

MIT Open Access Articles

Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Dal Bello, Martina, Lee, Hyunseok, Goyal, Akshit and Gore, Jeff. 2021. "Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism." *Nature Ecology & Evolution*, 5 (10).

As Published: 10.1038/S41559-021-01535-8

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/141887>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



1
2
3
4
5
6
7
8
9

Title:

Resource-diversity relationships in bacterial communities reflect the network structure of microbial metabolism

Authors: Martina Dal Bello*, Hyunseok Lee, Akshit Goyal, Jeff Gore*.

Affiliations:

Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, 02139 Cambridge, Massachusetts, United States.

*Correspondence to: gore@mit.edu, dalbello@mit.edu

10 **The relationship between the number of available nutrients and community diversity is a**
11 **central question for ecological research that remains unanswered. Here, we studied the**
12 **assembly of hundreds of soil-derived microbial communities on a wide range of well-**
13 **defined resource environments, from single carbon sources to combinations of up to 16. We**
14 **found that, while single resources supported multispecies communities varying from 8 to 40**
15 **taxa, mean community richness increased only one-by-one with additional resources.**
16 **Cross-feeding could reconcile these seemingly contrasting observations, with the metabolic**
17 **network seeded by the supplied resources explaining the changes in richness due to both**
18 **the identity and the number of resources, as well as the distribution of taxa across different**
19 **communities. By using a consumer-resource model incorporating the inferred cross-feeding**
20 **network, we provide further theoretical support to our observations and a framework to**
21 **link the type and number of environmental resources to microbial community diversity.**

22 Uncovering the determinants of community diversity is central in ecology¹⁻³ and microbiome
23 research⁴, posing unique challenges to microbial ecologists. Indeed, microbes are the most
24 abundant form of life on our planet⁵, the most ancient and the most phylogenetically diverse⁶.
25 Surveys of a variety of ecosystems, from oceans⁷ to the human body⁸, have revealed that
26 thousands of different taxa can stably coexist within the same community. Importantly, microbial
27 communities drive the bulk of global nutrient cycling⁹, sustain human health¹⁰ and modulate the
28 response of the biosphere to climate change¹¹. Hence, deepening the knowledge of the drivers of
29 microbial community diversity is pivotal to understand the functioning of Earth's ecosystems.

30 Several mechanisms contribute to the diversity of microbial communities, including the spatial
31 and temporal structure of the environment¹², dispersal and bacterial motility¹³, warfare^{14,15}, and
32 resource-mediated competition and cooperation¹⁶⁻¹⁸. With respect to resources, ecological theory
33 has mostly focused on the effect of the number of available resources on community diversity
34 rather than their identity¹⁹. In particular, according to the principle of competitive exclusion, the
35 number of stably coexisting species is predicted to be bounded by the number of available
36 resources²⁰⁻²². Despite the wealth of theoretical work on how resources can affect microbial
37 community diversity, empirical tests of resource-diversity relationships have been limited,
38 having been explored either in 2-3 species assemblages¹⁷ or in enriched cultures grown on 1-2
39 resources^{18,23-25}. Systematic experiments encompassing a range of resource combinations are still
40 lacking.

41 While an empirical test of the relationship between the number of available nutrients and
42 community diversity remained elusive, bottom-up experiments have implicated cross-feeding as
43 a major factor influencing the assembly of microbial communities, even in simple environments.
44 Cross-feeding, whereby metabolic byproducts of one taxa become resources for others²⁶, can
45 increase niche partitioning, ultimately allowing the coexistence of several taxa even when only a
46 single source of carbon is provided^{18,23,25,27}. There is also some evidence that the identity of the
47 supplied resource dictates community composition, as microbial taxa display different resource
48 preferences and patterns of metabolite excretion^{23,28}. Nevertheless, the manner in which cross-
49 feeding and niche partitioning systematically change with the identity and the number of
50 supplied resources is still unclear. This lack of knowledge impairs our ability to link variations in
51 resource availability with shifts in microbial community diversity.

52 Here, we used a high-throughput experimental protocol and 16S amplicon sequencing to explore
53 the relationship between microbial community diversity and resource availability in experimental

54 microcosms. By growing soil-derived communities in media containing different combinations
55 of carbon sources (from single resources up to 16), we discovered that community diversity was
56 high in single resources but then increased only modestly with additional nutrients. These
57 seemingly contrasting observations reflected the structure of the metabolic network seeded by
58 the supplied resources. Cross-fed byproducts predicted to originate from each resource via
59 microbial metabolism were coupled to the richness and composition of single resource
60 communities. Additionally, a consumer-resource model incorporating the inferred metabolic
61 network recapitulated the linear increase of community diversity with additional resources.

62 **Results**

63 In order to illuminate how the availability of resources, namely their number and identity, shape
64 the richness of microbiomes, we assayed the assembly of soil-derived bacterial communities in
65 laboratory microcosms^{16,18}. We started by inoculating a rich microbial suspension obtained from
66 a soil sample (Fig. S1) into 75 resource environments, each containing minimal media
67 supplemented with different combinations of carbon sources, ranging from one to 16 (Fig. 1a,
68 S2, Table S1). The 16 carbon sources represented a broad range of common soil compounds
69 (e.g., mannose, xylose, cellulose and hydroxyproline), encompassing both glycolytic (e.g.,
70 simple and complex sugars) and gluconeogenic substrata (e.g., organic acids). We adopted a
71 daily-dilution protocol, whereby at the end of each 24-hour growth cycle the bacterial cultures
72 were diluted 1/30x into fresh media. We observed that the majority of microcosms reached
73 stability after 3 days from the inoculum (Fig. S3). We continued the experiment until day 7 and
74 measured the final richness as the number of ASVs (amplicon sequence variants) observed
75 within each community (Fig. S4).

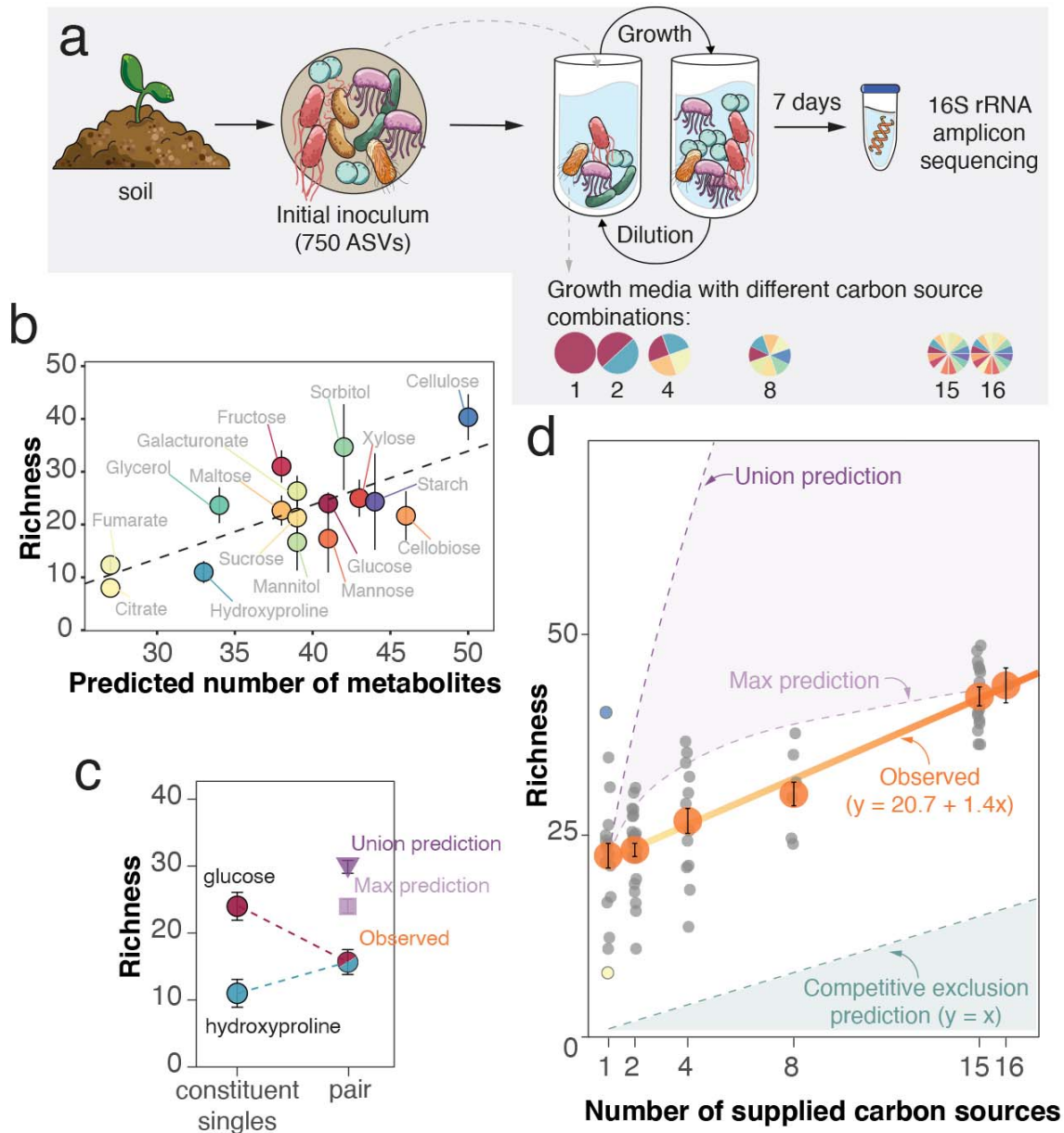
76 **Individually-supplied resources support complex multi-species communities**

77 Consistent with recent experimental studies^{18,24,29}, single-resource communities were remarkably
78 rich (mean richness = 23 ± 2 ASVs, Fig. 1b) and taxonomically diverse (Fig. S5). This is in
79 contrast with competitive exclusion predicting that the number of species cannot exceed the
80 number of resources^{20,30}—which, in single carbon sources, would result in no more than one
81 species surviving. Interestingly, the variability in richness among different resources was also
82 high—with the average number of ASVs ranging from 8 in citrate to 40 ± 4 in cellulose (Fig.
83 S6, Fig. 1b)—and larger than the variability among replicates of the same carbon source
84 (ANOVA test, $F_{resource} = 3.4339$, $p < 0.01$). Richness in single carbon sources therefore
85 depended on resource identity. Community richness did not correlate significantly with the
86 molecular weight of the supplied resource (Fig. S7), but did correlate with the predicted number
87 of metabolites which could be generated from the resource through intracellular biochemical
88 reactions and secreted in the environment (Fig. 1b, see Methods for the details on the prediction
89 of metabolites based on KEGG³¹ and MetaCyc³² databases). Notably, the lowest richness was
90 observed for gluconeogenic substrata (~10 for citrate, fumarate and hydroxyproline), which were
91 connected to the central metabolic pathway via the TCA cycle, hence resulting in the smallest
92 metabolite pools. Consistent with previous work, these results highlight the role of cross-feeding
93 in supporting community diversity^{33–35}. Moreover, they suggest that the extent of cross-feeding
94 may determine how many species can coexist on single resources.

95 Having found large numbers of coexisting species in single resources, we expected that
96 community diversity would increase rapidly if more resources were provided. As previously
97 observed in marine bacteria^{24,25}, community composition could potentially correspond to the sum
98 of the assemblages observed on each nutrient supplied in the mixture. To provide an example,
99 the expected richness of the community grown on glucose and hydroxyproline (Fig. 1c), each
100 alone supporting on average 24 and 11 ASVs, would be ~ 30 ASVs, i.e., the sum minus the
101 number of shared ASVs (union). Alternatively, niche overlap between the taxa found in the
102 single-resource media^{36,37} might bring the expected number of species down to the maximum
103 richness observed in the constituent singles; in the case of glucose + hydroxyproline, 24 ASVs.
104 However, when we measured the richness of the communities grown in a media supplied with
105 equal amounts of glucose and hydroxyproline, we found only ~16 ASVs on average, which is
106 significantly lower than both expectations (Fig. 1c). Yet, our observed richness was remarkably
107 similar to the mean richness measured in the two constituent single resources (17.5 ASVs), a
108 trend that was consistent across many two-resource communities (Fig. S8). Contradicting our
109 expectations based on previous results supporting additivity, we found that community richness
110 upon combining two carbon resources was approximately the average richness of constituent
111 single resource environments.

112 **Community diversity increases linearly with the number of supplied resources**

113 Next, we examined the full range of resource combinations included in the experiment. Again,
114 the richness predicted from the union of constituent singles significantly overestimated the
115 observed richness (Fig. 1d). The prediction based on the maximum of constituent singles gave an
116 increase with negative curvature that was not detected in our experiment (Fig. 1d). We found a
117 similar trend also when we estimated the number of metabolites generated from resource
118 combinations, with the same approach we used for single carbon sources (Fig. S9). Instead, the
119 observed average richness increased linearly with the number of supplied carbon sources, at the
120 constant rate of one to two ASVs for each new added resource (Fig. 1d, slope = 1.4 ± 0.1). As a
121 result, the richness of communities supported by 16 resources was roughly twice the average
122 richness of single-resource communities. The linear relationship was robust to the exclusion of
123 low-abundance ASVs—with the slope reduced to 1 ± 0.07 when ASVs with relative abundance
124 below 0.1% were excluded (Fig. S10a)—and coarse-graining at the family level (Fig. S10b). In
125 addition, as more resources were provided, communities became more even (see Methods and
126 Fig. S10c, d, S11), without changes in total biomass (Fig. S12). Despite confirming that the
127 number of supplied resources is an important driver of microbial diversity, the observed one-by-
128 one relation between richness and resource number was difficult to reconcile with the large
129 diversity found in single resources. Thus, we went back to the single-resource communities to
130 gain a better understanding of our observations.



131

132 **Figure 1. Microbial community diversity increases slowly with the number of resources despite individual**
 133 **resources supporting complex multispecies communities.** *a.* Layout of the experiment. We inoculated a rich
 134 microbial suspension obtained from a soil sample into 75 growth media, each supplemented with a different
 135 combination of carbon sources, from single compounds to a mix of 16, while keeping the total carbon
 136 concentration the same (0.1% w/v). Bacterial cultures were grown for 7 days under a regime of daily dilution
 137 and their composition assessed at the single nucleotide resolution using 16S rRNA amplicon sequencing.
 138 Community diversity is measured as richness, i.e., number of ASVs. *b.* Richness of microbial communities
 139 supported by single carbon sources correlates with the number of metabolites predicted to be generated from
 140 metabolic reactions mapped in the KEGG database (Pearson's correlation coefficient $r=0.75$ [95% CI: 0.4-
 141 0.91], $p<0.001$). Colored dots indicate, for each carbon source, the number of ASVs (mean \pm SEM, $N = 3$). *c.*
 142 A representative example of how observed richness in constituent single resources (mean \pm SEM, $N = 3$)
 143 compares to the observed richness in two-resource communities (mean \pm SEM, $N = 3$) and predictions
 144 calculated as the union (sum without overlapping ASVs, dark violet) or the maximum (light violet) of the
 145 richness in constituent singles (mean \pm SEM, N from permutations = 9). *d.* Observed average richness

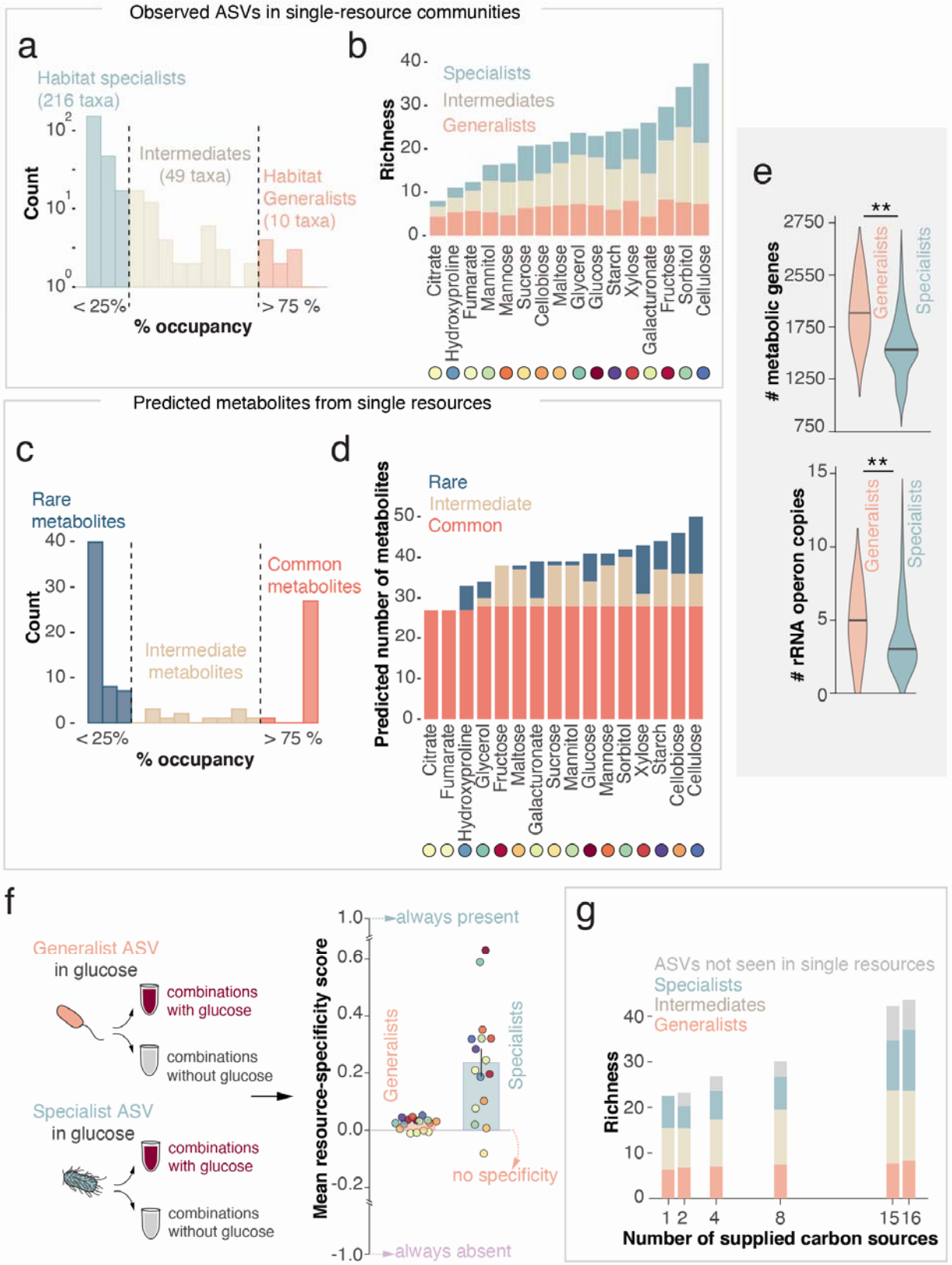
146 (orange dots, mean \pm SEM, $N = 16$ for single-resource, 24 for two-resource, 12 for four-resource, six for
147 eight-resource, 16 for 15-resource and 1 for 16-resource combinations) as a linear function of the number of
148 supplied carbon sources (solid orange line). Grey jittered dots indicate the average richness for each unique
149 combination of resources (mean \pm SEM, $N = 3$). Intercept = 20.7 ± 0.8 , slope 1.4 ± 0.1 ($p < 0.001$). In single
150 resources, the blue and yellow dots correspond to the highest and lowest average richness, measured in
151 cellulose and citrate, respectively. The predicted trajectory of richness based on the competitive exclusion
152 principle (dashed dark green line), the union (dashed dark violet line) and maximum (dashed light violet line)
153 estimates, as described for panel b, are shown for comparison.

154 **Communities are composed of generalists and variable numbers of specialists**

155 First, we measured the resource occupancy of the 275 ASVs observed in single resource media,
156 i.e., how many single-resource media a given ASV was found in (Fig. 2a, Fig. S13). Based on
157 resource occupancy, we considered habitat specialists the ASVs that were observed in less than
158 25% of single-resource media, and habitat generalists those that occupied more than 75% of
159 single-resource media (Fig. 2a). The majority of ASVs (216 out of 275) were specialists,
160 whereas very few of them (10) were generalists. Some ASVs (49) displayed an intermediate
161 occupancy, being present in between four to twelve media. This is reminiscent of natural
162 communities, in which few taxa are usually universally present across different habitats, while
163 the majority is found only under specific environmental conditions³⁸⁻⁴⁰. Importantly, previous
164 work has shown that variations in the proportion of generalist and specialists taxa within a
165 community impact its dynamics⁴¹⁻⁴³.

166 Next, we inspected the distribution of specialists and generalists within single resources. We
167 found that species-poor communities, grown on gluconeogenic substrates like citrate, were
168 dominated by generalist ASVs (often representing $> 50\%$ of observed taxa; Fig. 2b, S14a), while
169 species-rich communities were enriched in specialists (see cellulose in Fig. 2b, S14b). We
170 noticed that glycolytic substrates, which can produce a much larger metabolite pool before
171 connecting with the central carbon metabolism, supported communities where more specialists
172 coexisted with generalists. This suggested a link between the metabolite pool generated from
173 each supplied resource and its ability to sustain both generalists and specialists in the same
174 community.

175 Remarkably, just like ASVs, our predicted metabolic byproducts could also be broken up into
176 two broad classes. Based on how many single resources could trigger their production, we could
177 distinguish between common metabolites, present in association with the majority of single
178 resources (like generalist taxa), and rarer metabolites, present only in association with one or few
179 resources (like specialists) (Fig. 2c). Metabolites that were commonly produced constituted the
180 *core* intermediates of the central metabolic pathway, including the TCA cycle and lower
181 glycolysis. The rarely produced metabolites, instead, were the intermediates of *peripheral*
182 branches of the central pathway. For example, if either citrate or fumarate were provided, we
183 predicted that the central pathway proceeds in the gluconeogenic direction, generating only
184 byproducts belonging to the core pool. In contrast, individually-supplied sugars were predicted to
185 go through a series of reactions before entering the central pathway, ultimately generating both
186 core and peripheral metabolites (Fig. 2d). It appeared that the number of peripheral metabolites
187 varied with the position from which the resource entered the “metabolic map”. The parallelism in
188 the distribution of ASVs and predicted metabolites reinforces the idea that community structure
189 is coupled to the metabolite pool, and suggests a link between the resource occupancy and
190 metabolic capability of taxa.



192 **Figure 2. Experimental communities are composed of generalists and variable numbers of specialists, with**
193 **the latter driving the increase in community diversity. a.** The 275 ASVs found across all single-resource
194 communities were classified in generalist, specialists and intermediates depending on their resource
195 occupancy. The majority of ASVs exhibited a more specialized resource-utilization strategy. **b.** The richness in
196 single resource-media is displayed highlighting the mean number of generalist (pink), intermediate (beige) and
197 specialist (teal) ASVs (mean, $N=3$, error bars are omitted for clarity). **c.** The metabolites estimated to be
198 produced starting from the supplied single resources through cell reactions can be classified in common,
199 intermediate and rare, based on resource-occupancy as for ASVs. **d.** The total of metabolites estimated for
200 each single resource is displayed highlighting the number of common (red), intermediate (light brown) and
201 rare (blue) metabolites. **e.** Upper panel. The distribution of the number of metabolic genes retrieved for each
202 ASV in single resources (see Methods) differs between generalists and specialists ($p < 0.01$, from Kolmogorov-
203 Smirnov test). Lower panel. The distribution of rRNA operon copy numbers, calculated at the genus level, of
204 generalist ASVs differs from that of specialist ASVs ($p < 0.01$, from Kolmogorov-Smirnov test). **f.** The
205 specificity score is calculated, for each ASV found in a single resource (target resource), using the number of
206 multi-resource media containing the target resource in which the ASV was found (X) and the number of media
207 not containing the target resource in which the ASV is found (Y), as $(X - Y)/(X + Y)$. It ranges from 1,
208 indicating that the ASV is present only in a combination containing the resource, to -1, implying that the ASV
209 is always absent when the resource is supplied. A score of 0 is indicative of an ASV showing no specificity for
210 that particular resource. Bars indicate the mean specificity score \pm SEM for generalists (pink) and specialists
211 (teal) ($N = 16$). Colored dots indicate the mean score for each resource (SEM are omitted for clarity, N varies
212 for each resource, see Methods). **g.** The mean number of generalist (pink), intermediate (beige) and specialist
213 (teal) ASVs for media with the same number of resources is shown as stacked bars. The average number of
214 ASVs that were not detected in single-resource communities but appeared in other combinations is indicated in
215 grey. Error bars are omitted for clarity.

216 We next tested for systematic differences in the metabolic capabilities between generalist and
217 specialist taxa in our experimental microcosms. Generalist ASVs belonged to the most abundant
218 families, i.e., Pseudomonadaceae, Enterobacteriaceae and Micrococcaceae (Fig. S13) and
219 differed metabolically from specialists, e.g., taxa from Cellvibrionaceae. In particular, generalists
220 were estimated to harbor a larger number of metabolic genes (Fig. 2e upper panel, see Methods
221 for details on the estimation of gene content) and more copies of the 16S rRNA operon compared
222 to specialists (Fig. 2e lower panel, see Methods for details on the matching with the number of
223 copies of the rRNA operon), indicative of faster max growth rates⁴⁴. Both results are consistent
224 with studies showing the hallmarks of a generalist life style: flexible metabolism^{38,45} (indicated
225 by the number of metabolic genes) and capacity for fast growth (indicated by the 16S rRNA
226 copy number)^{46,47}. At the same time, several of the taxa classified as generalists are known to
227 show distinct resource preferences when grown in isolation. For example, *Pseudomonads* species
228 dominated in the communities sustained by organic acids, most likely because of their advantage
229 over other taxa preferring sugars⁴⁸, but were also present in all the media in which organic acids
230 could have been generated as byproducts of the glycolytic metabolism of sugars⁴⁹ (see Fig. S3).
231 This might indicate that generalists were present in all the communities because the substrates
232 that they utilize were always generated as byproducts of bacterial metabolism. Indeed, even
233 habitat generalists show resource preferences⁵⁰, such as *Pseudomonas* spp., which consumes
234 preferentially acetate and other organic acids²³. In contrast, since many sugars and their
235 intermediates could not be produced via gluconeogenic metabolism⁵¹, the survival of the taxa
236 specializing on them was prevented unless those sugars were externally supplied. Together these
237 observations are consistent with the idea of habitat generalists and specialists assembling in a
238 community in relation to the available supplied and cross-fed metabolites.

239 The coupling between metabolite pool and community structure observed in single resources
240 suggested that resource-ASVs associations would be maintained also in multi-resource
241 environments. In particular, we expected that generalists would be present in all communities,
242 while specialists would be mostly detected when the favorite substrate was provided or
243 metabolically generated. To verify these expectations, we calculated a resource-specificity score.
244 For each ASV present in a single resource (target resource), the resource specificity score was
245 calculated as the difference between the number of multi-resource media containing the target
246 resource in which the ASV was found and the number of media not containing the target
247 resource in which the ASV was found, divided by the total number of media in which the ASV
248 was found. The score ranged from 1, indicating that the ASV was present only when the target
249 resource was provided in a combination, to -1, implying that, although the ASV was found in the
250 single resource, it was always absent when that resource was supplied with others. A score of 0
251 indicated that an ASV showed no specificity for that resource (Fig. 2f). We found that
252 specialists' scores were on average positive across all resources (Fig. 2f, mean score = $0.24 \pm$
253 0.05), while generalists' scores were on average nearly zero (0.02 ± 0.01). Together, these
254 findings highlight that (specialist) taxa tend to show resource-specific associations, and that
255 single resource-ASV associations are maintained even in multi-resource environments.

256 To verify how resource-ASV associations impacted the resource-diversity relationship, we then
257 calculated the average number of specialists, generalists and intermediates (as defined based on
258 single resource occupancy) for each combination of carbon sources. We found that going from 1
259 to 16 resources, communities went from containing a balanced mixture of generalists and
260 specialists to being dominated by more specialized ASVs (both specialists and intermediates,
261 Fig. 2g). Overall, these results point to the consistent coexistence in our experimental
262 microcosms of distinct groups of bacteria, with more specialized taxa progressively favored by
263 the supply of additional resources. While this was in line with the expectation that specialists of
264 each resource should be favored by the higher chances to introduce a glycolytic compound as
265 more resources were added, it is important to note that, at the same time, several specialist ASVs
266 were lost and few new ASVs were introduced, especially going from one to two carbon sources
267 (grey bars in Fig. 2f, S15, these ASVs remained unclassified).

268 In summary, our experimental results revealed that 1) single resources were able to sustain
269 multispecies communities, 2) going from one to two resources, community richness did not
270 significantly increase; 3) overall, the resource-diversity relationship was linear and only
271 modestly increasing; 4) all experimental communities were composed of both habitat generalists
272 and specialists and their ratio changed with the number of supplied resources. We also show that
273 the structure of the metabolite pool, which is the result of the ensemble of metabolic reactions
274 fueled by the supplied and cross-fed resource(s), is the most likely driver of the observed
275 manifestations of the resource-diversity relationship. We next asked: can we recapitulate some of
276 the primary features of our experimental results by incorporating the metabolic network in a
277 resource-explicit modelling framework?

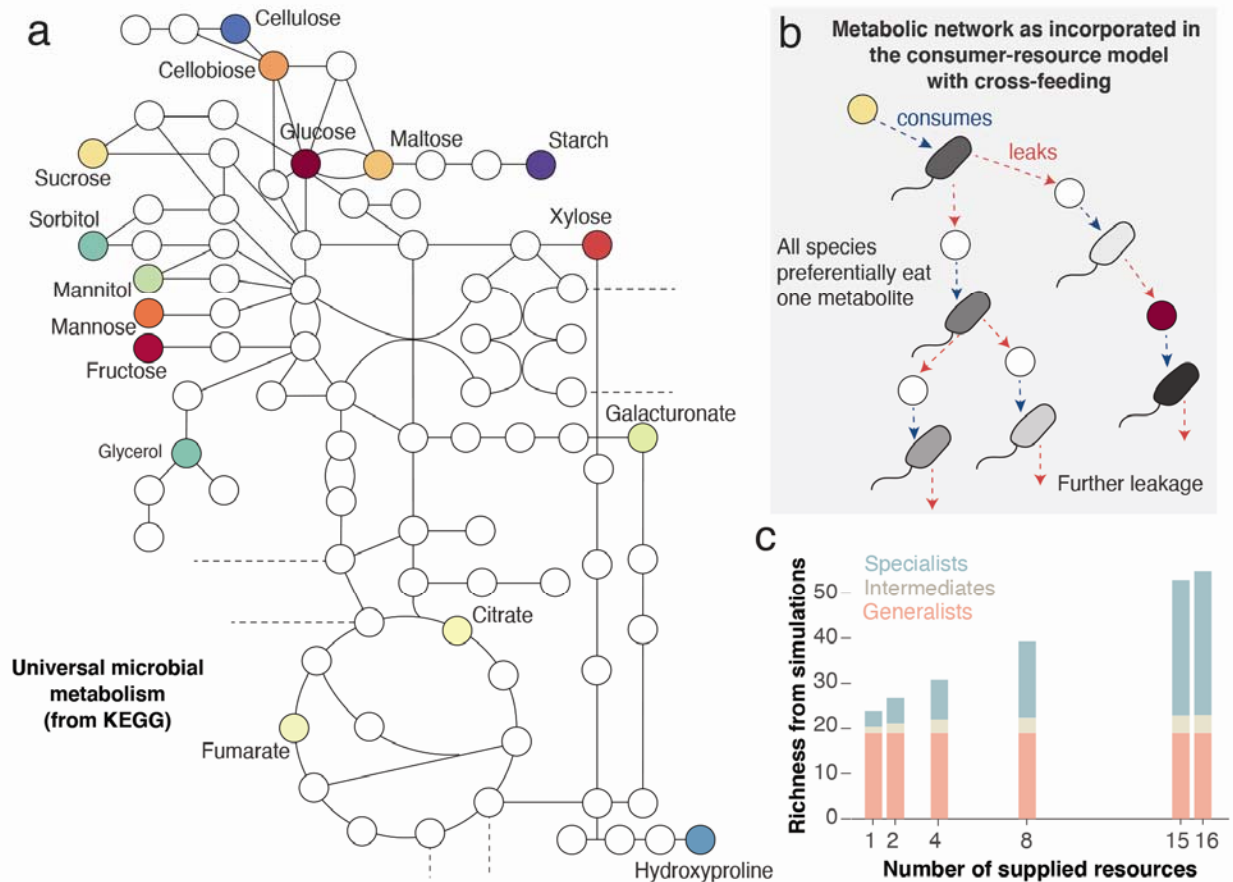
278 **A resource-explicit model incorporating a realistic metabolic network reproduces our** 279 **experimental results**

280 We hypothesized that the metabolic network seeded by the supplied carbon sources could
281 explain the observed resource-diversity relationship. To test this, we implemented the well-
282 known MacArthur consumer-resource model with cross-feeding^{19,21,35}. In contrast to other

283 implementations which used an abstract, randomly-generated cross-feeding network^{18,35}, we used
284 a realistic network inferred using KEGG and MetaCyc databases (Fig. 3a, see Methods for
285 details). Note that we used the same network to estimate the possible number of metabolic
286 byproducts generated from each of the 16 carbon sources in single-resource environments (Fig.
287 1b). In our model, for simplicity, each species consumed one resource to grow, to approximate
288 particular resource preferences in different species. It then leaked metabolic byproduct(s) into the
289 environment, each of which was one step downstream from the consumed metabolite according
290 to the cross-feeding network (see Methods). Other species could then consume these leaked
291 metabolites, in turn releasing new by-products into the environment (Fig. 3b). Importantly,
292 leaked byproducts always comprised a fixed fraction of the consumed resource, resulting in a
293 progressive decrease in the concentration of metabolic byproducts available to microbes
294 downstream^{34,35} (Fig. 3b). Finally, to account for metabolic overflow⁵¹⁻⁵³, we added a small
295 quantity of TCA intermediates and acetate to all simulated media. Overall, this model
296 incorporated ecological dynamics and cross-feeding in a realistic fashion while retaining
297 simplicity.

298 Simulations of this model reproduced our two most-prominent experimental observations. That
299 is, we could observe the stable coexistence of many species in single resources (between 19 and
300 28 species, Fig. S16) yet a modest linear increase in richness with the number of resources (slope
301 ~ 2 , Fig. 3c, Fig. S16). Since the estimated metabolic byproducts came from mapped metabolic
302 reactions, we already knew that their number increased non-linearly with supplied resources
303 (Fig. S9). So, how could we get a linear resource-diversity relationship? In our simulations, the
304 concentration of byproducts varied with two factors: their position in the metabolic network and
305 the initial concentration of the supplied resources. To mimic our experiment, we maintained a
306 constant total resource concentration. This resulted in the concentration of each supplied
307 resource decreasing with the total number of supplied resources in the medium (as $1/R$, R being
308 the number of supplied resources). As we provided more resources, a progressively larger
309 fraction of byproducts had their steady-state concentrations decrease non-linearly. Byproducts at
310 very low concentrations could no longer support microbial species, since their growth rates fell
311 below the dilution rate. Hence, even though the number of metabolites grew non-linearly with
312 the supplied resources, the metabolites that could support the growth of new species grew much
313 slower, resulting in a linear increase in diversity in our simulations.

314 By classifying the species in our simulations based on their resource occupancy (as in the
315 experiment), our model also predicted a constant number of generalists and an increasing number
316 of intermediates and specialists with additional resources (Fig. 3c). This is consistent with our
317 experimental observations (Fig. 2g), and further corroborates the idea that the generalists
318 observed in our communities were better at taking advantage of core metabolites, while the
319 specialists that survived were those that were better at competing for rarer metabolites.
320 Importantly, by implementing our model with a realistic network, we were able to simulate all
321 the possible combinations of our sixteen resources, even those that we did not experimentally
322 grow. These simulations showed the same relationship between richness and resource
323 availability (Fig. 3c), suggesting that the outcome of our experiment would not have changed if
324 we had included more/different resource combinations. We concluded that the realistic cross-
325 feeding network seeded by the pool of carbon sources in our experiment could explain the
326 observed relation between microbial community diversity and resource number.



327

328 **Figure 3. A resource explicit model incorporating a realistic metabolic network recapitulates our**
 329 **experimental results.** **a.** A simplified version of the metabolic map derived from KEGG and MetaCyc
 330 databases is shown, where the carbon sources used in the experiments are highlighted. The metabolic map is
 331 used to build the cross-feeding matrix used in the model. **b.** Schematic showing the flow of metabolic
 332 byproducts in our model. Colored circles indicate supplied resources; white circles indicate metabolic
 333 byproducts, i.e., metabolites that are downstream from the resource in the metabolic network; colored
 334 microbes indicate different microbial taxa; and arrows indicate leakage of metabolic byproducts, which serve
 335 as resources for other taxa. **c.** Richness obtained from simulations with all the possible combinations of 16
 336 resources (14,843 conditions in total) is plotted as stacked bars indicating the average number of species for
 337 each category: generalists (pink), specialists (teal) and intermediates (beige). Species in the model are
 338 classified based the number of “media” they survived in, analogously to the distinction applied in the
 339 experiment (Fig. 2). Error bars are omitted for clarity. The total richness increases linearly with the number of
 340 resources (intercept = 22.7, slope = 2).

341 Discussion

342 Understanding the relationship between available nutrients and community diversity is central to
 343 both theoretical and experimental ecology. Here, using a high-throughput culture enrichment
 344 approach amenable to mathematical modeling, we provide experimental and theoretical evidence
 345 of how the identity and the number of available resources modulate microbial community
 346 diversity via a network of metabolic cross-feeding interactions. We showed that the richness of
 347 communities grown on single sources of carbon can be predicted from the number of cross-fed
 348 byproducts generated using intracellular metabolic reactions fueled by those resources. In
 349 addition, using this realistic metabolic network as the cross-feeding network in a resource-

350 explicit model was sufficient to reproduce the observed linear, modest increase of richness with
351 the number of available resources.

352 Our results add to the wealth of studies stressing the importance of metabolic cross-feeding as a
353 pivotal driver of species coexistence⁵⁴ and its link to the identity of available resources^{18,27}. We
354 have observed multispecies communities on all compounds provided as single sources of carbon,
355 including mono-, di- and polysaccharides, sugars alcohols and organic acids. Importantly, we
356 were able to link the variability in community richness to the identity of the supplied resource by
357 mapping the metabolic pathways triggered by each resource and estimating the number of
358 byproducts potentially produced and leaked into the growth media. We are aware that these
359 predictions do not provide any information on the bio-availability of the metabolic byproducts
360 and might be biased towards well-characterized bacterial species. Nevertheless, they provide a
361 simple and tractable way to estimate byproducts using only the structure of an overall metabolic
362 network. Incidentally, community-scale flux balance simulations on the single-resource
363 communities in our experiment also predicted a correlation between the number of expected
364 byproducts and community richness (Fig. S17 and Methods). Further advances in linking the
365 available byproducts with richness could be provided by targeted metabolomics, a technique
366 which can assess the relative concentration of the metabolites in a medium, as in ref.⁵¹.

367 Together with the remarkable richness in single resources, the other striking characteristic of our
368 results was the modest linear increase in community diversity with the number of additional
369 nutrients. Both these features appeared to stem from the structure of the metabolic network
370 seeded by the pool of resources included in the experiment. Indeed, implementing this realistic
371 network in a resource-explicit model was sufficient to recapitulate both features. Another
372 important ingredient of the model was the concentration of metabolites, which in turn depended
373 on their position in the metabolic network and the concentration of the supplied resource they
374 were generated from. Despite its simplifying assumptions (e.g., we used species-independent
375 growth and leakage rates), our model captured the combined effect of dilution and resource
376 concentration that might have determined the diversity of our experimental communities. At the
377 same time, while simulations of our model recapitulate the observed relationship, its theoretical
378 bases still remain to be fully understood, including an extensive exploration of how the structure
379 of the metabolic network affects resource-diversity relationships. Other approaches to
380 complement such theoretical efforts might include experimentally testing the effect of resource
381 concentration and quantitatively modelling intracellular metabolism, thus also accounting for
382 metabolic fluxes and redox balances⁵⁵⁻⁵⁹.

383 The position from which a resource enters the central metabolism affects not only its availability
384 but also the direction in which metabolic reactions run (i.e., glycolytic or gluconeogenic). We
385 showed that whether a resource is glycolytic or gluconeogenic was an important predictor of the
386 diversity and structure of microbial communities, as it dictated the ratio between habitat
387 generalists and specialists. These results suggest, and a two-parameter regression supports (Fig.
388 S18), that adding gluconeogenic resources (e.g., organic acids) while keeping the total
389 concentration of carbon constant may not increase the community diversity. Overall, our results
390 add to the studies stressing that the position from which a resource enters the central metabolism
391 eventually determines its use, including diauxic shifts vs. co-utilization⁶⁰ and tradeoffs between
392 growth and lag in changing environments⁵¹.

393 A further indication of the role played by the metabolic network sustained by the supplied
394 resources came from the striking parallelism that we observed between the structure of
395 experimental communities and the architecture of the network itself. Just like metabolism
396 consists of shared (e.g., the TCA cycle) and unique reaction modules (i.e., specific to the
397 degradation of a particular resource), all experimental communities harbored a core group of
398 metabolically flexible, faster-growing habitat generalists and variable numbers of taxa associated
399 with a particular nutrient (habitat specialists). This suggested that the habitat generalists present
400 in our stabilized microbial communities were “specialists for common nutrients”, i.e., they
401 preferentially consumed substrates that are commonly produced during bacterial growth. In this
402 sense, generalists growing on downstream metabolites (e.g., TCA intermediates) depended on
403 specialists for the production of their favorite substrates. Consistent with this, community flux
404 balance simulations where we paired a generalist and specialist ASV showed that it was the
405 specialists that are likely to leak metabolic byproducts used by generalists, and not *vice versa*
406 (Fig. S19 and Methods).

407 Finally, in our experiments, habitat specialists outnumbered generalists on the whole, a pattern
408 that is commonly observed when natural communities from different locations are compared³⁸.
409 Surveys of microbiomes across different ecosystems have also highlighted a remarkable level of
410 determinism in the association between microbiome composition at coarse taxonomic resolutions
411 (e.g., at the family-level) and availability of nutrients. This feature is recapitulated by other
412 studies^{18,61}. Here we showed both the persistence of strong taxa-resource associations at the ASV
413 (Fig. 2f) and the family level, with the relative abundance of several families, comprising
414 prevalently either generalist or specialist taxa, changing as a function of the relative
415 concentration of specific resources (see Methods for how we established which resources
416 influenced the most each family). Specifically, we observed that the relative abundance of
417 several specialist families decreased drastically or went to zero when the relative concentration
418 of the “favorite resource” dropped by half (e.g., Cellvibrionaceae, Fig. S20), while the relative
419 abundance of generalist taxa increased non-linearly with the relative concentration of few
420 resources, e.g., Pseudomonadaceae with hydroxyproline and fumarate (Fig. S20). The fact that
421 empirically observed features of natural microbial communities emerge in controlled
422 experiments suggests that they might reflect the effects of deterministic processes linked to
423 nutrient availability rather than be generic emergent properties of complex multi-agent systems.

424 **Methods**

425 Growth media preparation

426 All the chemicals were purchased from Sigma-Aldrich unless otherwise stated.

427 All bacterial cultures were grown in M9 media (prepared from 5X M9 salts, 1X Trace Metal
428 Mixture (Teknova) and 1M stock solutions of MgSO₄ and CaCl₂) supplemented with 0.1 % w/v
429 of one of 75 carbon source combinations. These combinations include: 16 compounds commonly
430 available in soil that were provided as single carbon sources (D-(+)-glucose, D-(-)-fructose, D-
431 (+)-xylose, D-(+)-mannose, D-(+)-cellobiose, D-(+)-maltose monohydrate, sucrose, citric acid,
432 fumaric acid, D-(+)-galacturonic acid monohydrate, D-mannitol, D-sorbitol, glycerol, trans-4-
433 Hydroxy-D-proline, methyl cellulose, starch); 24 random combinations of two of these
434 resources; 12 random combinations of four resources; 6 random combinations of eight resources;
435 the 16 combinations containing 15 resources; and all the 16 resources together (see Table S1 for

436 the complete list and Fig. S2). The total concentration of carbon was kept the same and resources
437 were in all instances supplied in equal amounts, that was 100%, 50%, 25%, 12.5%, 6.7% and
438 6.25% each for single-, two-, four-, eight-, 15- and 16-resource combinations. All solutions were
439 filter-sterilized with a 0.22 μm filter and kept at 4°C for the duration of the experiment.

440 Collection of microbial communities from the environment

441 The soil from which the initial inoculum comes from was sampled from a lawn in Cambridge,
442 Massachusetts, at a depth of ~15 cm using a sterile corer and tweezers. Once in the lab, a total of
443 1.5 g of the collected soil was diluted in 20 mL phosphate buffered saline (PBS; Corning), then
444 vortex at intermediate speed for 30 s and incubated on a platform shaker (Innova 2000;
445 Eppendorf) at 250 r.p.m. at room temperature. After 1 hour, the sample was allowed to settle for
446 ~5min and the supernatant was filtered with a 100 μm cell strainer (Thermo Fisher Scientific)
447 and then directly used for inoculation. Both the original soil sample and the remaining
448 supernatant were stored at -80 °C for subsequent DNA extraction.

449 Experimental microcosms

450 Aliquots (7 μL) of the supernatant containing the soil microbial suspension were inoculated into
451 203 μL of growth media in 96-deepwell plates (Deepwell plate 96/500 μL ; Eppendorf), for a
452 total of 231 microcosms (3 replicates for each different resource combinations, except 16-
453 resource combinations that were replicated 9 times). Deepwell plates were covered with
454 AeraSeal adhesive sealing films (Excel Scientific). Bacterial cultures were grown at 30°C under
455 constant shaking at 1,350 r.p.m. (on Titramax shakers; Heidolph). To avoid evaporation, they
456 were incubated inside custom-built acrylic boxes.

457 Every 24 h, the cultures were thoroughly mixed by pipetting up and down 3 times using the
458 VIAFLO 96-well pipettor (Viaflo 96, Integra Biosciences; settings: pipette/mix program
459 aspirating 7 μL , mixing volume 10 μL , speed 6) and then diluted 1/30x into fresh media. We
460 applied a total of seven daily dilution cycles. At the end of every cultivation day we measured
461 the optical density (OD₆₀₀) using a Varioskan Flash (Thermo Fisher Scientific) plate reader. The
462 remaining bacterial culture was frozen at -80 °C for subsequent DNA extraction.

463 DNA extraction, 16S rRNA sequencing and analysis pipeline

464 DNA extraction was performed with the QIAGEN DNeasy PowerSoil HTP 96 Kit following the
465 provided protocol. The obtained DNA was used for 16S amplicon sequencing of the V4 region.
466 Library preparation and sequencing, which was done on an Illumina MiSeq platform, were
467 performed by the MIT BioMicroCenter (Cambridge, Massachusetts).

468 We used the R package DADA2 to obtain the amplicon sequence variants (ASVs)⁶² following
469 the workflow described in Callahan et al.⁶³. Taxonomic identities were assigned to ASVs using
470 the SILVA version 132 database⁶⁴. The phylogenetic tree (Fig. S4) was reconstructed using
471 Randomized Axelerated Maximum Likelihood (RAxML) using default parameters⁶⁵.

472 Data analysis

473 Analysis, unless otherwise stated were conducted in R, version 3.6.1⁶⁶.

474 Sequencing data was handled using the R package phyloseq⁶⁷. We obtained an average of 20,613
475 reads per sample. Sequencing depth did not affect our estimation of community diversity indexes

476 (Fig. S4). Richness was calculated as the number of ASVs with abundance larger than 0 found in
477 each sample. Community diversity was also measured by Shannon Diversity index and Shannon
478 Entropy index following^{68,69} (Fig. S10). The significance of differences in richness due to single
479 supplied resources was tested through ANOVA⁷⁰ using the package GAD.

480 *Richness predictions*

481 Predictions of how richness would grow with the number of supplied carbon sources were
482 computed using all the three replicated communities grown on a single resource and all the
483 possible combinations of single resources (120 combinations of two resources, 1,820
484 combinations of four, 12,870 combinations of eight, 16 combinations of 15 and one combination
485 of 16 resources) (Fig. 1D). As an example of the prediction based on the maximum of constituent
486 singles, the richness of the community grown in a medium containing glucose + hydroxyproline
487 was obtained by calculating the maximum richness over each couple of replicates (one
488 containing only glucose and the other containing only hydroxyproline) and subsequently
489 averaging across all the predicted maxima (in total 9 predicted values). The same procedure was
490 used for the average of constituent singles. Analogously, for the predictions based on the union
491 of constituent singles, the richness in glucose + hydroxyproline was predicted by calculating the
492 number of unique ASVs found in each couple of replicates of constituent singles (i.e., the total
493 number of ASVs minus the number of overlapping ASVs) and then averaging across all obtained
494 unions (9 values).

495 *Rank abundance distributions*

496 First, we computed abundance distributions (RADs) for each sample, i.e., each replicate
497 community grown on a unique combination of carbon sources, by sorting ASVs based on their
498 relative abundance. Then, we plotted the RADs in a log-linear fashion and fitted a regression line
499 in order to compare their slopes (Fig. S11A). The absolute value of the slope of the fitted
500 regression line informs on the abundance distribution of the ASVs in a community. More even
501 communities usually display smaller slopes (Fig. S11B). Since each community exhibited a
502 different richness, we normalized the RADs for richness (Fig. S11C) To do this, we used the
503 *RADnormalization_matrix* function in the RADanalysis package: from each RAD with an
504 observed richness, this function generates a “normalized RAD” with a richness corresponding to
505 the minimum richness observed in the experiment (7 ASVs) by randomly resampling the original
506 RADs for 10 times⁷¹. In this way, samples with different richness can be compared and changes
507 in evenness properly assessed.

508 *Definition of generalists and specialists based on single resource occupancy*

509 ASVs found in single resources were classified in three categories based on how many media
510 containing a single resource they were found in, i.e. they exhibited abundance larger than 0^{38,43}.
511 We considered specialists the ASVs that were observed in less than 25% of single-resource
512 media, i.e., in one, two or three resources. Generalists were those ASVs found in more than the
513 75% of media, i.e., in 13 or more resources. We defined intermediates the ASVs found between
514 four and twelve resources. These thresholds were chosen arbitrarily, but the resulted in about ~
515 4% generalists and 80% specialists, consistently with proportions of generalists and specialists
516 observed in natural communities^{38,39,43}. We chose this simple way of assigning ASVs to
517 generalist, intermediate and specialist categories over other methods, e.g. as in²⁴ in order to
518 leave aside their relative abundance, which was analyzed separately.

519 *Prediction of possible metabolic byproducts in resource environments*

520 We predicted the possible number of metabolic byproducts that could be produced using the
521 resources present in each medium using a curated metabolic network. The metabolic network
522 contained a large set of metabolic reactions encompassing carbohydrate, sugar and amino acid
523 metabolism extracted from the KEGG database³¹. We manually curated this large set of reactions
524 using the MetaCyc database³² in order to limit it to reactions possible by most microbial taxa
525 common to the soil, such as *Pseudomonas*. We used this network to estimate all the metabolic
526 compounds that could be produced as byproducts, starting from the carbon sources available in
527 each medium. We assumed that a small set of “currency” molecules, such as water, carbon
528 dioxide and ATP, were always available as reactants when required (full list of currency
529 molecules: phosphate, oxygen, carbon dioxide, water, H⁺, ATP, NAD(P)H, Acetyl-CoA, CoA).

530 To estimate the possible byproducts in each medium, we employed the well-known scope
531 expansion algorithm^{72–76}. Each reaction in our curated metabolic network consisted of a set of
532 reactants and resulting products. For each medium, we first asked which reactions could be
533 performed using only the carbon sources available in the medium (i.e., the current “scope” of the
534 medium). We assumed that the products of these reactions could be produced and added them to
535 the set of reactants – the new scope – for the next step. In the next step, we again asked which
536 reactions could be performed using the new scope. We added their products to the scope for the
537 next step. We continued this process, step by step, until we could add no new products to the
538 scope. The resulting final scope of metabolites, minus the currency molecules provided in the
539 medium, was our estimated set of possible metabolic byproducts producible in that medium.

540 Adding some amino acids as currency molecules, mimicking our experimental protocol, yielded
541 a larger set (~3x) of possible metabolic byproducts for each medium, including many amino
542 acids and anabolic products. This expanded set of metabolites for each medium was also
543 correlated with the observed average species richness in that medium (data not shown).

544 We also tried an alternative approach to estimate the number of metabolic byproducts in single
545 resource environments, using community-scale flux balance simulations. For each ASV observed
546 in a single resource environment, we first obtained the phylogenetically closest whole genome
547 sequence in NCBI’s RefSeq database. For this, we mapped the 16S sequence of each ASV to
548 complete genomes in the RefSeq database using BLAST⁷⁷. For each ASV, we chose the genome
549 that had the highest identity; when multiple genomes matched this criterion, we chose the longest
550 genome, following similar work⁷⁸. We then obtained all the mapped genome sequences and
551 constructed metabolic models for each of them using CarveMe⁵⁵; we gap-filled all models to
552 grow on M9 minimal medium supplemented with metal ions, such as iron and copper, which are
553 present in trace amounts in experimental bacterial growth media.

554 To estimate the number of metabolic byproducts in each single resource environment, we
555 performed community-scale metabolic simulations using the package MICOM⁵⁷. For each
556 community, we input all the metabolic models for all ASVs detected in that community, and
557 simulated their growth in the corresponding media. We then counted all metabolites which were
558 predicted to be exported by each community as the estimated number of byproducts for that
559 community. For each medium, we averaged the number of byproducts across all three replicate
560 communities; we used this as our estimated number of byproducts for that medium.

561 *Characterization of the structure of the metabolite pool*

562 Following the same logic that we used for ASVs, metabolites estimated to be produced through
563 metabolic reaction starting from single resources were classified in three categories based on the
564 number of resources that could be produced from. We considered *rare* metabolites those
565 observed in less than 25% of single-resource media, i.e., in one, two or three resources. In
566 contrast, *common* metabolites were those found in more than the 75% of media, i.e., in 13 or
567 more resources. Finally, *intermediate* metabolites were those present in between four and twelve
568 resources. The chosen thresholds separate the metabolites of the central metabolic pathway
569 (common metabolites) from the peripheral metabolites belonging to branches descending into the
570 central pathway (rare and intermediate metabolites).

571 *Inference of rRNA operon copy number for generalist and specialist taxa*

572 To test for signatures of different life-history strategies of the generalist and specialist taxa in our
573 study, we estimated their 16S rRNA operon copy numbers. We estimated rRNA copy numbers at
574 the level of both genus and family, separately for generalist and specialist taxa. For each genus
575 identified, we queried rrnDB⁷⁹—a database of rRNA operon copy number statistics—for the
576 median copy number corresponding to the genus. We used this as an estimate for the rRNA
577 operon copy number of that genus.

578 *Inference of number of metabolic genes for generalist and specialist taxa*

579 To test for metabolic differences between the generalist and specialist taxa in our study, we
580 estimated the number of metabolic genes in their genomes. Since we did not have either isolates
581 or assembled genomes corresponding to the observed taxa, we relied on a popular indirect
582 method of estimating gene content. Namely, for each ASV, we used the reference genome which
583 was phylogenetically closest to that ASV as a proxy for its genome. For this, we used
584 PICRUSt2⁸⁰ using default parameters; as an input to the tool, we provided the 16S rRNA
585 sequences of all 226 generalist and specialist taxa as well as their abundances in each sample.
586 After running PICRUSt2, we obtained a table of the predicted gene content for each ASV (i.e.,
587 presence/absence of a specific KO number in the KEGG database). We extracted all metabolic
588 genes from this table by only choosing those KO numbers which had at least one known
589 metabolic reaction corresponding to them. Doing so resulted in an estimated set of metabolic
590 genes for each ASV; we used this as an indirect estimate of the metabolic capabilities of each
591 ASV.

592 *Calculation of the resource-specificity score*

593 We used a resource-specificity score to test if the ASV-resource associations that we observed in
594 single resources were maintained when the single resource(s) in which the ASV was found was
595 combined with others. For each ASV present in a single resource (target resource), the resource
596 specificity score is calculated as the difference between the number of multi-resource media
597 containing the target resource in which the ASV is found and the number of media not
598 containing the target resource in which the ASV is found divided by the total number of media in
599 which the ASV is found (Fig. 2E). This is reminiscent of a preference index, which is a standard
600 measure in the behavioral sciences. Single resources are excluded from the count. The resource-
601 specificity score ranges from 1, indicating that the ASV is present only when the target resource
602 is provided, to -1, implying that the ASV is always absent when that resource is supplied with
603 other resources. A score of 0 is indicative of an ASV showing no specificity for that particular
604 resource (Fig. 2E). We calculated a score for each ASV-resource pair, such that each ASV had

605 as many scores as the number of single resources is found in. Then, we computed the average of
606 the scores obtained for each single resource, separating between scores belonging to generalist
607 and specialist ASVs (Fig. 2E).

608 *Inference of metabolic interactions between generalist and specialist taxa*

609 To estimate whether metabolic interactions between the generalist and specialist taxa in our
610 communities were likely to be unidirectional or bidirectional, we used SMETANA v.1.0⁵⁶, using
611 default settings. For each generalist-specialist pair that we experimentally detected in single
612 resource environments, we used SMETANA on a model community comprising both ASVs (a
613 generalist and a specialist) using the settings --flavor bigg --exclude inorganic.txt -d. We
614 explicitly disallowed inorganic molecules such as phosphates, carbon dioxide and metal ions
615 from being exchanged by using the --exclude option in SMETANA. To consider interaction
616 directionality, we looked at the donor and receiver of each exchanged metabolite. When there
617 was only one donor for every exchanged metabolite, we inferred the interaction as unidirectional,
618 with the direction going from the donor to the receiver of the metabolites.

619 *Detection of family-resource associations using an ensemble tree regression model*

620 We calculated the relative abundance of the most prevalent families (37) in the 75 replicated
621 bacterial communities and ran an ensemble tree regression model to detect significant patterns of
622 variations in family abundance due to changes in the relative concentration of resources.

623 We chose to coarse-grain the abundance data at the family level because, while several ASVs
624 were lost and others were gained going from one to 16 resources in the growth media, the
625 families found across all combinations of resources were mostly the same. In addition, we
626 distinguished between generalist families, i.e., those that contained at least one generalist ASV,
627 and specialist families, i.e., containing only specialist ASV. Consistent with ASV-level
628 definition, generalist families displayed higher mean rRNA operon copy number compared to
629 specialist families.

630 We employed XGBoost, a gradient boosting framework based on decision trees⁸¹. Specifically,
631 we implemented a regression model for each family in which the input was the relative resource
632 concentration and the output was the log-transformed relative family abundance. We trained the
633 model on two replicates by performing leave-one-out crossvalidation of the XGBoost parameters
634 “max_depth”, “n_estimators” and “learning_rate”⁸² and tested on the third one with average
635 mean-squared error across families of 6.05. We applied the Shapley Additive exPlanations
636 (SHAP)⁸³ to identify the resources that were more important in driving changes in the abundance
637 of each family. This analysis has been done using Python version 3.8.

638 Results of this analysis revealed that variations in the abundance all of the 37 families were
639 driven by one or multiple resources based on their dominant life strategy. To simplify the
640 visualization of the results we plotted the relative abundance of some representative families as a
641 function of the concentration of the resources identified by the analysis (Fig. S20). Families
642 mostly composed of specialist taxa, e.g., Cellvibrionaceae and Bacillaceae, showed abrupt
643 changes in their abundance with the concentration of the “favorite” resource (Fig. S20). By
644 contrast, more generalist families, e.g., Pseudomonadaceae and Enterobacteriaceae, exhibited
645 smooth trends in their abundance with the concentration of multiple resources.

646 *Resource-consumer model with cross-feeding and simulations*

647 The parallelism between species and metabolite distribution (see Fig. 2) that we observed in our
648 experiment highlighted that the cross-feeding network is key to understand microbial
649 communities under each combination of supplied carbon sources. To test this idea, we used a
650 model encompassing the metabolic network that we obtained from the analysis of KEGG and
651 MetaCyc databases. This was achieved by a consumer-resource model with cross-feeding^{18,21,35}.
652 In our consumer-resource model with the realistic metabolic network, we made the following
653 simplifying assumptions.

654 First, we assumed that every species consumes only one preferred metabolite. Upon this
655 assumption, competitive exclusion guarantees that only the best grower in each resource
656 survives; thus, we implemented only one species for each resource in our simulation as a post-
657 selection pool. This assumption reflected the resource-species association we observed (Fig. 2),
658 which suggested that the taxa identified as generalists may specialize on core metabolites that are
659 found everywhere. Also, while many species can consume multiple resources, they may still
660 grow much faster on the most preferred one. Metabolic strategies such as *diauxie* also highlight
661 that growth on the most preferred resource can be a dominant factor for community assembly⁵¹.

662 Second, growth rates, biomass yield, and leakage rates (these quantities are described below) are
663 universal, independent of species identity. This assumption led to the simplest implementation of
664 our metabolite network.

665 Third, we assumed that each species leaked out all the immediate metabolites of the metabolite it
666 consumes. The list of immediate metabolites that are produced from each metabolite was
667 obtained from scope expansion analysis. This information is encoded by a cross-feeding matrix
668 CF_{ij} , which is nonzero when i^{th} metabolite immediately leaks j^{th} metabolite and 0 otherwise.
669 For simplicity, the nonzero values of CF_{ij} are set to be $1/(\text{number of metabolites produced by } i^{th}$
670 $\text{metabolite})$.

671 The scope expansion analyses based on the metabolic reactions mapped in KEGG and MetaCyc
672 databases identified 96 metabolites that could be produced starting from the supplied carbon
673 sources. Thus, CF is a 96x96 matrix. The original scope expansion analysis included reactions
674 where multiple reactants were required to generate products. Since it is impossible to fully
675 capture such interdependences with a matrix, we assumed that reactions were activated as long
676 as one or more reactants were present. Also, to mimic the highly connected and cyclic structure
677 of TCA-cycle, we set each TCA intermediate to generate all other TCA intermediates.

678 Under these assumptions, we simulated the dynamics of the following model:

$$\begin{aligned} \dot{n}_i &= (1 - l)r c_i n_i - \delta n_i \\ \dot{c}_i &= -r n_i c_i + l r \sum_j CF_{ji} n_j c_j + \delta (c_{i0} - c_i) \end{aligned}$$

679 where N_i is the population of i^{th} species, and c_i is the concentration of i^{th} metabolite. l is the
680 leakage rate, r_i is the per-capita, per-resource growth rate of i^{th} species, delta is the dilution rate
681 of the chemostat-like environment. CF_{ij} tells whether i^{th} metabolite is leaked from j^{th}
682 metabolite based on the scope expansion analysis. c_{i0} is the supply resource concentration
683 corresponding to each combination of supplied carbon sources, controlled by overall scale c_0 .
684 For example, when glucose is supplied, $c_{i0} = c_0$ for i =glucose and 0 otherwise. And when a

685 combination of glucose and hydroxyproline is supplied, $c_{i0} = \frac{1}{2}c_0$ for i =glucose, hydroxyproline
686 and 0 otherwise. To simulate the effects of metabolic overflow⁵¹⁻⁵³, we supplied a small quantity
687 ($c = 0.2$) of TCA intermediates and acetate to all media.

688 The first equation models the dynamics of population. The first term tells that the growth rate of
689 each species is proportional to the concentration of the preferred resource. We also assumed that
690 species can only convert a fraction $1 - l$ of the preferred resource into biomass, since l is leaked
691 in the environment as by-product(s). The second term represents dilution as the main driver of
692 mortality in this chemostat-like system.

693 The second equation models the dynamics of resources (both supplied and cross-fed). The first
694 term represents the consumption of the resource by the specialized species. The second term
695 represents the leakage from upstream resources that cross-feed the i^{th} resource. The third term
696 represents the dilution and external supply of resource in the chemostat system.

697 We simulated the model dynamics under all possible combinations of 1, 2, 4, 8, 15, and 16
698 number of supplied resources (14843 combinations total). We chose the parameters $\delta = 0.1$,
699 which is comparable to the dilution we imposed in the experiment, $r = 1$, $c_0 = 100$, and
700 $l = 0.1$. In Fig. 3c we show the results of all combinations, while in Fig. S16 we plotted only the
701 combinations included in the experiment. The simulations were run for $1e^3$ unit time starting
702 from initial population set as e^{-7} , and communities reached equilibrium at the end of the
703 simulations. The population cutoff for survival was set as e^{-7} . Simulations were run in Python
704 version 3.7.4.

705 Data availability

706 Data files and analysis/simulation codes will be available via GitHub upon publication. 16S
707 Amplicon sequencing data and metadata files have been deposited in the NCBI SRA database
708 under NCBI BioProject ID PRJNA715195.

709 **References**

- 710 1. Hubbell, S. *The Unified Neutral Theory of Biodiversity and Biogeography*. (Princeton University Press,
711 2001). doi:10.1016/s0925-8574(01)00094-5
- 712 2. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science (80-.)*. **294**, 843–
713 845 (2001).
- 714 3. Hutchinson, G. E. *The Paradox of the Plankton*. *The American Naturalist* **95**,
- 715 4. Koskella, B., Hall, L. J. & Metcalf, C. J. E. The microbiome beyond the horizon of ecological and
716 evolutionary theory. *Nat. Ecol. Evol.* **1**, 1606–1615 (2017).
- 717 5. Flemming, H. C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev.*
718 *Microbiol.* **17**, 247–260 (2019).
- 719 6. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, (2016).
- 720 7. Salazar, G. & Sunagawa, S. Marine microbial diversity. *Current Biology* **27**, R489–R494 (2017).
- 721 8. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–
722 214 (2012).
- 723 9. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive earth's biogeochemical

- 724 cycles. *Science* **320**, 1034–1039 (2008).
- 725 10. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut
726 microbiome and the immune system. *Nature* **474**, 327–336 (2011).
- 727 11. Cavicchioli, R. *et al.* Scientists’ warning to humanity: microorganisms and climate change. *Nature Reviews*
728 *Microbiology* **17**, 569–586 (2019).
- 729 12. Blasche, S. *et al.* Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial
730 community. *Nat. Microbiol.* **6**, 196–208 (2021).
- 731 13. Gude, S. *et al.* Bacterial coexistence driven by motility and spatial competition. *Nature* **578**, 588–592
732 (2020).
- 733 14. Kommineni, S. *et al.* Bacteriocin production augments niche competition by enterococci in the mammalian
734 gastrointestinal tract. *Nature* **526**, 719–722 (2015).
- 735 15. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The Evolution and Ecology of Bacterial Warfare.
736 *Current Biology* **29**, R521–R537 (2019).
- 737 16. Ratzke, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in
738 microbial communities. *Strength species Interact. Determ. Biodivers. Stab. Microb. communities* 671008
739 (2019). doi:10.1101/671008
- 740 17. Hoek, T. A. *et al.* Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial
741 Cross-Feeding Mutualism. *PLOS Biol.* **14**, e1002540 (2016).
- 742 18. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science (80-.)*. **361**, 469–474
743 (2018).
- 744 19. Tilman, D. *Resource Competition and Community Structure. Resource Competition and Community*
745 *Structure. (MPB-17), Volume 17* (Princeton, NJ: Princeton University Press, 1982).
746 doi:10.2307/j.ctvx5wb72
- 747 20. Gause, G. F. *The struggle for existence.* (Hafner Press, New York, 1934).
- 748 21. MacArthur, R. Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11
749 (1970).
- 750 22. Levin, S. A. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle.
751 *Am. Nat.* **104**, 413–423 (1970).
- 752 23. Estrela, S. *et al.* Metabolic rules of microbial community assembly. *bioRxiv* 2020.03.09.984278 (2020).
753 doi:10.1101/2020.03.09.984278
- 754 24. Enke, T. N. *et al.* Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Curr.*
755 *Biol.* (2019). doi:10.1016/j.cub.2019.03.047
- 756 25. Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in
757 synthetic phycospheres. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3656–3662 (2020).
- 758 26. Gralka, M., Szabo, R., Stocker, R. & Cordero, O. X. Trophic Interactions and the Drivers of Microbial
759 Community Assembly. *Curr. Biol.* **30**, R1176–R1188 (2020).
- 760 27. Enke, T. N. *et al.* Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Curr.*
761 *Biol.* **29**, 1528–1535.e6 (2019).
- 762 28. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic
763 perspective. *Science (80-.)*. **350**, aac9323–aac9323 (2015).
- 764 29. Naylor, D. *et al.* Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules. *MBio*
765 **11**, (2020).

- 766 30. MacArthur, R. H. *Geographical Ecology. Patterns in the Distribution of Species*. (Harper & Row, 1972).
767 doi:10.2307/3393
- 768 31. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**,
769 27–30 (2000).
- 770 32. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633–
771 D639 (2018).
- 772 33. Wang, T., Goyal, A., Dubinkina, V. & Maslov, S. Evidence for a multi-level trophic organization of the
773 human gut microbiome. *PLOS Comput. Biol.* **15**, e1007524 (2019).
- 774 34. Goyal, A. & Maslov, S. Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial
775 Ecosystems. *Phys. Rev. Lett.* **120**, 158102 (2018).
- 776 35. Marsland, R. *et al.* Available energy fluxes drive a transition in the diversity, stability, and functional
777 structure of microbial communities. *PLoS Comput. Biol.* **15**, (2019).
- 778 36. Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species diversity.
779 *Nature* **461**, 254–257 (2009).
- 780 37. Tromas, N. *et al.* Niche Separation Increases With Genetic Distance Among Bloom-Forming Cyanobacteria.
781 *Front. Microbiol.* **9**, 438 (2018).
- 782 38. Sriswasdi, S., Yang, C. C. & Iwasaki, W. Generalist species drive microbial dispersion and evolution. *Nat.*
783 *Commun.* **8**, 1–8 (2017).
- 784 39. Logares, R. *et al.* Biogeography of bacterial communities exposed to progressive long-term environmental
785 change. *ISME J.* **7**, 937–948 (2013).
- 786 40. Monard, C., Gantner, S., Bertilsson, S., Hallin, S. & Stenlid, J. Habitat generalists and specialists in
787 microbial communities across a terrestrial-freshwater gradient. *Sci. Rep.* **6**, 1–10 (2016).
- 788 41. Székely, A. J. & Langenheder, S. The importance of species sorting differs between habitat generalists and
789 specialists in bacterial communities. *FEMS Microbiol. Ecol.* **87**, 102–112 (2014).
- 790 42. Pandit, S. N., Kolasa, J. & Cottenie, K. Contrasts between habitat generalists and specialists: An empirical
791 extension to the basic metacommunity framework. *Ecology* **90**, 2253–2262 (2009).
- 792 43. Muscarella, M. E., Boot, C. M., Broeckling, C. D. & Lennon, J. T. Resource heterogeneity structures aquatic
793 bacterial communities. *ISME J.* **13**, 2183–2195 (2019).
- 794 44. Roller, B. R. K., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to investigate
795 bacterial reproductive strategies. *Nat. Microbiol.* **1**, 1–7 (2016).
- 796 45. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic
797 perspective. *Science* **350**, (2015).
- 798 46. Goldfarb, K. C. *et al.* Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying
799 Chemical Recalcitrance. *Front. Microbiol.* **2**, 94 (2011).
- 800 47. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological
801 strategies of bacteria. *Appl. Environ. Microbiol.* **66**, 1328–1333 (2000).
- 802 48. Rojo, F. Carbon catabolite repression in *Pseudomonas*: Optimizing metabolic versatility and interactions
803 with the environment. *FEMS Microbiology Reviews* **34**, 658–684 (2010).
- 804 49. Mills, C. G., Allen, R. J. & Blythe, R. A. Resource spectrum engineering by specialist species can shift the
805 specialist-generalist balance. *Theor. Ecol.* **13**, 149–163 (2020).
- 806 50. Bajic, D. & Sanchez, A. The ecology and evolution of microbial metabolic strategies. *Current Opinion in*
807 *Biotechnology* **62**, 123–128 (2020).

- 808 51. Basan, M. *et al.* A universal trade-off between growth and lag in fluctuating environments. *Nature* **584**,
809 470–474 (2020).
- 810 52. Paczia, N. *et al.* Extensive exometabolome analysis reveals extended overflow metabolism in various
811 microorganisms. *Microb. Cell Fact.* **11**, 122 (2012).
- 812 53. Pinu, F. R. *et al.* Metabolite secretion in microorganisms: the theory of metabolic overflow put to the test.
813 *Metabolomics* **14**, (2018).
- 814 54. Douglas, A. E. The microbial exometabolome: Ecological resource and architect of microbial communities.
815 *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, (2020).
- 816 55. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale
817 metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
- 818 56. Zelezniak, A. *et al.* Metabolic dependencies drive species co-occurrence in diverse microbial communities.
819 *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6449–6454 (2015).
- 820 57. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer
821 Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, (2020).
- 822 58. Garza, D. R., Van Verk, M. C., Huynen, M. A. & Dutilh, B. E. Towards predicting the environmental
823 metabolome from metagenomics with a mechanistic model. *Nat. Microbiol.* **3**, 456–460 (2018).
- 824 59. Pacheco, A. R. & Segrè, D. The effects of environmental complexity on microbial community yield and
825 structure. *bioRxiv* 2020.07.09.195404 (2020). doi:10.1101/2020.07.09.195404
- 826 60. Wang, X., Xia, K., Yang, X. & Tang, C. Growth strategy of microbes on mixed carbon sources. *Nat.*
827 *Commun.* **10**, 1–7 (2019).
- 828 61. Louca, S. *et al.* High taxonomic variability despite stable functional structure across microbial communities.
829 *Nat. Ecol. Evol.* **1**, (2017).
- 830 62. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*
831 **13**, 581–583 (2016).
- 832 63. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor Workflow
833 for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research* **5**, 1492 (2016).
- 834 64. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based
835 tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
- 836 65. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.
837 *Bioinformatics* **30**, 1312–1313 (2014).
- 838 66. Core, R. & Computing. R: A Language and Environment for Statistical Computing. (2019).
- 839 67. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics
840 of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).
- 841 68. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation
842 in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
- 843 69. Chao, A., Chiu, C. H. & Jost, L. Phylogenetic diversity measures based on Hill numbers. *Philos. Trans. R.*
844 *Soc. B Biol. Sci.* **365**, 3599–3609 (2010).
- 845 70. Underwood, A. J. *Experiments in Ecology. Experiments in Ecology* (Cambridge University Press, 1997).
846 doi:10.1017/cbo9780511806407
- 847 71. Saeedghalati, M. *et al.* Quantitative Comparison of Abundance Structures of Generalized Communities:
848 From B-Cell Receptor Repertoires to Microbiomes. *PLoS Comput. Biol.* **13**, e1005362 (2017).

- 849 72. Goldford, J. E., Hartman, H., Smith, T. F. & Segrè, D. Remnants of an Ancient Metabolism without
850 Phosphate. *Cell* **168**, 1126–1134.e9 (2017).
- 851 73. Raymond, J. & Segrè, D. The effect of oxygen on biochemical networks and the evolution of complex life.
852 *Science* (80-.). **311**, 1764–1767 (2006).
- 853 74. Handorf, T., Ebenhoöh, O. E. & Heinrich, R. Expanding Metabolic Networks: Scopes of Compounds,
854 Robustness, and Evolution. doi:10.1007/s00239-005-0027-1
- 855 75. Ebenhoöh, O., Handorf, T. & Heinrich, R. *Structural Analysis of Expanding Metabolic Networks*. *Genome*
856 *Informatics* **15**, (2004).
- 857 76. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* **28**, 245–248
858 (2010).
- 859 77. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, 5–9 (2008).
- 860 78. Machado, D. *et al.* Polarization of microbial communities between competitive and cooperative metabolism.
861 *Nat. Ecol. Evol.* 1–9 (2021). doi:10.1038/s41559-020-01353-4
- 862 79. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. & Schmidt, T. M. rrnDB: Improved tools for
863 interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development.
864 *Nucleic Acids Res.* **43**, D593–D598 (2015).
- 865 80. Douglas, G. M. *et al.* PICRUSt2: An improved and extensible approach for metagenome inference. *bioRxiv*
866 (2019). doi:10.1101/672295
- 867 81. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of the ACM SIGKDD*
868 *International Conference on Knowledge Discovery and Data Mining* **13-17-August-2016**, 785–794
869 (Association for Computing Machinery, 2016).
- 870 82. Hastie, Trevor, Tibshirani, Robert, Friedman, J. *The Elements of Statistical Learning The Elements of*
871 *Statistical Learning Data Mining, Inference, and Prediction, Second Edition*. *Springer series in statistics*
872 (2009). doi:10.1007/978-0-387-84858-7
- 873 83. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat.*
874 *Mach. Intell.* **2**, 56–67 (2020).

875

876 **Acknowledgements.** The authors would like to thank Jacopo Grilli, Marco Costantino-
877 Lagomarsino, Mattia Corigliano and Matthieu Barbier for feedback on models, the members of
878 Gore Lab for comments on the manuscript, and Bartolomeo Stellato for the help with the
879 ensemble tree regression model. This work was supported by NIH and the Schmidt Foundation.
880 A.G. is supported by the Gordon and Betty Moore Foundation as a Physics of Living Systems
881 Fellow through grant number GBMF4513.

882 **Author contributions.** MDB and JG conceived the study. MDB performed the experiments and
883 the sequencing analysis. HL performed theoretical modeling. AG performed metabolic and
884 genomic analyses. All authors analyzed the data and wrote the manuscript.

885 **Competing interests.** The authors declare no competing interests.

886 **Materials & Correspondence.** Data files and analysis/simulation codes will be available via
887 GitHub upon publication and can be requested to dalbelo@mit.edu. 16S Amplicon sequencing
888 data and metadata files have been deposited in the NCBI SRA database under NCBI BioProject
889 ID PRJNA715195.