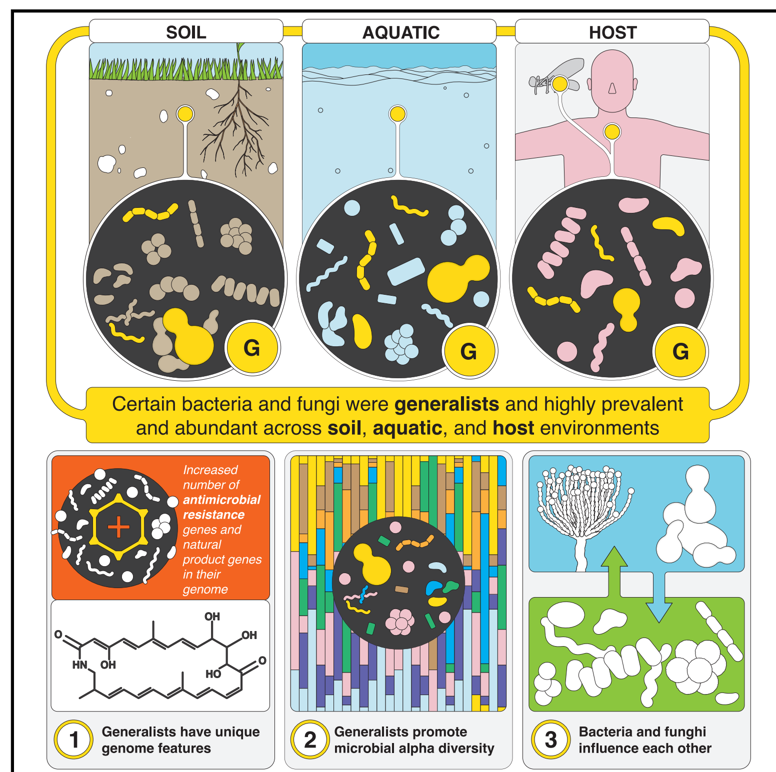
Cell Reports

Report

A global survey of host, aquatic, and soil microbiomes reveals shared abundance and genomic features between bacterial and fungal generalists

Graphical abstract



Highlights

- Bacterial and fungal generalists are widely distributed in aquatic, host, and soil biomes
- Generalists have larger genomes with more secondary metabolites and AMR genes
- Samples containing generalists show higher alpha diversity
- Generalists underpin cross-kingdom community structure

Authors

Daniel Loos, Ailton Pereira da Costa Filho, Bas E. Dutilh, Amelia E. Barber, Gianni Panagiotou

Correspondence

amelia.barber@uni-jena.de (A.E.B.),
gianni.panagiotou@leibniz-hki.de (G.P.)

In brief

Environmental change is reshaping microbial communities critical to Earth and human health. Loos, Pereira, et al. analyze 1,580 microbiome samples from aquatic, host, and soil biomes to identify characteristics of bacteria and fungi that are resilient to change and identify those that may be vulnerable to biodiversity loss.

Loos et al., 2024, Cell Reports *43*, 114046
April 23, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2024.114046

CellPress

Cell Reports

CellPress
OPEN ACCESS

Report

A global survey of host, aquatic, and soil microbiomes reveals shared abundance and genomic features between bacterial and fungal generalists

Daniel Loos,[1,7] Ailton Pereira da Costa Filho,[2,3,7] Bas E. Dutilh,[3,4,5] Amelia E. Barber,[2,3,*] and Gianni Panagiotou[1,3,6,8,*]
[1]Department of Microbiome Dynamics, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena, Germany
[2]Junior Research Group Fungal Informatics, Institute of Microbiology, Friedrich Schiller University, Jena, Germany
[3]Cluster of Excellence Balance of the Microverse, Friedrich Schiller University, Jena, Germany
[4]Institute of Biodiversity, Friedrich Schiller University, Jena, Germany
[5]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, the Netherlands
[6]Faculty of Biological Sciences, Friedrich Schiller University, Jena, Germany
[7]These authors contributed equally
[8]Lead contact
*Correspondence: amelia.barber@uni-jena.de (A.E.B.), gianni.panagiotou@leibniz-hki.de (G.P.)
https://doi.org/10.1016/j.celrep.2024.114046

SUMMARY

Environmental change, coupled with alteration in human lifestyles, is profoundly impacting the microbial communities critical to the health of the Earth and its inhabitants. To identify bacteria and fungi that are resistant and susceptible to habitat change, we analyze thousands of genera detected in 1,580 host, soil, and aquatic samples. This large-scale analysis identifies 48 bacterial and 4 fungal genera that are abundant across the three biomes, demonstrating fitness in diverse environmental conditions. Samples containing these generalists have significantly higher alpha diversity. These generalists play a significant role in shaping cross-kingdom community structure, boasting larger genomes with more secondary metabolism and antimicrobial resistance genes. Conversely, 30 bacterial and 19 fungal genera are only found in a single habitat, suggesting a limited ability to adapt to different and changing environments. These findings contribute to our understanding of microbial niche breadth and its consequences for global biodiversity loss.

INTRODUCTION

Environments, plants, and animals are colonized with communities of microbial organisms, termed the microbiome, which play critical roles in the function and health of their hosts and habitats. While understudied relative to bacteria, fungi play critical roles in environmental and host microbial communities, including important roles in carbon cycling and beneficial symbioses with plant and hosts.[1–3] Importantly, environmental change and alterations in host lifestyle are profoundly affecting microbial consortia. Westernized diets low in fiber and rich in saturated fats and sugars have decreased the abundance of beneficial microbes and been linked with myriad health conditions, including obesity, type 2 diabetes, and inflammatory bowel disease.[4–7] Changes in marine environments due to climate change have induced major shifts in marine food webs, primary productivity, and carbon export.[8–11] Additionally, anthropogenic climate change is resulting in net carbon loss in soil and changes in microbial community composition.[12]

Ecological theory predicts that generalists, or organisms that are fit across a wider range of conditions, will be more resilient to changing environmental conditions.[13–15] Conversely, specialists, or organisms that are adapted to thrive in very specific environments, will be less able to withstand perturbations to their habitat. As the Earth and its inhabitants are experiencing unprecedented changes to their health and habitats, it is crucial to understand the capacity of individual microbial taxa to adapt to changing environmental conditions. Those unable to change are susceptible to biodiversity loss, while generalists that can grow in a wider range of conditions may survive and flourish with unknown consequences. Pan-habitat meta-analyses of bacterial community data have identified ecological and evolutionary features of bacterial generalists and specialists, including differences in abundance and speciation rates.[16,17] However, the corresponding studies examining these features in fungi have only been performed on distinct habitats or taxonomic groups.[18–20] Moreover, bacterial and fungal kingdoms have not been considered together at the global scale despite substantial evidence from individual settings that bacteria and fungi commonly interact with each other with pronounced consequences.[21,22]

To this end, we performed a large-scale analysis of community sequencing datasets from host, soil, and aquatic environments with paired bacterial and fungal characterization to shed light

© 2024 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

on the ecological properties of the genera present and their putative resilience to change. We focused on three aspects: (1) the identification of bacteria and fungi that occurred in diverse environments (generalists) or were limited to highly specific environments (specialists); (2) the relative abundance of bacterial and fungal generalists and specialists as a marker for their fitness and competitive colonization potential; and (3) whether their presence in a habitat was associated with global changes in inter- and cross-kingdom population structure.

RESULTS

Environmental specificity of bacterial and fungal communities

For a global survey of bacteria and fungi across microbial communities, we analyzed paired 16S and internal transcribed spacer (ITS) rRNA amplicon sequence data from 1,580 samples deposited in public databases. Samples were collected from Europe, Asia, and the Americas between 2010 and 2018 (Figure 1A). For cross-biome comparisons, samples were classified as aquatic, host, or soil environments based on the habitat they were collected from. This broad grouping is supported by principal coordinate analysis based on Bray-Curtis dissimilarity showing that samples from each environment largely cluster with each other and are distinct from the other environments (Figure 1B)—a finding mirrored by a recent study of 22,700 bacterial microbiomes.[16]

Of the 1,580 samples that we analyzed, 871 originated from soils, 494 from hosts (including mammalian, non-mammalian, and plant hosts), and 215 from aquatic environments. The habitats that contributed the largest number of samples for each environment were temperate (n = 498) and conifer forests (n = 147) for the soil, gut (n = 287) and skin (n = 68) for the hosts, and large lakes (n = 87) and other freshwater (n = 71) for the aquatic environments (see Data S1 for details of all projects). Due to the limited availability of fungal community data in marine samples, this large habitat was only represented with 10 samples. Taxonomic profiling of bacterial and fungal communities was performed using the SILVA and UNITE databases, respectively. Rarefaction curves of each habitat indicated that most projects adequately captured the diversity of both the bacterial and fungal communities (Figure 1C). In total, 2,977 bacterial and 1,740 fungal genera were detected across all samples (Figure 1D). We next examined the overlap of genera between environments, where a genus was considered shared if it was detected in at least one habitat in each of the three different environments (host, aquatic, and soil). For bacteria, soil and aquatic environments had the highest number of shared genera (n = 1,662), followed by host-soil (n = 1,483) and host-aquatic (n = 1,226). The pattern was different for fungi, with host-soil sharing the most (n = 884), followed by aquatic-soil (n = 205) and host-aquatic (n = 189). These trends remained after controlling for the different number of samples across the three environments in 842 and 998 out of 1,000 random downsampled subsets for bacteria and fungi, respectively. Finally, we also confirmed that a similar degree of overlap between the environments was observed for different 16S and ITS amplicons, as well as significant correlations in the relative abundances of individual genera (Figures S1A–S1C).

While 40% of the total bacterial genera were found in all three environments, the percentage dropped to only 11% for fungal genera, indicating a higher degree of environmental specificity (Figures 1E and 1F). The most prevalent higher-order taxonomic ranks that were detected in all three environments were *Proteobacteria*, followed closely by *Firmicutes* for bacteria and *Ascomycota* for fungi. For both bacteria and fungi, soil was the environment with the highest percentage of uniquely detected genera (i.e., genera not detected in any sample from host or aquatic origin) with 23% and 38%, respectively, for each kingdom. While aquatic-specific bacteria accounted for 7% of the total number of detected genera, the percentage of unique fungi in aquatic samples was only 2% (Figures 1E and 1F). The opposite trend was observed for host-associated microbes, with only 3% and 8% of unique bacteria and fungi, respectively, in this environment.

We subsequently compared the relative abundance of genera that were found in all environments or were uniquely detected in soil-, host-, or aquatic-associated environments. Bacterial genera detected in all three environments were significantly more abundant (Wilcoxon rank-sum test, $p < 0.001$) than genera uniquely detected in one of the environments (Figure 1G). A similar pattern was observed with fungi. However, a notable exception was the relatively high abundance of fungi that were uniquely detected in aquatic samples. Genera of aquatic fungi were more abundant than either common genera and uniquely detected in soil- or host-associated environments (Figure 1H). This observation was also robust across the different 16S and ITS regions used in the dataset (Figure S1D). Taken together, we find that soil bacteria and fungi show a higher degree of biome specificity compared to aquatic and host environments and that genera detected in all three environments were also more abundant than those only detected in only one environment.

Bacterial and fungal generalists are more abundant than specialists and have distinct genomic features

Generalists and specialists play important yet distinct roles in ecosystems. However, objectively identifying them has proven challenging. To define multi-kingdom generalists and specialists, we set the following criteria: generalists are genera found with high prevalence (>40%) in at least one habitat (e.g., gut, boreal forest) from each of the three environments (host, aquatic, soil), where prevalence was defined as a relative abundance above 0.01% in >1 sample from the habitat. Conversely, specialists were genera with a high prevalence (>40%) in one habitat and low prevalence (<5%) in every other. Using this approach, we detected 48 bacterial generalists and 30 specialists (Table S1). To ensure that the generalists identified were legitimately present in the microbial community and not the result of reagent or sequencing contamination, we performed decontamination on low-biomass projects, including all aquatic projects and low-biomass host sites like the lung (see STAR Methods for details). There were no significant differences in the relative abundance of generalists before and after sequence data decontamination (Figure S2). As an additional control, we compared the DNA extraction methods/kits used, any DNA purification kits utilized, the library preparation kit, and the

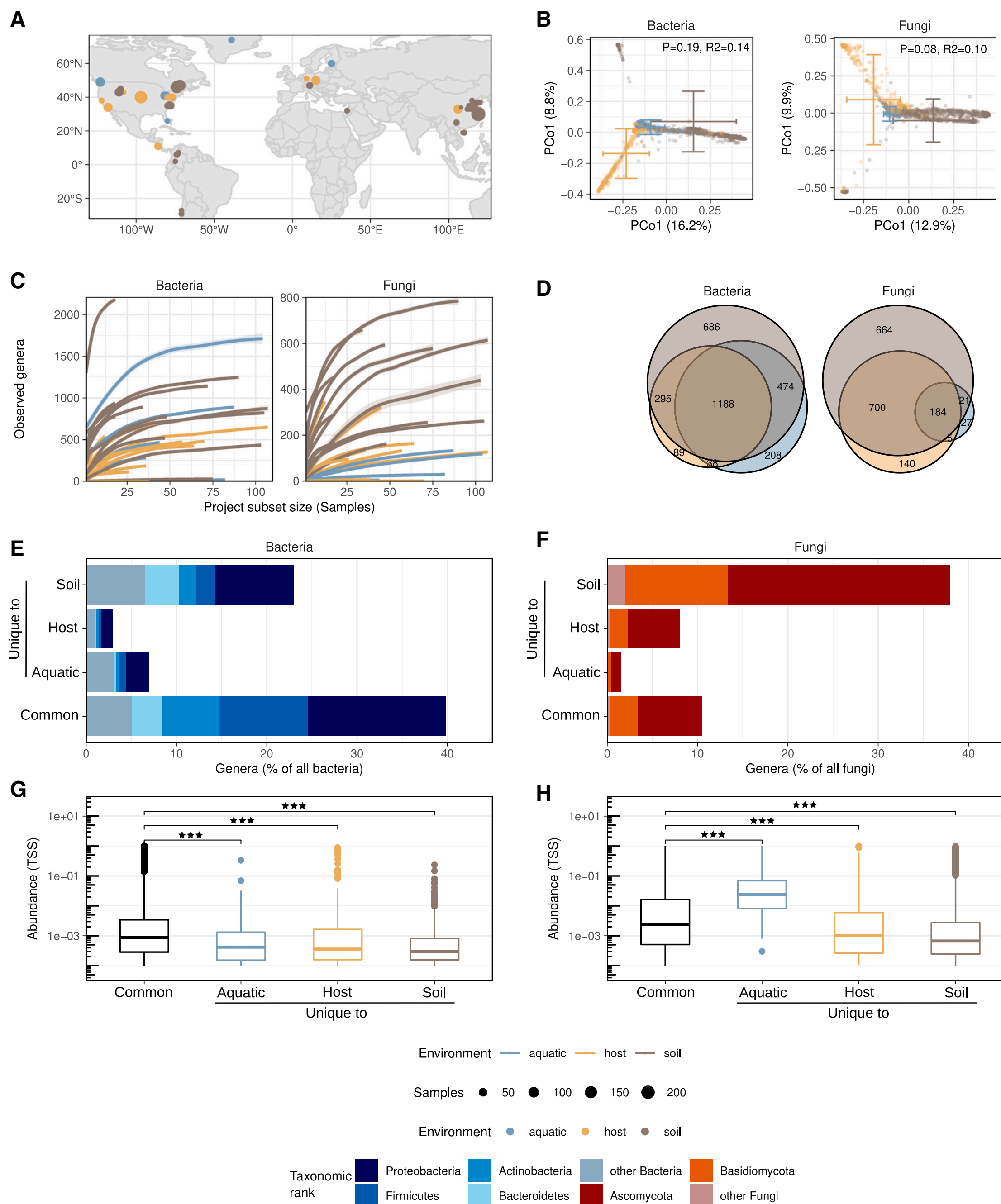


Figure 1. A global analysis of microbial communities reveals differences in environmental specificities between bacteria and fungi

(A) Distribution of samples used in this study (n = 1,580) by geographic location.

(B) Bray-Curtis dissimilarity between samples colored by environment. Crosshatches represent the mean ± SD for each environment. Significance calculated by PERMANOVA.

(C) Rarefaction curves of Shannon alpha diversity for each study demonstrate sufficient sampling depth. Curves are shown as LOESS regressions from 10 independent sampling trials at 10 given sampling subset sizes. Lines are colored by environment and are surrounded by ribbons indicating the 95% confidence interval across the trails.

(D) Intersection of bacterial and fungal genera found in at least one sample in each environment as Venn diagrams.

(E and F) Percentage of genera found in all three or only one environment.

(G and H) Abundance comparisons of common and unique genera by total sum scaling (TSS). A genus was considered present in a sample using a threshold of abundance >0.01%. Significance was determined by Wilcoxon rank-sum test; ***p < 0.001.

sequencing facility of the projects included in our study. We could not find appreciable overlap among the project studies that would imply that the generalists identified were the result of contamination via common methodology or reagents (Data S2).

To confirm our definition of generalists and specialists, we calculated Levins' niche breadth indices (B_n), which measure taxon distribution across environments and where higher values indicate even distribution across environments.[23] Generalists showed significantly higher B_n values than specialists (Wilcoxon rank-sum test, $p < 0.001$; Figure S3A). All specialists and all generalists, with the exception of the *Christensenellaceae R7* group, were above the detection limit and had a significant B_n after Benjamini-Hochberg adjustment.[23] As our criteria for defining generalists and specialists were reliant on human-defined biome annotations, we further validated our approach by comparing it to the recently developed social niche breadth (SNB) score.[16] By comparing the similarity or diversity of microbial communities where a given genus occurs, SNB provides a data-driven score independent of biome annotations based on an independent dataset of over 22,500 bacterial microbiomes.[16] Indeed, the generalists identified in our study had significantly higher SNB scores than the bacterial specialists we identified (Wilcoxon rank-sum test, $p < 0.001$; Figure S3B).

We observed multiple phylogenetic origins for both generalists or specialists (chi-squared test, $p > 0.05$), indicating that their roles as generalists and specialists evolved independently (Figure S4). Each of the top five bacterial generalists were detected in more than 50% of the 1,580 samples. Among them, the most prevalent was *Pseudomonas*, which was detected in 52%, 70%, and 89% of host, soil, and aquatic samples, respectively, followed by *Bacillus* (33%, 71%, 35%) and *Bradyrhizobium* (17%, 73%, 35%). The most extreme bacterial specialists came from the soil. While *Gryllotalpicola* and *Anaerovibrio* were found in >91% of biochar samples, the prevalence dropped to 0.1% on average for non-soil environments (Table S1). Specialists were also found in host- and aquatic-associated environments. For example, *Acetatifactor* was found in 80% of samples from the murine gut but had a prevalence of <3% in all other habitats. The genus *Leptospira* was found in 83% of samples of the Cuyahoga River but had a prevalence less than 2% in all other habitats. Interestingly, when comparing the relative abundance of generalists and specialists, we observed that both bacterial and fungal generalists had a significantly higher relative abundance (Figure 2A; Wilcoxon rank-sum test, $p < 0.001$). This pattern remained when we used stricter and looser thresholds to define generalists and specialists (Figure S5). This finding confirms the pattern observed above (Figures 1G and 1H), suggesting that independently of how groups are defined, genera that can colonize diverse environments are usually able to outcompete habitat-specific genera.

When looking at the fungal kingdom, the number of generalists was much lower, and only *Aspergillus*, *Malassezia*, *Aureobasidium*, and *Cortinarius* satisfied the criteria of a generalist (Table S1). Among these, *Aspergillus* had the highest overall prevalence among all samples with 38%, 52%, and 12% in the host, soil, and aquatic samples, respectively. From the 19 fungal specialists, *Chrysanthotrichum* and *Mycocentrospora* were the most habitat specific, with prevalences of 68% and 48% in temperate and conifer forests, respectively, but a mean prevalence of $\leq 0.1\%$ in all other habitats. Only two of the 19 fungal specialists (11%) originated from outside soil environments (*Vuilleminia* and *Seimatosporium* from plants). As with bacterial genera, the relative abundance of fungal generalists was significantly higher than that of fungal specialists (Figure 2A; Wilcoxon rank-sum test, $p < 0.001$).

To gain insight into how generalists achieve high relative abundance in diverse environments, we analyzed the genomes of generalists and specialist genera available on NCBI (see STAR Methods for details on genome selection). For bacteria, when analyzing the genomes of 2,328 generalists and 471 specialists (Data S3), the generalists had significantly larger genomes as measured by the number of coding sequences (CDSs), with a mean of 4,671 CDSs for generalists and 3,189 for specialists (permutation test, $p < 0.001$; Data S3). This trend remained after controlling for genome length, with generalists having a mean of 925 CDS/Mb genome compared to 921 CDS/Mb for specialists (permutation test, $p < 0.001$). As secondary metabolism genes are often used by microbes during competition for resources and as chemical warfare in crowded environments, we examined the genomes of generalists and specialists for the presence of biosynthetic gene clusters (BCGs). Strikingly, the genomes of bacterial generalists encoded significantly more BCGs with an average of 7.4 BGCs and 1.4 BGCs/Mb genome compared to 2.7 BGCs and 1.2 BGCs/Mb for specialists (Figure 2C; permutation test, $p < 0.001$). Further differentiating bacterial generalists, they also contained significantly more antimicrobial resistance (AMR) genes, with an average of 4.2 AMR genes and 0.39 AMR genes/Mb genome compared to 1.0 AMR genes and 0.36 AMR genes/Mb for specialists (Figure 2D; permutation test, $p = 0.001$). For fungi, no significant differences in either the number of CDS or BGCs was observed (Figures 2B and 2C), likely due to the severe underrepresentation of publicly available fungal specialist genomes ($n = 5$).

To explore intra- and inter-kingdom interaction patterns and to gain further insight into the downstream effects of the observed differences between generalists and specialists, we constructed individual coabundance networks for soil, host, and aquatic environments (see STAR Methods for details). As many taxa were not detected across all environments, networks were constructed only considering the 1,188 bacterial and 184 fungal genera commonly detected in all three environments. Despite starting with the same genera, the topological characteristics of the networks for each environment were highly distinct, as measured by significant differences in betweenness and Kleinberg's hub node centrality scores (Wilcoxon test, $p < 0.001$; Figure S6). In spite of the differences in topology, we could still compile subnetworks of inter- and intra-kingdom correlations found jointly among host-soil, host-aquatic, and/or soil-aquatic environments. Strikingly, 45 of the 48 bacterial generalists and all 4 fungal generalists were part of those subnetworks, which are characterized by a higher number of positive than negative edges (Figure 2E). The ratio of positive to negative edges was higher in correlations involving a generalist (2.5) compared to all other edges (2.2). When we looked for interactions between genera found in all three environments, we identified 43 such

CellPress
OPEN ACCESS

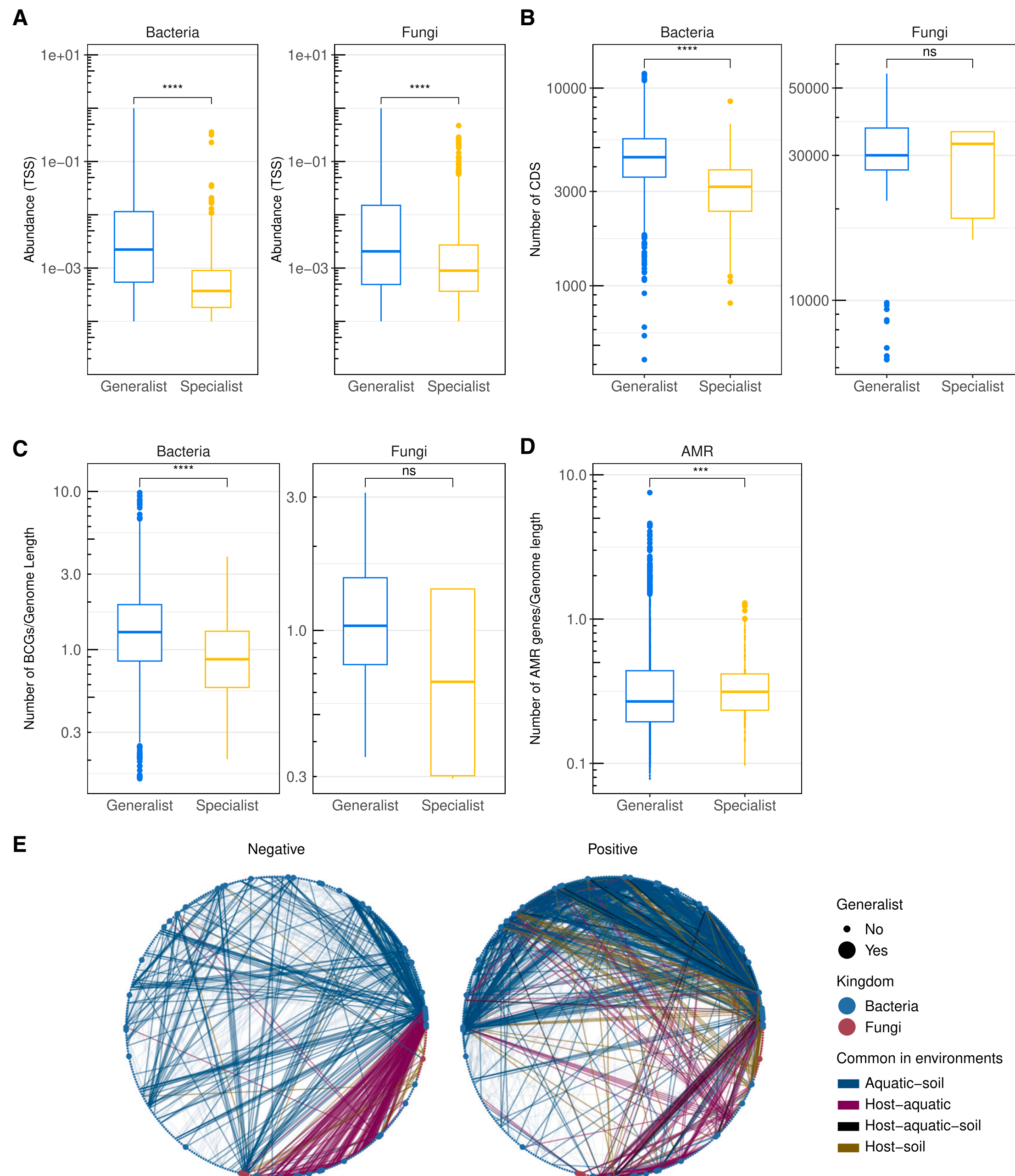


Figure 2. Generalists are more abundant and bacterial generalists have larger genomes with more biosynthetic gene clusters (BCGs) and antimicrobial resistance genes

(A) Relative abundances of bacterial and fungal generalists and specialists. Values were averaged by project to account for different cohort sizes. Statistical significance was calculated using Wilcox rank-sum test (****$p < 0.0001$).

(B and C) Number of coding sequences (CDSs) (B) and BCGs (C) in the genomes of generalists and specialists normalized to their genome length in bp. Data are from the genomes of 2,328 bacterial generalists, 117 fungal generalists, 471 bacterial specialists, and 5 fungal specialists. Statistical significance was calculated by permutation test (****$p < 0.0001$; ns denotes $p > 0.05$).

(D) Number of antimicrobial resistance (AMR) in the genomes of bacterial specialists normalized by genome length. Statistical significance calculated by Permutation test (****$p < 0.0001$).

(E) Networks of genera found in all three environments and significantly coabundant in the majority of environments (SparCC false discovery rate [FDR] $p < 0.05$, $|r| > 0.2$).

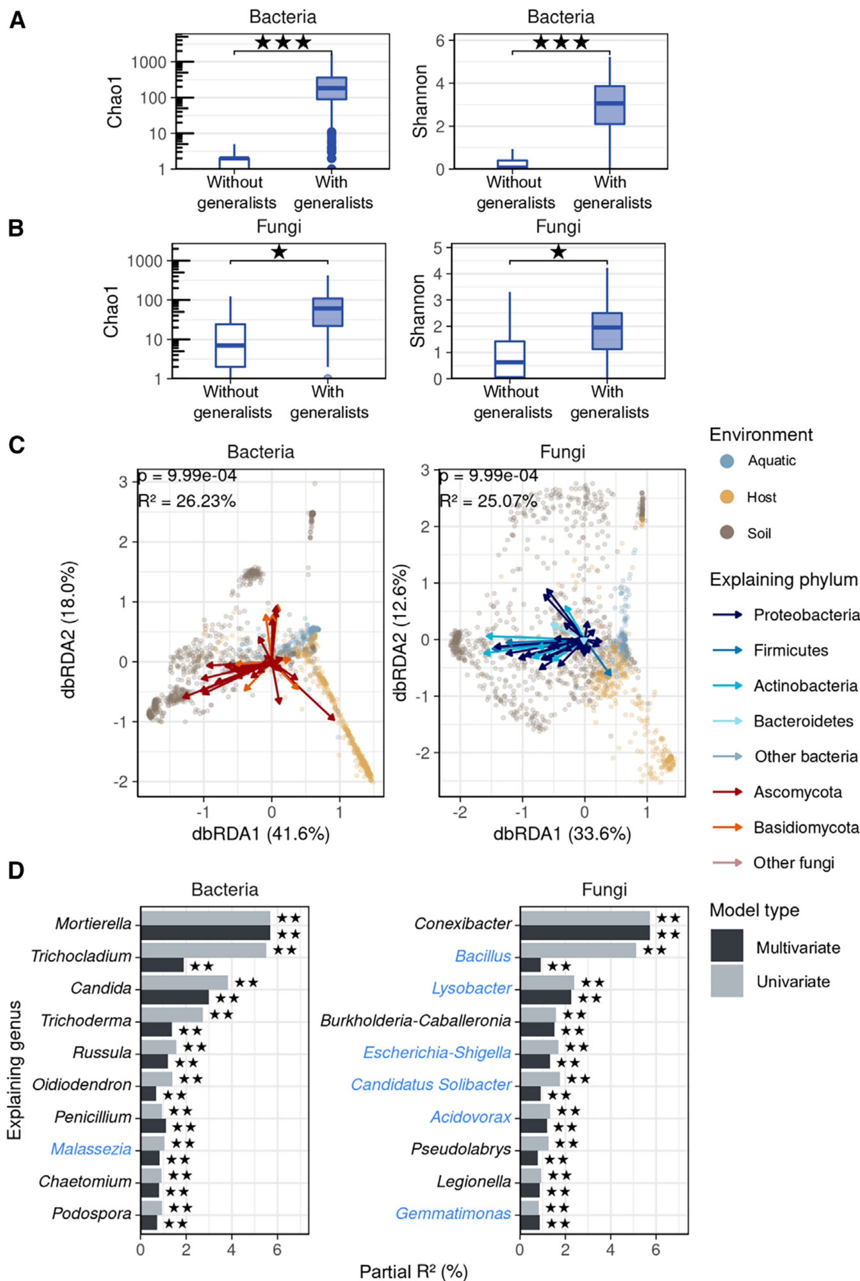


Figure 3. Generalists significantly impact diversity and cross-kingdom variation

(A and B) Shannon and Chao1 alpha diversity were calculated for bacteria (A) and fungi (B). Samples were grouped by whether they contained any generalist (genera with >40% prevalence in at least one habitat from every environment; abundance >0.01%; n = 1,500 for bacteria, n = 1,125 for fungi) or do not contain any generalist (n = 91 for bacteria; n = 466 for fungi). Significance bars indicate permutation test compared to samples without random taxa instead of generalists (*q < 0.05, **q < 0.01, and ***q < 0.001).

(C and D) Bacterial and fungal Bray-Curtis dissimilarities constrained by explanatory genus abundances of the other kingdom using distance-based redundancy analysis (dbRDA). (C) Explaining genera were selected using a feedforward approach. Significance determined by PERMANOVA. Effect size of most explanatory taxa is shown in (D) by multivariate model (as displayed in C) or univariate model containing only the taxon of interest. Generalists are indicated with blue text. Stars indicate significance by ANOVA (**p < 0.01).

edges that all represented positive interactions between bacteria and included 21 generalists. Together, these findings suggest that the success of generalists in colonizing diverse environments and achieving high relative abundances may be attributable to their ability to carve out a niche for themselves using secondary metabolism and AMR genes and by eliciting positive interactions with other highly prevalent genera.

Bacterial generalists exert a strong influence on the intra- and inter-kingdom community structure

We subsequently explored whether the presence of generalists and specialists had an impact on the diversity of a community. Interestingly, alpha diversity, as measured as Chao 1 and Shannon, was significantly lower in samples where no generalist was detected compared to samples with generalists present for both bacterial (Figure 3A) and fungal (Figure 3B) communities (permutation test of samples lacking any of the N generalists compared with samples lacking any N random taxa, 1×10^4 permutations, $p < 0.03$). Conversely, the impact of specialists on alpha diversity in their specific habitat was much less profound and varied by habitat without a clear trend (Figure S7).

We subsequently shifted our focus to inter-kingdom interactions, which are often overlooked in microbial ecology studies, and examined bacterial generalists for a role in shaping the mycobiome community structure, and vice versa. As expected, we observed a significant separation between the soil, host, and aquatic micro- and mycobiome beta diversity by Bray-Curtis dissimilarity (Figure 3C; PERMANOVA, $p < 0.001$ for both bacteria and fungi). Constrained ordination revealed a significant, linear relationship between bacterial Bray-Curtis dissimilarity and fungal community composition, and vice versa (distance-based redundancy analysis [dbRDA], ANOVA $p < 0.04$ for all explanatory genera of the other kingdom in a multivariate model). Bacterial genera could explain an extensive part of the mycobiome variation observed in the three environments with a partial R2 of 25% by dbRDA. Of the bacterial genera, *Conexibacter*, *Bacillus*, and *Lysobacter* had the highest explanatory power on mycobiome variation (Figure 3D). Interestingly, six out of the top ten explanatory bacterial

genera in the dbRDA were generalists. Similarly, fungal genera explained 26% of the microbiome variation between host, soil, and aquatic samples, with *Mortierella*, *Trichocladium*, and *Candida* having the highest explanatory power (Figure 3D). Among the top ten explanatory genera was one of the four fungal generalists—*Malassezia*. Altogether, our analysis indicates that bacterial and fungal generalists profoundly impact microbial communities by contributing positively to the taxonomic (alpha) diversity of their kingdom—an ecological characteristic often associated with healthy environments—and they can also contribute to shaping cross-kingdom microbial structures.

DISCUSSION

Recent global changes are profoundly affecting the health of our planet and its inhabitants.[24–26] As environmental and host-associated microbial communities are increasingly exposed to changing habitats, we still lack knowledge about the capacity of millions of bacterial and fungal species to cope with these shifts. With this in mind, we performed a large-scale global survey of host, aquatic, and soil microbiomes to reveal ecological and genomic properties of bacterial and fungal genera that may promote or limit their establishment in new environments and how they contribute to the richness and diversity of an environment. Analysis of 1,580 paired host, soil, and aquatic micro- and mycobiomes identified ~70 specialist genera whose limited distribution suggests they may struggle in different or changing habitats and identify ~50 widespread genera with high relative abundance across all environments examined, suggesting an ability to thrive across diverse habitats. Through this analysis, we also identified genomic and ecological properties associated with generalists, including their contributions to alpha diversity, and structuring the beta diversity in the other kingdom. Generalists also had larger genomes with more secondary metabolism genes, which serve as antimicrobial weapons and tools for nutrient scavenging, microbial communication, and influencing microbial community composition.[27–29] This suggests a mechanism for how generalists thrive in such diverse and often highly competitive habitats.

While the concept of generalists and specialists is not new to microbial ecology, it has mostly been applied in specific habitats.[30–34] Similarly, while studies of prokaryotic generalist and specialist microbes have been carried out,[16,17,31,35–39] they have rarely considered eukaryotic microorganisms such as fungi, despite the critical role fungi play in many habits.[40–45] We demonstrate that both bacterial and fungal generalists share ecological features including the ability to reach significantly higher relative abundances than specialists and contributing positively to the richness and diversity of their respective kingdoms. Moreover, six bacterial generalists, including *Bacillus*, *Lysobacter*, *Escherichia*, and *Gemmatimonas*, and one fungal generalist, *Malassezia*, harbor additional ecological properties and appear to play a significant role in shaping cross-kingdom microbial composition (Figure 3D).

These positive roles for generalists are somewhat at odds with previous work showing that generalists negatively impact ecosystems through homogenization.[46–48] One possible explanation is that the species- and strain-level diversity of the microbial world is enormous compared to higher eukaryotes, where many key studies have been conducted. The variable and pronounced effects of strain- and species-level diversity within generalist genera are highlighted in biocontrol agents. Strains of the generalist bacteria *Bacillus*, *Pseudomonas*, and *Streptomyces* are approved as biocontrol agents for soil-borne diseases in the European Union (source: EU Active Substance Pesticide Database, accessed December 2023). However, other strains of *Bacillus* and *Pseudomonas* are pathogens of crops.

Indeed, a challenge for the future will be to move the analysis of microbial generalists and specialists beyond taxonomic description to understanding the functional characteristics that distinguish them from other taxa. Our study was carried out at the genus level due to the limitations of accurate species-level classification of bacteria and fungi via amplicon metabarcoding.[49–51] For bacteria, one way forward may be deep functional characterization at the pathway and enzyme levels using shotgun metagenomics datasets. The functional characterization of fungal generalists and specialists will prove to be a much greater challenge. Their genomes are larger and more complex and their physiology less studied. Consequently, functional prediction tools and community-level modeling based on metagenomic data for fungi lag behind those for prokaryotic microorganisms.

Overall, the generalists and specialists identified cumulatively account for a small fraction of the total taxonomic diversity of each kingdom (<2.6% of total bacterial genera and <1.3% of fungal genera). However, we find that generalism is proportionally rarer in fungi, with 1.6% of bacterial genera in the dataset meeting the definition for generalism compared to 0.2% of fungi. This is likely influenced by the clear specialization of fungi for soil. Current estimates suggest that 90% of the Earth's fungal species live in soil compared to 43% for bacterial life.[52] However, the reasons for the observed difference in specificity and generalism present an intriguing question for further work. We posit that it may be due to differences in the genome dynamics between bacteria and fungi. Bacterial genomes are more flexible and shaped by horizontal gene transfer to a degree not seen in fungi. Alternatively, differences in dispersal may allow for bacterial genera to expand to different habitats more efficiently than fungi. Indeed, the higher overall population size of bacteria over fungi in soil increases the probability of their dispersal.[53,54]

Our finding that bacterial and fungal generalists have larger genomes is consistent with the expectation that fitness across diverse habitats requires a larger genomic repertoire and parallels the results from another metastudy of bacterial community data.[17] In our work, we were able to delineate some of the specific tools used by generalists, such as the acquisition of an enrichment of BCGs and antimicrobial resistance genes. However, it is likely that other genomic factors underlie their generalism; the relative contributions of these factors involved in inter-microbe communication, compared to other genomic features such as metabolic versatility, is an open question.

Finally, the observation that roughly a quarter of bacterial community composition could be explained by fungal relative abundance, and vice versa, strongly emphasizes the role of the multi-kingdom interactions in microbial communities and highlights the amount of information potentially missed by examining only

one kingdom—an important point for future microbial community studies. In conclusion, our global survey of bacterial-fungal communities has generated a valuable list of genera that may be vulnerable to biodiversity decline and even extinction under changing environmental threats.[55,56] Conversely, the generalist bacteria and fungi identified are highly resilient against environmental perturbations. However, their functional roles in ecosystems, especially at the species and strain levels, will benefit from approaches that combine large-scale computational analyses and laboratory experiments. Together, these interdisciplinary approaches can address the many open questions about microbial niche range and its consequences for microbial extinction and global biodiversity loss.

Limitations of the study

As this is a metastudy, there are limitations to it worth mentioning. One is that aquatic, and particularly marine, fungi are understudied. This led to the exclusion of important marine habitats due to the lack of paired, publicly available 16S-ITS samples at the time we assembled our dataset. When we analyzed several marine projects post hoc representing seawater, marine sediment, and coastal seagrass water,[57–59] we did not identify any bacteria that were not already classified as generalists or specialists compared to our original dataset. On the other hand, we identified *Puccinia* as a fungal marine specialist and *Cladosporium*, *Penicillium*, and *Chaetomium* as fungal generalists. The ecological and genomic features of these taxa will be interesting to examine in the future.

While we confirmed that the rRNA subregion amplified for microbial diversity characterization did not significantly impact genera abundance within an environment (Figure S1), it is possible that variation in the specific primers for the same region still lead to the underdetection of specific taxonomic groups.

An additional caveat is that the relationship between habitats, especially between a host and its environment, is complex. For example, plants have selective effects on microbial diversity in the rhizosphere.[60–62] Nevertheless, in our study and other global microbiome studies,[15,16] rhizosphere microbial communities more closely resemble soil than other host-associated habitats, such as the host-associated microbiomes of insects, birds, and mammals.

Rarefaction is polarizing in microbial community research.[63,64] A final point to our study is that we opted not to rarefy the sequence data analyzed in our study. We decided not to because it could have potentially introduced new biases depending on whether the rarefaction was done by BioProject, habitat, biome, or across the entire study dataset. The analyses that are potentially most affected by this are the diversity analyses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Sample selection
  - Generation of genus-level abundance profiles
  - Discovery of sample rRNA amplified region
  - Abundance correlation between varying rRNA amplicons
  - Decontamination of low biomass projects
  - Genome features of generalists and specialists
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Workflow and statistical analysis
  - Diversity
  - Co-abundance networks
  - Generalists and specialists

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114046.

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 20151 – Project-ID 390813860. B.E.D. is supported by the European Research Council Consolidator grant 865694 and the Alexander von Humboldt Foundation in the context of an Alexander von Humboldt-Professorship.

AUTHOR CONTRIBUTIONS

Conceptualization, A.E.B. and G.P.; methodology/validation, D.L., A.E.B., G.P., and B.E.D.; visualization, D.L. and A.P.d.C.F.; project administration/supervision, A.E.B. and G.P.; funding acquisition, A.E.B. and G.P.; writing – original draft, D.L., A.P.d.C.F., A.E.B., and G.P.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 28, 2023
Revised: December 22, 2023
Accepted: March 19, 2024

REFERENCES

1. Orsi, W.D., Vuillemin, A., Coskun, Ö.K., Rodriguez, P., Oertel, Y., Niggemann, J., Mohrholz, V., and Gomez-Saez, G.V. (2022). Carbon assimilating fungi from surface ocean to subseafloor revealed by coupled phylogenetic and stable isotope analysis. ISME J. *16*, 1245–1261. https://doi.org/10.1038/s41396-021-01169-5.
2. Zhang, B., Li, S., Chen, S., Ren, T., Yang, Z., Zhao, H., Liang, Y., and Han, X. (2016). Arbuscular mycorrhizal fungi regulate soil respiration and its response to precipitation change in a semiarid steppe. Sci. Rep. *6*, 19990. https://doi.org/10.1038/srep19990.
3. van der Heijden, M.G.A., de Bruin, S., Luckerhoff, L., van Logtestijn, R.S.P., and Schlaeppi, K. (2016). A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. ISME J. *10*, 389–399. https://doi.org/10.1038/ismej.2015.120.
4. Cani, P.D., Bibiloni, R., Knauf, C., Waget, A., Neyrinck, A.M., Delzenne, N.M., and Burcelin, R. (2008). Changes in Gut Microbiota Control Metabolic Endotoxemia-Induced Inflammation in High-Fat Diet–Induced

Obesity and Diabetes in Mice. Diabetes *57*, 1470–1481. https://doi.org/10.2337/db07-1403.

5. Sugihara, K., Morhardt, T.L., and Kamada, N. (2018). The Role of Dietary Nutrients in Inflammatory Bowel Disease. Front. Immunol. *9*, 3183. https://doi.org/10.3389/fimmu.2018.03183.
6. Kim, K.A., Gu, W., Lee, I.A., Joh, E.H., and Kim, D.H. (2012). High fat diet-induced gut microbiota exacerbates inflammation and obesity in mice via the TLR4 signaling pathway. PLoS One *7*, 1–11. https://doi.org/10.1371/journal.pone.0047713.
7. Li, W.-Z., Stirling, K., Yang, J.-J., and Zhang, L. (2020). Gut microbiota and diabetes: From correlation to causality and mechanism. World J. Diabetes *11*, 293–308. https://doi.org/10.4239/wjd.v11.i7.293.
8. Guo, X., Gao, Y., Zhang, S., Wu, L., Chang, P., Cai, W., Zscheischler, J., Leung, L.R., Small, J., Danabasoglu, G., et al. (2022). Threat by marine heatwaves to adaptive large marine ecosystems in an eddy-resolving model. Nat. Clim. Change *12*, 179–186. https://doi.org/10.1038/s41558-021-01266-5.
9. Ullah, H., Nagelkerken, I., Goldenberg, S.U., and Fordham, D.A. (2018). Climate change could drive marine food web collapse through altered trophic flows and cyanobacterial proliferation. PLoS Biol. *16*, 1–21. https://doi.org/10.1371/journal.pbio.2003446.
10. Brun, P., Stamieszkin, K., Visser, A.W., Licandro, P., Payne, M.R., and Kiørboe, T. (2019). Climate change has altered zooplankton-fuelled carbon export in the North Atlantic. Nat. Ecol. Evol. *3*, 416–423. https://doi.org/10.1038/s41559-018-0780-3.
11. Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T., et al. (2019). Scientists' warning to humanity: Microorganisms and climate change. Nat. Rev. Microbiol. *17*, 569–586. https://doi.org/10.1038/s41579-019-0222-5.
12. Melillo, J.M., Frey, S.D., DeAngelis, K.M., Werner, W.J., Bernard, M.J., Bowles, F.P., Pold, G., Knorr, M.A., and Grandy, A.S. (2017). Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. Science *358*, 101–105. https://doi.org/10.1126/science.aan2874.
13. Wang, S., and Dai, L. (2019). Evolving generalists in switching rugged landscapes. PLoS Comput. Biol. *15*, 1–21. https://doi.org/10.1371/journal.pcbi.1007320.
14. Gilchrist, G.W. (1995). Specialists and generalists in changing environments. I. Fitness landscapes of thermal sensitivity. Am. Nat. *146*, 252–270.
15. Hernandez, D.J., Kiesewetter, K.N., Almeida, B.K., Revillini, D., and Afkhami, M.E. (2023). Multidimensional specialization and generalization are pervasive in soil prokaryotes. Nat. Ecol. Evol. *7*, 1408–1418. https://doi.org/10.1038/s41559-023-02149-y.
16. von Meijenfeldt, F.A.B., Hogeweg, P., and Dutilh, B.E. (2023). A social niche breadth score reveals niche range strategies of generalists and specialists. Nature Ecology & Evolution *7*, 768–781. https://doi.org/10.1038/s41559-023-02027-7.
17. Sriswasdi, S., Yang, C.C., and Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. Nat. Commun. *8*, 1162. https://doi.org/10.1038/s41467-017-01265-1.
18. Maynard, D.S., Bradford, M.A., Covey, K.R., Lindner, D., Glaeser, J., Talbert, D.A., Tinker, P.J., Walker, D.M., and Crowther, T.W. (2019). Consistent trade-offs in fungal trait expression across broad spatial scales. Nat. Microbiol. *4*, 846–853. https://doi.org/10.1038/s41564-019-0361-5.
19. Davison, J., Moora, M., Semchenko, M., Adenan, S.B., Ahmed, T., Akhmetzhanova, A.A., Alatalo, J.M., Al-Quraishy, S., Andriyanova, E., Anslan, S., et al. (2021). Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. New Phytol. *231*, 763–776. https://doi.org/10.1111/nph.17240.
20. Chaloner, T.M., Gurr, S.J., and Bebber, D.P. (2020). Geometry and evolution of the ecological niche in plant-associated microbes. Nat. Commun. *11*, 2955. https://doi.org/10.1038/s41467-020-16778-5.
21. Frey-Klett, P., Burlinson, P., Deveau, A., Barret, M., Tarkka, M., and Sarniguet, A. (2011). Bacterial-fungal interactions: Hyphens between agricultural, clinical, environmental, and food microbiologists. Microbiol. Mol. Biol. Rev. *75*, 583–609. https://doi.org/10.1128/mmbr.00020-11.
22. Deveau, A., Bonito, G., Uehling, J., Paoletti, M., Becker, M., Bindschedler, S., Hacquard, S., Hervé, V., Labbé, J., Lastovetsky, O.A., et al. (2018). Bacterial–fungal interactions: ecology, mechanisms and challenges. FEMS Microbiol. Rev. *42*, 335–352. https://doi.org/10.1093/femsre/fuy008.
23. Feinsinger, P., Spears, E.E., and Poole, R.W. (1981). A simple measure of niche breadth. Ecology *62*, 27–32.
24. Myers, S.S., Gaffikin, L., Golden, C.D., Ostfeld, R.S., Redford, K.H., Ricketts, T.H., Turner, W.R., and Osofsky, S.A. (2013). Human health impacts of ecosystem alteration. Proc. Natl. Acad. Sci. USA *110*, 18753–18760. https://doi.org/10.1073/pnas.1218656110.
25. Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Villarreal Ruiz, L., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., et al. (2014). Global diversity and geography of soil fungi. Science *346*, 1256688. https://doi.org/10.1126/science.1256688.
26. Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Gray, C.L., Scharlemann, J.P.W., Börger, L., Phillips, H.R.P., Sheil, D., Lysenko, I., and Purvis, A. (2016). Global patterns of terrestrial assemblage turnover within and among land uses. Ecography *39*, 1151–1163. https://doi.org/10.1111/ecog.01932.
27. Keller, N.P. (2019). Fungal secondary metabolism: Regulation, function and drug discovery. Nat. Rev. Microbiol. *17*, 167–180. https://doi.org/10.1038/s41579-018-0121-1.
28. Chevrette, M.G., Thomas, C.S., Hurley, A., Rosario-Meléndez, N., Sankaran, K., Tu, Y., Hall, A., Magesh, S., and Handelsman, J. (2022). Microbiome composition modulates secondary metabolism in a multispecies bacterial community. Proc. Natl. Acad. Sci. USA *119*, e2212930119. https://doi.org/10.1073/pnas.2212930119.
29. Dahlstrom, K.M., McRose, D.L., and Newman, D.K. (2020). Keystone metabolites of crop rhizosphere microbiomes. Curr. Biol. *30*, R1131–R1137. https://doi.org/10.1016/j.cub.2020.08.005.
30. Liao, J., Cao, X., Zhao, L., Wang, J., Gao, Z., Wang, M.C., and Huang, Y. (2016). The importance of neutral and niche processes for bacterial community assembly differs between habitat generalists and specialists. FEMS Microbiol. Ecol. *92*, fiw174.
31. Liu, L., Wang, S., and Chen, J. (2021). Transformations from specialists to generalists cause bacterial communities are more stable than micro-eukaryotic communities under anthropogenic activity disturbance. Sci. Total Environ. *790*, 148141.
32. Garrison, C.E., and Field, E.K. (2020). Introducing a "core steel microbiome" and community functional analysis associated with microbially influenced corrosion. FEMS Microbiol. Ecol. *97*, fiaa237.
33. Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. Annu. Rev. Microbiol. *65*, 411–429.
34. Chen, Y.J., Leung, P.M., Wood, J.L., Bay, S.K., Hugenholtz, P., Kessler, A.J., Shelley, G., Waite, D.W., Franks, A.E., Cook, P.L.M., and Greening, C. (2021). Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. ISME J. *15*, 2986–3004.
35. Muller, E.E.L. (2019). Determining Microbial Niche Breadth in the Environment for Better Ecosystem Fate Predictions. mSystems *4*, 000800-19–e119. https://doi.org/10.1128/mSystems.00080-19.
36. Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., Olson, J.B., Erwin, P.M., López-Legentil, S., Luter, H., et al. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. Nat. Commun. *7*, 11870. https://doi.org/10.1038/ncomms11870.
37. Malard, L.A., Anwar, M.Z., Jacobsen, C.S., and Pearce, D.A. (2019). Biogeographical patterns in soil bacterial communities across the Arctic

region. FEMS Microbiol. Ecol. *95*, fiz128. https://doi.org/10.1093/femsec/fiz128.

38. Cobo-Simón, M., and Tamames, J. (2017). Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. BMC Genom. *18*, 499. https://doi.org/10.1186/s12864-017-3888-y.

39. Garcia-Garcera, M., and Rocha, E.P.C. (2020). Community diversity and habitat structure shape the repertoire of extracellular proteins in bacteria. Nat. Commun. *11*, 758. https://doi.org/10.1038/s41467-020-14572-x.

40. Liu, S., García-Palacios, P., Tedersoo, L., Guirado, E., Van Der Heijden, M.G.A., Wagg, C., Chen, D., Wang, Q., Wang, J., Singh, B.K., and Delgado-Baquerizo, M. (2022). Phylotype diversity within soil fungal functional groups drives ecosystem stability. Nat. Ecol. Evol. *6*, 900–909. https://doi.org/10.1038/s41559-022-01756-5.

41. Ortiz-Vera, M.P., Olchanheski, L.R., Da Silva, E.G., De Lima, F.R., Martinez, L.R.D.P.R., Sato, M.I.Z., Jaffé, R., Alves, R., Ichiwaki, S., Padilla, G., and Araújo, W.L. (2018). Influence of water quality on diversity and composition of fungal communities in a tropical river. Sci. Rep. *8*, 14799. https://doi.org/10.1038/s41598-018-33162-y.

42. Ost, K.S., and Round, J.L. (2023). Commensal fungi in intestinal health and disease. Nat. Rev. Gastroenterol. Hepatol. *20*, 723–734. https://doi.org/10.1038/s41575-023-00816-w.

43. Frąc, M., Hannula, S.E., Bełka, M., and Jędryczka, M. (2018). Fungal biodiversity and their role in soil health. Front. Microbiol. *9*, 707. https://doi.org/10.3389/fmicb.2018.00707.

44. Hall, R.A., and Noverr, M.C. (2017). Fungal interactions with the human host: Exploring the spectrum of symbiosis. Curr. Opin. Microbiol. *40*, 58–64. https://doi.org/10.1016/j.mib.2017.10.020.

45. Grossart, H.-P., Van Den Wyngaert, S., Kagami, M., Wurzbacher, C., Cunliffe, M., and Rojas-Jimenez, K. (2019). Fungi in aquatic ecosystems. Nat. Rev. Microbiol. *17*, 339–354. https://doi.org/10.1038/s41579-019-0175-8.

46. McKinney, M.L., and Lockwood, J.L. (1999). Biotic homogenization: A few winners replacing many losers in the next mass extinction. Trends Ecol. Evol. *14*, 450–453. https://doi.org/10.1016/S0169-5347(99)01679-1.

47. Clavel, J., Julliard, R., and Devictor, V. (2011). Worldwide decline of specialist species: Toward a global functional homogenization? Front. Ecol. Environ. *9*, 222–228. https://doi.org/10.1890/080216.

48. Gámez-Virués, S., Perović, D.J., Gossner, M.M., Börschig, C., Blüthgen, N., De Jong, H., Simons, N.K., Klein, A.-M., Krauss, J., Maier, G., et al. (2015). Landscape simplification filters species traits and drives biotic homogenization. Nat. Commun. *6*, 8568. https://doi.org/10.1038/ncomms9568.

49. Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. *10*, 5029. https://doi.org/10.1038/s41467-019-13036-1.

50. Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One *9*, 1–12. https://doi.org/10.1371/journal.pone.0093827.

51. Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., Anslan, S., and Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. Mol. Ecol. *31*, 2769–2795. https://doi.org/10.1111/mec.16460.

52. Anthony, M.A., Bender, S.F., and Van Der Heijden, M.G.A. (2023). Enumerating soil biodiversity. Proc. Natl. Acad. Sci. USA *120*, e2304663120. https://doi.org/10.1073/pnas.2304663120.

53. Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., et al. (2006). Microbial biogeography: Putting microorganisms on the map. Nat. Rev. Microbiol. *4*, 102–112. https://doi.org/10.1038/nrmicro1341.

54. Yu, K., Van Den Hoogen, J., Wang, Z., Averill, C., Routh, D., Smith, G.R., Drenovsky, R.E., Scow, K.M., Mo, F., Waldrop, M.P., et al. (2022). The biogeography of relative abundance of soil fungi versus bacteria in surface topsoil. Earth Syst. Sci. Data *14*, 4339–4350. https://doi.org/10.5194/essd-14-4339-2022.

55. Weinbauer, M.G., and Rassoulzadegan, F. (2007). Extinction of microbes: Evidence and potential consequences. Endanger. Species Res. *3*, 205–215.

56. Louca, S., Polz, M.F., Mazel, F., Albright, M.B.N., Huber, J.A., O'Connor, M.I., Ackermann, M., Hahn, A.S., Srivastava, D.S., Crowe, S.A., et al. (2018). Function and functional redundancy in microbial systems. Nat. Ecol. Evol. *2*, 936–943. https://doi.org/10.1038/s41559-018-0519-1.

57. Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., and Polz, M.F. (2018). High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nat. Commun. *9*, 266. https://doi.org/10.1038/s41467-017-02571-4.

58. Hurtado-McCormick, V., Kahlke, T., Petrou, K., Jeffries, T., Ralph, P.J., and Seymour, J.R. (2019). Regional and Microenvironmental Scale Characterization of the Zostera muelleri Seagrass Microbiome. Front. Microbiol. *10*, 1011. https://doi.org/10.3389/fmicb.2019.01011.

59. Perini, L., Gostinčar, C., and Gunde-Cimerman, N. (2019). Fungal and bacterial diversity of Svalbard subglacial ice. Sci. Rep. *9*, 20230. https://doi.org/10.1038/s41598-019-56290-5.

60. Ling, N., Wang, T., and Kuzyakov, Y. (2022). Rhizosphere bacteriome structure and functions. Nat. Commun. *13*, 836. https://doi.org/10.1038/s41467-022-28448-9.

61. Berg, G., and Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol. Ecol. *68*, 1–13. https://doi.org/10.1111/j.1574-6941.2009.00654.x.

62. Costa, R., Götz, M., Mrotzek, N., Lottmann, J., Berg, G., and Smalla, K. (2006). Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. FEMS Microbiol. Ecol. *56*, 236–249. https://doi.org/10.1111/j.1574-6941.2005.00026.x.

63. Schloss, P.D. (2024). Waste not, want not: Revisiting the analysis that called into question the practice of rarefaction. mSphere *9*, e00355-23. https://doi.org/10.1128/msphere.00355-23.

64. McMurdie, P.J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput. Biol. *10*, e1003531. https://doi.org/10.1371/journal.pcbi.1003531.

65. Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. Bioscience *51*, 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

66. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat. Biotechnol. *35*, 316–319.

67. Gonzalez, A., Navas-Molina, J.A., Kosciolek, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., DeReus, J., Janssen, S., Swafford, A.D., Orchanian, S.B., et al. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. Nat. Methods *15*, 796–798.

68. Taylor, L.J., Abbas, A., and Bushman, F.D. (2020). grabseqs: simple downloading of reads and metadata from multiple next-generation sequencing data repositories. Bioinformatics *36*, 3607–3609.

69. Gaspar, J.M. (2018). NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinf. *19*, 536.

70. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

71. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

CellPress
OPEN ACCESS

72. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857.
73. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.
74. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596.
75. Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., et al. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. *47*, D259–D264.
76. Ebou, A., Koua, D., and Zeze, A. (2021). HyperEx: A tool to extract hypervariable regions from 16S rRNA sequencing data. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.03.455391.
77. Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., et al. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. *4*, 914–919. https://doi.org/10.1111/2041-210X.12073.
78. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Preprint at bioRxiv. https://doi.org/10.1101/221499.
79. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., and Weber, T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. *49*, W29–W35. https://doi.org/10.1093/nar/gkab335.
80. Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J.G., Haendiges, J., Haft, D.H., Hoffmann, M., Pettengill, J.B., Prasad, A.B., Tillman, G.E., et al. (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci. Rep. *11*, 12728. https://doi.org/10.1038/s41598-021-91456-0.
81. Michael Landau, W. (2018). The drake r package: A pipeline toolkit for reproducibility and high-performance computing. J. Open Source Softw. *3*, 550. https://doi.org/10.21105/joss.00550.
82. Dixon, P. (2003). VEGAN, a package of r functions for community ecology. J. Veg. Sci. *14*, 927–930.
83. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217.
84. Vieira-Silva, S., Falony, G., Belda, E., Nielsen, T., Aron-Wisnewsky, J., Chakaroun, R., Forslund, S.K., Assmann, K., Valles-Colomer, M., Nguyen, T.T.D., et al. (2020). Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. Nature *581*, 310–315.
85. Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. *8*, e1002687.
86. Watts, S.C., Ritchie, S.C., Inouye, M., and Holt, K.E. (2019). FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics *35*, 1064–1066.
87. Finn, D. (2020). MicroNiche: Microbial Niche Measurements.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
abind	RSPM (R 4.0.3)	1.4.5
ade4	RSPM (R 4.0.0)	1.7.15
ape	RSPM (R 4.0.2)	5.4.1
assertthat	RSPM (R 4.0.5)	0.2.1
backports	RSPM (R 4.0.2)	1.1.10
base64enc	RSPM (R 4.0.3)	0.1.3
base64url	RSPM (R 4.0.0)	1.4
BayesLogit	RSPM (R 4.0.0)	2.1
BBmisc	RSPM (R 4.0.3)	1.11
Biobase	Bioconductor	2.50.0
BiocGenerics	Bioconductor	0.36.1
biomformat	Bioconductor	1.18.0
Biostrings	Bioconductor	2.58.0
bit	RSPM (R 4.0.5)	4.0.4
bit64	RSPM (R 4.0.5)	4.0.5
blob	RSPM (R 4.0.3)	1.2.1
boot	CRAN (R 4.0.2)	1.3.25
broom	RSPM (R 4.0.2)	0.7.1
car	RSPM (R 4.0.3)	3.0.10
carData	RSPM (R 4.0.3)	3.0.4
cellranger	RSPM (R 4.0.3)	1.1.0
checkmate	RSPM (R 4.0.3)	2.0.0
circlize	RSPM (R 4.0.0)	0.4.10
class	CRAN (R 4.0.2)	7.3.17
ade4	RSPM (R 4.0.0)	1.7.15
ape	RSPM (R 4.0.2)	5.4.1
assertthat	RSPM (R 4.0.5)	0.2.1
backports	RSPM (R 4.0.2)	1.1.10
base64enc	RSPM (R 4.0.3)	0.1.3
base64url	RSPM (R 4.0.0)	1.4
BayesLogit	RSPM (R 4.0.0)	2.1
BBmisc	RSPM (R 4.0.3)	1.11
Biobase	Bioconductor	2.50.0
BiocGenerics	Bioconductor	0.36.1
biomformat	Bioconductor	1.18.0
Biostrings	Bioconductor	2.58.0
bit	RSPM (R 4.0.5)	4.0.4
bit64	RSPM (R 4.0.5)	4.0.5
blob	RSPM (R 4.0.3)	1.2.1
boot	CRAN (R 4.0.2)	1.3.25
broom	RSPM (R 4.0.2)	0.7.1
car	RSPM (R 4.0.3)	3.0.10
carData	RSPM (R 4.0.3)	3.0.4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cellranger	RSPM (R 4.0.3)	1.1.0
checkmate	RSPM (R 4.0.3)	2.0.0
circlize	RSPM (R 4.0.0)	0.4.10
class	CRAN (R 4.0.2)	7.3.17
classInt	RSPM (R 4.0.3)	0.4.3
cli	CRAN (R 4.0.2)	3.6.1
clue	RSPM (R 4.0.0)	0.3.57
cluster	CRAN (R 4.0.2)	2.1.0
clustermq	RSPM (R 4.0.2)	0.8.95.1
coda	RSPM (R 4.0.3)	0.19.4
codetools	CRAN (R 4.0.2)	0.2.16
colorspace	RSPM (R 4.0.0)	1.4.1
ComplexHeatmap	bioc_git2r (@08b4129444508fdea2e510e38068d69db452fb38)	2.4.3
CompQuadForm	RSPM (R 4.0.3)	1.4.3
conquer	RSPM (R 4.0.2)	1.0.2
crayon	RSPM (R 4.0.0)	1.3.4
curl	RSPM (R 4.0.3)	4.3
data.table	RSPM (R 4.0.2)	1.13.0
DBI	RSPM (R 4.0.0)	1.1.0
dbplyr	RSPM (R 4.0.0)	1.4.4
decontam	Bioconductor	1.20.0
DEoptimR	RSPM (R 4.0.3)	1.0.8
dials	RSPM (R 4.0.3)	0.0.9
DiceDesign	RSPM (R 4.0.3)	1.8.1
digest	RSPM (R 4.0.0)	0.6.25
diptest	RSPM (R 4.0.3)	0.75.7
doParallel	RSPM (R 4.0.0)	1.0.15
dotCall64	RSPM (R 4.0.3)	1.0.0
dplyr	RSPM (R 4.0.2)	1.0.2
drake	git2r (https://github.com/ropensci/drake@aa2fee497a449b0f9e849aff887c28aabcd9f062)	7.13.4
e1071	RSPM (R 4.0.0)	1.7.3
ellipsis	RSPM (R 4.0.3)	0.3.1
farver	RSPM (R 4.0.3)	2.0.3
fastDummies	RSPM (R 4.0.2)	1.6.2
fastmatch	RSPM (R 4.0.3)	1.1.0
fields	RSPM (R 4.0.2)	11.5
filelock	RSPM (R 4.0.0)	1.0.2
flexmix	RSPM (R 4.0.0)	2.3.15
FNN	RSPM (R 4.0.3)	1.1.3
forcats	RSPM (R 4.0.0)	0.5.0
foreach	RSPM (R 4.0.0)	1.5.0
foreign	CRAN (R 4.0.2)	0.8.80
Formula	RSPM (R 4.0.0)	1.2.3
fpc	RSPM (R 4.0.2)	2.2.8
fs	RSPM (R 4.0.3)	1.5.0
furrr	RSPM (R 4.0.0)	0.1.0
future	RSPM (R 4.0.2)	1.19.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
generics	RSPM (R 4.0.0)	0.0.2
getopt	RSPM (R 4.0.0)	1.20.3
GetoptLong	RSPM (R 4.0.2)	1.0.3
ggalluvial	RSPM (R 4.0.2)	0.12.5
ggfittext	RSPM (R 4.0.2)	0.10.2
ggforce	RSPM (R 4.0.2)	0.4.2
ggfortify	RSPM (R 4.0.2)	0.4.11
ggnewscale	RSPM (R 4.0.2)	0.4.3
ggplot2	RSPM (R 4.0.1)	3.3.2
ggpubr	RSPM (R 4.0.3)	0.4.0
ggraph	RSPM (R 4.0.2)	2.0.3
ggrepel	RSPM (R 4.0.2)	0.8.2
ggsci	RSPM (R 4.0.3)	2.9
ggsignif	RSPM (R 4.0.3)	0.6.0
ggtext	RSPM (R 4.0.0)	0.1.0
ggvenn	git2r (https://github.com/yanlinlin82/ggvenn@b7ff54baf91e2355432b3a9e05bef80690ace706)	0.1.9
glmnet	RSPM (R 4.0.1)	4.0.2
GlobalOptions	RSPM (R 4.0.3)	0.1.2
globals	RSPM (R 4.0.2)	0.13.0
glue	RSPM (R 4.0.5)	1.4.2
gower	RSPM (R 4.0.3)	0.2.2
GPfit	RSPM (R 4.0.3)	1.0.8
graphlayouts	RSPM (R 4.0.2)	0.7.0
gridExtra	RSPM (R 4.0.3)	2.3
gridtext	RSPM (R 4.0.2)	0.1.1
gtable	RSPM (R 4.0.3)	0.3.0
haven	RSPM (R 4.0.2)	2.3.1
here	RSPM (R 4.0.0)	0.1
Hmisc	RSPM (R 4.0.2)	4.4.1
hms	RSPM (R 4.0.0)	0.5.3
Hmsc	RSPM (R 4.0.0)	3.0.6
htmlTable	RSPM (R 4.0.3)	2.1.0
htmltools	RSPM (R 4.0.1)	0.5.0
htmlwidgets	RSPM (R 4.0.2)	1.5.2
httr	RSPM (R 4.0.3)	1.4.2
huge	RSPM (R 4.0.2)	1.3.4.1
igraph	RSPM (R 4.0.3)	1.2.6
infer	RSPM (R 4.0.2)	0.5.3
ipred	RSPM (R 4.0.3)	0.9.9
IRanges	Bioconductor	2.24.1
iterators	RSPM (R 4.0.0)	1.0.12
jpeg	RSPM (R 4.0.3)	0.1.8.1
jsonlite	RSPM (R 4.0.2)	1.7.1
kernlab	RSPM (R 4.0.3)	0.9.29
KernSmooth	CRAN (R 4.0.2)	2.23.17
kknn	RSPM (R 4.0.0)	1.3.1
knitr	RSPM (R 4.0.2)	1.3

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
latex2exp	RSPM (R 4.0.0)	0.4.0
lattice	CRAN (R 4.0.2)	0.20.41
latticeExtra	RSPM (R 4.0.3)	0.6.29
lava	RSPM (R 4.0.2)	1.6.8
lhs	RSPM (R 4.0.2)	1.1.1
life cycle	RSPM (R 4.0.3)	0.2.0
listenv	RSPM (R 4.0.3)	0.8.0
lme4	RSPM (R 4.0.2)	1.1.23
lubridate	RSPM (R 4.0.2)	1.7.9
magrittr	RSPM (R 4.0.0)	1.5
maps	RSPM (R 4.0.3)	3.3.0
MASS	CRAN (R 4.0.2)	7.3.51.6
Matrix	CRAN (R 4.0.2)	1.2.18
MatrixModels	RSPM (R 4.0.3)	0.4.1
matrixStats	RSPM (R 4.0.2)	0.57.0
mclust	RSPM (R 4.0.0)	5.4.6
mcmc	RSPM (R 4.0.3)	0.9.7
MCMCpack	RSPM (R 4.0.2)	1.4.9
memoise	RSPM (R 4.0.0)	1.1.0
meta	RSPM (R 4.0.2)	4.15.1
metafor	RSPM (R 4.0.3)	2.4.0
mgcv	CRAN (R 4.0.2)	1.8.31
minqa	RSPM (R 4.0.2)	1.2.4
mlr	RSPM (R 4.0.2)	2.18.0
modeldata	RSPM (R 4.0.2)	0.0.2
modelr	RSPM (R 4.0.3)	0.1.8
modeltools	RSPM (R 4.0.3)	0.2.23
multtest	Bioconductor	2.46.0
munsell	RSPM (R 4.0.3)	0.5.0
mvtnorm	RSPM (R 4.0.3)	1.1.1
nlme	CRAN (R 4.0.2)	3.1.148
nloptr	RSPM (R 4.0.3)	1.2.2.2
nnet	CRAN (R 4.0.2)	7.3.14
openxlsx	RSPM (R 4.0.2)	4.2.2
optparse	RSPM (R 4.0.0)	1.6.6
pander	CRAN (R 4.0.2)	0.6.5
parallelMap	RSPM (R 4.0.0)	1.5.0
ParamHelpers	RSPM (R 4.0.3)	1.14
parsnip	RSPM (R 4.0.2)	0.1.3
patchwork	RSPM (R 4.0.2)	1.0.1
pbmcapply	RSPM (R 4.0.3)	1.5.0
pdist	RSPM (R 4.0.3)	1.2
permute	RSPM (R 4.0.3)	0.9.5
phyloseq	bioc_git2r (@7829e59a5052b2dafd1a34036d67e32e09fd76b0)	1.32.0
pillar	RSPM (R 4.0.2)	1.4.6
pkgconfig	RSPM (R 4.0.3)	2.0.3
plyr	RSPM (R 4.0.2)	1.8.6
png	RSPM (R 4.0.3)	0.1.7

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
polyclip	RSPM (R 4.0.3)	1.10.0
prabclus	RSPM (R 4.0.3)	2.3.2
prettyunits	RSPM (R 4.0.3)	1.1.1
pROC	RSPM (R 4.0.2)	1.16.2
prodlim	RSPM (R 4.0.2)	2019.11.13
progress	RSPM (R 4.0.3)	1.2.2
pulsar	RSPM (R 4.0.2)	0.3.7
purrr	RSPM (R 4.0.5)	0.3.4
qs	RSPM (R 4.0.2)	0.23.3
quantreg	RSPM (R 4.0.2)	5.73
R6	RSPM (R 4.0.0)	2.4.1
ranger	RSPM (R 4.0.2)	0.12.1
RANN	RSPM (R 4.0.3)	2.6.1
RApiSerialize	RSPM (R 4.0.0)	0.1.0
RColorBrewer	RSPM (R 4.0.3)	1.1.2
Rcpp	RSPM (R 4.0.2)	1.0.5
RcppParallel	RSPM (R 4.0.3)	5.0.2
readr	RSPM (R 4.0.2)	1.4.0
readxl	RSPM (R 4.0.2)	1.3.1
recipes	RSPM (R 4.0.2)	0.1.13
reprex	RSPM (R 4.0.0)	0.3.0
reshape2	RSPM (R 4.0.2)	1.4.4
rhdf5	Bioconductor	2.34.0
rhdf5filters	Bioconductor	1.2.1
Rhdf5lib	Bioconductor	1.12.1
RhpcBLASctl	RSPM (R 4.0.3)	0.20.137
rio	RSPM (R 4.0.3)	0.5.16
rjson	RSPM (R 4.0.3)	0.2.20
rlang	RSPM (R 4.0.2)	0.4.8
rnaturalearth	RSPM (R 4.0.0)	0.1.0
rnaturalearthdata	RSPM (R 4.0.0)	0.1.0
robustbase	RSPM (R 4.0.0)	0.93.6
ROSE	RSPM (R 4.0.0)	0.0.3
rpart	CRAN (R 4.0.2)	4.1.15
rprojroot	RSPM (R 4.0.0)	1.3.2
rsample	RSPM (R 4.0.3)	0.0.8
RSQLite	RSPM (R 4.0.2)	2.2.1
rstatix	RSPM (R 4.0.3)	0.6.0
rstudioapi	RSPM (R 4.0.0)	0.11
rvest	RSPM (R 4.0.3)	0.3.6
S4Vectors	Bioconductor	0.28.1
scales	RSPM (R 4.0.3)	1.1.1
sessioninfo	CRAN (R 4.0.2)	1.2.2
sf	RSPM (R 4.0.2)	0.9.6
shape	RSPM (R 4.0.3)	1.4.5
sp	RSPM (R 4.0.2)	1.4.4
spam	RSPM (R 4.0.0)	2.5.1
SparseM	RSPM (R 4.0.3)	1.78

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SpiecEasi	git2r (https://github.com/zdk123/SpiecEasi@c463727a51d0df34db0c670d3b170195bb3d4eba)	1.1.2
statmod	RSPM (R 4.0.0)	1.4.34
storr	RSPM (R 4.0.0)	1.2.1
stringfish	RSPM (R 4.0.2)	0.14.2
stringi	RSPM (R 4.0.3)	1.5.3
stringr	RSPM (R 4.0.3)	1.4.0
survival	CRAN (R 4.0.2)	3.1.12
themis	RSPM (R 4.0.2)	0.1.2
tibble	RSPM (R 4.0.2)	3.0.3
tidygraph	RSPM (R 4.0.2)	1.2.0
tidymodels	RSPM (R 4.0.2)	0.1.1
tidyr	RSPM (R 4.0.2)	1.1.2
tidyselect	RSPM (R 4.0.3)	1.1.0
tidyverse	RSPM (R 4.0.3)	1.3.0
timeDate	RSPM (R 4.0.3)	3043.102
truncnorm	RSPM (R 4.0.3)	1.0.8
tune	RSPM (R 4.0.2)	0.1.1
tweenr	RSPM (R 4.0.2)	1.0.1
txtq	RSPM (R 4.0.2)	0.2.3
unbalanced	RSPM (R 4.0.0)	2
units	RSPM (R 4.0.2)	0.6.7
vctrs	CRAN (R 4.0.2)	0.3.6
vegan	RSPM (R 4.0.0)	2.5.6
VGAM	RSPM (R 4.0.0)	1.1.3
viridis	RSPM (R 4.0.3)	0.5.1
viridisLite	RSPM (R 4.0.3)	0.3.0
withr	RSPM (R 4.0.2)	2.3.0
workflows	RSPM (R 4.0.3)	0.2.1
writexl	RSPM (R 4.0.2)	1.3.1
xfun	RSPM (R 4.0.2)	0.18
xml2	RSPM (R 4.0.3)	1.3.2
XVector	Bioconductor	0.30.0
yaml	RSPM (R 4.0.3)	2.2.1
yardstick	RSPM (R 4.0.3)	0.0.7
zip	RSPM (R 4.0.3)	2.1.1
zlibbioc	Bioconductor	1.36.0
multiqc	https://multiqc.info/	1.8
fastqc	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	0.11.9
ngmerge	https://github.com/jsh58/NGmerge	0.3
grabseqs	https://github.com/louiejtaylor/grabseqs	0.7
trimmomatic	http://www.usadellab.org/cms/?page=trimmomatic	0.39
qiime2	https://qiime2.org/	2020.8
snakemake	https://snakemake.readthedocs.io/en/stable/	7.18.1
grabseqs	https://github.com/louiejtaylor/grabseqs	0.7
sra-tools	https://github.com/ncbi/sra-tools	3.0.0
ITSx	https://microbiology.se/software/itsx/	1.1.3
BBTools	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/	39.01

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
BWA	https://bio-bwa.sourceforge.net/	0.7
SeqKit	https://bioinf.shenwei.me/seqkit/	2.3.1
bedtools	https://bedtools.readthedocs.io/en/latest/	2.30.0
NCBI Datasets	https://github.com/ncbi/datasets	14.6.1
antiSMASH	https://antismash.secondarymetabolites.org/	6.1.1
AMRFinderPlus	https://github.com/ncbi/amr	3.11.8
in silico pcr tool	https://github.com/egonozer/in_silico_pcr	0.5.1
Other		
Silva	https://www.arb-silva.de/	132
Unite	https://unite.ut.ee/	8.2 dynamic

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Gianni Panagiotou (gianni.panagiotou@leibniz-hki.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- 16S and ITS data were obtained from public databases and are available under the accession numbers listed in the Data S1.
- All original code has been deposited at https://github.com/bioinformatics-leibniz-hki/its-16s and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Sample selection

Included studies were retrieved by querying NCBI BioProject with the terms 'bacteria' and 'fungi' in any field. Only BioSamples with both 16S rRNA and ITS amplicon sequencing data were considered for the concurrent analysis of both kingdoms. We also required that 16S and ITS sequences were deposited under a unified BioSample ID to definitely link patterns in bacterial and fungal diversity. This excluded some additional projects, as the 16S and ITS sequences were deposited under different BioSample IDs. We used both the identifier and attributes of the BioSample, such as aliases and library names, to map fungal and bacterial read files to a sample using a custom script. Samples were associated to an environment (aquatic, host, or soil) using manual curation of associated publications and BioSample attributes provided by the depositor. The three environments were further subdivided into 17 habitat groups based on the body part and/or the ecoregion of the sampling location for host and other samples, respectively.[65] Habitats with less than five samples were pooled together.

Generation of genus-level abundance profiles

Genus-level abundance profiles were calculated using a custom nextflow pipeline.[66] Briefly, reads were downloaded from NCBI SRA using grabseqs, except for the American Gut Project, which was downloaded from Qiita.[67,68] Paired-end reads were merged using NGmerge.[69] Quality Control (QC) and adapter removal was performed using trimmomatic with a minimum Phread quality of 20 and a minimal read length of 100.[70] Quality was assessed using FastQC and MultiQC.[71] Subsequent steps were performed using QIIME2.[72] Reads were dereplicated following closed-reference OTU picking for both kingdoms separately using VSEARCH with a 97% identity threshold.[73] For taxonomic annotation, SILVA 132 97% consensus and UNITE 8.2 dynamic databases were used for bacteria and fungi, respectively.[74,75] As detection of archaea was highly variable across the 16S datasets, any counts assigned to archaea were removed prior to downstream analyses. Relatedly, co-amplifying plant and non-fungal microbial eukaryote sequences were excluded from analysis as we used a version of the UNITE database that only included fungal sequences. Following quality control, a total of 1,580 samples were selected for downstream analyses.

Discovery of sample rRNA amplified region

Multiple rRNA regions were used to characterize microbial diversity as the study dataset is composed of many sequencing projects. When available, the specific rRNA region amplified was obtained from deposited metadata or linked publication. For BioProjects where this information was not available, the following was performed. As the SILVA database (v138.1) contains full length bacterial rRNA sequence, the hypervariable regions (e.g., V1-V3, V4-V5) from each taxon was extracted using the in silico pcr tool (https://github.com/egonozer/in_silico_pcr) with primers described in.[76]Amplicon sequence data from each project was then aligned to each variable region using BWA-MEM v.0.7 and contig coverage quantified using BBTools v.39.01. The 16S variable region with the highest percent coverage was taken as the region amplified in the study. For the ITS amplicon data, ITSx 1.1.13[77] was used to extract the ITS1 and/or ITS2 consensus from sequence reads. The BioProject primers identified through this analysis, as well as those retrieved from association publications is listed in Data S1.

Abundance correlation between varying rRNA amplicons

To calculate the correlation in genus abundances between the differing rRNA regions amplified, genera that were detected in all three environments were considered and samples aggregated into whether they included sequence from the V1-V4 regions or V4-V5 regions for bacteria and ITS1 or ITS2 for fungi. For each rRNA category, Pearson's correlation coefficients were calculated for genus abundance in each environment. The similarity between the correlation matrices (V1-V4 and V4-V5 for bacteria and ITS1 and ITS2 for fungi) was then calculated by transforming the upper triangle of each correlation matrix into a vector and calculating the correlation coefficient between the two.

Decontamination of low biomass projects

Low biomass samples are susceptible to amplification of low-level contaminating sequences in extraction kits and other reagents. To control for this, low biomass projects, including all aquatic project and low biomass host habitats like the long were decontaminated using the decontam R package.[78] The majority of low biomass projects contained blanks or negative controls and were decontaminated using the ''prevalence'' method of decontam. In this method, OTUs observed in the non-sample control samples are labeled as contaminants and removed from the abundance table. One low biomass project did not have sequencing controls but provided us with the input DNA concentrations used for library preparation, as determined by Qubit (Invitrogen). This allowed the samples from this project to be decontamined using the ''frequency'' mode of decontam. In this mode, contaminating OTUs are identified and removed from downstream analyses as their abundance inversely correlates with the input DNA concentration, rather than independent of it. The mode of deconam used for each project is indicated in Figure 2.

Genome features of generalists and specialists

As amplicon sequence data is based on maker genes, deposited genomes were used to characterize functional traits associated with the genomes of generalists and specialists. The generalists and specialist genera were queried in NCBI RefSeq. Of the resulting genome list, all genomes or up to 60 randomly selected genomes if more were available were selected for each genus. This resulted in genomes for 2,328 bacterial generalists, 117 fungal generalists, 471 bacterial specialists, and 5 fungal specialists. Genome size and number of coding regions were obtained from the NCBI metadata. For the calculation of the number and type of biosynthetic gene clusters in each genome, AntiSMASH v6.1.1 was used.[79] Antimicrobial and stress resistance genes were predicted in bacterial genomes using AMRFinderPlus.[80] Counts were divided by genome length for normalization.

QUANTIFICATION AND STATISTICAL ANALYSIS

Workflow and statistical analysis

Analyses were performed using a custom drake pipeline[81] built using the programming language R 4.0.2. Briefly, abundances obtained from OTU profiling were total-sum-scaled (TSS) and pooled at genus rank. All tools were used with default parameters if not explicitly specified.

Diversity

Alpha diversity was estimated using Shannon and Chao1 metrics with the phyloseq and vegan packages.[82,83] To quantify the contributions of a bacterial community profile with the fungal one and vice versa, we used linear and unsupervised Canonical Correlation Analysis, as implemented in the function CCorA of the vegan R package.[82] P-values were obtained using blocked permutations to control for the habitat and to reduce assumptions of the test. Supervised constrained ordination was performed using stepwise Distance-based Redundancy Analysis (dbRDA) adapted from.[84] This analysis shows linear relationships between bacterial dissimilarities and abundances of selected explanatory fungal genera (and vice versa). An optimal subset of up to 50 explanatory genera of the other kingdom was computed using a stepwise feedforward approach, as implemented in the ordistep function of the vegan R package.[82]

Co-abundance networks

SparCC, as implemented in FastSpar, was used to assess correlation between taxa pairs for each environment separately.[85,86] Both kingdoms were pooled together, allowing for the identification of interkingdom correlations. Only genera found in all

three environments were considered for pairwise correlation. Node topology metrics were calculated using the R package igraph.

Generalists and specialists

Genera were defined as generalists if they were found in at least 40% of samples in at least one habitat from each environment (host, soil, aquatic) with a relative abundance of at least 0.01%. Conversely, genera were defined as specialists if they were found in at least 40% of samples in one habitat and less than 5% of samples in all other habitats using the same abundance threshold as for generalists. Levins' niche breadth index was calculated as implemented in the R package MicroNiche.[87] Social niche breadth (SNB) was calculated as in von Meijenfeldt et al., Nature Ecology and Evolution (2023) using the data from the MGNIFY database analyzed in this study.