Title: A simple linear relationship between resource availability and microbial community diversity

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

Abstract: Microbial community diversity is pivotal for the functioning of our planet, but its drivers are still unclear, in particular the role of resource number and identity. To fill this gap, we studied the assembly of hundreds of soil-derived microbial communities on a wide range of well-defined resource environments, from single carbon sources to combinations of up to 16. We found a remarkable diversity in single resources but a linear one-by-one increase in the number of species with the number of additional resources. We show, both experimentally and theoretically, that both these observations could originate from generalist and specialist taxa interacting in a modular fashion within the community. Since generalists and specialists are ubiquitous in natural microbiomes, our results might apply to a variety of different ecological settings, providing a framework to predict how community diversity responds to changes in resource availability.

One Sentence Summary: While many species coexist in single resources, community diversity only increases one-by-one as more resources are added.

Main Text:

Microbial communities are extremely diverse and contribute to the health and function of every ecosystem (1). Although photosynthetic bacteria are among the most abundant organisms on Earth (2), many microbial communities, including those living in the human gut and in the soil, rely on externally supplied carbon sources to survive and thrive (3). A major question is thus: do larger numbers of available resources mean more coexisting species? This is still largely unresolved because, traditionally, different facets of resource availability have been tackled separately, with theory focusing on the effects of the number of resources on species coexistence, e.g. (*4–7*), and experiments concentrating on the impact of resource identity on community composition and function (*8–10*). Yet, understanding how the identity and number of available resources shape microbial diversity is key to anticipating the response of ecosystem to alterations in the environment, including global climate change and evolving human diets (*11*).

In order to illuminate how resource number and identity shape the richness of microbiomes, we assayed the assembly of soil-derived bacterial communities in laboratory microcosms (*8*, *12*). We started by inoculating a rich microbial suspension obtained from a soil sample (Fig. S1) into 75 resource environments, each containing minimal media supplemented with different combinations of carbon sources, ranging from one to 16 (Fig. 1A, S2, Table S1). The 16 carbon sources represented a broad range of common soil compounds (e.g. mannose, xylose, cellulose and hydroxyproline), encompassing simple sugars as well as complex polymers. We adopted a daily-dilution protocol, whereby at the end of each 24-hour growth cycle the bacterial cultures were diluted 1/30x into fresh media. We observed that the majority of microcosms reached stability after 3 days from the inoculum (Fig. S3). We continued the experiment until day 7 and measured the final richness as the number of ASVs (amplicon sequence variants) observed within each community (Fig. S4).

Consistent with recent experimental studies (*8*, *9*, *13*), single-resource communities were remarkably rich (mean richness = 22.5 $\pm$ 2 ASVs, Fig. 1B) and taxonomically diverse (Fig. S5). This is in contrast with competitive exclusion predicting that the number of species cannot exceed the number of resources (*14*, *15*)—which, in single resources, would result into no more than one species surviving. Interestingly, the variability in richness among different resources was also high—with the average number of ASVs ranging from eight, in citrate, to ~40, in cellulose (Fig. 1B)—and larger than the variability among replicates of the same carbon source (ANOVA test, $F_{resource} = 3.4339$, $p < 0.01$). Richness in single carbon sources therefore depended on resource identity. Community richness did not correlate significantly with molecular weight (Fig. S6), but did correlate with the number of metabolites predicted to be generated from biochemical reactions within the cell (Fig. S7), consistent with the idea that extensive cross-feeding allows for the stable coexistence of many species on single resources (*16–18*).

The large numbers of coexisting species in single resources generated the expectation that community diversity would increase rapidly if more resources were provided. As previously observed in marine bacteria (*9*, *19*), community composition could be roughly predicted as the sum of the assemblages observed on each nutrient supplied in the mixture. To provide an example, the expected richness of the community grown on glucose and hydroxyproline (Fig. 1C), each alone supporting on average 24 and 11 ASVs, would be ~ 30 ASVs, i.e. the sum minus the number of shared ASVs (union). Alternatively, niche overlap between the taxa found in the single-resource media (*20*, *21*) might bring the expected number of species down to the

maximum richness observed in the constituent singles; in the case of glucose + hydroxyproline, 24 ASVs. However, when we measured the richness of the communities grown in a media supplied with equal amounts of glucose and hydroxyproline, we found only ~16 ASVs on average, which is significantly lower than both expectations (Fig. 1C). Yet, our observed richness is remarkably similar to the mean richness measured in the two constituent single resources (17.5 ASVs), a trend that was consistent across many two-resource communities (Fig. S8). Contradicting our expectations based on previous results supporting additivity, we found that the community richness upon combining two carbon resources was approximately the average richness of single resource environments.

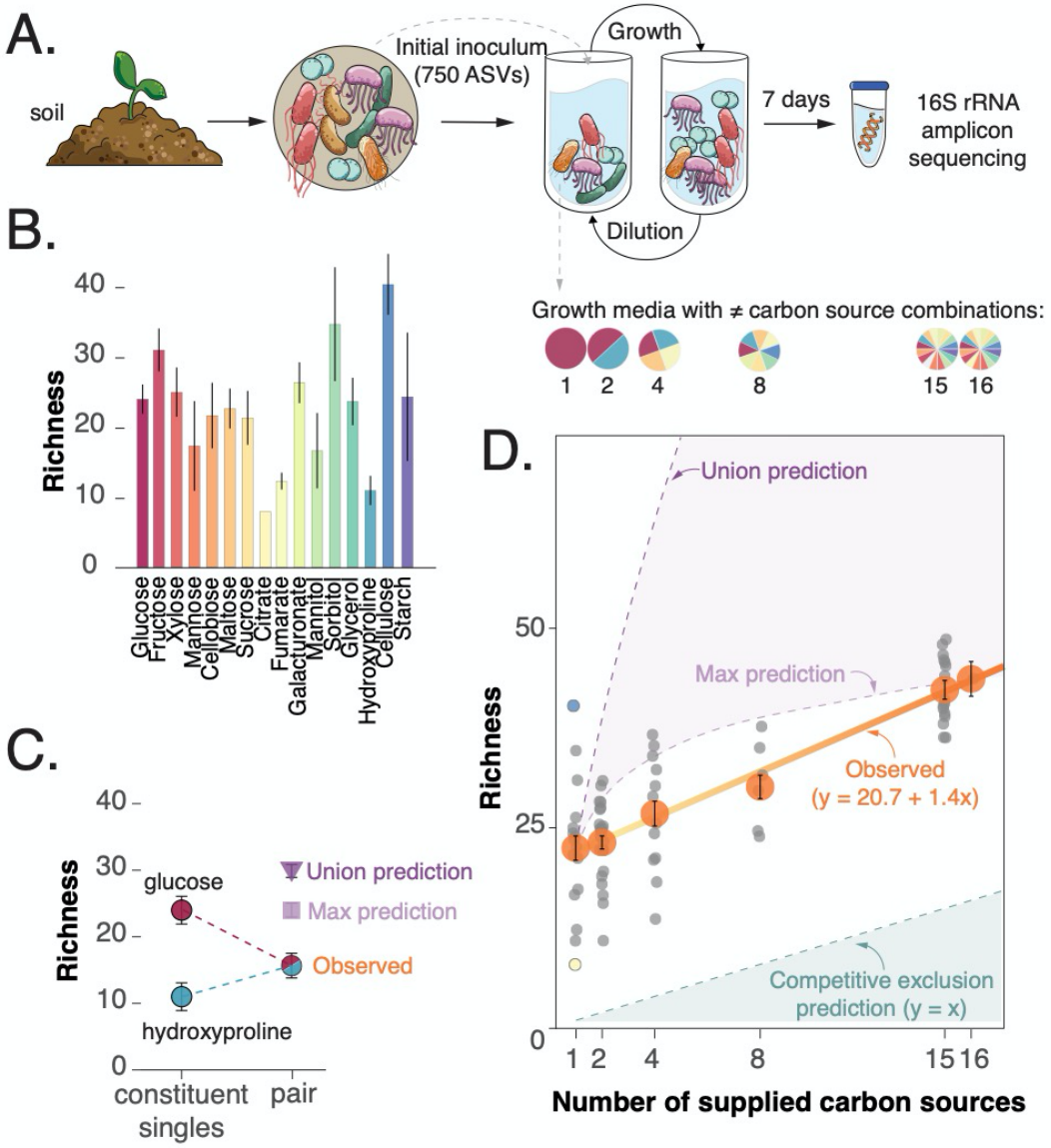

Fig. 1. Richness increases weakly and linearly with the number of supplied carbon sources, despite the high richness observed in single resources. A. Layout of the experiment. We inoculated a rich microbial

suspension obtained from a soil sample into 75 growth media, each supplemented with a different combination of carbon sources, from single compounds to a mix of 16, while keeping the total carbon concentration the same (0.1% w/v). Bacterial cultures were grown for 7 days under a regime of daily dilution and their composition assessed at the single nucleotide resolution using 16S rRNA amplicon sequencing. **B**. Richness of microbial communities supported by single carbon sources. Bars indicate, for each carbon source, the number of ASVs (mean ± SEM, N = 3). **C**. A representative example of how observed richness in constituent single resources (mean ± SEM, N = 3) compares to the observed richness in two-resource communities (mean ± SEM, N = 3) and predictions calculated as the union (sum minus shared ASVs, dark violet) or the maximum (light violet) of the richness in constituent singles (mean ± SEM, N from permutations = 9). **D**. Observed average richness (orange dots, mean ± SEM, N = 16 for single-resource, 24 for two-resource, 12 for four-resource, six for eight-resource, 16 for 15-resource and 1 for 16-resource combinations as a linear function of the number of supplied carbon sources (solid orange line). Grey jittered dots indicate the average richness for each unique combination of resources (mean ± SEM, N = 3). In single resources, the blue and yellow dots correspond to the highest and lowest average richness, measured in cellulose and citrate, respectively. The predicted trajectory of richness based on the competitive exclusion principle (dashed dark green line), the union (dashed dark violet line) and maximum (dashed light violet line) estimates, as described for panel B, are shown for comparison.

Next, we examined the full range of resource combinations included in the experiment. We found that the richness predicted from the union of constituent singles significantly overestimated the observed richness (Fig. 1D). The prediction based on the maximum of constituent singles gave an increase with negative curvature that was not detected in our experiment (Fig. 1D). A similar trend originated also when we estimated the number of metabolites generated from resource combinations, with the same approach we used for single carbon sources (Fig. S9). Instead, the observed average richness increased linearly with the number of supplied carbon sources, at the constant rate of roughly one ASV for each new added resource (Fig. 1D, slope = 1.4). The linear trend was robust to the exclusion of low-abundance ASVs (Fig. S10A) and, when other indices to measure microbial diversity were used, the increase with the number of resources was still approximately linear (Fig. S10B, C). In addition, as more resources were provided, communities became more even (Materials and Methods and Fig. S11), without changes in biomass (Fig. S12). Despite confirming that the number of supplied resources is an important driver of microbial diversity, the observed one-by-one relation between richness and resource number was difficult to reconcile with the large diversity found in single resources. Thus, we went back to the single-resource communities to gain a better understanding of our observations.

A key feature that we did not consider while making predictions is that, in natural and experimental communities, bacterial taxa exhibit a spectrum of resource-utilization strategies (*22*, *23*), ranging from generalists, for which resources are largely substitutable, to specialists, showing specific resource requirements (*24*, *25*). Importantly, these features can alter the structure of the interactions among community members (*26*, *27*). In light of this, we measured the resource occupancy of the 275 ASVs observed in single resources, i.e. how many media a given ASV was found in, to find a range of different resource-utilization strategies (Fig. 2A, Fig. S13).

Based on resource occupancy (*23*, *26*), we considered specialists the ASVs that were observed in less than 25% of single-resource media (Fig. 2A), and generalists those that occupied the vast majority of single-resource media (more than 75% of them) (Fig. 2A). As a result of this distinction, the majority of ASVs (216 out of 275) were specialists, whereas few of them (10)

were generalists, reminiscent of natural communities (*23*, *28*). Some ASVs (49) displayed an intermediate occupancy, being present in between four to twelve media. Generalist ASVs belonged to the most abundant families, i.e. Pseudomonadaceae, Enterobacteriaceae and Micrococcaceae (Fig. S13), and exhibited higher rRNA copy numbers compared to specialists, e.g. taxa from Cellvibrionaceae (Fig. 2B, the median was 5 copies for generalists and 3 for specialists, calculated at the genus level and Fig. S14), indicative of faster max growth rates (*29*). This is consistent with studies showing that a hallmark of a generalist life style is fast growth usually underpinned by multiple copies of the rRNA operon (*30*, *31*). Notably, low-richness communities were composed of a greater proportion of generalist ASVs, which could constitute more than 50% of observed taxa (Fig. 2C, S15A), while more diverse communities, were enriched with specialists (Fig. 2C, S15B).

To understand how the ratio between resource-utilization strategies changed in the communities as we added more carbon sources, we calculated the average number of specialists, generalists and intermediates (as defined based on single resource occupancy) for each combination of resources. Going from one to 16 resources, communities went from containing a balanced mixture of generalists and specialists to being dominated by more specialized ASVs (both specialists and intermediates, Fig. 2D). Importantly, specialists maintained their resource specificity when the occupied single resource was provided in a combination, indicating a remarkable degree of determinism in community assembly, even on complex resource mixtures. This was revealed by the resource-specificity score—reflecting the "preference" for a particular resource exhibited by an ASV—being positive, on average, for specialist taxa across all resources (Fig. 2E, mean score for generalists was 0.02 ± 0.01, for specialists 0.24 ± 0.05, see Materials and methods). Overall, these findings indicate that the increase in richness with the number of resources was driven by more specialized ASVs and that resource identity played an important role in community assembly.

It is important to note that, in the communities supported by combinations of resources, several specialist ASVs were lost (Fig. 2D, S16) and few new ASVs were introduced (grey bars in Fig. 2D). These observations suggest the existence of strong competitive interactions within specialist taxa and might be interpreted as a signature that niche overlap is modular, i.e. happening among taxa with similar habitat/resource utilization strategies (*13*, *19*, *32*). To verify this intuition, we sought a modelling framework in which both a weak linear relation between diversity and resource number and large coexistence in single resources could emerge just from modularity in the structure of bacterial competitive interactions.

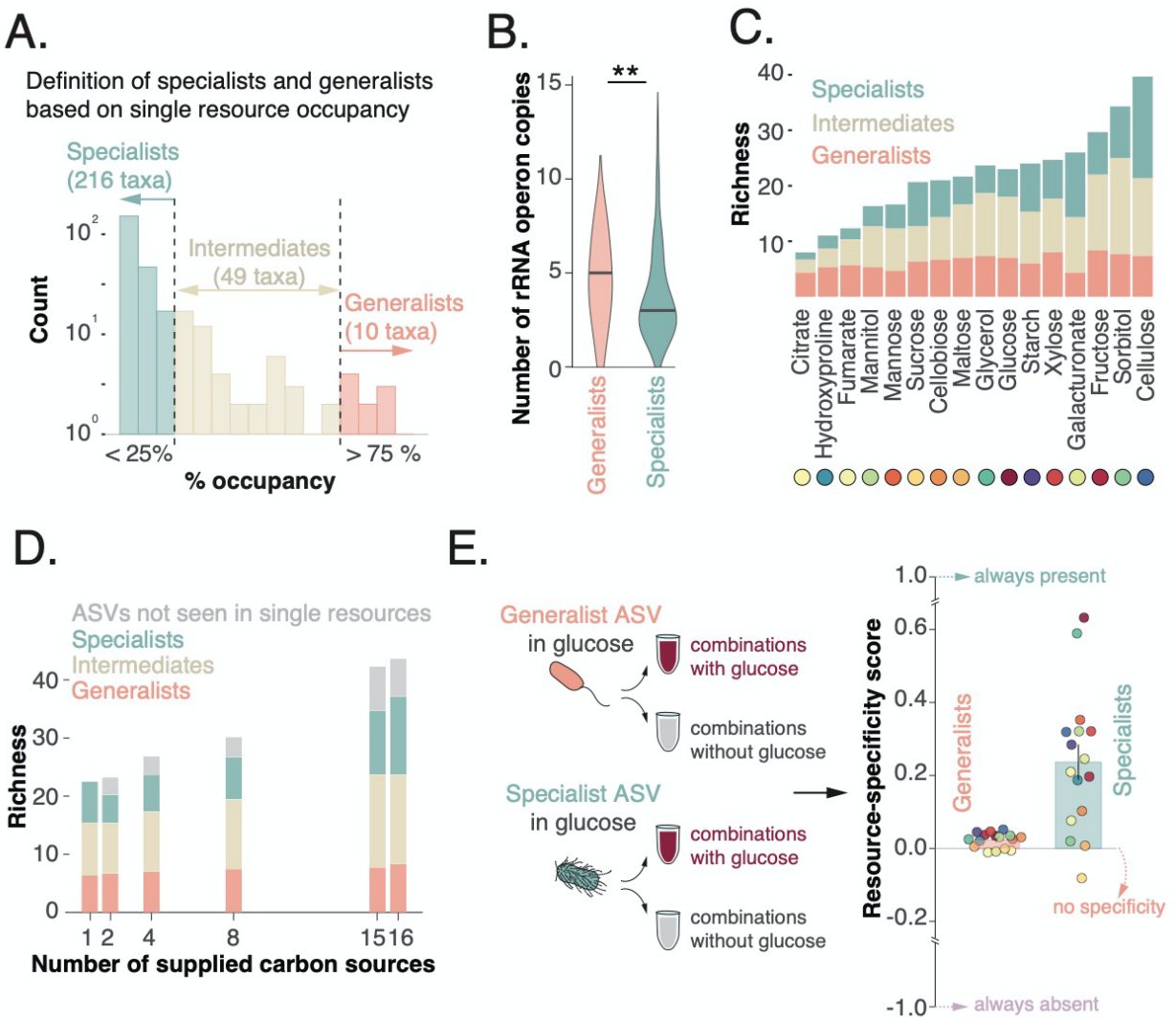

Fig. 2. As more resources are provided, the number of specialist taxa increases while the number of generalist taxa stays constant. A. The 275 ASVs found across all single-resource communities were classified in generalist, specialists and intermediates depending on their resource occupancy. As a result of this distinction, the majority of ASVs exhibited a more specialized resource-utilization strategy. **B**. The distribution of rRNA operon copy numbers, calculated at the genus level, of generalist ASVs differed from that of specialist ASVs ($p < 0.01$, from Kolmogorov-Smirnov test). **C**. Going from one to 16 resources, the number of more specialized taxa increases compared to the number of generalist taxa. The mean number of generalist (pink), intermediate (beige) and specialist (teal) ASVs for media with the same number of resources is shown as stacked bars. The average number of ASVs that were not detected in single-resource communities but appeared in other combinations is indicated in grey. Error bars are omitted for clarity. **D.** The specificity score is calculated, for each ASV found in a single resource (target resource), using the number of multi-resource media containing the target resource in which the ASV was found (X) and the number of media not containing the target resource in which the ASV is found (Y), as (X – Y)/(X + Y). It ranges from 1, indicating that the ASV is always present in a combination containing the resource, to -1, implying that the ASV is always absent when the resource is supplied. A score of 0 is indicative of an ASV showing no specificity for that particular resource. Bars indicate the mean specificity score ± SEM for generalists (pink) and specialists (teal) (N = 16). Colored dots indicate the mean score for each resource (SEM are omitted for clarity, N varies for each resource, see Materials and methods).

We implemented a version of the Lotka-Volterra model modified to account for resource specialists and generalists (*33*). Briefly, we let the size of the species pool increase with the number of resources under the constraint that specialists could grow only when the resource they specialize on was present, while generalists could always grow (Fig. 3A, B). We tested different interaction structures (Materials and methods, Table S2, Fig. S17) to find that the one where competition between generalists and specialists was weak, but within generalists and specialists was strong (Fig. 3B) was sufficient to reproduce our experimental results. In our simulations, we could observe the stable coexistence of several species in single resources and an almost one-by-one relation between richness and number of resources (Fig. 3C), mostly driven by specialists (Fig. 3D). We concluded that the existence of a modular interaction structure linked to resource-utilization strategies of community members could underlie the observed relation between microbial community diversity and resource number.

Compared to resource-explicit models, phenomenological models like Lotka-Volterra require fewer assumptions and parameters, making it easier to identify the most relevant conditions underlying our experimental observations (*34*). Since others have suggested that resource-explicit models could be mapped to Lotka-Volterra models (*6*), these conditions might eventually lead to equivalent resource-explicit models. However, while our minimal model is sufficient to reproduce the observed patterns, it contains some simplifying assumptions. First, it treated all species as either true specialists or generalists: there are not species with intermediate resource specialization. In addition, it only considered resource presence or absence and ignored changes in the relative concentration as more resources were added. This could have a significant impact on microbial assembly (*12*) and potentially explain the loss of specialistic taxa in two-resource media (Fig. 2D, S6B, S16). In support of this, we observed that the relative abundance of several specialist taxa, coarse-grained at the family level, decreased drastically or went to zero when the relative concentration of the "favorite resource" dropped by half (see Materials and methods and Fig. S18).

Overall, our results provide experimental and theoretical support that both the number and the identity of available resources are important drivers of microbial community diversity, ultimately showing that community assembly is largely driven by deterministic factors linked to resource availability rather than stochastic processes (*35*). In addition, they show that the combined effect of resource identity and number is better understood when resource-utilization strategies of community members are taken into account. Given that the coexistence of specialist and generalist bacterial taxa is a ubiquitous feature of natural microbiomes, our results are likely generalizable to a variety of communities strongly depending on externally supplied resources, including the gut microbiome and degraders communities in the ocean. Moreover, our results provide a framework to anticipate how community diversity might change in response to modifications in the resource environment, a crucial step in light of the essential role bacteria play in the functioning of the biosphere.

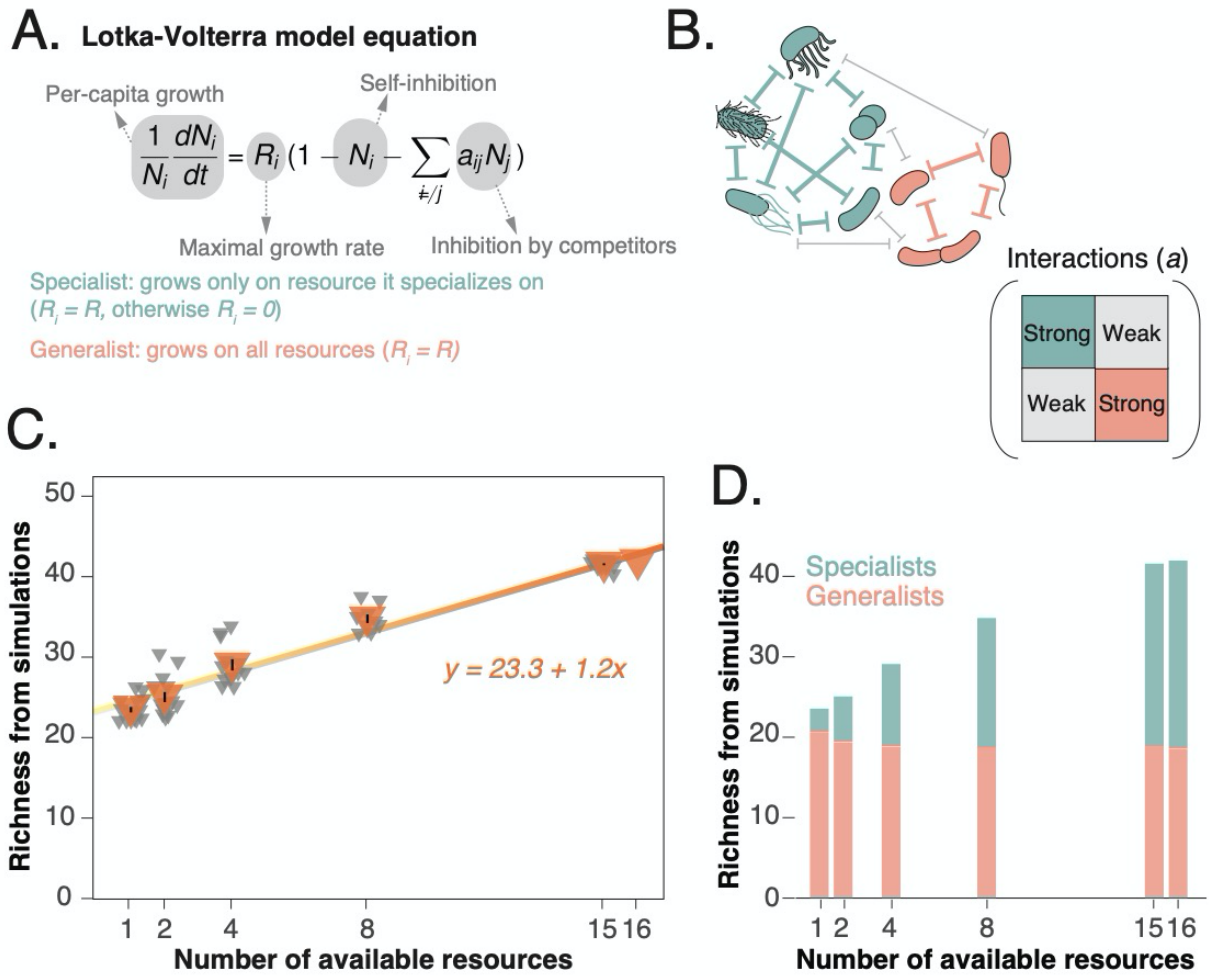

Fig. 3. **A simple Lotka-Volterra model modified to incorporate specialists and generalists recapitulates experimental results. A.** The classical Lotka-Volterra competition equation is shown. In our version, specialists have nonzero growth rates only when the resource they specialize on is available, while generalists always have nonzero growth rates. **B.** Schematic of the modular interaction structure implemented in the model. Specialist ASVs (teal bugs) compete strongly against each other, but weakly against generalist ASVs (pink bugs). Competitive interactions withing generalists are also strong. **C**. Richness obtained from simulations of the modified Lotka-Volterra model (orange triangles, mean $\pm$ SEM, N ranges from 16 to 1) as a linear function of the number of supplied carbon sources (solid orange line). Grey jittered triangles indicate the richness of communities grown on a particular resource combination (mean, 50 random initial conditions and parameters, SEM are not indicated for clarity) **D**. The average number of specialists (teal bars) increases as more resources are available while the average number of generalists (pink bars) remains the same. Error bars are omitted for clarity.

228 **Materials and Methods**

229 Growth media preparation

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

231 All bacterial cultures have been grown in M9 media (prepared from 5X M9 salts, 1X Trace
232 Metal Mixture (Teknova) and 1M stock solutions of $MgSO_4$ and $CaCl_2$) supplemented with 0.1
233 % w/v of one of 75 carbon source combinations. These combinations include: 16 compounds
234 commonly available in soil that were provided as single carbon sources (D-(+)-glucose, D-(–)-
235 fructose, D-(+)-xylose, D-(+)-mannose, D-(+)-cellobiose, D-(+)-maltose monohydrate, sucrose,
236 citric acid, fumaric acid, D-(+)-galacturonic acid monohydrate, D-mannitol, D-sorbitol, glycerol,
237 trans-4-Hydroxy-D-proline, methyl cellulose, starch); 24 random combinations of two of these
238 resources; 12 random combinations of four resources; 6 random combinations of eight resources;
239 the 16 combinations containing 15 resources; and all the 16 resources together (see Table S1 for
240 the complete list and Fig. S2). The total concentration of carbon was kept the same and resources
241 were in all instances supplied in equal amounts, that was 100%, 50%, 25%, 12.5%, 6.7% and
242 6.25% each for single-, two-, four-, eight-, 15- and 16-resource combinations. All solutions were
243 filter-sterilized with a 0.22 μm filter and kept at 4°C for the duration of the experiment.

244 Collection of microbial communities from the environment

245 The soil from which the initial inoculum comes from has been sampled from a lawn in
246 Cambridge, Massachusetts, at a depth of ~15 cm using a sterile corer and tweezers. Once in the
247 lab, a total of 1.5 g of the collected soil was diluted in 20 mL phosphate buffered saline (PBS;
248 Corning), then vortex at intermediate speed for 30 s and incubated on a platform shaker (Innova
249 2000; Eppendorf) at 250 r.p.m. at room temperature. After 1 hour, the sample was allowed to
250 settle for ~5min and the supernatant was filtered with a 100 μm cell strainer (Thermo Fisher
251 Scientific) and then directly used for inoculation. Both the original soil sample and the remaining
252 supernatant were stored at -80 °C for subsequent DNA extraction.

253 Experimental microcosms

254 Aliquots (7μL) of the supernatant containing the soil microbial suspension were inoculated into
255 203 μL of growth media in 96-deepwell plates (Deepwell plate 96/500 μL; Eppendorf), for a
256 total of 231 microcosms (3 replicates for each different resource combinations, except 16-
257 resource combinations that were replicated 9 times). Deepwell plates were covered with
258 AeraSeal adhesive sealing films (Excel Scientific). Bacterial cultures were grown at 30°C under
259 constant shaking at 1,350 r.p.m. (on Titramax shakers; Heidolph). To avoid evaporation, they
260 were incubated inside custom-built acrylic boxes.

261 Every 24 h, the cultures were thoroughly mixed by pipetting up and down 3 times using the
262 VIAFLO 96-well pipettor (Viaflo 96, Integra Biosciences; settings: pipette/mix program
263 aspirating 7 μL, mixing volume 10 μL, speed 6) and then diluted 1/30x into fresh media. We
264 applied a total of seven daily dilution cycles. At the end of every cultivation day we measured
265 the optical density (OD_{600}) using a Varioskan Flash (Thermo Fisher Scientific) plate reader. The
266 remaining bacterial culture was frozen at -80 °C for subsequent DNA extraction.

267 DNA extraction, 16S rRNA sequencing and analysis pipeline

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

We used the R package DADA2 to obtain the amplicon sequence variances (ASVs)(*36*) following the workflow described in Callahan et al. (*37*). Taxonomic identities were assigned to ASVs using the SILVA version 132 database (*38*). The phylogenetic tree (Fig. S4) was reconstructed using Randomized Axelerated Maximum Likelihood (RAxML) using default parameters (*39*).

Data analysis

Analysis, unless otherwise stated have been conducted in R, version 3.6.1 (*40*).

Sequencing data has been handled using the R package phyloseq (*41*). We obtained an average of 20,613 reads per sample. Sequencing depth did not affect our estimation of community diversity indexes (Fig. S4). Richness was calculated as the number of ASVs with abundance larger than 0 found in each sample. Community diversity was also measured by Shannon Diversity index and Shannon Entropy index following (*42*, *43*) (Fig. S10). The significance of differences in richness due to single supplied resources was tested through ANOVA(*44*) using the package GAD.

Richness predictions

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

Rank abundance distributions

First, we computed abundance distributions (RADs) for each sample, i.e. each replicate community grown on a unique combination of carbon sources, by sorting ASVs based on their relative abundance. Then, we plotted the RADs in a log-linear fashion and fitted a regression line in order to compare their slopes (Fig. S11A). The absolute value of the slope of the fitted regression line informs on the abundance distribution of the ASVs in a community. More even communities usually display smaller slopes (Fig. S11B). Since each community exhibited a different richness, we normalized the RADS for richness (Fig. S11C) To do this, we used the *RADnormalization_matrix* function in the RADanalysis package: from each RAD with an observed richness, this function generates a “normalized RAD” with a richness corresponding to

310 the minimum richness observed in the experiment (7 ASVs) by randomly resampling the original
311 RADs for 10 times (*45*) . In this way, samples with different richness can be compared and
312 changes in evenness properly assessed.

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

314 ASVs found in single resources were classified in three categories based on how many media
315 containing a single resource they were found in, i.e. they exhibited abundance larger than 0 (*23*,
316 *26*). We considered specialists the ASVs that were observed in less than 25% of single-resource
317 media, i.e. in one, two or three resources. Generalists were those ASVs found in more than the
318 75% of media, i.e. in 13 or more resources. We defined intermediates the ASVs found between
319 four and twelve resources. These thresholds were chosen arbitrarily, but the resulted in about ~
320 4% generalists and 80% specialists, consistently with proportions of generalists and specialists
321 observed in natural communities (*23*, *26*, *28*). We chose this simple way of assigning ASVs to
322 generalist, intermediate and specialist categories over other methods, e.g. as in (*9*) in order to
323 leave aside their relative abundance, which has been analyzed separately.

324 *Prediction of possible metabolic byproducts in resource environments*

325 We predicted the possible number of metabolic byproducts that could be produced using the
326 resources present in each medium using a curated metabolic network. The metabolic network
327 contained a large set of metabolic reactions encompassing carbohydrate, sugar and amino acid
328 metabolism extracted from the KEGG database (*46*). We manually curated this large set of
329 reactions using the MetaCyc database (*47*) in order to limit it to reactions possible by most
330 microbial taxa common to the soil, such as *Pseudomonas*. We used this network to estimate all
331 the metabolic compounds that could be produced as byproducts, starting from the carbon sources
332 available in each medium. We assumed that a small set of "currency" molecules, such as water,
333 carbon dioxide and ATP, were always available as reactants when required (full list of currency
334 molecules: phosphate, oxygen, carbon dioxide, water, H^+, ATP, NAD(P)H, Acetyl-CoA, CoA).

335 To estimate the possible byproducts in each medium, we employed the well-known scope
336 expansion algorithm (*48–52*). Each reaction in our curated metabolic network consisted of a set
337 of reactants and resulting products. For each medium, we first asked which reactions could be
338 performed using only the carbon sources available in the medium (i.e., the current "scope" of the
339 medium). We assumed that the products of these reactions could be produced and added them to
340 the set of reactants – the new scope – for the next step. In the next step, we again asked which
341 reactions could be performed using the new scope. We added their products to the scope for the
342 next step. We continued this process, step by step, until we could add no new products to the
343 scope. The resulting final scope of metabolites, minus the currency molecules provided in the
344 medium, was our estimated set of possible metabolic byproducts producible in that medium.

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

346 To test for signatures of different life-history strategies of the generalist and specialist taxa in our
347 study, we estimated their 16S rRNA operon copy numbers. We estimated rRNA copy numbers at
348 the level of both genus and family, separately for generalist and specialist taxa. For each genus
349 identified, we queried rrnDB (*53*)—a database of rRNA operon copy number statistics—for the
350 median copy number corresponding to the genus. We used this as an estimate for the rRNA

351 operon copy number of that genus. For each family identified, we calculated the median copy
352 number of all genera in our dataset belonging to that family.

353 *Calculation of the resource-specificity score*

354 We used a resource-specificity score to test if the ASV-resource associations that we observed in
355 single resources were maintained when the single resource(s) in which the ASV was found was
356 combined with others. For each ASV present in a single resource (target resource), the resource
357 specificity score is calculated as the difference between the number of multi-resource media
358 containing the target resource in which the ASV is found and the number of media not
359 containing the target resource in which the ASV is found divided by the total number of media in
360 which the ASV is found (Fig. 2E). This is reminiscent of a preference index, which is a standard
361 measure in the behavioral sciences. Single resources are excluded from the count. The resource-
362 specificity score ranges from 1, indicating that the ASV is always present when the target
363 resource is provided, to -1, implying that the ASV is always absent when that resource is
364 supplied. A score of 0 is indicative of an ASV showing no specificity for that particular resource
365 (Fig. 2E). We calculated a score for each ASV-resource pair, such that each ASV had as many
366 scores as the number of single resources is found in. Then, we computed the average of the
367 scores obtained for each single resource, separating between scores belonging to generalist and
368 specialist ASVs (Fig. 2E).

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

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

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

380 We employed XGBoost, a gradient boosting framework based on decision trees (*54*).
381 Specifically, we implemented a regression model for each family in which the input was the
382 relative resource concentration and the output was the log-transformed relative family
383 abundance. We trained the model on two replicates by performing leave-one-out crossvalidation
384 of the XGBoost parameters "max_depth", "n_estimators" and "learning_rate" (*55*) and tested on
385 the third one with average mean-squared error across families of 6.05. We applied the Shapley
386 Additive exPlanations (SHAP) (*56*) to identify the resources that were more important in driving
387 changes in the abundance of each family. This analysis has been done using Python version 3.8.

388 Results of this analysis revealed that variations in the abundance all of the 37 families were
389 driven by one or multiple resources based on their dominant life strategy. To simplify the
390 visualization of the results we plotted the relative abundance of some representative families as a
391 function of the concentration of the resources identified by the analysis (Fig. S18). Families
392 mostly composed of specialist taxa, e.g. Cellvibrionaceae and Bacillaceae, showed abrupt

changes in their abundance with the concentration of the "favorite" resource (Fig. S18). By contrast, more generalist families, e.g. Pseudomonadaceae and Enterobacteriaceae, exhibited smooth trends in their abundance with the concentration of multiple resources.

Lotka-Volterra model implementation and simulations

In the main text, we discussed the observed linear trend of richness with respect to the number of supplied resources. Here we discuss in detail how we reproduced this linear trend with a simple model consisting of resource generalists and specialists.

We hypothesized that modularity in the interaction structure could explain the observed linear trend between richness and number of carbon sources. With modularity we mean that taxa displaying similar resource strategies interact strongly against each other and weakly with taxa showing a different resource strategy. To test this hypothesis, we decided to use a phenomenological model that characterizes only the essential variables of our system identified through the experiments. While more explicit formulations of resource dynamics could be a natural way to model the observed trend, any resource-explicit model which can capture the high diversity under single supplied resource requires cross-feeding and different resource consumption strategies. This increases a lot the number of parameters compared to Lotka-Volterra models (*33*), making more difficult the identification of the sufficient condition able to produce the linear increase in richness with the number of resources. Rather, starting from a minimal phenomenological model, we were able to identify a feasible explanation of our experimental results, which also hints on how some classes of resource explicit models might also be able to do the same.

We used a modified version of the Lotka-Volterra model:

$$\frac{1}{n_i}\frac{dn_i}{dt} = R_i\left(1 - n_i - \sum_{j\neq i} a_{ij}n_j\right)$$

If specialist: $R_i = R$ if specialized resource is provided, 0 otherwise
If generalist: $R_i = R$

n_i is the population size of the *i*-th species. R_i is the per capita growth rate of the *i*-th species. The interaction coefficient a_{ij} indicates the strength of the competitive inhibition on the *i*-th species by the *j*-th species. Each population n_i is normalized by the carrying capacity of each species. The interaction coefficient a_{ij} is normalized as well, with $a_{ij} = 1$ for the interspecific competition between the *i*-th and the *j*-th species and the same strength as the intraspecific competition within the *i*-th species.

We assumed that a specialist species grows only when the resource it specializes on is provided either as the sole supplied resource in the media or in combination with other resources. In contrast, a generalist species grows on any resource. Although in the experiment we observed ASVs that displayed an intermediate occupancy of single resources, for simplicity, we modelled only two representative versions of specialists and generalists. We also assumed that the per capita growth rate R in the model is the same for any species in any environment where it can grow. Another assumption was the absence of dilution in the system, despite the fact that we

applied a daily dilution protocol in our experiments. However, we have found that the addition of a dilution rate in the model does not qualitatively affect the results (results not shown). Finally, we assumed that all interactions are antagonistic and that positive interactions, such as cross-feeding, do not enable specialists to grow when the supplied resources do not include the one that they can grow on.

We implemented a modular interaction structure by differentiating among *intra-group* pairwise interactions, i.e. within generalists and within specialists, and *inter-group* pairwise interactions, i.e. between generalists and specialists. We tested the four interaction structures obtained by keeping the strength of inter-group competitive interactions constant and varying the relative strength of intra-group competitive interactions (Table S2). The interaction structure that we expected to reproduce the linear one-by-one increase in richness with the number of supplied resources is the one in which competition within generalists and within specialists are both stronger than the competition between generalists and specialists.

We found that a model with the hypothesized modular interaction structure recapitulated our experimental results, yielding a one-by-one linear increase of richness with the number of supplied resources and a constant generalist population across different combinations of resources (Fig. 3C, D, Fig. S17A). In contrast, when competitive interactions within specialists were stronger than the inter-group interactions, while competitive interactions within generalists were weaker than the inter-group interactions, generalists always took over the community and the richness remained constant as more resources are provided (Fig. S17B). When the opposite was true—competitive interactions within specialists were weaker than the inter-group interactions, while competitive interactions within generalists were stronger than the inter-group interactions—specialists took over as generalists went extinct when many resources were available. This resulted in a U-shaped richness trend with the number of available resources (Fig. S17C). Finally, when both intra-group competitive interactions were weaker than intergroup competition, generalists outcompeted specialists when few resources were available and, in turn, specialists took over when many resources were available (Fig. S17D).

The inter-group interaction coefficients were drawn from uniform distributions with range (0.2, 0.5). The intra-group interaction coefficients, instead, were drawn from uniform distributions with range (0.1, 0.4) in case of weaker competition within groups than between groups and range (0.3, 0.6) in the case of stronger competition within groups than between groups (Table S2). The observed linear trend with weak competition between generalists and specialists and strong competition within generalists and within specialists is robust to the choice of parameter values. The implemented ones were those quantitatively reproducing the richness values observed in the experiment.

We simulated communities with 32 generalists and 48 specialists. The ratio between generalists and specialists is different from what we observed in the experimental data (10 generalists, 219 specialists), but this parameter choice, which quantitatively reproduced the observed richness values while qualitatively unaffecting results, reflected our choice of not implementing intermediates. The simulations were run for 2e4 unit time with R=1 starting from initial population drawn from uniform distribution of (1e-6, 1e-7), and communities reached equilibrium at the end of the simulations. Simulation were run in Python version 3.7.4.

In conclusion, our model suggests that weak inter-group interactions maintain the coexistence of generalists and specialists in the communities. This is reminiscent of the conditions required for

the coexistence of two species in the Lotka-Volterra model. Namely, when interspecific competition between two species is weaker than intraspecific competition within each species, the two species coexist. In contrast, when interspecific competition is stronger than intraspecific interactions, the stronger competitor cannot be invaded. This coexistence underpins a linear trend in richness, consisting of a constant number of generalist species and a linear increase driven by the specialist population.

References and Notes:

1. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive earth's biogeochemical cycles. *Science (80-.).* **320** (2008), pp. 1034–1039.
2. S. J. Biller, P. M. Berube, D. Lindell, S. W. Chisholm, Prochlorococcus: The structure and function of collective diversity. *Nat. Rev. Microbiol.* **13** (2015), pp. 13–27.
3. K. Zengler, L. S. Zaramela, The social network of microorganisms-how auxotrophies shape complex communities HHS Public Access. *Nat Rev Microbiol*. **16**, 383–390 (2018).
4. S. Allesina, J. M. Levine, A competitive network theory of species diversity. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5638–5642 (2011).
5. A. R. Pacheco, M. Moel, D. Segrè, Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* **10**, 1–12 (2019).
6. M. Barbier, J. F. Arnoldi, G. Bunin, M. Loreau, Generic assembly patterns in complex ecological communities. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2156–2161 (2018).
7. W. Cui, R. Marsland, P. Mehta, Effect of Resource Dynamics on Species Packing in Diverse Ecosystems. *Phys. Rev. Lett.* **125**, 048101 (2020).
8. J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta, A. Sanchez, Emergent simplicity in microbial community assembly. *Science (80-.).* **361**, 469–474 (2018).
9. T. N. Enke, M. S. Datta, J. Schwartzman, N. Cermak, D. Schmitz, J. Barrere, A. Pascual-García, O. X. Cordero, Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Curr. Biol.* (2019), doi:10.1016/j.cub.2019.03.047.
10. D. Naylor, S. Fansler, C. Brislawn, W. C. Nelson, K. S. Hofmockel, J. K. Jansson, R. McClure, Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules. *MBio*. **11** (2020), doi:10.1128/mBio.01349-20.
11. R. Cavicchioli, W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, T. W. Crowther, R. Danovaro, C. M. Foreman, J. Huisman, D. A. Hutchins, J. K. Jansson, D. M. Karl, B. Koskella, D. B. Mark Welch, J. B. H. Martiny, M. A. Moran, V. J. Orphan, D. S. Reay, J. V. Remais, V. I. Rich, B. K. Singh, L. Y. Stein, F. J. Stewart, M. B. Sullivan, M. J. H. van Oppen, S. C. Weaver, E. A. Webb, N. S. Webster, Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17** (2019), pp. 569–586.
12. C. Ratzke, J. Barrere, J. Gore, Strength of species interactions determines biodiversity and stability in microbial communities, doi:10.1101/671008.
13. D. Naylor, S. Fansler, C. Brislawn, W. C. Nelson, K. S. Hofmockel, J. K. Jansson, R. McClure, Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules. *MBio*. **11** (2020), doi:10.1128/mBio.01349-20.
14. G. F. Gause, *The struggle for existence* (Hafner Press, New York, 1934).
15. R. H. MacArthur, *Geographical Ecology. Patterns in the Distribution of Species* (Harper & Row, 1972).
16. T. Wang, A. Goyal, V. Dubinkina, S. Maslov, Evidence for a multi-level trophic organization of the human gut microbiome. *PLOS Comput. Biol.* **15**, e1007524 (2019).
17. A. Goyal, S. Maslov, Diversity, Stability, and Reproducibility in Stochastically Assembled Microbial Ecosystems. *Phys. Rev. Lett.* **120**, 158102 (2018).
18. R. Marsland, W. Cui, J. Goldford, A. Sanchez, K. Korolev, P. Mehta, Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Comput. Biol.* **15** (2019), doi:10.1371/journal.pcbi.1006793.
19. H. Fu, M. Uchimiya, J. Gore, M. A. Moran, Ecological drivers of bacterial community assembly in synthetic

phycospheres. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3656–3662 (2020).
20. J. M. Levine, J. HilleRisLambers, The importance of niches for the maintenance of species diversity. *Nature*. **461**, 254–257 (2009).
21. N. Tromas, Z. E. Taranu, B. D. Martin, A. Willis, N. Fortin, C. W. Greer, B. J. Shapiro, Niche Separation Increases With Genetic Distance Among Bloom-Forming Cyanobacteria. *Front. Microbiol.* **9**, 438 (2018).
22. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A phylogenetic perspective. *Science (80-.).* **350**, aac9323–aac9323 (2015).
23. S. Sriswasdi, C. C. Yang, W. Iwasaki, Generalist species drive microbial dispersion and evolution. *Nat. Commun.* **8**, 1–8 (2017).
24. N. Fierer, M. A. Bradford, R. B. Jackson, Toward an ecological classification of soil bacteria. *Ecology*. **88**, 1354–1364 (2007).
25. J. T. Lennon, Z. T. Aanderud, B. K. Lehmkuhl, D. R. Schoolmaster, Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology*. **93**, 1867–1879 (2012).
26. M. E. Muscarella, C. M. Boot, C. D. Broeckling, J. T. Lennon, Resource heterogeneity structures aquatic bacterial communities. *ISME J.* **13**, 2183–2195 (2019).
27. A. J. Székely, S. Langenheder, The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiol. Ecol.* **87**, 102–112 (2014).
28. R. Logares, E. S. Lindström, S. Langenheder, J. B. Logue, H. Paterson, J. Laybourn-Parry, K. Rengefors, L. Tranvik, S. Bertilsson, Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J.* **7**, 937–948 (2013).
29. B. R. K. Roller, S. F. Stoddard, T. M. Schmidt, Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 1–7 (2016).
30. K. C. Goldfarb, U. Karaoz, C. A. Hanson, C. A. Santee, M. A. Bradford, K. K. Treseder, M. D. Wallenstein, E. L. Brodie, Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical Recalcitrance. *Front. Microbiol.* **2**, 94 (2011).
31. J. A. Klappenbach, J. M. Dunbar, T. M. Schmidt, rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**, 1328–1333 (2000).
32. J. Friedman, L. M. Higgins, J. Gore, Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 0109 (2017).
33. A. Hastings, in *Population Biology* (Springer New York, New York, NY, 1997; http://link.springer.com/10.1007/978-1-4757-2731-9_7), pp. 129–149.
34. C. I. Abreu, J. Friedman, V. L. Andersen Woltz, J. Gore, Mortality causes universal changes in microbial community composition. *Nat. Commun.* **10**, 1–9 (2019).
35. S. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).
36. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. **13**, 581–583 (2016).
37. B. J. Callahan, K. Sankaran, J. A. Fukuyama, P. J. McMurdie, S. P. Holmes, Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research*. **5**, 1492 (2016).
38. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
39. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
40. R. Core, Computing, R: A Language and Environment for Statistical Computing (2019), (available at https://www.r-project.org/).
41. P. J. McMurdie, S. Holmes, phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*. **8**, e61217 (2013).
42. A. Chao, N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, A. M. Ellison, Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
43. A. Chao, C. H. Chiu, L. Jost, Phylogenetic diversity measures based on Hill numbers. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 3599–3609 (2010).
44. A. J. Underwood, *Experiments in Ecology* (Cambridge University Press, 1997; /core/books/experiments-in-ecology/DCF3663D5E7C9923D19B5ECE88167780).
45. M. Saeedghalati, F. Farahpour, B. Budeus, A. Lange, A. M. Westendorf, M. Seifert, R. Küppers, D.

Hoffmann, Quantitative Comparison of Abundance Structures of Generalized Communities: From B-Cell Receptor Repertoires to Microbiomes. *PLoS Comput. Biol.* **13**, e1005362 (2017).

46. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28** (2000), pp. 27–30.
47. R. Caspi, R. Billington, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, P. E. Midford, Q. Ong, W. K. Ong, S. Paley, P. Subhraveti, P. D. Karp, The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633–D639 (2018).
48. J. E. Goldford, H. Hartman, T. F. Smith, D. Segrè, Remnants of an Ancient Metabolism without Phosphate. *Cell*. **168**, 1126-1134.e9 (2017).
49. J. Raymond, D. Segrè, The effect of oxygen on biochemical networks and the evolution of complex life. *Science (80-.).* **311**, 1764–1767 (2006).
50. T. Handorf, O. E. Ebenho¨h, R. Heinrich, Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution, doi:10.1007/s00239-005-0027-1.
51. O. Ebenhoh, T. Handorf, R. Heinrich, "Structural Analysis of Expanding Metabolic Networks" (2004).
52. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28** (2010), pp. 245–248.
53. S. F. Stoddard, B. J. Smith, R. Hein, B. R. K. Roller, T. M. Schmidt, rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **43**, D593–D598 (2015).
54. T. Chen, C. Guestrin, in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, NY, USA, 2016; https://dl.acm.org/doi/10.1145/2939672.2939785), vols. 13-17-August-2016, pp. 785–794.
55. J. Hastie, Trevor, Tibshirani, Robert, Friedman, *The Elements of Statistical Learning The Elements of Statistical LearningData Mining, Inference, and Prediction, Second Edition* (2009).
56. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).

Acknowledgements: The authors would like to thank the members of Gore Lab for comments on the manuscript and Bartolomeo Stellato for the help with the ensemble tree regression model. **Funding:** This work was supported by NIH and the Schmidt Foundation. A.G. is supported by the Gordon and Betty Moore Foundation as a Physics of Living Systems Fellow through grant number GBMF4513. **Authors contributions:** MDB and JG conceived the study. MDB performed the experiments and the sequencing analysis. HL performed theoretical modeling. AG performed the scope expansion analysis. All authors analyzed the data and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All materials, data and codes are available from the authors upon request, and will be made publicly available upon publication.

Supplementary Materials

Fig. S1-S18

Table S1-S2

References 35-56