



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

Martina Dal Bello  , Hyunseok Lee , Akshit Goyal and Jeff Gore  

**The relationship between the number of available nutrients and community diversity is a central question in ecological research that remains unanswered. Here 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 that, while single resources supported multispecies communities varying from 8 to 40 taxa, mean community richness increased only one-by-one with additional resources. Cross-feeding could reconcile these seemingly contrasting observations, with the metabolic network seeded by the supplied resources explaining the changes in richness due to both the identity and the number of resources, as well as the distribution of taxa across different communities. By using a consumer–resource model incorporating the inferred cross-feeding network, we provide further theoretical support to our observations and a framework to link the type and number of environmental resources to microbial community diversity.**

Uncovering the determinants of community diversity is central in ecology<sup>1–3</sup> and microbiome research<sup>4</sup>, posing unique challenges to microbial ecologists. Indeed, microbes are the most abundant form of life on our planet<sup>5</sup>, as well as the most ancient and most phylogenetically diverse<sup>6</sup>. Surveys of a variety of ecosystems, from oceans<sup>7</sup> to the human body<sup>8</sup>, have revealed that thousands of different taxa can stably coexist within the same community. Importantly, microbial communities drive the bulk of global nutrient cycling<sup>9</sup>, sustain human health<sup>10</sup> and modulate the response of the biosphere to climate change<sup>11</sup>. Hence, deepening our knowledge of the drivers of microbial community diversity is pivotal to understanding the functioning of Earth's ecosystems.

Several mechanisms contribute to the diversity of microbial communities, including the spatial and temporal structure of the environment<sup>12</sup>, dispersal and bacterial motility<sup>13</sup>, warfare<sup>14,15</sup>, and resource-mediated competition and cooperation<sup>16–18</sup>. With respect to resources, ecological theory has mostly focused on the effect of the number of available resources, rather than their identity, on community diversity<sup>19</sup>. In particular, according to the principle of competitive exclusion, the number of stably coexisting species is predicted to be bounded by the number of available resources<sup>20–22</sup>. Despite the wealth of theoretical work on how resources can affect microbial community diversity, empirical tests of resource–diversity relationships have been limited, having been explored either in two to three species assemblages<sup>17</sup> or in enriched cultures grown on one to two resources<sup>18,23–25</sup>. Systematic experiments encompassing a range of resource combinations are still lacking.

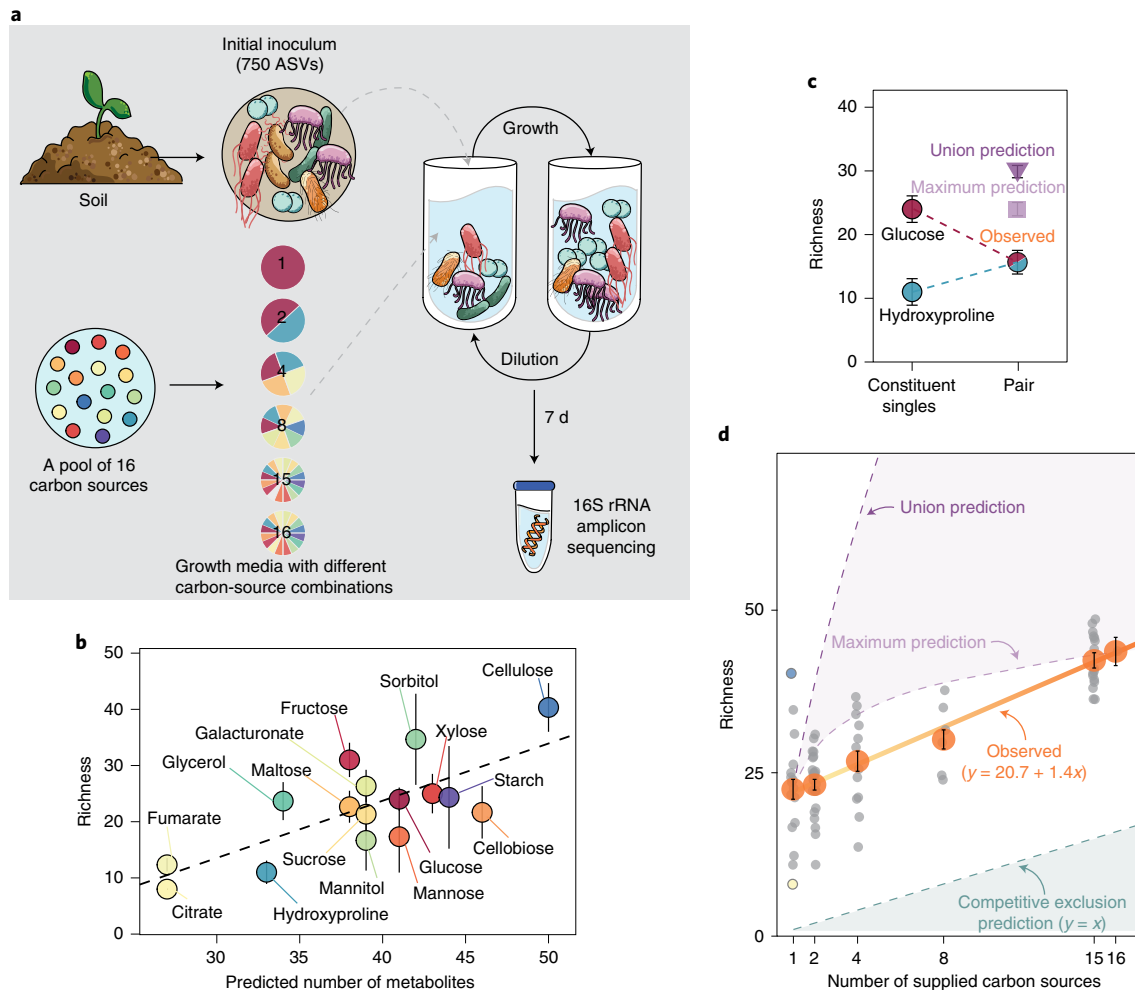
While an empirical test of the relationship between the number of available nutrients and community diversity remains elusive, bottom-up experiments have implicated cross-feeding as a major factor influencing the assembly of microbial communities, even in simple environments. Cross-feeding, whereby metabolic by-products of one taxa become resources for others<sup>26</sup>, can increase niche partitioning, ultimately allowing the coexistence of several taxa even when only a single source of carbon is provided<sup>18,23,25,27</sup>.

There is also some evidence that the identity of the supplied resource dictates community composition, as microbial taxa display different resource preferences and patterns of metabolite excretion<sup>23,28</sup>. Nevertheless, the manner in which cross-feeding and niche partitioning systematically change with the identity and the number of supplied resources is still unclear. This lack of knowledge impairs our ability to link variations in resource availability with shifts in microbial community diversity.

## Results

To illuminate how the availability of resources, namely their number and identity, shape the richness of microbiomes, we assayed the assembly of soil-derived bacterial communities in laboratory microcosms<sup>16,18</sup>. We started by inoculating a rich microbial suspension obtained from a soil sample (Supplementary Fig. 1) into 75 resource environments, each containing minimal media supplemented with a different combination of carbon sources, ranging from single compounds to a mix of 16 compounds (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). The 16 carbon sources represented a broad range of common soil compounds (for example, mannose, xylose, cellulose and hydroxyproline), encompassing both glycolytic (for example, simple and complex sugars) and gluconeogenic substrata (for example, organic acids). We adopted a daily dilution protocol, whereby at the end of each 24 h growth cycle, the bacterial cultures were diluted 1/30× into fresh media. We observed that the majority of microcosms reached stability 3 d after inoculation (Extended Data Fig. 1). We continued the experiment until day 7 and measured the final richness as the number of amplicon sequence variants (ASVs) observed within each community (Supplementary Fig. 3).

**Single resources support multispecies communities.** Consistent with recent experimental studies<sup>18,24,29</sup>, single-resource communities were remarkably rich (mean richness =  $23 \pm 2$  ASVs, Fig. 1b) and taxonomically diverse (Extended Data Fig. 2). This is



**Fig. 1 | Microbial community diversity increases slowly with the number of resources despite individual resources supporting rich communities.**

**a**, Experimental layout. We inoculated a 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 compounds (total carbon concentration, 0.1% w/v). Bacterial cultures were grown for 7 d under a regime of daily dilution and their composition assessed at the single nucleotide resolution using 16S rRNA amplicon sequencing. Community diversity was measured as richness (number of ASVs). **b**, Richness of microbial communities in single carbon sources (mean  $\pm$  s.e.m.,  $N = 3$ ) correlates with the number of metabolites predicted to be generated from metabolic reactions mapped in the KEGG database. Pearson's correlation coefficient  $r = 0.75$ ; 95% confidence interval, 0.4–0.91;  $P < 0.001$ . **c**, A representative example of how observed richness in constituent single resources (mean  $\pm$  s.e.m.,  $N = 3$ ) compares with the observed richness in two-resource communities (mean  $\pm$  s.e.m.,  $N = 3$ ) and predictions calculated as the union (sum without overlapping ASVs, dark violet) or the maximum (light violet) of the richness in constituent singles (mean  $\pm$  s.e.m.,  $N$  from permutations = 9). **d**, Observed average richness (orange dots, mean  $\pm$  s.e.m.,  $N = 16$  for 1-resource, 24 for 2-resource, 12 for 4-resource, 6 for 8-resource, 16 for 15-resource, 1 for 16-resource combinations) as a linear function of the number of supplied carbon sources (solid orange line). Grey dots indicate the average richness for each unique combination of resources (mean  $\pm$  s.e.m.,  $N = 3$ , intercept,  $20.7 \pm 0.8$ , slope,  $1.4 \pm 0.1$ ,  $P < 0.001$ ). 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 **b**, are shown.

in contrast with competitive exclusion predicting that the number of species cannot exceed the number of resources<sup>20,30</sup>—which, in single carbon sources, would result in only one species surviving. Interestingly, the variability in richness among different resources was also high—with the average number of ASVs ranging from 8 in citrate to  $40 \pm 4$  in cellulose (Extended Data Fig. 3 and Fig. 1b)—and higher than the variability among replicates of the same carbon source (analysis of variance (ANOVA),  $F_{\text{resource}} = 3.4339$ ,  $P < 0.01$ ). Richness in single carbon sources therefore depended on resource identity. Community richness did not correlate significantly with the molecular weight of the supplied resource (Supplementary Fig. 4), but correlated with the predicted number

of metabolites that could be generated from the resource through intracellular biochemical reactions and secreted in the environment (Fig. 1b, see Methods for details on the prediction of metabolites on the basis of the KEGG<sup>31</sup> and MetaCyc<sup>32</sup> databases). Notably, the lowest richness was observed for gluconeogenic substrata (~10 for citrate, fumarate and hydroxyproline), which were connected to the central metabolic pathway via the TCA (tricarboxylic acid) cycle, hence resulting in the smallest metabolite pools. Consistent with previous work, these results highlight the role of cross-feeding in supporting community diversity<sup>33–35</sup>. Moreover, they suggest that the extent of cross-feeding may determine how many species can coexist on single resources.

Having found large numbers of coexisting species in single resources, we expected that community diversity would increase rapidly if more resources were provided. As previously observed in marine bacteria<sup>24,25</sup>, community composition could potentially correspond to 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, that is, the sum minus the number of shared ASVs (union). Alternatively, niche overlap between the taxa found in the single-resource media<sup>36,37</sup> 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 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 was 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 (Extended Data Fig. 4). Contradicting our expectations based on previous results supporting additivity, we found that community richness upon combining two carbon resources was approximately the average richness of constituent single-resource environments.

#### Community diversity increases slowly with resource number.

Next we examined the full range of resource combinations included in the experiment. Again, 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). We also found a similar trend when we estimated the number of metabolites generated from resource combinations, with the same approach that we used for single carbon sources (Extended Data Fig. 5). Instead, the observed average richness increased linearly with the number of supplied carbon sources, at the constant rate of one to two ASVs for each new added resource (Fig. 1d; slope,  $1.4 \pm 0.1$ ). As a result, the richness of communities supported by 16 resources was roughly twice the average richness of single-resource communities. The linear relationship was robust to the exclusion of low-abundance ASVs—with the slope reduced to  $1 \pm 0.07$  when ASVs with relative abundance below 0.1% were excluded (Extended Data Fig. 6a)—and coarse-graining at the family level (Extended Data Fig. 6b). In addition, as more resources were provided, communities became more even (Methods and Extended Data Figs. 6c,d and 7), without changes in total biomass (Supplementary Fig. 5). Despite confirming that the number of supplied resources is an important driver of microbial diversity, the observed one-by-one relationship between richness and resource number was difficult to reconcile with the high diversity found in single resources. Thus, we went back to the single-resource communities to gain a better understanding of our observations.

#### Communities harbour both habitat generalists and specialists.

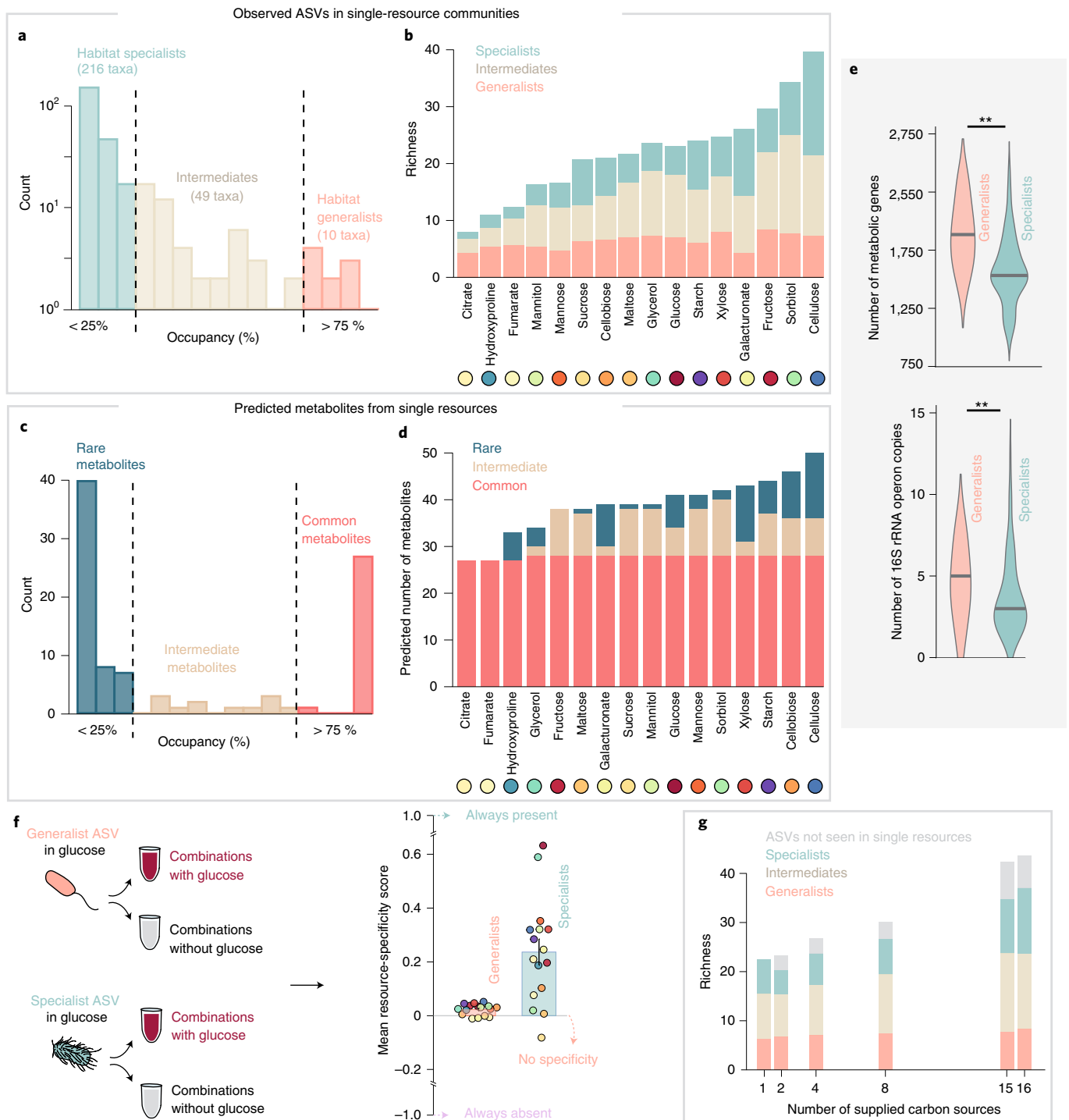
First, we measured the resource occupancy of the 275 ASVs observed in single-resource media, that is, how many single-resource media a given ASV was found in (Fig. 2a and Extended Data Fig. 8). On the basis of resource occupancy, we considered habitat specialists as ASVs that were observed in less than 25% of single-resource media, and habitat generalists as those that occupied more than 75% of single-resource media (Fig. 2a). Most ASVs (216 out of 275) were specialists, whereas very few (10) were generalists. Some ASVs (49) displayed an intermediate occupancy, being present in 4 to 12 media. This is reminiscent of natural communities, in which few taxa are usually universally present across different habitats, while the majority is found only under specific environmental condi-

tions<sup>38–40</sup>. Importantly, previous work has shown that variations in the proportion of generalist to specialist taxa within a community impact its dynamics<sup>41–43</sup>.

Next we inspected the distribution of specialists and generalists within single resources. We found that species-poor communities grown on gluconeogenic substrates such as citrate, were dominated by generalist ASVs (often representing more than 50% of observed taxa; Fig. 2b and Extended Data Fig. 9a), while species-rich communities were enriched in specialists (see cellulose in Fig. 2b and Extended Data Fig. 9b). We noticed that glycolytic substrates, which can produce a much larger metabolite pool before connecting with the central carbon metabolism, supported communities where more specialists coexisted with generalists. This suggested a link between the metabolite pool generated from each supplied resource and its ability to sustain both generalists and specialists in the same community.

**Parallelism in taxa and metabolite distributions.** Remarkably, similar to ASVs, our predicted metabolic by-products could also be broken up into two broad classes. On the basis of the number of single resources that could trigger their production, we could distinguish between common metabolites, present in association with most of the single resources (similar to generalist taxa), and rarer metabolites, present only in association with one or few resources (similar to specialists) (Fig. 2c). Metabolites that were commonly produced constituted the ‘core’ intermediates of the central metabolic pathway, including the TCA cycle and lower glycolysis, whereas the rarely produced metabolites were the intermediates of ‘peripheral’ branches of the central pathway. For example, if either citrate or fumarate were provided, we predicted that the central pathway proceeds in the gluconeogenic direction, generating only by-products belonging to the core pool. In contrast, individually supplied sugars were predicted to go through a series of reactions before entering the central pathway, ultimately generating both core and peripheral metabolites (Fig. 2d). It appeared that the number of peripheral metabolites varied with the position from which the resource entered the ‘metabolic map’. The parallelism in the distribution of ASVs and predicted metabolites reinforces the idea that community structure is coupled to the metabolite pool, and suggests a link between the resource occupancy and metabolic capability of taxa.

We next tested for systematic differences in the metabolic capabilities between generalist and specialist taxa in our experimental microcosms. Generalist ASVs belonged to the most abundant families, that is, Pseudomonadaceae, Enterobacteriaceae and Micrococcaceae (Extended Data Fig. 8) and differed metabolically from specialists, for example, taxa from Cellvibrionaceae. In particular, generalists were estimated to harbour a greater number of metabolic genes (Fig. 2e, top; see Methods for details on the estimation of gene content) and more copies of the 16S ribosomal RNA operon compared with specialists (Fig. 2e, bottom; see Methods for details on the matching with the number of copies of the rRNA operon), indicative of faster maximum growth rates<sup>44</sup>. Both results are consistent with studies showing the hallmarks of a generalist life style: flexible metabolism<sup>38</sup> (indicated by the number of metabolic genes) and capacity for fast growth (indicated by the 16S rRNA operon copy number)<sup>45,46</sup>. Nonetheless, several of the taxa classified as generalists are known to show distinct resource preferences when grown in isolation. For example, *Pseudomonas* species dominated in the communities sustained by organic acids, most probably because of their advantage over other taxa preferring sugars<sup>47</sup>, but were also present in all media in which organic acids could have been generated as by-products of the glycolytic metabolism of sugars<sup>48</sup> (Extended Data Fig. 1). This may indicate that generalists were present in all the communities because the substrates that they utilize were always generated as by-products of bacterial metabolism.



**Fig. 2 | Experimental communities are composed of generalists and variable numbers of specialists, with the latter driving the increase in community diversity.** **a**, The 275 ASVs found across single-resource communities were classified as generalists, specialists and intermediates depending on their resource occupancy. **b**, Mean number of generalist (pink), intermediate (beige) and specialist (teal) ASVs ( $N=3$ , error bars are omitted for clarity) in single resources. **c**, The metabolites predicted to be produced from the supplied single resources through cell reactions were classified as common, intermediate and rare, on the basis of resource occupancy, as with ASVs. **d**, Number of common (red), intermediate (light brown) and rare (blue) metabolites for single resources. **e**, Top: the distribution of the number of metabolic genes retrieved for each ASV in single resources (Methods) differs between generalists and specialists (\*\* $P < 0.01$ , Kolmogorov-Smirnov test). Bottom: the distribution of 16S rRNA operon copy numbers, calculated at the genus level, for generalist ASVs differs from that of specialist ASVs (\*\* $P < 0.01$ , Kolmogorov-Smirnov test). **f**, Mean resource-specificity score  $\pm$  s.e.m. for generalists (pink) and specialists (teal) ( $N=16$ ). Coloured dots indicate the mean score for each resource (s.e.m. are omitted for clarity,  $N$  varies for each resource). The score was calculated for each ASV found in a single resource (target resource) as  $(X - Y)/(X + Y)$ , where  $X$  is the probability for multiresource media including the target resource to harbour the ASV and  $Y$  is the probability for media not including the target resource to harbour the ASV (Methods). **g**, Mean number of generalist (pink), intermediate (beige) and specialist (teal) ASVs for media with equal number of supplied resources. The average number of ASVs that were not detected in single resources but appeared in other combinations is shown in grey. Error bars are omitted for clarity.



Indeed, even habitat generalists show resource preferences<sup>49</sup>, such as *Pseudomonas* spp., which preferentially consume acetate and other organic acids<sup>23</sup>. In contrast, since many sugars and their intermediates could not be produced via gluconeogenic metabolism<sup>50</sup>, the survival of the taxa specializing on them was prevented unless those sugars were externally supplied. Together, these observations are consistent with the idea that habitat generalists and specialists assemble in a community in relation to the available supplied and cross-fed metabolites.

#### Habitat specialists drive the increase in community diversity.

The coupling between metabolite pool and community structure observed in single resources suggest that resource–ASV associations would also be maintained in multiresource environments. In particular, we expected that generalists would be present in all communities, while specialists would be mostly detected when the favourite substrate was provided or metabolically generated. To verify these expectations, we calculated a resource-specificity score for each ASV, as its tendency to be present only along with a particular resource (Methods). The score ranged from 1, indicating that the ASV was present only when a resource was provided in a combination, to  $-1$ , implying that, although the ASV was found in the single resource, it was always absent when that resource was supplied with others. A score of 0 indicated that an ASV showed no specificity for that resource (Fig. 2f). We found that specialists' scores were on average positive across all resources (Fig. 2f; mean score,  $0.24 \pm 0.05$ ), while generalists' scores were on average nearly zero ( $0.02 \pm 0.01$ ). Together, these findings emphasize that (specialist) taxa tend to show resource-specific associations, and that single-resource–ASV associations are maintained even in multiresource environments.

To verify how resource–ASV associations impacted the resource–diversity relationship, we then calculated the average number of specialists, generalists and intermediates (as defined on the basis of single-resource occupancy) for each combination of carbon sources. We found that going from 1 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. 2g). Overall, these results point to the consistent coexistence of distinct groups of bacteria in our experimental microcosms, with more specialized taxa progressively being favoured by the supply of additional resources. While this was in line with the expectation that specialists of each resource should be favoured by the higher chances to introduce a glycolytic compound as more resources were added, it is important to note that, at the same time, several specialist ASVs were lost and few new ASVs were introduced, especially going from one to two carbon sources (grey bars in Fig. 2f; Supplementary Fig. 6; these ASVs remained unclassified).

#### A resource-explicit model reproduces experimental results.

We next asked: can we recapitulate some of the primary features of our experimental results by incorporating the metabolic network in a resource-explicit modelling framework? We hypothesized that the metabolic network seeded by the supplied carbon sources could explain the observed resource–diversity relationship. To test this, we implemented the well-known MacArthur consumer–resource model with cross-feeding<sup>19,21,35</sup>. In contrast to other implementations that used an abstract, randomly generated cross-feeding network<sup>18,35</sup>, we used the realistic network inferred using the KEGG and MetaCyc databases (Fig. 3a and Methods). In our model, for simplicity, each species consumed one resource to grow, to approximate specific resource preferences in different species. It then leaked metabolic byproduct(s) into the environment, each of which was one step downstream from the consumed metabolite according to the cross-feeding network (Methods). Other species could

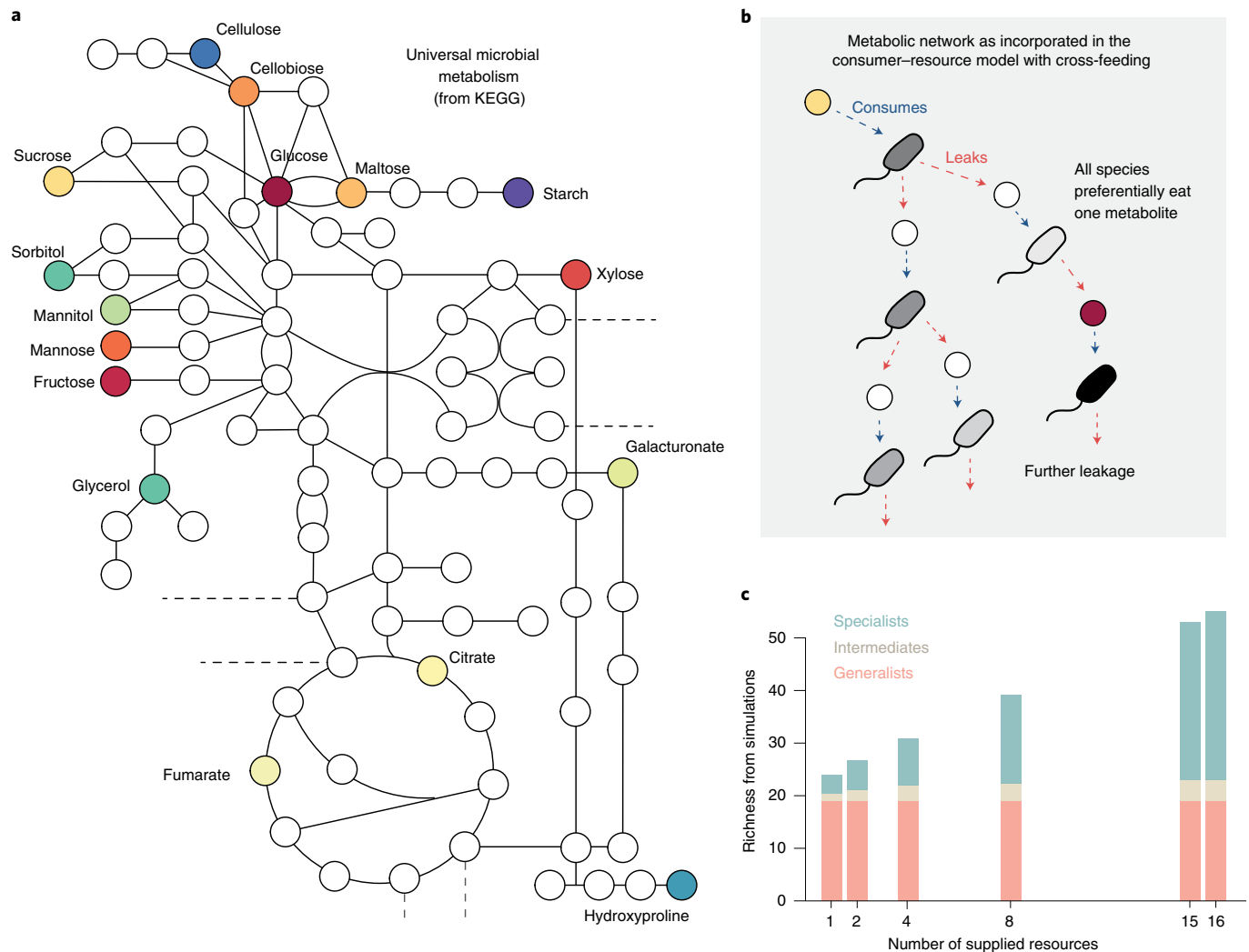
then consume these leaked metabolites, in turn releasing new by-products into the environment (Fig. 3b). Importantly, leaked by-products always comprised a fixed fraction of the consumed resource, resulting in a progressive decrease in the concentration of metabolic by-products available to microbes downstream<sup>34,35</sup> (Fig. 3b). Finally, to account for metabolic overflow<sup>50–52</sup>, we added a small quantity of TCA intermediates and acetate to all simulated media. Overall, this model incorporated ecological dynamics and cross-feeding in a realistic fashion while retaining simplicity.

Simulations of this model reproduced our two most-prominent experimental observations. That is, we could observe the stable coexistence of many species in single resources (between 19 and 28 species, Extended Data Fig. 10), yet also noted a modest linear increase in richness with the number of resources (slope  $\sim 2$ , Fig. 3c and Extended Data Fig. 10). Since the estimated metabolic by-products came from mapped metabolic reactions, we already knew that their number increased nonlinearly with supplied resources (Extended Data Fig. 5). So, how could we have gotten a linear resource–diversity relationship? In our simulations, the concentration of by-products varied with two factors: their position in the metabolic network and the initial concentration of the supplied resources. To mimic our experiment, we maintained a constant total resource concentration. This resulted in the concentration of each supplied resource decreasing with the total number of supplied resources in the medium (as  $1/R$ ,  $R$  being the number of supplied resources). As we provided more resources, a progressively larger fraction of by-products had their steady-state concentrations decrease nonlinearly. By-products at very low concentrations could no longer support microbial species, since their growth rates fell below the dilution rate. Hence, even though the number of metabolites grew nonlinearly with the supplied resources, the metabolites that could support the growth of new species grew much slower, resulting in a linear increase in diversity in our simulations.

By classifying the species in our simulations based on their resource occupancy (as in the experiment), our model also predicted a constant number of generalists and an increasing number of intermediates and specialists with additional resources (Fig. 3c). This is consistent with our experimental observations (Fig. 2g), and further corroborates the idea that the generalists observed in our communities were better at taking advantage of core metabolites, while the specialists that survived were those that were better at competing for rarer metabolites. Importantly, by implementing our model with a realistic network, we were able to simulate all the possible combinations of our 16 resources, even those that we did not experimentally grow. These simulations showed the same relationship between richness and resource availability (Fig. 3c), suggesting that the outcome of our experiment would not have changed if we had included more/different resource combinations. We conclude that the realistic cross-feeding network seeded by the pool of carbon sources in our experiment could explain the observed relationship between microbial community diversity and resource number.

#### Discussion

Understanding the relationship between available nutrients and community diversity is central to both theoretical and experimental ecology. Our results add to the wealth of studies stressing the importance of metabolic cross-feeding as a pivotal driver of species coexistence<sup>53</sup> and its link to the identity of available resources<sup>18,27</sup>. We have observed multispecies communities on all compounds provided as single sources of carbon, including mono-, di- and polysaccharides, sugars, alcohols and organic acids. Importantly, we were able to link the variability in community richness to the identity of the supplied resource by mapping the metabolic pathways triggered by each resource and estimating the number of by-products potentially produced and leaked into the growth media. We are aware that these predictions do not provide any



**Fig. 3 | A resource-explicit model incorporating a realistic metabolic network recapitulates our experimental results. a**, A simplified version of the metabolic map derived from the KEGG and MetaCyc databases is shown, where the carbon sources used in the experiments are highlighted. The metabolic map was used to build the cross-feeding matrix in the model. **b**, Schematic showing the flow of metabolic by-products in our model. Coloured circles indicate supplied resources; white circles indicate metabolic by-products, that is, metabolites that are downstream from the resource in the metabolic network; coloured microbes indicate different microbial taxa; arrows indicate leakage of metabolic by-products, which serve as resources for other taxa. **c**, Richness obtained from simulations with all the possible combinations of 16 resources (14,843 conditions in total) is plotted as stacked bars indicating the average number of species for each category: generalists (pink), specialists (teal) and intermediates (beige). Species in the model were classified on the basis of the number of ‘media’ they survived in, analogous to the distinction applied in the experiment (Fig. 2). Error bars are omitted for clarity. The total richness increased linearly with the number of resources (intercept, 22.7, slope, 2).

information on the bio-availability of the metabolic by-products and might be biased towards well-characterized bacterial species. Nevertheless, they provide a simple and tractable way to estimate by-products using only the structure of an overall metabolic network. Incidentally, community-scale flux balance simulations on the single-resource communities in our experiment also predicted a correlation between the number of expected by-products and community richness (Supplementary Fig. 7 and Methods). Further advances in linking the available by-products with richness could be provided by targeted metabolomics, a technique that can assess the relative concentration of the metabolites in a medium, as in ref. 50.

Our modelling approach, despite its simplifying assumptions (for example, we used species-independent growth and leakage rates), recapitulated our experimental results, capturing the combined effect of dilution and resource concentration that might have determined the diversity of experimental communities. At the same time, its theoretical bases still remain to be fully understood, including

an extensive exploration of how the structure of the metabolic network affects resource–diversity relationships. Other approaches to complement such theoretical efforts might include experimentally testing the effect of resource concentration and quantitatively modelling intracellular metabolism, thus also accounting for metabolic fluxes and redox balances<sup>54–58</sup>.

The position from which a resource enters central metabolism affects not only its availability but also the direction in which metabolic reactions run (that is, glycolytic or gluconeogenic). We showed that whether a resource is glycolytic or gluconeogenic was an important predictor of the diversity and structure of microbial communities, as it dictated the ratio between habitat generalists and specialists. These results suggest, and a two-parameter regression supports (Supplementary Fig. 8), that adding gluconeogenic resources (for example, organic acids) while keeping the total concentration of carbon constant may not increase community diversity. Overall, our results add to the studies stressing that the position

from which a resource enters central metabolism eventually determines its use, including diauxic shifts versus co-utilization<sup>59</sup>, and trade-offs between growth and lag in changing environments<sup>50</sup>.

A further indication of the role played by the metabolic network sustained by the supplied resources came from the striking parallelism that we observed between the structure of experimental communities and the architecture of the network itself. Just as metabolism consists of shared (for example, the TCA cycle) and unique reaction modules (that is, specific to the degradation of a particular resource), all experimental communities harboured a core group of metabolically flexible, faster-growing habitat generalists and variable numbers of taxa associated with a particular nutrient (habitat specialists). This suggests that the habitat generalists present in our stabilized microbial communities were 'specialists for common nutrients', that is, they preferentially consumed substrates that are commonly produced during bacterial growth. In this sense, generalists growing on downstream metabolites (for example, TCA intermediates) depended on specialists for the production of their favourite substrates. Consistent with this, community flux balance simulations where we paired a generalist and specialist ASVs showed that it was the specialists that are likely to leak metabolic by-products used by generalists, and not vice versa (Supplementary Fig. 9 and Methods).

Finally, in our experiments, habitat specialists outnumbered generalists overall, a pattern that is commonly observed when natural communities from different locations are compared<sup>38</sup>. Surveys of microbiomes across different ecosystems have also highlighted a remarkable level of determinism in the association between microbiome composition at coarse taxonomic resolutions (for example, at the family level) and availability of nutrients<sup>18,60</sup>. Here we showed the persistence of strong taxa–resource associations both at the ASV (Fig. 2f) and the family levels (Supplementary Fig. 10), with the relative abundance of several families, comprising prevalently either generalist or specialist taxa, changing as a function of the relative concentration of specific resources (see Methods for how we established which resources influenced each family the most). Specifically, we observed that the relative abundance of several specialist families decreased drastically or dropped to zero when the relative concentration of the 'favourite resource' dropped by half (for example, Cellvibrionaceae, Supplementary Fig. 10), while the relative abundance of generalist taxa increased nonlinearly with the relative concentration of few resources (for example, Pseudomonadaceae with hydroxyproline and fumarate, Supplementary Fig. 10). The fact that empirically observed features of natural microbial communities emerge in controlled experiments suggests that they might reflect the effects of deterministic processes linked to nutrient availability rather than being generic emergent properties of complex multi-agent systems.

## Methods

**Growth media preparation.** All the chemicals were purchased from Sigma-Aldrich unless otherwise stated.

All bacterial cultures were grown in M9 media (prepared from 5X M9 salts, 1X trace metal mixture (Teknova) and 1 M stock solutions of MgSO<sub>4</sub> and CaCl<sub>2</sub>) supplemented with 0.1% w/v of one of 75 carbon-source combinations. These combinations include: 16 compounds commonly available in soil, which were provided as single carbon sources (D-(+)-glucose, D-(+)-fructose, D-(+)-xylose, D-(+)-mannose, D-(+)-cellobiose, D-(+)-maltose monohydrate, sucrose, citric acid, fumaric acid, D-(+)-galacturonic acid monohydrate, D-mannitol, D-sorbitol, glycerol, *trans*-4-hydroxy-D-proline, methyl cellulose, starch); 24 random combinations of 2 of these resources; 12 random combinations of 4 resources; 6 random combinations of 8 resources; the 16 combinations containing 15 resources; and all the 16 resources together (see Supplementary Table 1 for the complete list and Supplementary Fig. 2). The total concentration of carbon was kept the same and resources were, in all instances, supplied in equal amounts, that is, 100, 50, 25, 12.5, 6.7 and 6.25% for 1-, 2-, 4-, 8-, 15- and 16-resource combinations, respectively. All solutions were filter-sterilized with a 0.22 µm filter and kept at 4°C for the duration of the experiment.

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

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

Every 24 h, the cultures were thoroughly mixed by pipetting up and down three times using the VIAFLO 96-well pipettor (Vialflo 96, Integra Biosciences; settings: pipette/mix program aspirating 7 µl, mixing volume 10 µl, speed 6) and then diluted 1/30X into fresh media. We applied a total of 7 daily dilution cycles. At the end of every cultivation day, we measured the optical density (OD<sub>600</sub>) using a Varioskan Flash (Thermo Fisher Scientific) plate reader. The remaining bacterial culture was frozen at –20°C for subsequent DNA extraction.

**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 Massachusetts Institute of Technology BioMicroCenter.

We used the R package DADA2 to obtain the amplicon sequence variants (ASVs)<sup>61</sup> following the workflow described in Callahan et al.<sup>62</sup>. Taxonomic identities were assigned to ASVs using the SILVA version 132 database<sup>63</sup>. The phylogenetic tree (Supplementary Fig. 4) was reconstructed using randomized accelerated maximum likelihood (RAxML) using default parameters<sup>64</sup>.

**Data analysis.** Unless otherwise stated, analyses were conducted in R version 3.6.1<sup>65</sup>.

Sequencing data was handled using the R package phyloseq<sup>66</sup>. We obtained an average of 20,613 reads per sample. Sequencing depth did not affect our estimation of community diversity indices (Supplementary Fig. 3). Richness was calculated as the number of ASVs with abundance higher than 0 in each sample. Community diversity was also measured using the Shannon diversity index and the Shannon entropy index following refs. <sup>67,68</sup> (Extended Data Fig. 6). The significance of differences in richness due to single supplied resources was tested through ANOVA<sup>69</sup> using the package GAD.

**Richness predictions.** Predictions of how richness would grow with the number of supplied carbon sources were computed using all three replicated communities grown on a single resource and all the possible combinations of single resources (120 combinations of 2 resources, 1,820 combinations of 4, 12,870 combinations of 8, 16 combinations of 15 and 1 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 a medium containing glucose + hydroxyproline was obtained by calculating the maximum richness over each couple of replicates (one containing only glucose and the other containing only hydroxyproline) and subsequently averaging across all the predicted maxima (nine predicted values in total). The same procedure was used for the average of constituent singles. Analogously, for the predictions based on the union of constituent singles, the richness in glucose + hydroxyproline was predicted by calculating the number of unique ASVs found in each couple of replicates of constituent singles (that is, the total number of ASVs minus the number of overlapping ASVs) and then averaging across all obtained unions (nine values).

**Rank abundance distributions.** First, we computed rank abundance distributions (RADs) for each sample, that is, each replicate community grown on a unique combination of carbon sources, by sorting ASVs on the basis of their relative abundance. Then we plotted the RADs in a log-linear fashion and fitted a regression line to compare their slopes (Extended Data Fig. 7a). The absolute value of the slope of the fitted regression line provides information on the abundance distribution of the ASVs in a community. More even communities usually display smaller slopes (Extended Data Fig. 7b). Since each community exhibited a different richness, we normalized the RADs for richness (Extended Data Fig. 7c) using 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 the minimum richness observed in the experiment



(seven ASVs) by randomly resampling the original RADs ten times<sup>70</sup>. In this way, samples with different richness can be compared and changes in evenness properly assessed.

**Definition of generalists and specialists based on single-resource occupancy.** ASVs found in single resources were classified into three categories on the basis of the number of media containing a single resource they were found in, that is, they exhibited abundance higher than 0 (refs. <sup>38,43</sup>). We considered as specialists those ASVs that were observed in less than 25% of single-resource media, that is, in 1, 2 or 3 resources. Generalists were those ASVs found in more than 75% of media, that is, in 13 or more resources. We defined as intermediates those ASVs that were found in 4 to 12 resources. These thresholds were chosen arbitrarily, but they resulted in ~4% generalists and 80% specialists, consistent with proportions of generalists and specialists observed in natural communities<sup>38,39,43</sup>. We chose this simple way of assigning ASVs to generalist, intermediate and specialist categories over other methods, for example, as in ref. <sup>24</sup>, to leave aside their relative abundance, which was analysed separately.

**Prediction of possible metabolic by-products in resource environments.** We predicted the possible number of metabolic by-products that could be produced with the resources present in each medium using a curated metabolic network. The metabolic network contained a large set of metabolic reactions encompassing carbohydrate, sugar and amino acid metabolism extracted from the KEGG database<sup>31</sup>. We manually curated this large set of reactions using the MetaCyc database<sup>32</sup> to limit it to reactions possible in most microbial taxa common in the soil, such as *Pseudomonas*. We used this network to estimate all the metabolic compounds that could be produced as by-products, starting from the carbon sources available in each medium. We assumed that a small set of 'currency' molecules, such as water, carbon dioxide and ATP, were always available as reactants when required (full list of currency molecules: phosphate, oxygen, carbon dioxide, water, H<sup>+</sup>, ATP, NAD(P)H, acetyl-CoA and CoA).

To estimate the possible by-products in each medium, we employed the well-known scope expansion algorithm<sup>71–75</sup>. Each reaction in our curated metabolic network consisted of a set of reactants and resulting products. For each medium, we first asked which reactions could be performed using only the carbon sources available in the medium (that is, the current 'scope' of the medium). We assumed that the products of these reactions could be produced and added them to the set of reactants—the new scope—for the next step. In the next step, we again asked which reactions could be performed using the new scope. We added their products to the scope for the next step. We continued this process, step by step, until we could no longer add new products to the scope. The resulting final scope of metabolites, minus the currency molecules provided in the medium, was our estimated set of possible metabolic by-products producible in that medium.

Adding some amino acids as currency molecules, mimicking our experimental protocol, yielded a larger set (~3×) of possible metabolic by-products for each medium, including many amino acids and anabolic products. This expanded set of metabolites for each medium was also correlated with the observed average species richness in that medium.

We also tried an alternative approach to estimate the number of metabolic by-products in single-resource environments, using community-scale flux balance simulations. For each ASV observed in a single-resource environment, we first obtained the phylogenetically closest whole-genome sequence in the RefSeq database of the National Center for Biotechnology Information (NCBI). For this, we mapped the 16S sequence of each ASV to complete genomes in the RefSeq database using BLAST<sup>76</sup>. For each ASV, we chose the genome that had the highest identity; when multiple genomes matched this criterion, we chose the longest genome, following similar work<sup>77</sup>. We then obtained all the mapped genome sequences and constructed metabolic models for each of them using CarveMe<sup>54</sup>; we gap-filled all models to grow on M9 minimal medium supplemented with metal ions, such as iron and copper, which were present in trace amounts in experimental bacterial growth media.

To estimate the number of metabolic by-products in each single-resource environment, we performed community-scale metabolic simulations using the package MICOM<sup>56</sup>. For each community, we input all the metabolic models for all ASVs detected in that community, and simulated their growth in the corresponding media. We then counted all metabolites that were predicted to be exported by each community as the estimated number of by-products for that community. For each medium, we averaged the number of by-products across all three replicate communities; we used this as our estimated number of by-products for that medium.

**Characterization of the structure of the metabolite pool.** Following the same logic that we used for ASVs, metabolites estimated to be produced through metabolic reaction starting from single resources were classified into three categories on the basis of the number of resources they could be produced from. We classified 'rare' metabolites as those observed in less than 25% of single-resource media, that is, in 1, 2 or 3 resources. In contrast, 'common' metabolites were those found in more than 75% of media, that is, in 13 or more resources. Finally, 'intermediate' metabolites were those present in 4 to 12 resources. The chosen thresholds separate

the metabolites of the central metabolic pathway (common metabolites) from the peripheral metabolites belonging to branches descending into the central pathway (rare and intermediate metabolites).

**Inference of rRNA operon copy number for generalist and specialist taxa.** To test for signatures of different life-history strategies of the generalist and specialist taxa in our study, we estimated their 16S rRNA operon copy numbers. We estimated rRNA copy numbers at the level of both genus and family, separately for generalist and specialist taxa. For each genus identified, we queried rrnDB<sup>78</sup>—a database of rRNA operon copy number statistics—for the median copy number corresponding to the genus. We used this as an estimate of the rRNA operon copy number for that genus.

**Inference of number of metabolic genes for generalist and specialist taxa.** To test for metabolic differences between the generalist and specialist taxa in our study, we estimated the number of metabolic genes in their genomes. Since we did not have either isolates or assembled genomes corresponding to the observed taxa, we relied on a popular indirect method of estimating gene content. Namely, for each ASV, we used the reference genome which was phylogenetically closest to that ASV as a proxy for its genome. For this, we used PICRUSt2<sup>79</sup> using default parameters; as an input to the tool, we provided the 16S rRNA sequences of all 226 generalist and specialist taxa as well as their abundances in each sample. After running PICRUSt2, we obtained a table of the predicted gene content for each ASV (that is, presence/absence of a specific KO number in the KEGG database). We extracted all metabolic genes from this table by only choosing those KO numbers which had at least one known metabolic reaction corresponding to them. Doing so resulted in an estimated set of metabolic genes for each ASV; we used this as an indirect estimate of the metabolic capabilities of each ASV.

**Calculation of the resource-specificity score.** We used a resource-specificity score to test whether the ASV–resource associations that we observed in single resources were maintained when the single resource(s) in which the ASV was found was combined with others. For each ASV present in a single resource (target resource), the resource-specificity score was calculated as the difference between the probability for multiresource media including the target resource to harbour the ASV and the probability for media not including the target resource to harbour the ASV, divided by the sum of these two probabilities (Fig. 2f). This is reminiscent of a preference index, which is a standard measure in the behavioural sciences. Single resources are excluded from the count. The resource-specificity score ranges from 1, indicating that the ASV is present only when the target resource is provided, to –1, implying that the ASV is always absent when that resource is supplied with other resources. A score of 0 is indicative of an ASV showing no specificity for that particular resource (Fig. 2f). We calculated a score for each ASV–resource pair, such that each ASV had as many scores as the number of single resources it is found in. Then we computed the average of the scores obtained for each single resource, separating between scores belonging to generalist and specialist ASVs (Fig. 2f).

**Inference of metabolic interactions between generalist and specialist taxa.** To estimate whether metabolic interactions between the generalist and specialist taxa in our communities were likely to be unidirectional or bidirectional, we used SMETANA v.1.0<sup>55</sup>, using default settings. For each generalist–specialist pair that we experimentally detected in single-resource environments, we used SMETANA on a model community comprising both ASVs (a generalist and a specialist) using the settings: flavour bigg–exclude inorganic.txtt -d. We explicitly disallowed inorganic molecules such as phosphates, carbon dioxide and metal ions from being exchanged by using the –exclude option in SMETANA. To consider interaction directionality, we looked at the donor and receiver of each exchanged metabolite. When there was only one donor for every exchanged metabolite, we inferred the interaction was unidirectional, with the direction going from the donor to the receiver of the metabolites.

**Detection of family–resource associations using an ensemble tree regression model.** We calculated the relative abundance of the most prevalent families (37) in the 75 replicated bacterial communities and ran an ensemble tree regression model to detect significant patterns of variations in family abundance due to changes in the relative concentration of resources.

We chose to coarse-grain the abundance data at the family level because, while several ASVs were lost and others were gained when going from 1 to 16 resources in the growth media, the families found across all combinations of resources were mostly the same. In addition, we distinguished between generalist families, that is, those that contained at least one generalist ASV, and specialist families, that is, those containing only specialist ASV. Consistent with ASV-level definition, generalist families displayed higher mean rRNA operon copy number compared with specialist families.

We employed XGBoost, a gradient-boosting framework based on decision trees<sup>80</sup>. Specifically, we implemented a regression model for each family in which the input was the relative resource concentration and the output was the log-transformed relative family abundance. We trained the model on two



replicates by performing leave-one-out cross-validation of the XGBoost parameters 'max\_depth', 'n\_estimators' and 'learning\_rate'<sup>81</sup>, and tested on the third replicate with average mean-squared error of 6.05 across families. We applied the Shapley additive explanations (SHAP)<sup>82</sup> to identify the resources that were more important in driving changes in the abundance of each family. This analysis was done using Python version 3.8.

Results of this analysis revealed that variations in the abundance of all the 37 families were driven by one or multiple resources based on their dominant life strategy. To simplify the visualization of the results, we plotted the relative abundance of some representative families as a function of the concentration of the resources identified by the analysis (Supplementary Fig. 10). Families mostly composed of specialist taxa, for example, Cellvibrionaceae and Bacillaceae, showed abrupt changes in their abundance with the concentration of the 'favourite' resource (Supplementary Fig. 10). By contrast, more generalist families, for example, Pseudomonadaceae and Enterobacteriaceae, exhibited smooth trends in their abundance with the concentration of multiple resources.

**Resource–consumer model with cross-feeding and simulations.** The parallelism between species and metabolite distribution (Fig. 2) that we observed in our experiment emphasized that the cross-feeding network is key to understanding microbial communities under each combination of supplied carbon sources. To test this idea, we used a model encompassing the metabolic network that we obtained from the analysis of the KEGG and MetaCyc databases. This was achieved by a consumer–resource model with cross-feeding<sup>18,21,35</sup>. In our consumer–resource model with the realistic metabolic network, we made the following simplifying assumptions.

First, we assumed that every species consumes only one preferred metabolite. On this assumption, competitive exclusion guarantees that only the best grower in each resource survives; thus, we implemented only one species for each resource in our simulation as a post-selection pool. This assumption reflected the resource–species association we observed (Fig. 2), which suggested that the taxa identified as generalists may specialize on core metabolites that are found everywhere. Also, while many species can consume multiple resources, they may still grow much faster on the most preferred one. Metabolic strategies such as 'diauxie' also emphasize that growth on the most preferred resource can be a dominant factor for community assembly<sup>50</sup>.

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

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

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

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

$$\dot{n}_i = (1 - l)r_i c_i n_i - \delta n_i$$

$$\dot{c}_i = -r_i n_i c_i + l r \sum_j CF_{ij} n_j c_j + \delta (c_{i0} - c_i)$$

where  $n_i$  is the population of the  $i$ th species and  $\dot{n}$  its time derivative;  $c_i$  is the concentration of the  $i$ th metabolite and  $\dot{c}$  its time derivative;  $l$  is the leakage rate;  $r_i$  is the per-capita, per-resource growth rate of the  $i$ th species;  $\delta$  is the dilution rate of the chemostat-like environment.  $CF_{ij}$  indicates whether the  $i$ th metabolite is leaked from the  $j$ th metabolite on the basis of the scope expansion analysis.  $c_{i0}$  is the supply–resource concentration corresponding to each combination of supplied carbon sources, controlled by overall scale  $c_0$ . For example, when glucose is supplied,  $c_{i0} = c_0$  for  $i = \text{glucose}$ , and 0 otherwise. When a combination of glucose and hydroxyproline is supplied,  $c_{i0} = \frac{1}{2} c_0$  for  $i = \text{glucose, hydroxyproline}$ , and 0 otherwise. To simulate the effects of metabolic overflow<sup>50–52</sup>, we supplied a small quantity ( $c = 0.2$ ) of TCA intermediates and acetate to all media.

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

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

We simulated the model dynamics under all possible combinations of 1, 2, 4, 8, 15 and 16 supplied resources (14,843 combinations in total). We chose the parameters  $\delta = 0.1$ , which is comparable to the dilution we imposed in the experiment,  $r = 1$ ,  $c_0 = 100$  and  $l = 0.1$ . In Fig. 3c we show the results of all combinations, while in Extended Data Fig. 10 we plotted only the combinations included in the experiment. The simulations were run for  $10^3$  unit times starting from an initial population set at  $10^{-7}$ , and communities reached equilibrium at the end of the simulations. The population cutoff for survival was set at  $10^{-7}$ . Simulations were run in Python version 3.7.4.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data on 16S amplicon sequencing and metadata files have been deposited in the NCBI SRA database under NCBI BioProject ID [PRJNA715195](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA715195).

## Code availability

Data files and analysis/simulation codes to reproduce all figures are available at <https://github.com/mdalbello/Resource-diversity-relationship>.

Received: 10 December 2020; Accepted: 14 July 2021;

Published online: 19 August 2021

## References

- Hubbell, S. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
- Tilman, D. et al. Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
- Hutchinson, G. E. The paradox of the plankton. *Am. Nat.* **95**, 137–145 (1961).
- Koskella, B., Hall, L. J. & Metcalf, C. J. E. The microbiome beyond the horizon of ecological and evolutionary theory. *Nat. Ecol. Evol.* **1**, 1606–1615 (2017).
- Flemming, H. C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* **17**, 247–260 (2019).
- Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
- Salazar, G. & Sunagawa, S. Marine microbial diversity. *Curr. Biol.* **27**, R489–R494 (2017).
- Huttenhower, C. et al. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
- Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
- Cavicchioli, R. et al. Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
- Blasche, S. et al. Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nat. Microbiol.* **6**, 196–208 (2021).
- Gude, S. et al. Bacterial coexistence driven by motility and spatial competition. *Nature* **578**, 588–592 (2020).
- Kommineni, S. et al. Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* **526**, 719–722 (2015).
- Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The evolution and ecology of bacterial warfare. *Curr. Biol.* **29**, R521–R537 (2019).
- Ratzke, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nat. Ecol. Evol.* **4**, 376–383 (2020).
- Hoek, T. A. et al. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biol.* **14**, e1002540 (2016).
- Goldford, J. E. et al. Emergent simplicity in microbial community assembly. *Science* **361**, 469–474 (2018).
- Tilman, D. *Resource Competition and Community Structure* Vol. 17 (Princeton Univ. Press, 1982).
- Gause, G. F. *The Struggle for Existence* (Hafner Press, 1934).
- MacArthur, R. Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970).
- Levin, S. A. Community equilibria and stability, and an extension of the competitive exclusion principle. *Am. Nat.* **104**, 413–423 (1970).

23. Estrela, S. et al. Metabolic rules of microbial community assembly. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.03.09.984278> (2020).
24. Enke, T. N. et al. Modular assembly of polysaccharide-degrading marine microbial communities. *Curr. Biol.* **29**, 1528–1535 (2019).
25. Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in synthetic phycospheres. *Proc. Natl Acad. Sci. USA* **117**, 3656–3662 (2020).
26. Gralka, M., Szabo, R., Stocker, R. & Cordero, O. X. Trophic interactions and the drivers of microbial community assembly. *Curr. Biol.* **30**, R1176–R1188 (2020).
27. Enke, T. N. et al. Modular assembly of polysaccharide-degrading marine microbial communities. *Curr. Biol.* **29**, 1528–1535.e6 (2019).
28. Martiny, J. B. H., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: a phylogenetic perspective. *Science* **350**, aac9323 (2015).
29. Naylor, D. et al. Deconstructing the soil microbiome into reduced-complexity functional modules. *mBio* **11**, e01349–20 (2020).
30. MacArthur, R. H. *Geographical Ecology. Patterns in the Distribution of Species* (Harper & Row, 1972).
31. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
32. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **46**, D633–D639 (2018).
33. Wang, T., Goyal, A., Dubinkina, V. & Maslov, S. Evidence for a multi-level trophic organization of the human gut microbiome. *PLoS Comput. Biol.* **15**, e1007524 (2019).
34. Goyal, A. & Maslov, S. Diversity, stability, and reproducibility in stochastically assembled microbial ecosystems. *Phys. Rev. Lett.* **120**, 158102 (2018).
35. Marsland, R. et al. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Comput. Biol.* **15**, e1006793 (2019).
36. Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species diversity. *Nature* **461**, 254–257 (2009).
37. Tomas, N. et al. Niche separation increases with genetic distance among bloom-forming Cyanobacteria. *Front. Microbiol.* **9**, 438 (2018).
38. Sriswasdi, S., Yang, C. C. & Iwasaki, W. Generalist species drive microbial dispersion and evolution. *Nat. Commun.* **8**, 1162 (2017).
39. Logares, R. et al. Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J.* **7**, 937–948 (2013).
40. Monard, C., Gantner, S., Bertilsson, S., Hallin, S. & Stenlid, J. Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Sci. Rep.* **6**, 37719 (2016).
41. Székely, A. J. & Langenheder, S. The importance of species sorting differs between habitat generalists and specialists in bacterial communities. *FEMS Microbiol. Ecol.* **87**, 102–112 (2014).
42. Pandit, S. N., Kolasa, J. & Cottenie, K. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* **90**, 2253–2262 (2009).
43. Muscarella, M. E., Boot, C. M., Broeckling, C. D. & Lennon, J. T. Resource heterogeneity structures aquatic bacterial communities. *ISME J.* **13**, 2183–2195 (2019).
44. Roller, B. R. K., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 16160 (2016).
45. Goldfarb, K. C. et al. Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front. Microbiol.* **2**, 94 (2011).
46. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**, 1328–1333 (2000).
47. Rojo, F. Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol. Rev.* **34**, 658–684 (2010).
48. Mills, C. G., Allen, R. J. & Blythe, R. A. Resource spectrum engineering by specialist species can shift the specialist-generalist balance. *Theor. Ecol.* **13**, 149–163 (2020).
49. Bajic, D. & Sanchez, A. The ecology and evolution of microbial metabolic strategies. *Curr. Opin. Biotechnol.* **62**, 123–128 (2020).
50. Basan, M. et al. A universal trade-off between growth and lag in fluctuating environments. *Nature* **584**, 470–474 (2020).
51. Paczia, N. et al. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microb. Cell Fact.* **11**, 122 (2012).
52. Pinu, F. R. et al. Metabolite secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14**, 43 (2018).
53. Douglas, A. E. The microbial exometabolome: ecological resource and architect of microbial communities. *Phil. Trans. R. Soc. B* **375**, 20190250 (2020).
54. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
55. Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl Acad. Sci. USA* **112**, 6449–6454 (2015).
56. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *mSystems* **5**, e00606–19 (2020).
57. Garza, D. R., van Verk, M. C., Huynen, M. A. & Dutilh, B. E. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nat. Microbiol.* **3**, 456–460 (2018).
58. Pacheco, A. R., Osborne, M. L. & Segrè, D. Non-additive microbial community responses to environmental complexity. *Nat. Commun.* **12**, 2365 (2021).
59. Wang, X., Xia, K., Yang, X. & Tang, C. Growth strategy of microbes on mixed carbon sources. *Nat. Commun.* **10**, 1279 (2019).
60. Louca, S. et al. High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* **1**, 0015 (2017).
61. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
62. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research* **5**, 1492 (2016).
63. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
64. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
65. R Core Team R: *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
66. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
67. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
68. Chao, A., Chiu, C. H. & Jost, L. Phylogenetic diversity measures based on Hill numbers. *Phil. Trans. R. Soc. B* **365**, 3599–3609 (2010).
69. Underwood, A. J. *Experiments in Ecology* (Cambridge Univ. Press, 1996); <https://doi.org/10.1017/cbo9780511806407>
70. Saeedghalati, M. et al. Quantitative comparison of abundance structures of generalized communities: from B-cell receptor repertoires to microbiomes. *PLoS Comput. Biol.* **13**, e1005362 (2017).
71. Goldford, J. E., Hartman, H., Smith, T. F. & Segrè, D. Remnants of an ancient metabolism without phosphate. *Cell* **168**, 1126–1134.e9 (2017).
72. Raymond, J. & Segrè, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311**, 1764–1767 (2006).
73. Handorf, T., Ebenhoh, O. E. & Heinrich, R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **61**, 498–512 (2005).
74. Ebenhoh, O., Handorf, T. & Heinrich, R. Structural analysis of expanding metabolic networks. *Genome Inform.* **15**, 35–45 (2004).
75. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
76. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, 5–9 (2008).
77. Machado, D. et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* **5**, 195–203 (2021).
78. Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K. & Schmidt, T. M. 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).
79. Douglas, G. M. et al. PICRUSt2: an improved and extensible approach for metagenome inference. Preprint at *bioRxiv* <https://doi.org/10.1101/672295> (2019).
80. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016).
81. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn. (Springer Series in Statistics, Springer, 2009).
82. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).

## Acknowledgements

We thank J. Grilli, M. Costantino-Lagomarsino, M. Corigliano and M. Barbier for feedback on models, the members of the Gore Lab for comments on the manuscript, and B. Stellato for help with the ensemble tree regression model. This work was

supported by the Simons Collaboration: Principles of Microbial Ecosystems (PriME) award number 542395 and NIH (R01-GM102311). A.G. was supported by the Gordon and Betty Moore Foundation as a Physics of Living Systems Fellow through grant number GBMF4513.

### Author contributions

M.D.B. and J.G. conceived the study. M.D.B. performed the experiments and the sequencing analysis. H.L. performed theoretical modelling. A.G. performed metabolic and genomic analyses. All authors analysed the data and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-021-01535-8>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-021-01535-8>.

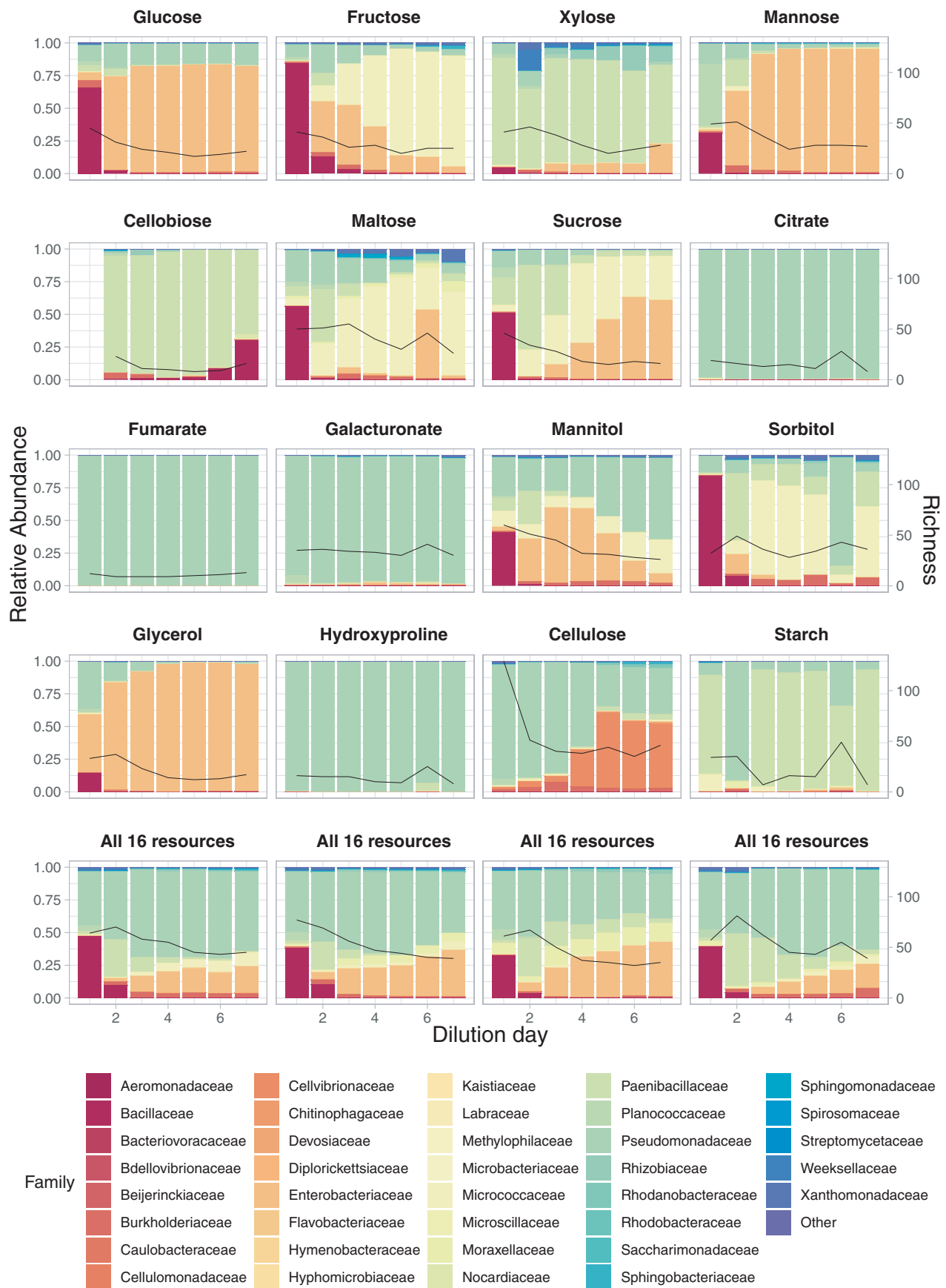
**Correspondence and requests for materials** should be addressed to M.D.B. or J.G.

**Peer review information** *Nature Ecology & Evolution* thanks Marc Garcia-Garcera and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021



Extended Data Fig. 1 | See next page for caption.

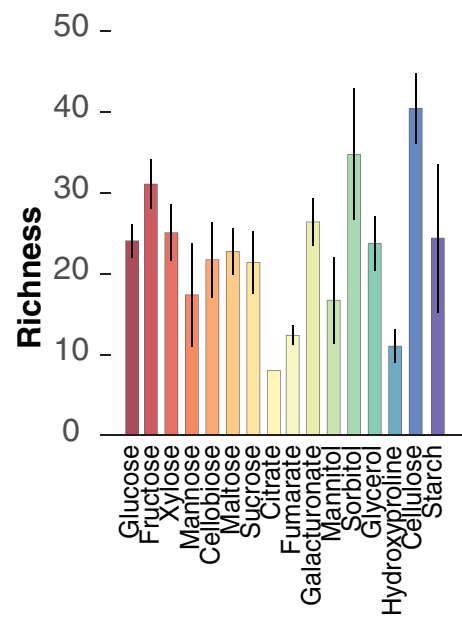


**Extended Data Fig. 1 | The majority of communities reached equilibrium before the end of the experiment.** Each panel shows the temporal trajectories of the composition of one community at the family level. The most prevalent 37 families are included. The first 16 plots depict communities grown on a single carbon source. The last four plots depict replicated communities grown on a media containing all the 16 carbon sources. The solid black line indicates the observed community richness.



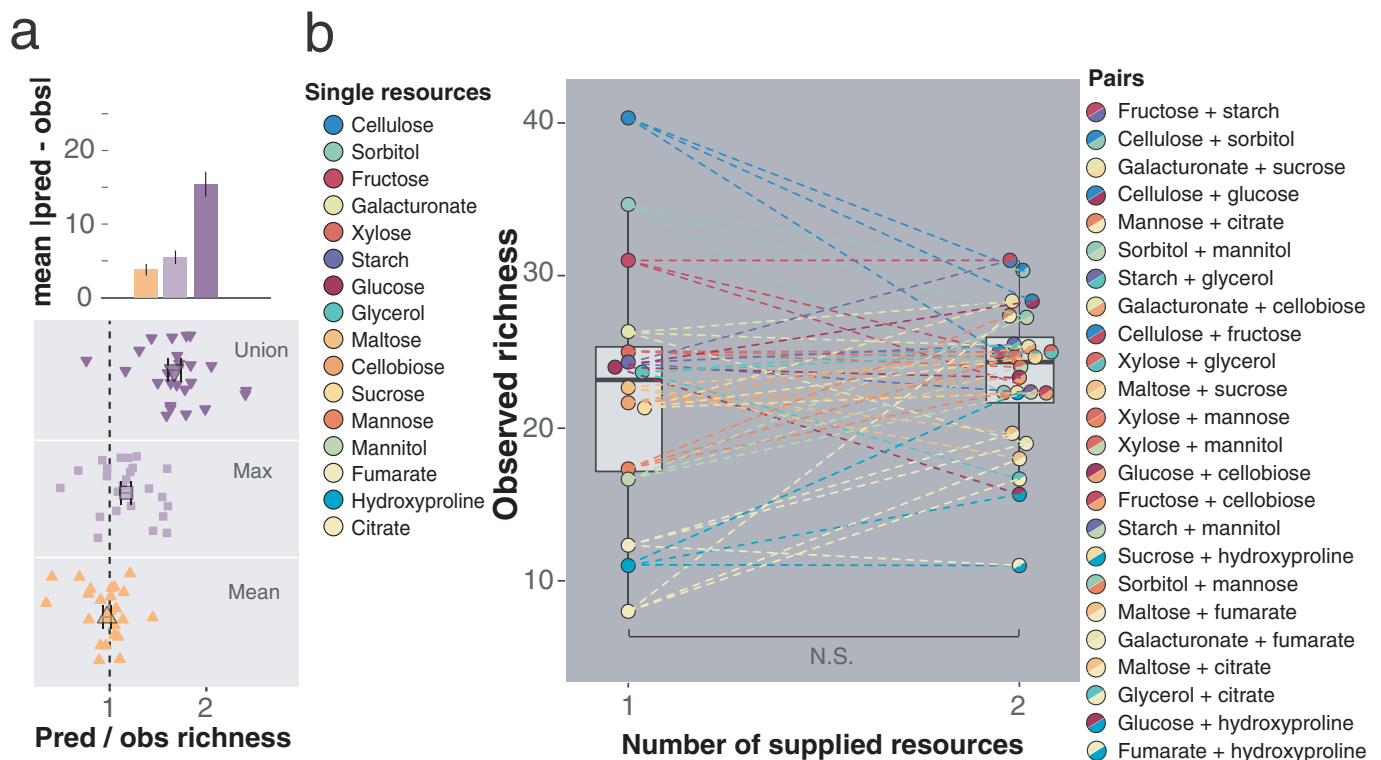
Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Single carbon sources support microbial assemblages spanning a wide phylogenetic diversity.** The pool of ASVs found across all media supplied with single resources is phylogenetically diverse, encompassing 7 classes, indicated by colored lines on the right side of the plot (black lines indicate ASVs that could not be identified at any taxonomical level except the Domain, Bacteria). Families are indicated on the left side of the phylogenetic tree. Colored tiles indicate the media in which the ASV is found (for each carbon source, there are three replicated microcosms for a total of 48 communities.) Carbon sources are ordered based on the average richness they support.



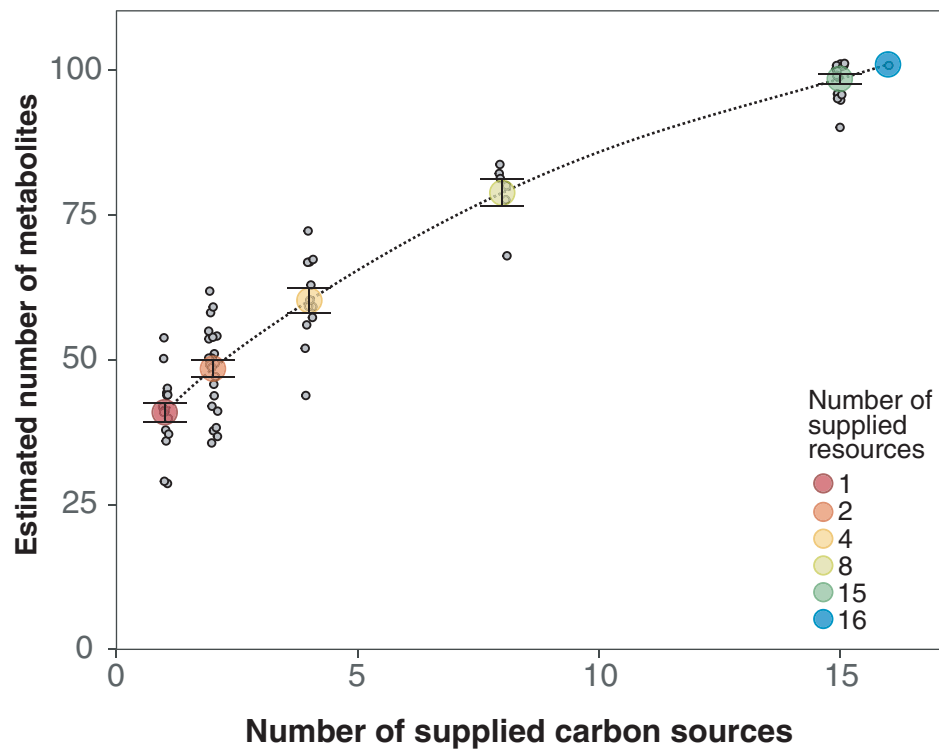
**Extended Data Fig. 3 | All single resources supported multispecies communities, but richness varied with the identity of the resource.** Bars indicate, for each carbon source, the number of ASVs (mean  $\pm$  s.e.m.,  $N = 3$ ).



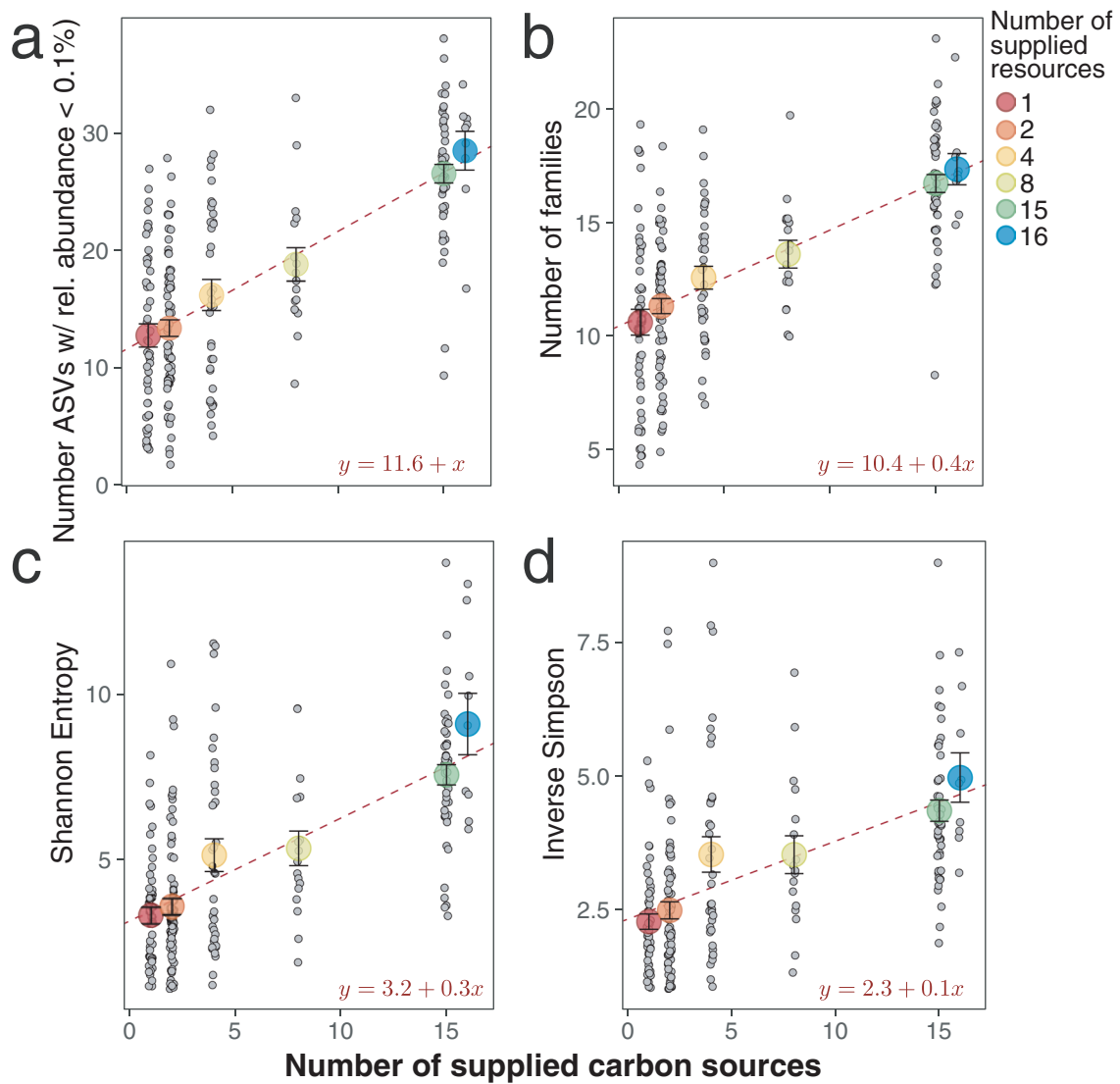


**Extended Data Fig. 4 | Richness of two-resource communities is approximately the average richness of constituent single-resource communities.**

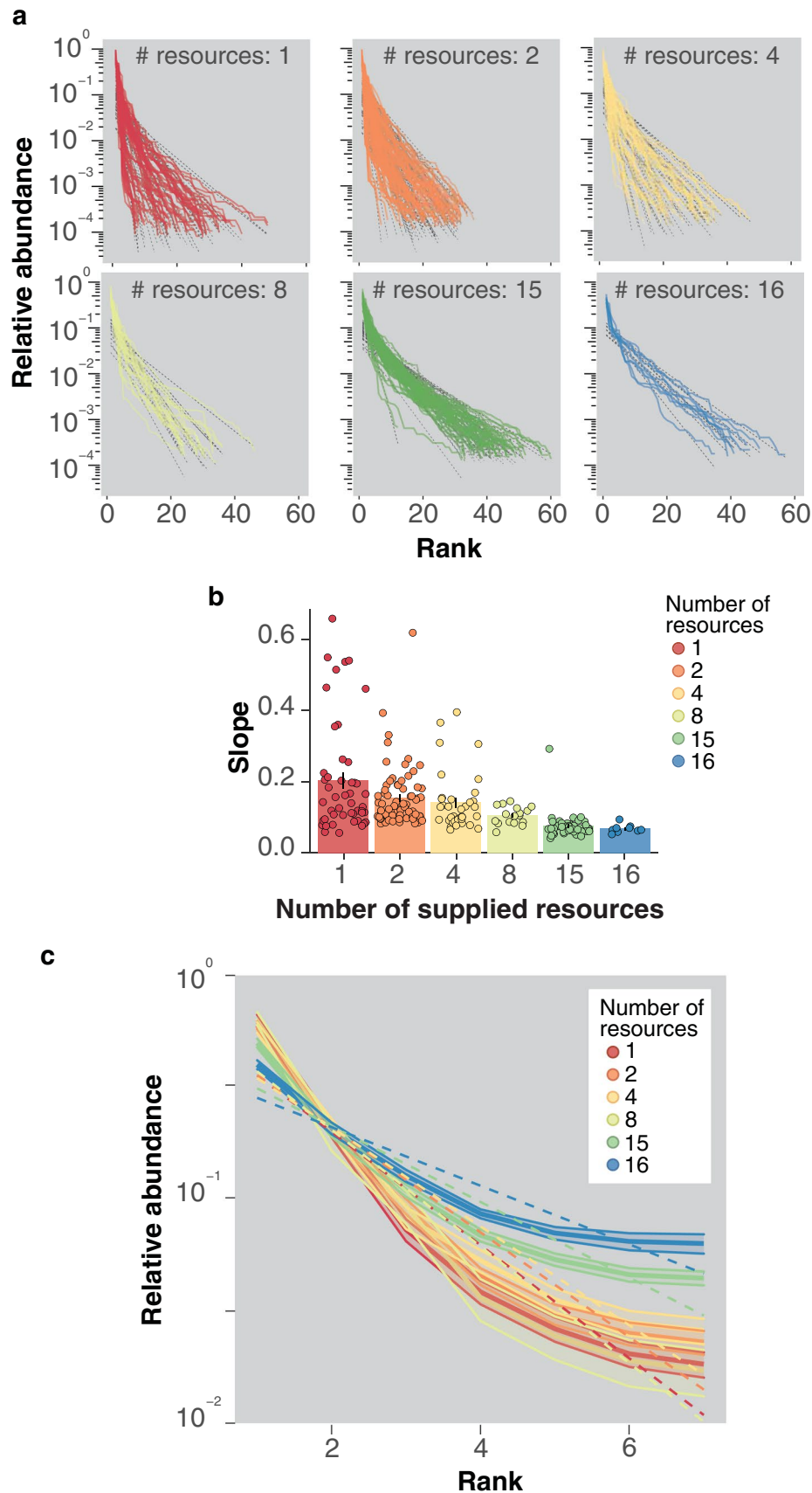
**a.** Observed richness of each two-resource community is best approximated by the average richness of constituent single resources, compared to the maximum and the union. Both the average error for the three predictions, calculated from the absolute values of predicted minus observed richness, and the ratio between predicted and observed richness are shown. **b.** Average richness of two-resource communities does not differ from the average richness of single-resource communities (boxplots with median and 95 % confidence interval; number of 1-resource media = 16, number of 2-resource media = 24; each dot is obtained from the mean of 3 replicates, s.e.m. are not shown for clarity). The color of the dots indicates the supporting resource(s).



**Extended Data Fig. 5 | The estimated number of metabolites for combinations of carbon sources increases fast and tends to saturate with the number of supplied resources.** The number of metabolites has been computed from KEGG and MetaCyc databases (see Methods). Large colored dots indicate the average number of metabolites for each number of supplied resources (mean  $\pm$  s.e.m.) while small grey dots indicate the average richness in each media containing a combination of resources (16 for single-resource, 24 for two-resource, 12 for four-resource, six for eight-resource, 16 for 15-resource and one for 16-resource combinations). Error bars are omitted for clarity. The dotted line was obtained by fitting a spline to the points.



**Extended Data Fig. 6 | The observed linear trend is sufficiently robust to the exclusion of low-abundance ASVs, coarse-graining at the family level, and the index used to measure microbial community diversity. a.** Richness was calculated as the number of ASVs after the exclusion of those with relative abundance lower than 0.1%. **b.** Richness was calculated as the number of unique families in the media. **c, d.** The increase in diversity, measured as Shannon Entropy and Inverse Simpson Index, with the number of carbon sources can still be approximated by a line. These indices give progressively more weight to abundant species, accounting, in this way, for the evenness of the communities. In each panel, large colored dots indicate the mean  $\pm$  s.e.m. while small grey dots indicate the average richness in each media containing a combination of resources (16 for single-resource, 24 for two-resource, 12 for four-resource, six for eight-resource, 16 for 15-resource and one for 16-resource combinations). Error bars are omitted for clarity.



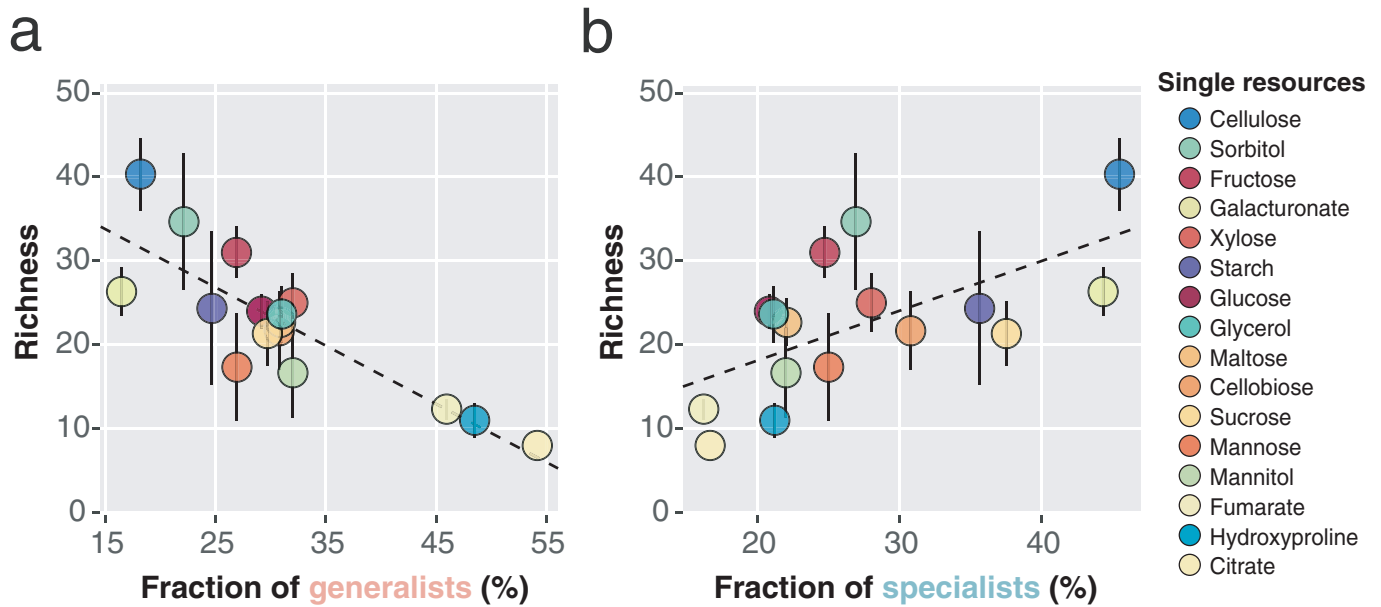
Extended Data Fig. 7 | See next page for caption.



**Extended Data Fig. 7 | The evenness of the microbial communities increases with the number of supplied carbon sources.** **a.** Log-linear rank abundance distributions (RADs) are shown for all the experimental microbial microcosms (48 for single-resource, 72 for two-resource, 36 for four-resource, 18 for eight-resource, 16 for 15-resource and nine for 16-resource combinations) together with the fitted regression lines (black dashed lines). Going from one to 16 resources, RADs exhibit heavier tails. **b.** The average absolute value of the slope (bars indicate mean  $\pm$  s.e.m. across replicates with the same number of supplied resources, while jittered dots represent the slope for each individual replicate) decreases with the number of supplied resources. **c.** Changes in evenness are independent from changes in richness, as revealed by RADs normalized for richness (mean RADs, dashed colored lines,  $\pm$  SD, shaded colored ribbon, for each number of supplied resources).

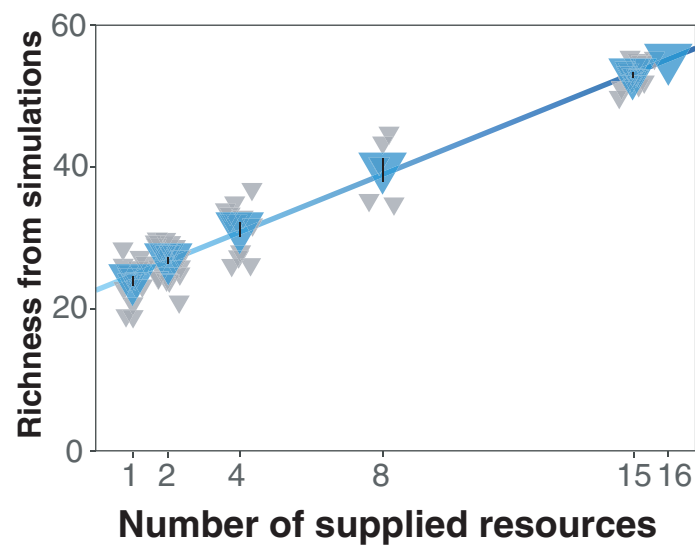


**Extended Data Fig. 8 | Resource occupancy of the 275 ASVs found in media containing a single carbon source.** The histogram shows the number of single resources in which each ASV is found. Bars are colored depending on whether the ASV has been classified as a generalist (pink), a specialist (teal) or an intermediate (beige). The families to which ASVs belong are reported.



**Extended Data Fig. 9 | The fraction of habitat generalist decreases while the fraction of habitat specialists increases with community richness.** A. Percentage fraction of habitat generalists (mean,  $N=3$ ) as a function of community richness (mean  $\pm$  s.e.m.,  $N=3$ ). B. Percentage fraction of habitat specialists (mean,  $N=3$ ) as a function of community richness (mean  $\pm$  s.e.m.,  $N=3$ ). Fitted linear regression lines (black, dashed) are shown.





**Extended Data Fig. 10 | Richness from simulations grows modestly with the number of resources.** Richness obtained from simulations of the consumer-resource model with cross-feeding (blue triangles, mean  $\pm$  s.e.m.,  $N=16$  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 available resources (solid blue line, intercept = 22.7, slope = 2). Grey jittered triangles indicate the richness of communities grown on a particular resource combination.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	16S amplicon sequencing raw data was processed with dada2 package in R. Taxonomic identities were assigned to ASVs using the SILVA version 132 database. The phylogenetic tree was reconstructed using Randomized Accelerated Maximum Likelihood (RAxML) using default parameters. The number and identity of metabolites for each combination of resources were extracted from KEGG database and manually curated using MetaCyc database. The metabolic gene content was estimated from 16S data using PICRUSt2 with default parameters. The number of copies of the rRNA operon for each genus from obtained from rrnDB database. Details and references are provided in the Methods section.
Data analysis	Analysis were conducted in R version 3.6.1 and Python version 3.7+. Details on specific packages and libraries are provided in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability

16S Amplicon sequencing data and metadata files have been deposited in the NCBI SRA database under NCBI BioProject ID PRJNA715195.

## Code availability

Data files and analysis/simulation codes to reproduce all figures are available at <https://github.com/mdalbello/Resource-diversity-relationship>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No calculation was performed to decide the sample size. However, the experiment has a tournament structure in which we randomly grouped the 16 selected resources across media with the same total number of resources, from 1 to 16 (experimental conditions). In total we had 75 conditions (16 single-resource, 24 two-resource, 12 four-resource, 6 eight-resource, 16 fifteen-resource and 1 sixteen-resource media - see Fig. S2). This ensured that we had at least 5 conditions for each number of resources.
Data exclusions	We excluded few replicates (4) for which read coverage was too low. Those replicates were flagged also by the company that performed both library preparation and sequencing.
Replication	All the experimental conditions (combinations of resources) were in triplicate (excluded replicates belonged to different conditions so that we never had less than 2 replicates per condition). In the condition with 16 resources we had 9 replicates.
Randomization	All experimental cultures were inoculated with the same starting bacterial suspension and differed in the resources provided in the growth media.
Blinding	All data has been obtained from 16S sequencing and therefore did not require blinding.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging