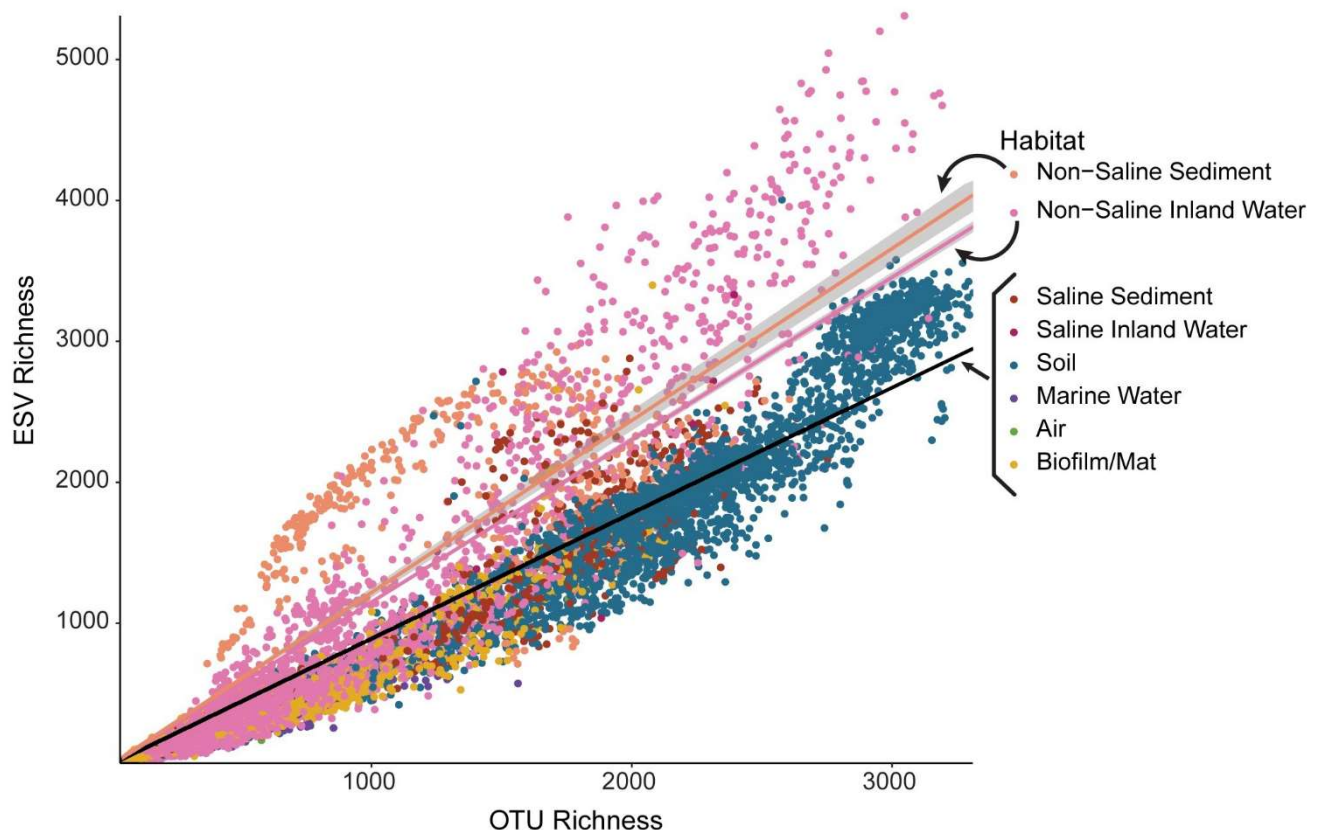


Fig 4. Comparison of OTU and ESV richness. On a sample-by-sample basis, OTU richness and ESV richness are highly correlated ($r^2 = 0.933$, $P < 0.0001$). On average, for any given sample, ESV richness is equal to 95.26% of OTU richness. However, not all habitats showed the exact same relationship between OTU and ESV richness. Non-saline sediment and inland water samples have significantly higher ESV richness given their relative OTU richness. All samples except for non-saline sediment and inland water have a regression line with a slope of 0.891 ($P < 0.0001$, $r^2 = 0.971$), shown as the black regression line on the graph. Non-saline sediment samples have a regression line with a slope of 1.220 ($P < 0.0001$, $r^2 = 0.895$), shown as light orange on the graph. Non-saline inland water samples have a regression line with a slope of 1.152 ($P < 0.0001$, $r^2 = 0.902$), shown as light pink on the graph.

<https://doi.org/10.1371/journal.pone.0233872.g004>

water exhibited 17% lower beta-diversity (PERMDISP: $F = 10.7$, $P = 0.001$; Fig 2C). To test that these patterns were not influenced by unequal sampling (number of geoclusters) of the habitats, we subsampled the habitats (to 20 geoclusters per habitat) and retested the patterns. All 100 subsamplings produced the same beta-diversity rankings, and all models were significant, indicating that unequal sampling did not influence within-habitat beta-diversity. Within the soil habitat, beta-diversity did not differ by biome ($P = 0.526$; Fig 2D).

Overall, bacterial community composition differed significantly by habitat (PERMANOVA: $P = 0.001$, Pseudo- $F = 9.8601$, $r^2 = 0.210$; Fig 2E) and by biome for soils ($P = 0.001$, Pseudo- $F = 4.221$, $r^2 = 0.227$; Fig 2F). Because salinity influenced the community composition for both sediments ($P = 0.002$, Pseudo- $F = 2.0168$, $r^2 = 0.075$) and inland water ($P = 0.02$, Pseudo- $F = 1.7702$, $r^2 = 0.085$), we tested whether salinity likewise influenced beta-diversity within these habitats. Beta-diversity did not differ between saline and non-saline samples within sediments (PERMDISP: $P = 0.21$, $F = 2.4404$) or inland water ($P = 0.849$, $F = 0.27406$; S4 Fig).

The above results did not depend on the beta-diversity metric used. On a sample by sample basis, nine beta-diversity indices, including Bray-Curtis, were correlated with each other ($r = 0.435$ – 1.00 , $P = 0.001$, mean $r = 0.88$; S5A Fig). Raup-Crick has been suggested as a more appropriate metric when comparing groups with different alpha-diversity levels [40]. Because

Raup-Crick assumes that all communities are part of the same regional species pool, we calculated the mean and standard error within each habitat (excluding between habitat comparisons) and compared the trend to that generated by the Bray-Curtis metric. Both Bray-Curtis and Raup-Crick metrics showed the same trend of beta-diversity among habitats (S5B Fig).

Gamma-diversity

Considering the accumulation of taxon richness across geoclusters, sediments exhibited the highest gamma-diversity of any habitat, followed by soils and inland water (Fig 5). The sediment rarefaction curve showed little sign of flattening out, indicating that most taxa are yet to be sampled. In contrast, the soil curve noticeably leveled off, even at a similar level of sampling. The gamma-diversity of marine water, biofilms/mats, and air were not statistically distinguishable from one another but, as a group, exhibited lower gamma-diversity than inland water, soils, and sediments.

Discussion

Here, we tested which habitat contains the most bacterial taxa within a single sample (alpha-diversity), which exhibits the most variation among samples (beta-diversity), and which contains the most taxa across all samples (gamma-diversity). We show that a single sample of soil on average contained higher bacterial alpha-diversity than any other habitat, including sediment (Fig 2A). However, sediment had higher gamma-diversity, with much of its diversity yet to be sampled (Fig 5). Within soils, we found that agricultural soils had among the highest richness and exhibited just as much compositional variation (beta-diversity) as other biomes.

Although both sediments and soil were previously known to be highly diverse microbial habitats, previous studies demonstrated conflicting results about their relative ranking [16,18–20]. Using 11,680 samples and minimizing geographic biases, this analysis suggests that soil contains higher alpha-diversity than sediments. In contrast, marine water, inland water, air, and biofilms/mats contain the lowest alpha-diversity (Fig 2A).

Within-habitat heterogeneity is a known driver of plant and animal diversity [41,42] and has been correlated with microbial communities as well [43,44]. Here, we show that the alpha-diversity patterns are consistent with the idea that habitat heterogeneity may drive bacterial diversity at a single sample. The highly mixed water and air environments harbor lower diversity, consistent with previous smaller-scale studies [19,45,46]. While both sediments and soil are not as well mixed, sediments contain higher water content than soils. Water content increases connectivity and thus reduces environmental heterogeneity and promotes dispersal, both of which can result in lower diversity [9,47–49]. At the same time, biofilms and mats, despite being spatially structured, also displayed low alpha-diversity [50]. However, the biofilm/mat samples from the EMP dataset encompassed samples with the highest pH and temperature (S2 Fig). We therefore speculate that these abiotic extremes contribute to low alpha-diversity [51–53]. In fact, diversity in many habitats is lowest at extreme temperatures [14,28,29,51–55]. Yet ultimately, little is known about the environmental conditions, and their heterogeneity, at the spatial scale that matters for microorganisms [56]. To test the importance of within-sample heterogeneity on microbial diversity directly, finer-scale data are needed.

Our analysis is the first to quantify bacterial beta-diversity among habitats across many parts of the globe. While soils contained the highest alpha-diversity within a single sample, sediments displayed higher beta-diversity among geoclustered samples within a habitat (Fig 2C). Sediment beta-diversity was also similar to that of inland water and biofilms/mats. Additional environmental data associated with the individual samples would be needed to distinguish

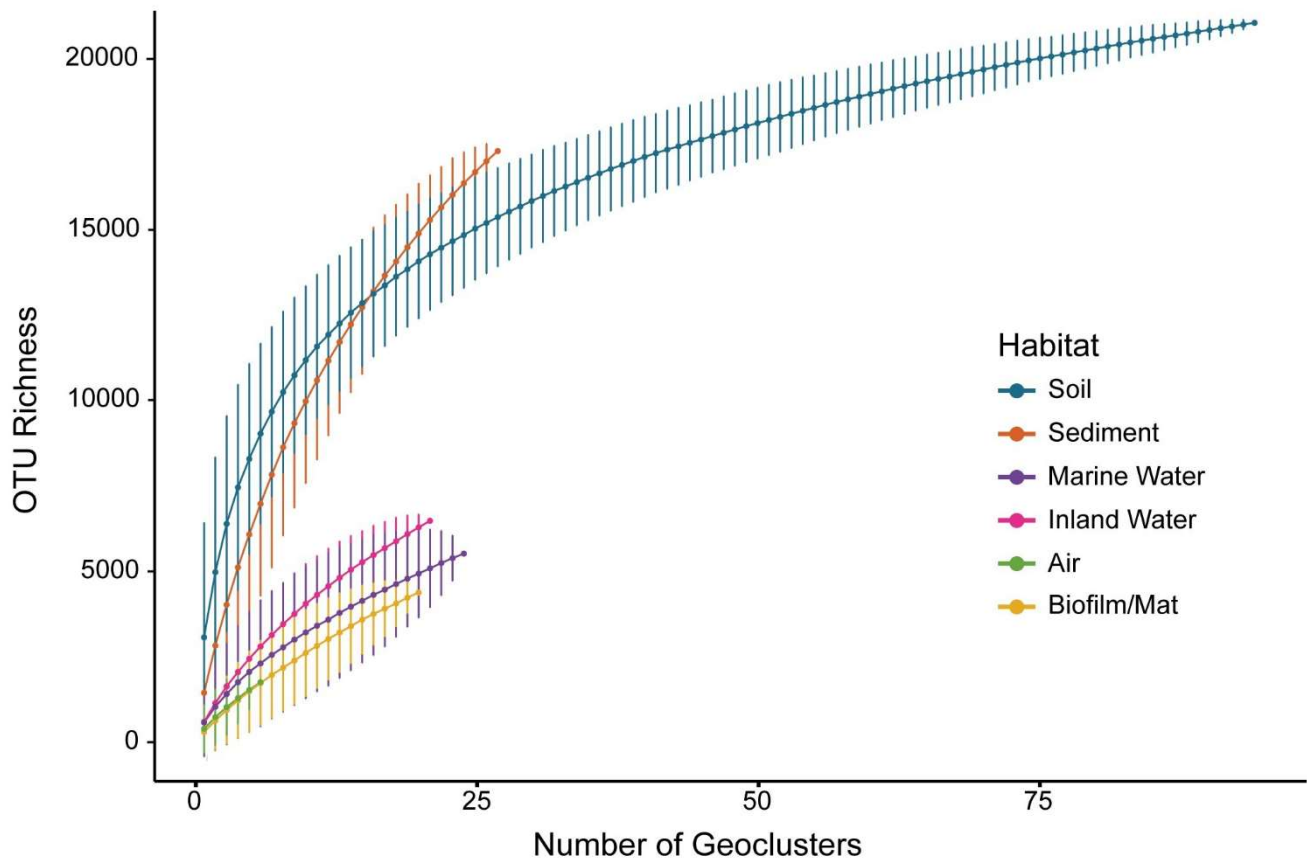


Fig 5. Geocluster accumulation curves. Geocluster accumulation curves (gamma-diversity) for mean OTU richness from a random sampling of geoclusters (permutations = 999) with 95% confidence intervals drawn for each habitat.

<https://doi.org/10.1371/journal.pone.0233872.g005>

whether these beta-diversity patterns might be driven by dispersal limitation [9,57] or spatial variation in environmental conditions [6,45].

Given their high alpha- and beta-diversity, it is not surprising that sediments are also estimated to contain the highest gamma-diversity (Fig 5). While extracellular DNA (eDNA) may be particularly prevalent in ocean sediments [58], evidence thus far suggests that eDNA has minimal effect on sediment diversity estimates from sequencing surveys [59]. The taxa accumulation curves also suggest that, while we may have observed most bacterial taxa in soil (at least from highly sampled continents), there is much more diversity to discover in sediments. Similarly, other than soil, the accumulation curves suggest that air, water habitats, and biofilms/mats remain undersampled as well. Of course, the number and localities of samples available will influence the diversity estimates. Thus, a limitation to these conclusions is that the samples are highly concentrated in North America and Europe (Fig 1), and continued sampling is needed to test the robustness of these diversity patterns.

Because plant diversity and productivity are shown to impact microbial communities [23,24], we further characterized alpha- and beta-diversity trends among biomes from which the soil samples were collected. Agricultural soils contained among the highest alpha-diversity, as previously noted in smaller scale studies [27,60]. Indeed, some agricultural practices, such as application of manure, are known to increase bacterial diversity [61,62]. Even more notable, however, is that agricultural soils encompassed similar levels of beta-diversity to those of other biomes. While some agricultural practices have been shown to homogenize communities

within a single field [60], not all practices have a homogenizing effect [63]. Further, the diversity of agricultural practices around the world [61,64] seems to select for as much variation in bacterial composition (beta-diversity) as different types forests or deserts.

Contrary to our hypothesis, biomes with higher plant diversity or productivity, such as forest or shrubland soils, were no more diverse within a sample than other biomes (Fig 2B). These results support previous findings that grasslands contain more bacterial diversity than forests [65] and, overall, plant and soil diversity are uncoupled [66]. We therefore propose that abiotic factors may be more important for soil bacterial alpha-diversity than plant biomes. Biomes differed significantly in pH and temperature, and soil alpha-diversity was strongly correlated with both factors (S3 Fig). These results are consistent with previous studies [14,28,29] that find bacterial richness in soils peaks at a neutral pH and at mid-temperatures (around 10°C). Of course, other unmeasured environmental factors and/or ecological interactions are likely influencing soil diversity.

Finally, these diversity patterns appear to be robust to two key methodological issues. First, diversity trends did not depend on the particular alpha- or beta-diversity metrics used (Figs 3A and S5A). Air, as the only exception, was more even than expected, given its richness (Fig 3B). We speculate that the movement of air contributes to its evenness as air likely picks up a sampling of bacteria from many different habitats [67,68]. Second, the results were robust to the degree of clustering of the amplicon sequences (Fig 4). Both the 97% OTU and the ESV datasets yielded the same alpha-diversity trends, as previously noted in a smaller scale study [69]. Most of those outliers in this analysis came from non-saline sediment or inland water samples, which had higher ESV richness than 97% OTU richness. While this pattern could suggest higher finer-scale diversity within these habitats, we caution that these samples originated from only four geoclusters. Overall, while ESVs can be useful for resolving finer diversity among specific taxonomic groups [70], broad-scale alpha-diversity patterns do not seem to be altered by these particular operational definitions.

With the largest dataset created with consistent methodology and a geographically wide-spread sampling effort, we show that soils support the highest diversity within a single sample (alpha-diversity) and that sediments are more variable in composition among locations (beta-diversity) and likely support the most bacterial taxa at a larger spatial scale (gamma-diversity). Within soils, we find biome type impacts soil alpha-diversity but not beta-diversity. Many of these results appear consistent with the idea that spatial heterogeneity and dispersal limitation promote bacterial diversity. These baseline patterns set the stage for new research on the mechanisms driving the generation and maintenance of bacterial diversity.

Supporting information

S1 Fig. The influence of salinity on alpha-diversity. Taxon richness did not differ between saline and non-saline samples from inland water or sediment habitats. (TIF)

S2 Fig. The influence of abiotic factors on taxon richness. (A) pH significantly impacts taxon richness ($P < 0.0001$). Each point represents an individual EMP sample (not a geocluster) and is colored by habitat. (B) pH differs among habitats (ANOVA, $p < 0.0001$). Out of all the EMP environmental samples used in this study, 30.7% had associated pH metadata: 0% of air samples, 60.0% of inland water samples, 19.6% of sediment samples, 6.5% of marine water samples, 5.1% of biofilm/mat samples, and 23.4% of soil samples. (C) Temperature significantly influences taxon richness ($P < 0.0001$). Each point represents an individual EMP sample (not a geocluster) and is colored by habitat. (D) Temperature recorded when samples were collected (EMP metadata) differs among habitats (ANOVA, $p < 0.0001$). Out of all the EMP

environmental samples used in this study, 47.3% had associated temperature metadata: 9.0% of air samples, 77.0% of inland water samples, 17.5% of sediment samples, 48.8% of marine water samples, 81.0% of biofilm/mat samples, and 0.6% of soil samples.

(TIF)

S3 Fig. Abiotic factors influence taxon richness in soil. (A) pH significantly impacts taxon richness ($P = 0.001$). Each point represents an individual EMP soil sample (not a geocluster) and is colored by biome. (B) pH differs among biomes (ANOVA, $p < 0.0001$). (C) Mean annual temperature significantly impacts taxon richness ($P < 0.0001$). Each point represents an individual EMP soil sample (not a geocluster) and is colored by biome. Temperature data was retrieved from WorldClim, a publicly available data source. (D) Temperature differs among biomes (ANOVA, $p < 0.0001$).

(TIF)

S4 Fig. The influence of salinity on beta-diversity. The level of beta-diversity within sediments and inland water is not driven by the combination of saline and non-saline samples within a single habitat. (A) Mean beta-diversity (distance from centroid) $\pm$ standard error of the six habitats, ranking habitats from highest to lowest beta-diversity. (B) Mean beta-diversity $\pm$ standard error of the six habitats with sediment and inland water habitats split into saline and non-saline samples.

(TIF)

S5 Fig. Comparison of beta-diversity metrics. Beta-diversity patterns are not dependent on the beta-diversity metric used. (A) Heatmap shows degree of correlation (r from a Spearman's mantel test with all EMP samples used in analysis). Dendrogram shows relatedness of metrics based on their correlation strength. (B) Mean Raup-Crick dissimilarity (distance from centroid) $\pm$ standard error. The patterns shown by Raup-Crick analysis match those demonstrated by Bray-Curtis metric.

(TIF)

S1 Appendix. Sample locations.

(DOCX)

Acknowledgments

We would like to thank Luke Thompson, Jack Gilbert, and the Earth Microbiome Project team for their vision and hard work in assembling this dataset and making it widely available. We also thank Steve Allison and Cascade Sorte for their input on analyses and methods, and Michaeline Albright and Alex Chase for help with statistical analyses and computational methods. We further thank Alex Chase, Cynthia Rodriguez, Sarai Finks, Claudia Weihe, Joia Capocchi, Kazuo Isobe, and Pauline Nguyen for their guidance with earlier drafts.

Author Contributions

Conceptualization: Kendra E. Walters, Jennifer B. H. Martiny.

Data curation: Kendra E. Walters.

Formal analysis: Kendra E. Walters.

Funding acquisition: Jennifer B. H. Martiny.

Methodology: Kendra E. Walters, Jennifer B. H. Martiny.

Project administration: Jennifer B. H. Martiny.

Resources: Jennifer B. H. Martiny.

Supervision: Jennifer B. H. Martiny.

Validation: Kendra E. Walters, Jennifer B. H. Martiny.

Visualization: Kendra E. Walters.

Writing – original draft: Kendra E. Walters, Jennifer B. H. Martiny.

References

- Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. *Proc Natl Acad Sci*. 1998 Jun 9; 95(12):6578–83. <https://doi.org/10.1073/pnas.95.12.6578> PMID: 9618454
- Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. The contribution of species richness and composition to bacterial services. *Nature*. 2005; 436(7054):1157–60. <https://doi.org/10.1038/nature03891> PMID: 16121181
- Martiny JBH, Martiny AC, Weihe C, Lu Y, Berlemont R, Brodie EL, et al. Microbial legacies alter decomposition in response to simulated global change. *ISME J*. 2017; 11(2):1–10.
- Wagg C, Bender SF, Widmer F, van der Heijden MGA. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc Natl Acad Sci*. 2014 Apr 8; 111(14):5266–70. <https://doi.org/10.1073/pnas.1320054111> PMID: 24639507
- Strickland MS, Lauber C, Fierer N, Bradford MA. Testing the functional significance of microbial community composition. *Ecology*. 2009 Feb 1; 90(2):441–51. <https://doi.org/10.1890/08-0296.1> PMID: 19323228
- Barberán A, Casamayor E. Global phylogenetic community structure and β -diversity patterns in surface bacterioplankton metacommunities. *Aquat Microb Ecol*. 2010 Mar 11; 59(1):1–10.
- Philippot L, Spor A, Hénault C, Bru D, Bizouard F, Jones CM, et al. Loss in microbial diversity affects nitrogen cycling in soil. *ISME J*. 2013; 7(8):1609–19. <https://doi.org/10.1038/ismej.2013.34> PMID: 23466702
- Mittelbach GG, Schemske DW, Cornell H V., Allen AP, Brown JM, Bush MB, et al. Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. Vol. 10, *Ecology Letters*. John Wiley & Sons, Ltd (10.1111); 2007. p. 315–31. <https://doi.org/10.1111/j.1461-0248.2007.01020.x> PMID: 17355570
- Carson JK, Gonzalez-Quir6nes V, Murphy D V, Hinz C, Shaw JA, Gleeson DB. Low pore connectivity increases bacterial diversity in soil. *Appl Environ Microbiol*. 2010 Jun 15; 76(12):3936–42. <https://doi.org/10.1128/AEM.03085-09> PMID: 20418420
- Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, Green JL. Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci*. 2008; 105(Supplement 1):11505–11.
- Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberød AK, Schmidt TSB, et al. Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome*. 2020 Apr 20; 8(1):1–17. <https://doi.org/10.1186/s40168-019-0777-4> PMID: 31901242
- Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, et al. Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell*. 2019; 179(5):1084–97. <https://doi.org/10.1016/j.cell.2019.10.008> PMID: 31730851
- Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci*. 2012 Dec 26; 109(52):21390–5. <https://doi.org/10.1073/pnas.1215210110> PMID: 23236140
- Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A*. 2006; 103(3):626–31. <https://doi.org/10.1073/pnas.0507535103> PMID: 16407148
- Neufeld JD, Mohn WW. Unexpectedly High Bacterial Diversity in Arctic Tundra Relative to Boreal Forest Soils, Revealed by Serial Analysis of Ribosomal Sequence Tags Unexpectedly High Bacterial Diversity in Arctic Tundra Relative to Boreal Forest Soils, Revealed by Serial Ana. *Appl Environ Microbiol*. 2005; 71(10):5710–8. <https://doi.org/10.1128/AEM.71.10.5710-5718.2005> PMID: 16204479
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017; 551(7681):457–63. <https://doi.org/10.1038/nature24621> PMID: 29088705
- Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol*. 2015 Jun 4; 75(2):129–37.

18. Lozupone CA, Knight R. Global patterns in bacterial diversity. *Proc Natl Acad Sci*. 2007 Jul 3; 104(27):11436–40. <https://doi.org/10.1073/pnas.0611525104> PMID: 17592124
19. Torsvik V, Øvreås L, Thingstad TF. Prokaryotic diversity—magnitude, dynamics, and controlling factors. *Science*. 2002 May 10; 296(5570):1064–6. <https://doi.org/10.1126/science.1071698> PMID: 12004116
20. Fierer N, Lennon JT. The generation and maintenance of diversity in microbial communities. *Am J Bot*. 2011; 98(3):439–48. <https://doi.org/10.3732/ajb.1000498> PMID: 21613137
21. Anderson MJ, Ellingsen KE, McArdle BH. Multivariate dispersion as a measure of beta diversity. *Ecol Lett*. 2006 Jun; 9(6):683–93. <https://doi.org/10.1111/j.1461-0248.2006.00926.x> PMID: 16706913
22. Kier G, Mutke J, Dinerstein E, Ricketts TH, Küper W, Kreft H, et al. Global patterns of plant diversity and floristic knowledge. *J Biogeogr*. 2005 Jun 2; 32(7):1107–16.
23. Zak DR, Holmes WE, White DC, Peacock AD, Tilman D. Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology*. 2003 Aug 1; 84(8):2042–50.
24. Lamb EG, Kennedy N, Siciliano SD. Effects of plant species richness and evenness on soil microbial community diversity and function. *Plant Soil*. 2011 Jan 4; 338(1):483–95.
25. Holdridge LR. Life zone ecology. Life zone ecology. Tropical Science Center; 1967.
26. Jangid K, Williams MA, Franzluebbers AJ, Sanderlin JS, Reeves JH, Jenkins MB, et al. Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biol Biochem*. 2008 Nov 1; 40(11):2843–53.
27. Upchurch R, Chi CY, Everett K, Dyszynski G, Coleman DC, Whitman WB. Differences in the composition and diversity of bacterial communities from agricultural and forest soils. *Soil Biol Biochem*. 2008; 40(6):1294–305.
28. Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. The bacterial biogeography of British soils. *Environ Microbiol*. 2011 Jun 1; 13(6):1642–54. <https://doi.org/10.1111/j.1462-2920.2011.02480.x> PMID: 21507180
29. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, et al. Structure and function of the global topsoil microbiome. Vol. 560, *Nature*. Nature Publishing Group; 2018. p. 233–7. <https://doi.org/10.1038/s41586-018-0386-6> PMID: 30069051
30. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010 May 11; 7(5):335–6. <https://doi.org/10.1038/nmeth.f.303> PMID: 20383131
31. R Core T. A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2017.
32. Nychka D, Furrer R, Paige J, Sain S. fields: Tools for spatial data. 2015.
33. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017 Oct 1; 37(12):4302–15.
34. Hijmans RJ. raster: Geographic Data Analysis and Modeling. 2018.
35. Magurran A. Ecological diversity and its measurement. Princeton university press; 1988.
36. Martiny JBH, Walters KE. Towards a Natural History of Soil Bacterial Communities. Vol. 26, *Trends in Microbiology*. Elsevier Ltd; 2018. p. 250–2. <https://doi.org/10.1016/j.tim.2018.02.010> PMID: 29523393
37. Clarke KR, Gorley RN. Primer v6: User Manual/Tutorial. 2006.
38. Anderson MJ, Gorley RN, Clarke KR. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. 2008.
39. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. In: R package version 25–1. 2018.
40. Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD. Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere*. 2011 Feb 1; 2(2):art24.
41. Tews J, Brose U, Grimm V, Tielbörger K, Wichmann MC, Schwager M, et al. Animal species diversity driven by habitat heterogeneity/diversity: The importance of keystone structures. Vol. 31, *Journal of Biogeography*. Wiley/Blackwell (10.1111); 2004. p. 79–92.
42. Johnson MP, Simberloff DS. Environmental determinants of island species numbers in the british Isles. *J Biogeogr*. 1974; 1(3):149–54.
43. Curd EE, Martiny JBH, Li H, Smith TB. Bacterial diversity is positively correlated with soil heterogeneity. *Ecosphere*. 2018;
44. Cheeke TE, Schütte UM, Hemmerich CM, Cruzan MB, Rosenstiel TN, Bever JD. Spatial soil heterogeneity has a greater effect on symbiotic arbuscular mycorrhizal fungal communities and plant growth than genetic modification with *Bacillus thuringiensis* toxin genes. *Mol Ecol*. 2015;

45. Asakawa S, Kimura M. Comparison of bacterial community structures at main habitats in paddy field ecosystem based on DGGE analysis. *Soil Biol Biochem*. 2008 Jun 1; 40(6):1322–9.
46. Crump BC, Amaral-Zettler LA, Kling GW. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J*. 2012 Sep 1; 6(9):1629–39. <https://doi.org/10.1038/ismej.2012.9> PMID: 22378536
47. Freestone AL, Inouye BD. Dispersal Limitation and Environmental Heterogeneity Shape Scale-Dependent Diversity Patterns in Plant Communities. *Ecology*. 2006 Oct 1; 87(10):2425–32. [https://doi.org/10.1890/0012-9658\(2006\)87\[2425:dlaehs\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[2425:dlaehs]2.0.co;2) PMID: 17089651
48. Bickel S, Chen X, Papritz A, Or D. A hierarchy of environmental covariates control the global biogeography of soil bacterial richness. *Sci Rep*. 2019 Dec 1; 9(1):1–10. <https://doi.org/10.1038/s41598-018-37186-2> PMID: 30626917
49. Vellend M. Conceptual Synthesis in Community Ecology. *Q Rev Biol*. 2010; 85(2):183–206. <https://doi.org/10.1086/652373> PMID: 20565040
50. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: From the natural environment to infectious diseases. Vol. 2, *Nature Reviews Microbiology*. Nature Publishing Group; 2004. p. 95–108. <https://doi.org/10.1038/nrmicro821> PMID: 15040259
51. Miller SR, Strong AL, Jones KL, Ungerer MC. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl Environ Microbiol*. 2009 Jul 1; 75(13):4565–72. <https://doi.org/10.1128/AEM.02792-08> PMID: 19429553
52. Sharp CE, Brady AL, Sharp GH, Grasby SE, Stott MB, Dunfield PF. Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J*. 2014 Jun 16; 8(6):1166–74. <https://doi.org/10.1038/ismej.2013.237> PMID: 24430481
53. Campbell BJ, Kirchman DL. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J*. 2013 Jan 16; 7(1):210–20. <https://doi.org/10.1038/ismej.2012.93> PMID: 22895159
54. Raes EJ, Bodrossy L, van de Kamp J, Bissett A, Waite AM. Marine bacterial richness increases towards higher latitudes in the eastern Indian Ocean. *Limnol Oceanogr Lett*. 2018 Feb 1; 3(1):10–9.
55. Milici M, Tomasch J, Wos-Oxley ML, Wang H, Jáuregui R, Camarinha-Silva A, et al. Low diversity of planktonic bacteria in the tropical ocean. *Sci Rep*. 2016 Jan 11; 6(1):1–9. <https://doi.org/10.1038/s41598-016-0001-8> PMID: 28442746
56. Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. Micro-scale determinants of bacterial diversity in soil. Vol. 37, *FEMS Microbiology Reviews*. 2013. p. 936–54. <https://doi.org/10.1111/1574-6976.12023> PMID: 23550883
57. Evans S, Martiny JBH, Allison SD. Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J*. 2016; 11(1):1–10. <https://doi.org/10.1038/ismej.2016.118> PMID: 27636395
58. Dell'Anno A, Danovaro R. Ecology: Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* (80-). 2005 Sep 30; 309(5744):2179.
59. Ramírez GA, Jørgensen SL, Zhao R, D'Hondt S. Minimal Influence of Extracellular DNA on Molecular Surveys of Marine Sedimentary Communities. *Front Microbiol*. 2018 Dec 4; 9:2969. <https://doi.org/10.3389/fmicb.2018.02969> PMID: 30564217
60. Rodrigues JLM, Pellizari VH, Mueller R, Baek K, Jesus E d. C, Paula FS, et al. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proc Natl Acad Sci*. 2013; 110(3):988–93. <https://doi.org/10.1073/pnas.1220608110> PMID: 23271810
61. Sharma SK, Ramesh A, Sharma MP, Joshi OP, Govaerts B, Steenwerth KL, et al. Microbial Community Structure and Diversity as Indicators for Evaluating Soil Quality. In: *Biodiversity, Biofuels, Agroforestry and Conservation Agriculture*. Springer, Dordrecht; 2010. p. 317–58.
62. Ding J, Jiang X, Ma M, Zhou B, Guan D, Zhao B, et al. Effect of 35 years inorganic fertilizer and manure amendment on structure of bacterial and archaeal communities in black soil of northeast China. *Appl Soil Ecol*. 2016 Sep 1; 105:187–95.
63. O'Brien SL, Gibbons SM, Owens SM, Hampton-Marcell J, Johnston ER, Jastrow JD, et al. Spatial scale drives patterns in soil bacterial diversity. *Environ Microbiol*. 2016 Jun 1; 18(6):2039–51. <https://doi.org/10.1111/1462-2920.13231> PMID: 26914164
64. Soman C, Li D, Wander MM, Kent AD. Long-term fertilizer and crop-rotation treatments differentially affect soil bacterial community structure. *Plant Soil*. 2017; 413(1–2):145–59.
65. Delgado-Baquerizo M, Eldridge DJ. Cross-Biome Drivers of Soil Bacterial Alpha Diversity on a World-wide Scale. *Ecosystems*. 2019 Sep 1; 22(6):1220–31.

66. Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, et al. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Klironomos J, editor. *Ecol Lett*. 2015 Jan 1; 18(1):85–95. <https://doi.org/10.1111/ele.12381> PMID: 25430889
67. Rintala H, Pitkaranta M, Toivola M, Paulin L, Nevalainen A. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiol*. 2008 Apr 8; 8(1):56.
68. Miletto M, Lindow SE. Relative and contextual contribution of different sources to the composition and abundance of indoor air bacteria in residences. *Microbiome*. 2015 Dec 10; 3(1):61.
69. Glassman SI, Martiny JBH. BROADSCALE Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. 2018;
70. Needham DM, Sachdeva R, Fuhrman JA. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J*. 2017 Jul 11; 11(7):1614–29. <https://doi.org/10.1038/ismej.2017.29> PMID: 28398348