A Global Survey of Host, Aquatic, and Soil Microbiomes Reveals Ecological Properties Shared between Bacterial and Fungal Generalists

Daniel Loos (1), Ailton Pereira da Costa Filho (2),
Amelia E. Barber (2), Gianni Panagiotou (1,3,4)

(1) Research Group Systems Biology and Bioinformatics, 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) Institute of Microbiology, Friedrich Schiller University. Jena, Germany.

The State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong. Hong Kong SAR, China.

(4) Department of Medicine, The University of Hong Kong. Hong Kong SAR, China.

correspondence: gianni.panagiotou@leibniz-hki.de or amelia.barber@uni-jena.de

Abstract

Microbiome engineering is a fast-evolving area with relevance for human health, agriculture, and climate management solutions. Despite significant efforts in engineering microbiomes to repair dysbiotic communities, new microbes often fail to establish and/or alter ecosystem function. To identify bacterial and fungal genera with the desired ability to stably incorporate in diverse microbial communities, we retrieved paired 16S and ITS rRNA amplicon sequence data from 1,580 host, soil, and aquatic samples and explored the ecological patterns of the 2,977 bacteria and 1,740 fungal genera detected across all samples. Through this large-scale analysis, we revealed that a small number of bacterial and fungal generalists with high

prevalence across all environments positively contribute to the taxonomic diversity of their respective kingdom and explain a large percentage of the variation in the cross-kingdom community structure. We also observed that bacterial and fungal generalists have a significantly higher abundance compared to specialists, or genera whose prevalence was strongly associated with a single habitat - possibly due to their ability to stimulate positive associations with other highly prevalent genera. These findings can streamline existing strategies to identify bacterial and fungal inoculants with a higher probability to establish in recipient ecosystems and confer noticeable changes in their structure and function.

Introduction

With only a handful of exceptions, environments, plants, and animals are colonized with communities of microbial organisms, termed the microbiome. These microbial consortia play critical roles in the function and health of their hosts and habitats. For example, the microbiota contributes remarkably to the earth's biomass and plays an important role in shaping soil and oceanic biogeochemical processes.[1–3] Microbial communities also play critical roles in host's health. In humans, the microbiome helps digest food, produces essential vitamins, educates, and regulates the immune system, and protects against more harmful microbes.[4–11] Similarly, in plants, mycorrhizal fungi are critical symbionts for growth[12] and mediate disease-resistance.[13] However, these microbial communities can be altered or disturbed, leading to functional changes and dysbiosis.

An exciting new frontier in microbiome research is the rational editing of disbalanced microbial communities. This has been demonstrated in humans through both fecal microbiota transplants that cause global changes in the microbiome structure[14] and more targeted approaches such as the administration of probiotics that shift the gut microbiome towards a beneficial state.[15] Similar strategies are also being explored in agriculture through biological control agents and the direct application of non-pathogenic microbes to fight plant disease caused by *Pythium ultimum*, *Gaeumannomyces graminis*, and *Fusarium oxysporum*,[16] as well as the application of bioactive signaling compounds that attract beneficial microbes such as

56 *Bacillus* and *Lysinibacillus*.[17] However, the selection of species to be introduced
57 directly or serve as chassis strains to modify the structure and function of a dysbiotic
58 microbiome is challenging and in many cases based largely on their ability to be
59 easily cultivated and known phenotypic properties. Moreover, their ecological
60 properties are rarely considered in bacteria and are completely ignored in fungi.

61 To this end, we performed a large-scale meta-analysis of community sequencing
62 data sets from host, soil, and aquatic environments with paired bacterial and fungal
63 characterization to shed light on the ecological properties of the genera present and
64 their potential use as global or habitat-specific modifiers of microbial communities.
65 We focused specifically on three main properties: (i) the capacity of bacterial and
66 fungal genera to adapt to diverse environments (generalists) or specific
67 environments (specialists), which may be important for long-term establishment
68 when introducing new species into an environment; (ii) the relative abundance of
69 bacterial and fungal generalists and specialists which reflects their competitive
70 colonization potential and is a critical factor for both native and chassis strains to
71 make a functional impact on the ecosystem; and (iii) whether their presence in a
72 habitat might trigger global changes in inter- and cross-kingdom population
73 structure. These global changes can have the benefit of completely restructuring a
74 disbalanced community when this is the desired outcome. However, they also pose
75 the risk of establishing a completely new microbiome state with unpredictable
76 consequences.

77 Results

78 Environmental specificity of bacterial and fungal communities

79 To perform a cross-habitat comparison of bacterial-fungal communities, we
80 analyzed paired 16S and ITS rRNA amplicon sequence data from 1,580 aquatic, host-
81 , and soil - associated samples originating from 27 habitats deposited in NCBI
82 databases.

83 Samples were collected from Europe, Asia, and the Americas between 2010 and
84 2018 (Figure 1A & B). Of the 1,580 samples, 871 originated from host- (including

both mammalian and non-mammalian hosts, such as insects), 494 from soil-, and 215 from aquatic-associated environments. The habitats that contributed largest number of samples for each environment were gut (N=287) and skin (N=68) for the host, temperate (N=498) and conifer forests (N=147) for the soil and large lakes (N=87) followed by other freshwater (N=71) for the aquatic (see Supplementary table 1 for details of all projects). Taxonomic annotation of bacterial and fungal communities was performed using SILVA and UNITE databases, respectively. Rarefaction curves of each habitat indicated that the majority of habitats adequately capture the diversity of both the bacterial and fungal communities (Figure 1C). In total 2,977 bacteria 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). Soil and aquatic environments had the highest number of shared bacterial genera (N=1,662), followed by host-soil (N=1,483) and host-aquatic (1,226). The pattern was different for shared fungal genera, with host-soil sharing the most (N=884), followed by aquatic-soil (N=205) and host-aquatic (N=189). This trend remained after controlling for the different number of samples across the three environments in 842 and 998 out of 1000 random down sampled subsets for bacteria and fungi, respectively.

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 (Figure 1E & F). The most prevalent phyla 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 two 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% (Figure 1 E & F). The opposite trend was observed for host-associated microbes, with 3% and 8% of unique bacteria and fungi, respectively in this environment.

116 We subsequently compared the relative abundance of genera that were found in all
117 environments or were uniquely detected in soil-, host- or aquatic-associated
118 environments. Bacterial genera detected in all three environments were significantly
119 more abundant (Wilcoxon Rank Sum test, $p < 0.001$) than genera uniquely detected
120 in one of the environments (Figure 1G). A similar pattern was observed with fungi.
121 However, a notable exception was the relative abundance of fungi uniquely detected
122 in aquatic samples, which was the highest among both the common genera as well as
123 those uniquely detected in soil- or host-associated environments (Figure 1H). Taken
124 together, we find that soil bacteria and fungi show a higher degree of specificity and
125 genera commonly detected in all environments were also more abundant in these
126 microbial communities.

127 Bacterial and fungal generalists are more abundant than specialists

128 Although a classical concept in ecology, generalists and specialists have remained
129 challenging to objectively identify and quantify. To define multi-kingdom generalists
130 and specialists, we set the following criteria: generalists are genera found with high
131 prevalence (>40%) in at least one habitat from each of the three environments (host,
132 aquatic, soil). Conversely, specialists are genera with a high prevalence (>40%) in
133 one habitat and low prevalence (<5%) in every other. A relative abundance (RA) cut-
134 off of 0.01% was used for determining whether a genus was present in a sample.
135 Using this approach, we detected 48 bacterial generalists and 30 specialists (Figure
136 2A). To confirm our definition of generalists and specialists, we calculated Levins'
137 niche breadth indices. Generalists showed significantly higher Bn values compared
138 to specialists (Wilcoxon Rank Sum test, $p < 0.001$; Supplementary Figure 1). All
139 specialists and all generalists, with the exception of Christensenellaceae R7, were
140 above the detection limit and had a significant niche breadth signal after Benjamini-
141 Hochberg adjustment.[18]

142 We did not observe a clear phylogenetic origin for generalists or specialists (Chi
143 squared test, $p > 0.05$), indicating their roles as generalists and specialists evolved
144 independently. Each of the top five bacterial generalists were detected in more than
145 50% of the 1,580 samples. Among them, the most prevalent was *Pseudomonas* which
146 was detected in 52%, 70% and 89% of host, soil and aquatic samples respectively,

147 followed by *Bacillus* (33%, 71%, 35%) and *Bradyrhizobium* (17%, 73%, 35%)
148 (Supplementary table 2). Concerning the specialists, the most extreme bacterial
149 genera came from the soil. While *Gryllotalpicola* and *Anaerovibrio* were found in
150 >91% of biochar samples, the prevalence dropped to 0.1% on average for non-soil
151 environments (Supplementary table 2). Specialists were also found in host- and
152 aquatic-associated environments. For example, *Acetatifactor* was found in 80% of
153 samples from the murine gut, but had a prevalence less than 3% in all other habitats.
154 The genus *Leptospira* was found in 83% of samples of the Cuyahoga River, but had a
155 prevalence less than 2% in all other habitats. Interestingly, when comparing the
156 relative abundance of generalists and specialists, we observed that both bacterial
157 and fungal generalists had a significantly higher abundance (Figure 2B, Wilcoxon
158 Rank Sum test, $p < 0.001$). This pattern remained even when we used stricter and
159 looser thresholds to define generalists and specialists (Supplementary Figure 2).
160 This finding confirms the pattern observed above (Figure 1G & H), suggesting that
161 independently of how groups are defined, genera that can colonize diverse
162 environments are usually able to outcompete niche-specific genera.

163 When looking at the fungal kingdom, the number of generalists was much lower and
164 only *Aspergillus*, *Malassezia*, *Aureobasidium*, and *Cortinarius* satisfied the criteria of a
165 generalist (Figure 2A). Among these, *Aspergillus* had the highest overall prevalence
166 among all samples with 38%, 52% and 12% in the host, soil, and aquatic samples,
167 respectively. From the 19 fungal specialists, *Chrysanthotrichum* and
168 *Mycocentrospora* were the most niche-specific, with prevalences of 68% and 48% in
169 temperate and conifer forests, but a mean prevalence of ≤0.1% in all other habitats.
170 Only two of the 19 fungal specialists (11%) originated from outside soil
171 environments (*Vuilleminia* and *Seimatosporium*). As with bacterial genera, the
172 relative abundance of fungal generalists was significantly higher than that of fungal
173 specialists (Figure 2C, Wilcoxon Rank Sum test, $p < 0.001$).

174 To gain insight into the ecological lifestyle that allows generalists to dominate
175 microbial communities in abundance, even in diverse environments, we further
176 explored their intra and inter-kingdom interaction patterns. Using the 1,188 bacteria
177 and 184 fungal genera detected in all three environments, we constructed individual
178 co-abundance networks for soil, host, and aquatic environments (see Methods for

179 details). Despite only using genera commonly detected in all three environments, the
180 topological characteristics of the networks were highly distinct, as measured by
181 significant differences in betweenness and Kleinberg's hub node centrality scores
182 (Wilcoxon test, $p < 0.00$; Supplementary Figure 3). Despite the differences in
183 topology, we could still compile subnetworks of inter- and intra-kingdom
184 correlations found jointly among host-soil, host-aquatic and/or soil-aquatic
185 environments. Strikingly, 45 of the 48 bacterial generalists and all 4 fungal
186 generalists were part of those subnetworks, which are characterized by a higher
187 number of positive than negative edges (Figure 2D). The ratio of positive to negative
188 edges was higher in correlations involving a generalist (2.5) compared to all other
189 edges (2.2). When we looked for interactions between genera found in all three
190 environments, we identified 43 such edges that all represented positive interactions
191 between bacteria and included 21 generalists. These findings suggest that the
192 success of generalists in colonizing diverse environments and achieving high
193 abundances may be attributable to their ability to avoid competitive and elicit
194 positive interactions with other highly prevalent genera.

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

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

206 We subsequently shifted our focus to inter-kingdom interactions which are often
207 overlooked in microbial ecology studies and examined bacterial generalists for a role
208 in shaping the mycobiome community structure and *vice versa*. As expected, we
209 observed a significant separation between the soil, host, and aquatic micro- and

210 mycobiome beta diversity by Bray-Curtis dissimilarity (Figure 3C, PERMANOVA, $p <$
211 0.001 for both bacteria and fungi). Constrained ordination revealed a significant,
212 linear relationship between bacterial Bray-Curtis dissimilarity and fungal
213 community composition and *vice versa* (dbRDA, ANOVA $p < 0.04$ for all explanatory
214 genera of the other kingdom in a multivariate model). Bacteria genera could explain
215 an extensive part of the mycobiome variation observed in the three environments
216 with a partial R2 of 25% by dbRDA. Of the bacterial genera, *Conexibacter*, *Bacillus*,
217 and *Lysobacter* had the highest explanatory power on mycobiome variation (Fig 3D).
218 Interestingly, six out of the top ten explanatory bacteria genera in the dbRDA were
219 generalists. Similarly, fungal genera explained 26% of the microbiome variation
220 between host, soil, and aquatic samples, with *Mortierella*, *Trichocladium*, and
221 *Candida* having the highest explanatory power (Fig 3D). Among the top ten
222 explanatory genera was one of the four fungal generalists - *Malassezia*.

223 Altogether, our analysis indicates that bacterial and fungal generalists are attractive
224 candidates for re-structuring microbiome communities given that they can reach
225 high abundances, they contribute positively to the taxonomic (alpha) diversity of
226 their kingdom- an ecological characteristic often associated with healthy
227 environments, and they can also contribute to shaping cross-kingdom microbial
228 structures.

229 Discussion

230 Microbiome engineering is a fast-growing field that aims to improve ecosystem
231 function through the manipulation of its microbial composition. Microbiome
232 engineering has had several significant successes in human health, including in the
233 infant gut,[19,20] bioremediation,[21] wastewater,[22] and agriculture.[23] However, the
234 discipline still faces marked challenges - namely that inoculants often fail to
235 establish or modify ecosystem functioning over extended periods of time.[24] These
236 failures are likely due to microbial dispersal, as well as environmental and biotic
237 constraints to organism establishment.[25] A widely-accepted ecological framework
238 posits that inoculant success is determined by its dispersal potential, compatibility
239 with the new environment, and the range of interactions that occur between the

240 introduced and resident organisms. Given the broad range of potential barriers,
241 complementing existing microbiome engineering solutions with an ecology-
242 motivated decision tree that prioritizes inoculants is an urgent priority.[25]

243 With this in mind, we performed a global survey of host, aquatic, and soil
244 microbiomes to reveal ecological properties of bacterial and fungal genera that may
245 promote or deter their establishment in an environment and/or restore the richness
246 and diversity of microbial communities in diseased ecosystems. The metagenomic
247 analysis of >1,580 paired host, soil, and aquatic microbiomes and mycobiomes
248 identified approximately 3,000 bacterial and 1,700 fungal genera – highlighting the
249 daunting task of selecting organisms as potential engineering solutions. From this,
250 we used the concept of generalists and specialists to limit the selection space for
251 microbial community engineering.

252 While the concept of generalists and specialists in ecology is not new, it has mostly
253 been applied in specific habitats[26–30] and not on a global scale such as this. Moreover,
254 it has rarely included eukaryotic organisms such as fungi[27] and has not been used
255 towards a cross-kingdom understanding of the biotic interactions shaping microbial
256 communities. We demonstrate that many of the important barriers in microbiome
257 engineering are more likely to be overcome when selecting inoculants from a short
258 list of 48 bacterial and 4 fungal generalists. Besides their global presence, these
259 genera reach significantly higher abundances and contribute positively to the
260 richness and diversity of their respective kingdom. Moreover, six bacterial
261 generalists, including *Bacillus*, *Lysobacter*, *Escherichia* and *Gemmatimonas* and one
262 fungal generalist, *Malassezia*, harbor additional ecological properties and appear to
263 play a significant role in shaping cross-kingdom microbial composition (Figure 3D).

264 In summary, we put forward a strategy to improve microbiome engineering, with a
265 focus on microbial establishment and community diversity in both the bacterial and
266 fungal kingdoms. Our findings on the environment-agnostic ability of specific
267 bacterial and fungal genera to reach sufficient colonization levels to impact the
268 ecosystem function can be integrated with other properties (e.g. ease of cultivation,
269 genetic engineering) in a decision tree framework to accelerate microbiome
270 engineering solutions. The challenge ahead will be understanding the functional

271 characteristics that differentiate generalists from other microbes and dictates their
272 ability to flourish in highly diverse environmental niches. While for bacteria
273 generalists this can be likely explored by deep functional characterization at the
274 pathway and enzyme level using shotgun metagenomics datasets, the functional
275 characterization of fungal generalists will be a much greater challenge. Additionally,
276 care will need to be used when selecting microbes as the ability of generalists to
277 thrive in diverse environments is also a hallmark of many opportunistic
278 pathogens.[31,32] Nevertheless, we believe large-scale computational analyses
279 combined with multifactorial experiments in cross-disciplinary approaches will be
280 able to overcome these challenges and accelerate microbiome engineering solutions.

281 # Materials and Methods

282 ## Sample Selection

283 Included studies were retrieved by querying NCBI BioProject with the terms
284 'bacteria' and 'fungi' in any field. Only Biosamples with both 16S rRNA and ITS
285 amplicon sequencing data were considered for the concurrent analysis of both
286 kingdoms. We used both the identifier and attributes of the biosample, such as
287 aliases and library names, to map fungal and bacterial read files to a sample using a
288 custom script. Samples were associated to an environment (aquatic, host, or soil)
289 using manual curation of associated publications and biosample attributes provided
290 by the depositor. The three environments were further subdivided into 17 habitat
291 groups based on the body part and/or the ecoregion of the sampling location for
292 host and other samples, respectively (Olson et al., 2001). Habitats with less than five
293 samples were pooled together.

294 ## Generation of Genus-level Abundance Profiles

295 Genus-level abundance profiles were calculated using a custom nextflow pipeline (Di
296 Tommaso et al., 2017). Briefly, reads were downloaded from NCBI SRA using
297 grabseqs, except for the American Gut Project, which was downloaded from Qiita
298 (Gonzalez et al., 2018; Taylor et al., 2020). Paired-end reads were merged using
299 NGmerge (Gaspar, 2018). Quality Control (QC) and adapter removal was performed

300 using trimmomatic with a minimum Phread quality of 20 and a minimal read length
301 of 100 (Bolger et al., 2014). Quality was assessed using FastQC and MultiQC (Ewels
302 et al., 2016). Subsequent steps were performed using QIIME2 (Bolyen et al., 2019).
303 Reads were dereplicated following closed-reference OTU picking for both kingdoms
304 separately using VSEARCH with a 97% identity threshold (Rognes et al., 2016). For
305 taxonomic annotation, SILVA 132 97% consensus and UNITE 8.2 dynamic databases
306 were used for bacteria and fungi, respectively (Nilsson et al., 2019; Quast et al.,
307 2013). Following quality control, a total of 1,580 samples were selected for
308 downstream analyses.

309 Workflow and Statistical Analysis

310 Analyses were performed using a custom drake pipeline (Landau, 2018) built using
311 the programming language R 4.0.2. Briefly, abundances obtained from OTU profiling
312 were total-sum-scaled (TSS) and pooled at genus rank. All tools were used with
313 default parameters if not explicitly specified.

314 Diversity

315 Alpha diversity was estimated using Shannon and Chao1 metrics with the phyloseq
316 and vegan packages (Dixon, 2003; McMurdie and Holmes, 2013). To quantify the
317 contributions of a bacterial community profile with the fungal one and *vice versa*, we
318 used linear and unsupervised Canonical Correlation Analysis, as implemented in the
319 function CCorA of the vegan R package (Dixon, 2003). P-values were obtained using
320 blocked permutations to control for the habitat and to reduce assumptions of the
321 test. Supervised constrained ordination was performed using stepwise Distance-
322 based Redundancy Analysis (dbRDA) adapted from (Vieira-Silva et al., 2020). This
323 analysis shows linear relationships between bacterial dissimilarities and
324 abundances of selected explanatory fungal genera (and *vice versa*). An optimal
325 subset of up to 50 explanatory genera of the other kingdom was computed using a
326 stepwise feed-forward approach, as implemented in the ordistep function of the
327 vegan R package (Dixon, 2003).

Co-abundance Networks

SparCC, as implemented in FastSpar, was used to assess correlation between taxa pairs for each environment separately.[33,34] 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

Taxa 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%. Complementary, taxa 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.

Code and Data Availability

Scripts created for data processing and statistical analysis are available at https://github.com/bioinformatics-leibniz-hki/its-16s. Raw sequence data can be downloaded from any International Nucleotide Sequence Database Collaboration (INSDC) server using accessions as provided in Supplemental Table 2 and the git repository.

References

1. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A* **115**, 6506–6511 (2018).
2. Widder, S. *et al.* Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* **10**, 2557–2568 (2016).
3. Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
4. Oak, S. J. & Jha, R. The effects of probiotics in lactose intolerance: A systematic review. *Crit Rev Food Sci Nutr* **59**, 1675–1683 (2019).

356 5. Mohajeri, M. H. *et al.* The role of the microbiome for human health: from basic
357 science to clinical applications. *Eur J Nutr* **57**, 1–14 (2018).
358 6. Holscher, H. D. Dietary fiber and prebiotics and the gastrointestinal
359 microbiota. *Gut Microbes* **8**, 172–184 (2017).
360 7. Fang, H., Kang, J. & Zhang, D. : a review and future perspectives. *Microb Cell*
361 *Fact* **16**, 15 (2017).
362 8. Rossi, M., Amaretti, A. & Raimondi, S. Folate production by probiotic bacteria.
363 *Nutrients* **3**, 118–134 (2011).
364 9. Conly, J. M. & Stein, K. The production of menaquinones (vitamin K2) by
365 intestinal bacteria and their role in maintaining coagulation homeostasis. *Prog Food*
366 *Nutr Sci* **16**, 307–343 (1992).
367 10. Souak, D. *et al.* Challenging Cosmetic Innovation: The Skin Microbiota and
368 Probiotics Protect the Skin from UV-Induced Damage. *Microorganisms* **9**, (2021).
369 11. Singh, R. *et al.* Donor feces infusion for eradication of Extended Spectrum
370 beta-Lactamase producing Escherichia coli in a patient with end stage renal disease.
371 *Clin Microbiol Infect* **20**, O977–978 (2014).
372 12. Heijden, M. G. A. van der, Martin, F. M., Selosse, M. A. & Sanders, I. R.
373 Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol*
374 **205**, 1406–1423 (2015).
375 13. Vannier, N., Agler, M. & Hacquard, S. Microbiota-mediated disease resistance
376 in plants. *PLoS Pathog* **15**, e1007740 (2019).
377 14. Khoruts, A. & Sadowsky, M. J. Understanding the mechanisms of faecal
378 microbiota transplantation. *Nat Rev Gastroenterol Hepatol* **13**, 508–516 (2016).
379 15. Li, J. *et al.* Probiotics modulated gut microbiota suppresses hepatocellular
380 carcinoma growth in mice. *Proc Natl Acad Sci U S A* **113**, E1306–1315 (2016).
381 16. Oku, S., Komatsu, A., Nakashimada, Y., Tajima, T. & Kato, J. Identification of
382 Pseudomonas fluorescens chemotaxis sensory proteins for malate, succinate, and
383 fumarate, and their involvement in root colonization. *Microbes Environ* **29**, 413–419
384 (2014).
385 17. Carvalhais, L. C. *et al.* Activation of the jasmonic acid plant defence pathway
386 alters the composition of rhizosphere bacterial communities. *PLoS One* **8**, e56457
387 (2013).

388 18. Feinsinger, P., Spears, E. E. & Poole, R. W. A simple measure of niche breadth.
389 *Ecology* **62**, 27–32 (1981).
390 19. Frese, S. A. *et al.* EVC001 in Breastfed Infants. *mSphere* **2**, (2017).
391 20. Panigrahi, P. *et al.* A randomized synbiotic trial to prevent sepsis among
392 infants in rural India. *Nature* **548**, 407–412 (2017).
393 21. Azubuike, C. C., Chikere, C. B. & Okpokwasili, G. C. Bioremediation techniques-
394 classification based on site of application: principles, advantages, limitations and
395 prospects. *World J Microbiol Biotechnol* **32**, 180 (2016).
396 22. Henze, M., Gujer, W., Mino, T. & Loosdrecht, M. C. van. Activated sludge
397 models ASM1, ASM2, ASM2d and ASM3. *IWA Scientific and Technical Report* **9**,
398 (2000).
399 23. Orozco-Mosqueda, M. D. C., Rocha-Granados, M. D. C., Glick, B. R. & Santoyo, G.
400 Microbiome engineering to improve biocontrol and plant growth-promoting
401 mechanisms. *Microbiol Res* **208**, 25–31 (2018).
402 24. Kaminsky, L. M., Trexler, R. V., Malik, R. J., Hockett, K. L. & Bell, T. H. The
403 Inherent Conflicts in Developing Soil Microbial Inoculants. *Trends Biotechnol* **37**,
404 140–151 (2019).
405 25. Albright, M. B. N. *et al.* Solutions in microbiome engineering: prioritizing
406 barriers to organism establishment. *ISME J* **16**, 331–338 (2022).
407 26. Liao, J. *et al.* The importance of neutral and niche processes for bacterial
408 community assembly differs between habitat generalists and specialists. *FEMS*
409 *Microbiol Ecol* **92**, (2016).
410 27. Liu, L., Wang, S. & Chen, J. Transformations from specialists to generalists
411 cause bacterial communities are more stable than micro-eukaryotic communities
412 under anthropogenic activity disturbance. *Sci Total Environ* **790**, 148141 (2021).
413 28. Garrison, C. E. & Field, E. K. Introducing a "core steel microbiome" and
414 community functional analysis associated with microbially influenced corrosion.
415 *FEMS Microbiol Ecol* **97**, (2020).
416 29. Walter, J. & Ley, R. The human gut microbiome: ecology and recent
417 evolutionary changes. *Annu Rev Microbiol* **65**, 411–429 (2011).
418 30. Chen, Y. J. *et al.* Metabolic flexibility allows bacterial habitat generalists to
419 become dominant in a frequently disturbed ecosystem. *ISME J* **15**, 2986–3004
420 (2021).

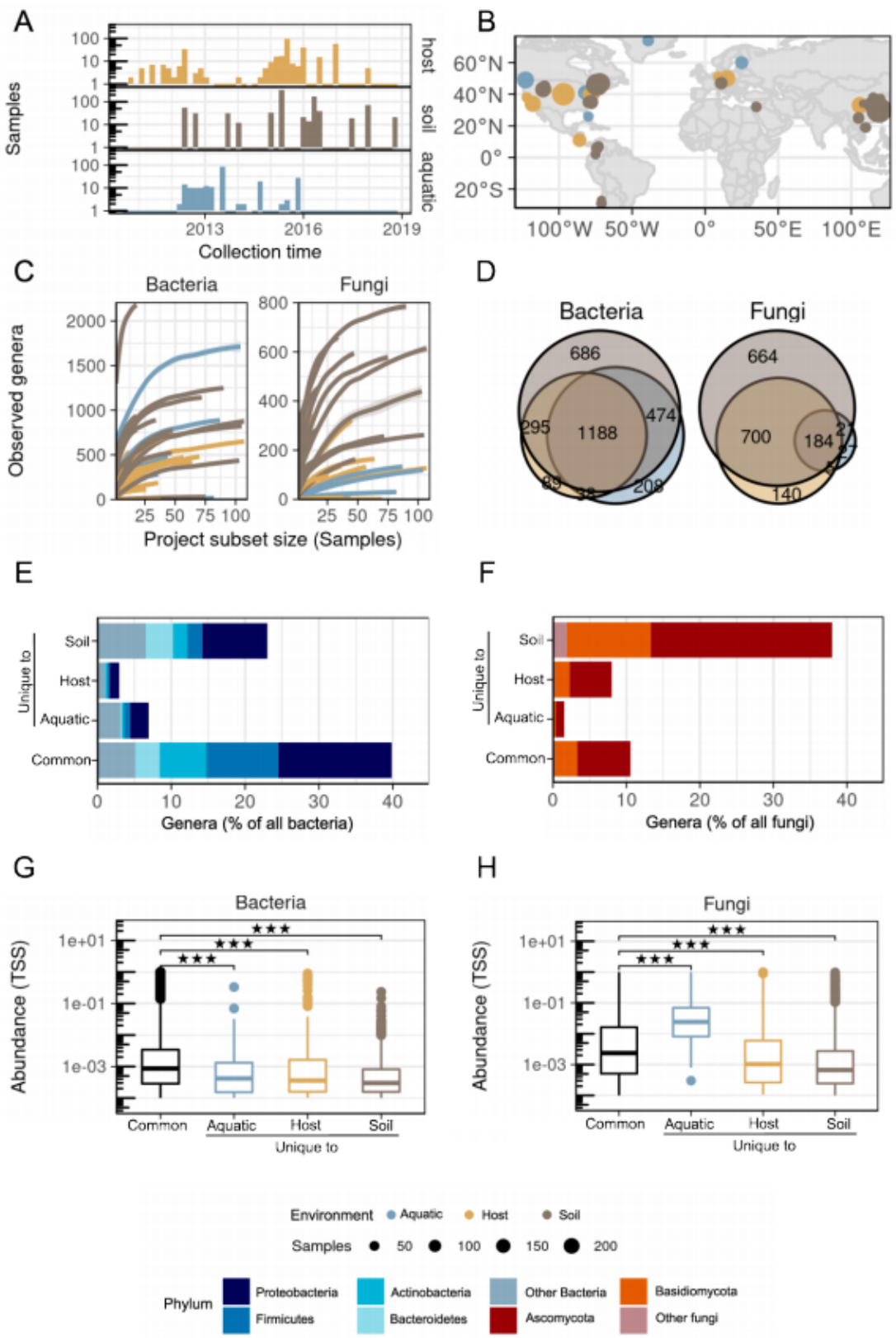
31. Zajc, J., Gostinčar, C., Černoša, A. & Gunde-Cimerman, N. Stress-Tolerant Yeasts: Opportunistic Pathogenicity Versus Biocontrol Potential. *Genes (Basel)* **10**, (2019).

32. Gostinčar, C. *et al.* Fungi between extremotolerance and opportunistic pathogenicity on humans. *Fungal diversity* **93**, 195–213 (2018).

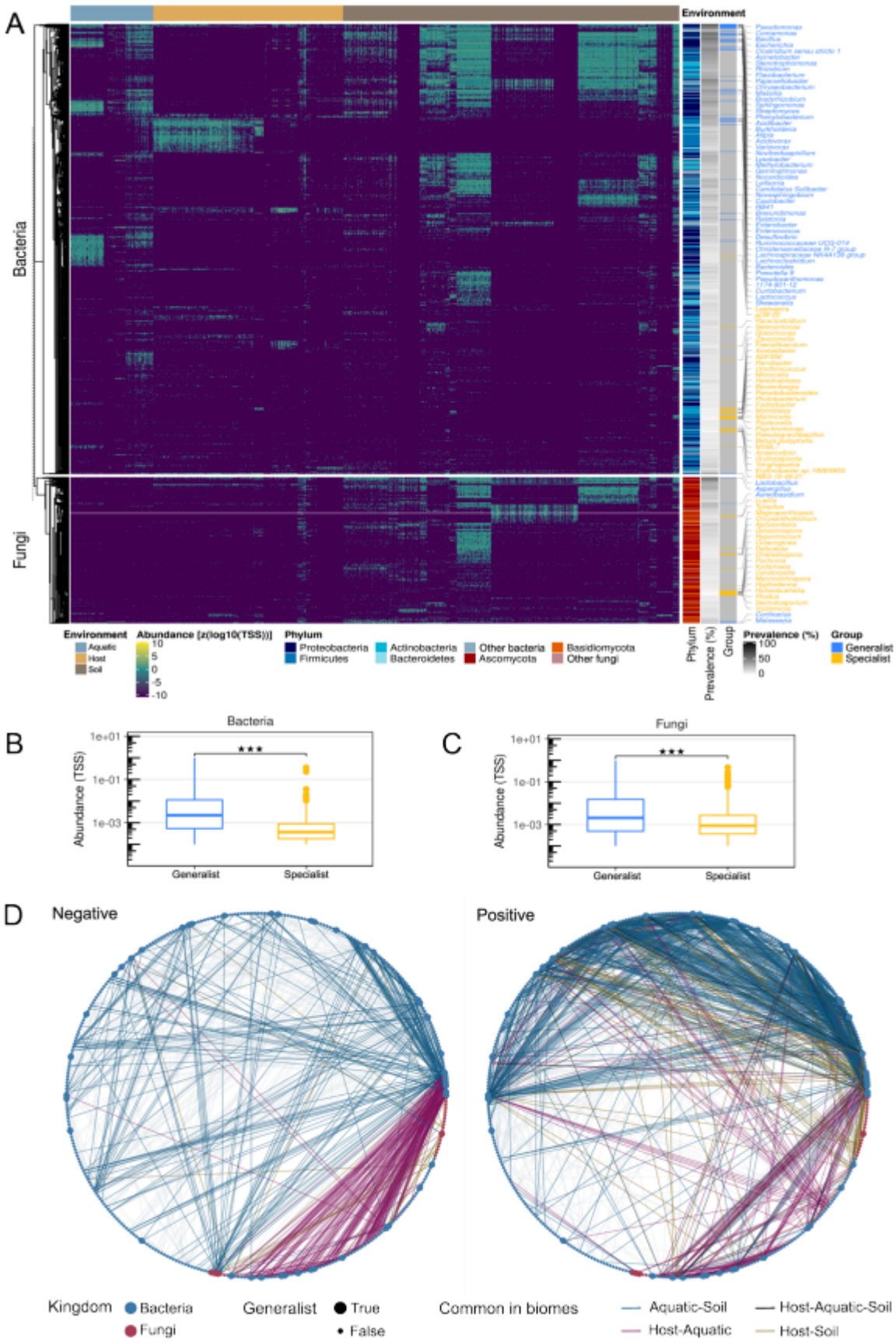
33. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* **8**, e1002687 (2012).

34. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* **35**, 1064–1066 (2019).

Figures

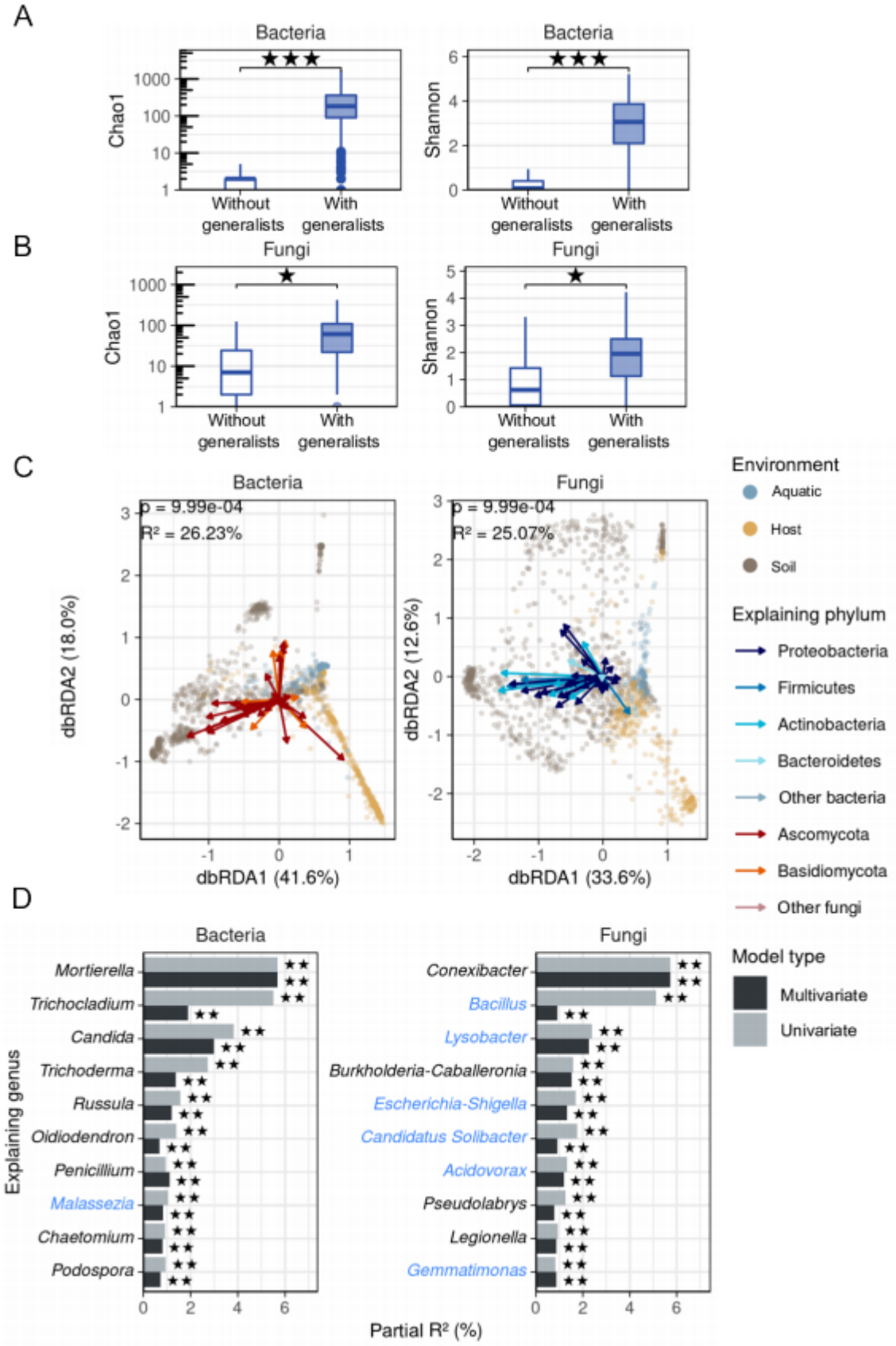

A
Samples
host
soil
aquatic
Collection time
B
C
Observed genera
Bacteria
Fungi
Project subset size (Samples)
D
Bacteria
Fungi
686
1188
474
295
208
664
700
184
140
E
Unique to
Soil
Host
Aquatic
Common
Genera (% of all bacteria)
F
Genera (% of all fungi)
G
Bacteria
Abundance (TSS)
Common
Aquatic
Host
Soil
Unique to
H
Fungi
Environment
Aquatic
Host
Soil
Samples
50
100
150
200
Phylum
Proteobacteria
Firmicutes
Actinobacteria
Bacteroidetes
Other Bacteria
Ascomycota
Basidiomycota
Other fungi

431 ***Figure 1: A global analysis of microbial communities reveals differences in***
432 ***environmental specificities between bacteria and fungi.*** *Distribution of samples*
433 *used in this study (N=1,580) shown over time (**A**) and geographic location (**B**). **C:***
434 *Rarefaction curves of Shannon alpha diversity for each study demonstrate sufficient*
435 *sampling depth. Curves are shown as LOESS regressions from 10 independent sampling*
436 *trials at 10 given sampling subset sizes. Lines are colored by environment and are*
437 *surrounded by ribbons indicating the 95% confidence interval across the trails. **D**:*
438 *Intersection of bacterial and fungal genera found in at least one sample in each*
439 *environment as Venn diagrams. **E, F**: Percentage of genera found in all three or only*
440 *one environment. **G, H**: Abundance comparisons of common and unique genera. A*
441 *genus was considered present in a sample using a threshold of abundance > .01%.*
442 *Significance determined by Wilcoxon rank sum test; *** denotes p < 0.001.*


A
Environment
Bacteria
Fungi
Environment
Aquatic
Host
Soil
Abundance [z(log10(TSS))]
10
5
0
-5
-10
Phylum
Proteobacteria
Firmicutes
Actinobacteria
Bacteroidetes
Other bacteria
Ascomycota
Basidiomycota
Other fungi
Phylum
Prevalence (%)
Group
Prevalence (%)
100
50
0
Group
Generalist
Specialist
B
Bacteria
Abundance (TSS)
1e+01
1e-01
1e-03
***
Generalist
Specialist
C
Fungi
Abundance (TSS)
1e+01
1e-01
1e-03
***
Generalist
Specialist
D
Negative
Positive
Kingdom
Bacteria
Fungi
Generalist
True
False
Common in biomes
Aquatic-Soil
Host-Aquatic
Host-Aquatic-Soil
Host-Soil

443

Figure 2: Abundance and co-abundance of generalists and specialists. A: Abundance heatmap of bacteria and fungal genera. Samples are arranged by environment (top bar) with only abundance values >0.01% displayed. Right bars indicate phylum and prevalence among all samples, where only taxa with >5% prevalence are displayed. Generalists (>40% prevalence in at least one habitat in all three environments) and specialists (>40% prevalence in one habitat and <5% prevalence in all others) are annotated in blue and yellow, respectively. The dendrogram was calculated using average linkage of Manhattan distances between taxa abundance profiles. **B, C**: Abundances of bacterial (**B**) and fungal (**C**) generalists and specialists. Values were averaged by project to account for different cohort sizes. Statistical significance calculated using Wilcox rank-sum test (*** denotes $p<0.001$). **D**: Networks of genera found in all three environments and significantly co-abundant in the majority of environments (SparCC FDR $p < 0.05$, $|r| > 0.2$)


A
Bacteria
Chao1
Shannon
Without generalists
With generalists
B
Fungi
C
Bacteria
Fungi
p = 9.99e-04
R² = 26.23%
R² = 25.07%
dbRDA1 (41.6%)
dbRDA2 (18.0%)
dbRDA1 (33.6%)
dbRDA2 (12.6%)
Environment
Aquatic
Host
Soil
Explaining phylum
Proteobacteria
Firmicutes
Actinobacteria
Bacteroidetes
Other bacteria
Ascomycota
Basidiomycota
Other fungi
D
Explaining genus
Mortierella
Trichocladium
Candida
Trichoderma
Russula
Oidiodendron
Penicillium
Malassezia
Chaetomium
Podospora
Conexibacter
Bacillus
Lysobacter
Burkholderia-Caballeronia
Escherichia-Shigella
Candidatus Solibacter
Acidovorax
Pseudolabrys
Legionella
Gemmatimonas
Model type
Multivariate
Univariate
Partial R² (%)

457

458 ***Figure 3: Generalists significantly impact diversity and cross-kingdom variation.***
459 ***A,B****: Shannon and Chao1 alpha diversity were calculated for bacteria (**A**) and fungi (**B**).*
460 *Samples were grouped by whether they contained any generalist (genera with >40%*
461 *prevalence in at least one habitat from every environment; abundance > 0.01%), or*
462 *not. Significance bars indicate permutation test compared to samples without random*
463 *taxa instead of generalists (* q<0.05, ** q<0.01, *** q<0.001). **C-D**: Bacterial and fungal*
464 *Bray-Curtis dissimilarities constrained by explanatory genus abundances of the other*
465 *kingdom using distance-based Redundancy Analysis (dbRDA). **C**: Explaining genera*
466 *were selected using a feedforward approach. Effect size of most explanatory taxa is*
467 *shown in **D** by multivariate model (as displayed in C) or univariate model containing*
468 *only the taxon of interest. Generalists are indicated with blue text. Stars indicate*
469 *significance by ANOVA (** p<0.01).*