The effects of environmental complexity on microbial community yield and structure

Alan R. Pacheco[1] and Daniel Segrè*[1,2]

[1] Graduate Program in Bioinformatics and Biological Design Center, Boston University, Boston, MA 02215, USA

[2] Department of Biology, Department of Biomedical Engineering and Department of Physics, Boston University, Boston, MA 02215, USA

* To whom correspondence should be addressed: dsegre@bu.edu

Abstract

In nature, microbial communities are exposed to a myriad of growth-supporting substrates. While different environmental molecules can modulate the productivity and taxonomic structure of a community, only a few conflicting reports exist on how these properties depend on environmental complexity (i.e. the number of different available substrates). Here, we combine high-throughput experiments and mathematical modeling to quantify the effects of increasing environmental complexity on synthetic bacterial communities containing up to 13 species. By systematically culturing these consortia in combinatorial assortments of 32 carbon sources, we generated a mapping of how these assortments contribute to overall community growth and patterns of coexistence and diversity. We found that increased environmental complexity promotes community yield under some circumstances, though it does not necessarily lead to greater average taxonomic diversity. In fact, diversity appears to counterintuitively saturate to low values as environmental complexity increases, partially due to a nonlinear behavior akin to diminishing returns epistasis. By complementing these experiments with stoichiometric and consumer resource models, we discovered that this low diversity is consistently induced by the uneven distribution of organism-specific resource utilization capabilities. These results hint at new principles that connect environmental complexity to ecosystem properties, with implications for understanding microbiomes and engineering synthetic communities.

Keywords

Environmental complexity, synthetic microbial communities, metabolic interactions, biodiversity, microbial ecology, cross-feeding

Introduction

Microbial communities form the basis for an enormous range of biological processes, from cycling of nutrients in the ocean to regulation of human health[1–5]. In nature, these communities are exposed to a multitude of different nutrients that collectively shape their structure and function[6,7]. Despite our growing knowledge of community compositions in various biomes[8–10] and of the role of microbial diversity in global biogeochemical processes[11–13], relatively little is known about how the nutrient complexity of an environment modulates community properties. In the gut microbiome, for example, recent work has highlighted how community composition can depend on the diversity of available substrates[14–16]. However, conflicting reports exist on how this environmental complexity modulates growth yields and taxonomic diversity[17–19], raising questions as to how these relationships vary within and across communities, and hindering nascent efforts to engineer microbiomes with defined functions[20–22].

Classical ecological theories based on competitive exclusion and niche partitioning suggest that there would be greater opportunities for biodiversity in environments with a greater breadth of nutrient types[23,24]. Though this is an intuitive hypothesis, factors such as organism-specific nutrient use capabilities[25,26] and the effects of interspecies interactions[27–29] can lead to significant deviations from this expectation. Additionally, previous computational[30,31] and experimental[32,33] work has shed light on how the presence of multiple nutrients can jointly increase the potential for growth in a variety of microbial systems, though their effects on biodiversity were less clear. Together, these studies suggest that the relationship between environmental complexity and

community phenotype is highly consequential, but not straightforward, motivating the need for systematic experiments in well-controlled conditions.

In this work, we study the effect of increasing environmental complexity on the yield, growth rate, and taxonomic diversity of synthetic bacterial communities. By culturing multispecies consortia of varying sizes in well-defined environments containing up to 32 unique carbon sources, we quantify the degree to which the diversity of available nutrients impacts these community properties. We also show how specific environments can reproducibly generate communities with compositions that cannot be inferred in a straightforward way from those in simpler environments. Lastly, we couple our experimental data with dynamical and stoichiometric models to mechanistically quantify the role of nutrient utilization preferences and metabolic exchange on community structure.

Results

Assembly of multispecies communities and combinatorial nutrient conditions

To test the effects of environmental complexity on community phenotype, we employed a bottom-up approach for assembling microbial consortia and combinatorial nutrient conditions (Figure 1). These microcosms, designed with varying degrees of initial taxonomic and environmental complexity, were assembled from an initial set of 15 bacterial organisms (Supplementary Table 1, Supplementary Figure 1a,b) and 95 carbon sources (Supplementary Figure 1c, Supplementary Table 2). We first profiled the nutrient utilization preferences of our organisms using a plate-based phenotypic assay (Methods), which revealed considerable variability in the distribution of growth yields achieved on single carbon sources (Supplementary Figure 2). A subset of 32 carbon sources, chosen to maximize taxonomic variability across different nutrient combinations, was selected for all subsequent experiments (Methods). Microcosms were assembled by combining organism cultures at equal ratios into 96-deep well plates containing combinations of these 32 limiting nutrients (Supplementary Table 3, Supplementary Table 4). Each environment was set to contain the same total amount of carbon (Figure 1b), so that the total biomass produced (yield) could be compared across experiments. These communities were grown and diluted into fresh nutrient media at rates informed by pilot time-course experiments (Supplementary Figure 3, Supplementary Figure 4b) in order to maximize the chance that each community had reached a stable taxonomic composition and had consumed the provided carbon sources. Community yields were quantified by measuring absorbance at OD600, and composition was measured by plate counting or 16s amplicon sequencing depending on the number of organisms involved (Methods). In total, variants of this procedure were applied to generate 266 unique community-environment pairings (Supplementary Table 5).

A null model for community yield under increased environmental complexity

To establish an expectation of how the yield of a multispecies community might change with increased environmental complexity, we used a consumer resource model based on a previously-established computational framework[34]. This modeling method simulates the abundances of each organism in a community as a function of resource availability and each organism's ability to metabolize each resource (Methods), enabling us to explore a wide space of theoretical organism metabolic preferences and efficiencies. We modeled *in silico* communities made up of 10 organisms with randomly-sampled nutrient preferences in environments containing up to 256 carbon sources at equal ratios, for a total of 512 habitats.

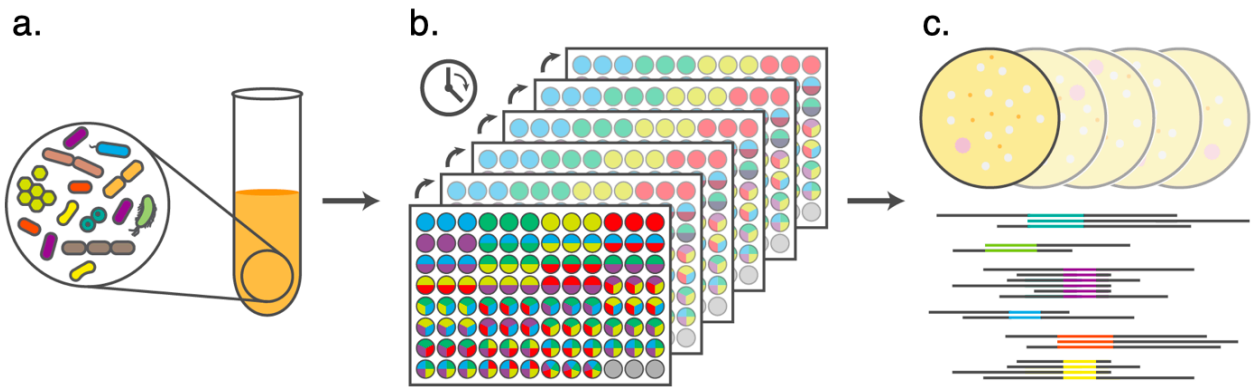


Figure 1. Experimental schematic for testing microbial community responses to environmental complexity. **(a).** Communities were assembled by combining a defined number of organisms at equal ratios. **(b).** These mixed cultures were then inoculated into deep-well plates containing a minimal medium plus equimolar combinations of up to 32 carbon sources. The communities were either grown in batch or diluted into fresh media over the course of several days. **(c).** The biomass yields of the communities were then assessed using a spectrophotometer and their composition was determined using either agar plating or 16s sequencing (Methods, Supplementary Table 5).

Our modeling indicated that the mean final yields of the simulated communities did not change significantly with increasing environmental complexity (Supplementary Figure 5a). To quantify possible deviations from this overall effect, we established a metric based on the differences in community growth between combined- and single-nutrient conditions. This metric is defined as the community yield under a combination of n nutrients $\{j_1, j_2, \dots, j_n\}$ minus the scaled sum of community yields under each of the corresponding individual nutrients j_i:

$$\partial_{\{j_1,j_2,\dots,j_n\}} = Biomass_{\{j_1,j_2,\dots,j_n\}} - \frac{1}{n}\sum_{i}^{n} Biomass_{\{j_i\}}$$

This metric (hereafter abbreviated as ∂_n for any given combination of n nutrients) quantifies how the yield under any nutrient combination deviates from the expected yield on its corresponding individual nutrients, similar to previous analyses of nonlinearities (epistasis) in genetic[35] or environmental[31] perturbations. Though our simulations revealed a number of individual nutrient combinations that deviate from this expectation (Supplementary Figure 6), the model did not predict a statistically-significant increase in yields to be sustained with increased environmental complexity (Supplementary Figure 5b).

Measuring synthetic community yield as a function of environmental complexity

Having established a basic computational model and expectation for community-level biomass yield, we experimentally measured the yields of synthetic microbial communities at increasing levels of environmental complexity. We first assayed simple three- and four-species consortia consisting of *Bacillus subtilis*, *Methylobacterium extorquens*, and *Shewanella oneidensis* (hereafter referred to as com3) as well as *Pseudomonas aeruginosa* (com4) in combinations of up to 32 carbon sources (Methods, Supplementary Table 5). When plotting the yield differences ∂_n against environmental complexity at the end of the growth passages, we observed an increasing curve that appeared to plateau between OD600 0.1-0.15 (Figure 2a,b). These curves differed substantially from our modeling predictions, as the yield differences ∂_n for each level of

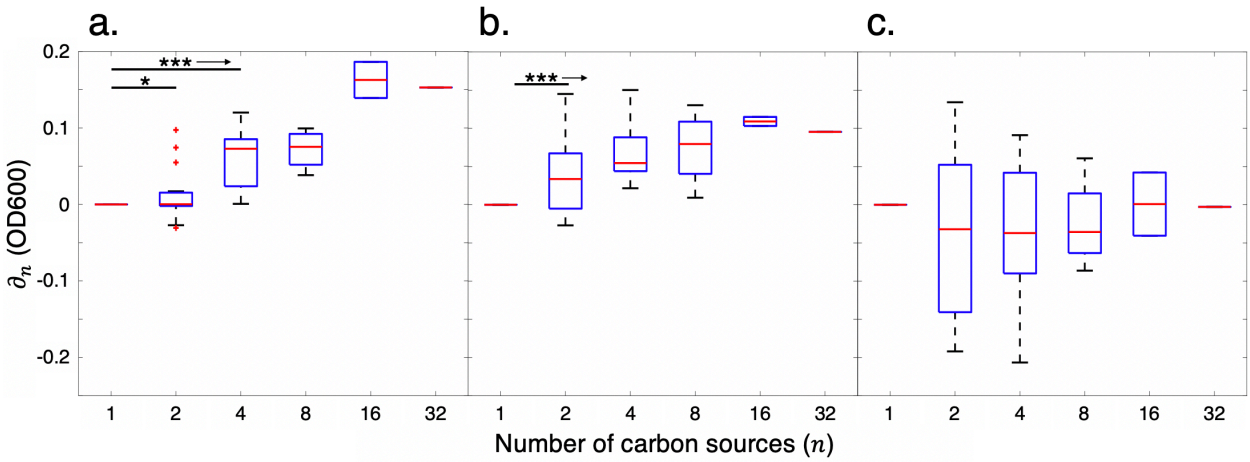


Figure 2. Changes in community yield in response to environmental complexity. **(a-c).** Endpoint differences ∂_n between community yields on n carbon sources and the scaled sum of community yields on the corresponding single carbon sources for **(a)** com3, **(b)** com4, and **(c)** com13. Here, as with all subsequent boxplots, the central mark indicates the median, the top and bottom box edges indicate the 25th and 75th percentiles, respectively, the whiskers extend to the most extreme points not considered outliers, and the red '+' symbols indicate outliers plotted individually. Significance levels are calculated between values for single- and multiple-carbon source conditions and are indicated by (*) $p < 0.05$, (**) $p < 0.01$, and (***) $p < 0.001$. Arrows indicate maintenance of highest significance level for all subsequent relationships relative to $n = 1$.

environmental complexity were significantly higher than in the single-nutrient cases. We nonetheless observed that this trend was less pronounced for com4, perhaps due to the fact that this consortium can grow on many more single-nutrient conditions owing to the generalist lifestyle of *P. aeruginosa* (Supplementary Figure 7, Supplementary Figure 8a,b). This result prompted us to explore whether single organisms could exhibit similar trajectories of ∂_n in monocultures. We explicitly tested this possibility by individually culturing each of the com4 organisms in the same 32-nutrient combination scheme for 48 hours, observing that all four displayed significantly higher than expected yields under increasing environmental complexity (Supplementary Figure 9).

As the above pattern deviated from our model's predictions, we reasoned that a better approximation of the statistical nature of our model would require repeating this experiment in a larger community. This community, consisting of 13 organisms and denoted com13, was cultured in the same manner as com3 and com4 (Supplementary Table 5) and represented a broader sampling of nutrient use capabilities. At the end of the experiment, the trajectory of ∂_n for com13 did not show any statistically-significant average increases as a function of environmental complexity (Figure 2c), consistent with model predictions. Interestingly, however, before reaching this point (at the 48-hour mark), the community displayed significantly increasing ∂_n values with higher environmental complexity, similar to the ones observed for com3 and com4 (Supplementary Figure 10). In spite of the flat distribution of the average ∂_n for com13 at the endpoint, however, individual nutrient combinations did give rise to communities with highly variable yields and significant deviations from $\partial_n = 0$ (Supplementary Figure 7a, Supplementary Figure 8c). One can see, for example, instances in which this community displayed lower than expected growth on pairs of nutrients relative to the single-nutrient conditions (Supplementary Figure 8d). In such cases, organisms might be displaying reduced efficiencies in using a resource in the presence of others, representing a type of 'resource interference' previously observed experimentally[32,36].

The average increases in ∂_n we observed in our smaller communities, as well as in com13 initially, may stem from multiple factors. One possible explanation is that, as the number of nutrients increases, more organisms can access their preferred carbon source while also having a reduced chance of competing for common resources. This effect may be enhanced by the presence of metabolic secretions, which would amplify the space of available nutrients and result in increased taxonomic diversity. Though we observed diauxic shifts as possible evidence of byproduct consumption in single-nutrient batch culture community experiments (com14, Supplementary Table 5, Supplementary Figure 3a, Supplementary Figure 11a-c), computational stoichiometric models indicated that the number of secreted metabolites quickly reach a plateau as the number of resources increases (Methods, Supplementary Figure 12). Furthermore, as shown later, greater taxonomic diversity was not observed at increasing levels of environmental complexity, suggesting that the trend in yield differences we observed was not solely due to an increase in provided or secreted molecules.

A second possibility is that there are basic nonlinear effects in the ways bacteria transform nutrients into biomass. In our experiments, upon increasing the number of carbon sources, we proportionately decrease their concentration in order to maintain a fixed total amount of carbon (Figure 1b). We would therefore expect the observed increasing trend in ∂_n if bacteria grew less efficiently on increasing amounts of an individual carbon source. While classical work points to final yields that are proportional to initial amounts of carbon[37,38], we identified clear deviations from this linearity for communities grown on different concentrations of individual carbon sources (Supplementary Figure 11d,e). In addition to this possibility, it may be that combinations of nutrients produce nonlinear synergistic effects within organisms' metabolic networks, in line with previous reports of enhanced growth potential in more complex media[31,32]. We found that all four of our organisms experienced cases of higher than expected yields on double vs. single nutrients (Supplementary Figure 13), which may reflect capabilities of the organisms to more efficiently fulfill metabolic requirements using multiple nutrients. Our observation parallels previous instances of this effect in microbial growth rates[33,39,40], which we also corroborated using a time-course experiment (Supplementary Figure 14). Thus, it seems plausible that the increases in ∂_n we observed with increasing environmental complexity are at least in part due to the nonlinear nature of metabolism itself, along with ecological effects.

Communities display deterministic compositions across all levels of resource complexity

Having examined the effects of environmental complexity on community growth, we sought to analyze its impact on taxonomic composition. We first used 16s amplicon sequencing to measure the endpoint taxonomic distributions of two 13-species communities (com13, com13a, Supplementary Table 5) in combinations of up to 32 and 5 carbon sources, respectively (Methods). Both communities displayed considerable variation in taxonomic composition across different nutrient conditions (Figure 3a, Supplementary Figure 15). Since all environmental conditions were tested in biological triplicate, we quantified the variability between replicates to estimate the degree of determinism involved in the assembly of our communities. We found that the vast majority of inter-replicate coefficients of variation fell under 1 (93.7% and 96.8% for com13 and com13a, respectively) (Supplementary Figure 16), and we encountered general consistency in community composition between comparable conditions of the two experiments (Supplementary Figure 17). These similarities suggest that the assembly patterns of these communities are largely deterministic based on environmental composition, even in relatively complex environments.

Resource competition drives community taxonomy, even in more complex environments
We next aimed to identify possible sources of the compositional variability we observed across environments. Though the overall distributions of species abundances were similar to those found in natural communities[41–43] (Supplementary Figure 18), we noticed that very few organisms out of the original 13 remained in our communities at the endpoint (Figure 3a, Supplementary Figure 15). We therefore sought to determine whether these losses in diversity were driven by basic niche selection or interspecies competition. Towards this goal, we first examined only the single-nutrient environments: we compared the organisms that persisted from the initial community in these conditions to those that grew individually in our initial phenotypic assay. We found that a majority of organisms did not appear in the communities despite being able to metabolize a substrate in monoculture (Figure 3b). This discrepancy was particularly striking in the D-glucose condition, in which, despite all but one organism being able to metabolize this carbon source, only *P. aeruginosa* remained. In addition, organisms that displayed a wide range of metabolic capabilities in monoculture, such as *B. licheniformis* and *V. natriegens*, were not observed at all in the single-nutrient communities, suggesting that the structure of our communities was largely driven by competition. In addition, we found generally poor correlations between the yields reached by the organisms in monoculture and in com13 (Supplementary Figure 19), indicating that monoculture resource utilization patterns are not necessarily predictive of how an organism will behave in a community.

Despite the apparent prevalence of interspecies competition, there were rare instances in which organisms that were not able to metabolize nutrients by themselves in the phenotypic assay did survive in the same individual nutrients in a community context (Figure 3b). An example of this effect was the community grown on D-ribose, which contained *P. putida* despite its inability to utilize it (Supplementary Figure 2a). We hypothesized that utilization of D-ribose by the other community members may be yielding metabolic byproducts that enable the survival of *P. putida*. A possible explanation is that *P. fluorescens*, which is present in high abundance in this community, secretes organic acids after catabolizing D-ribose into glyceraldehyde 3-phosphate via the pentose phosphate pathway. Indeed, strains of *P. fluorescens* that share the relevant genes with our strain[44] have been shown to engage in these metabolic transformations[45,46]. Such byproduct exchanges may be present in the remaining examples containing unexpected organisms, especially given that all of these cases contained at least two different bacterial species.

In order to extend the analysis of taxonomic diversity from single-nutrient to multi-nutrient conditions, we calculated the species richness S and Shannon entropy H values of our communities with respect to the number of provided carbon sources. In contrast to the intuitive expectation based on the availability of multiple niches, we found no statistically-significant relationships for the average of either quantity as a function of environmental complexity (Figure 3c,d). In fact, while we did not observe any instances of the lowest S and H values at the highest levels of environmental complexity, some single-carbon source conditions resulted in higher values than those of all other conditions. Additionally, the species co-occurrences we observed in single-nutrient environments were not preserved in more complex settings (Supplementary Figure 20). These results suggest that the number of substrates is not a key determining factor of taxonomic diversity or balance for our communities.

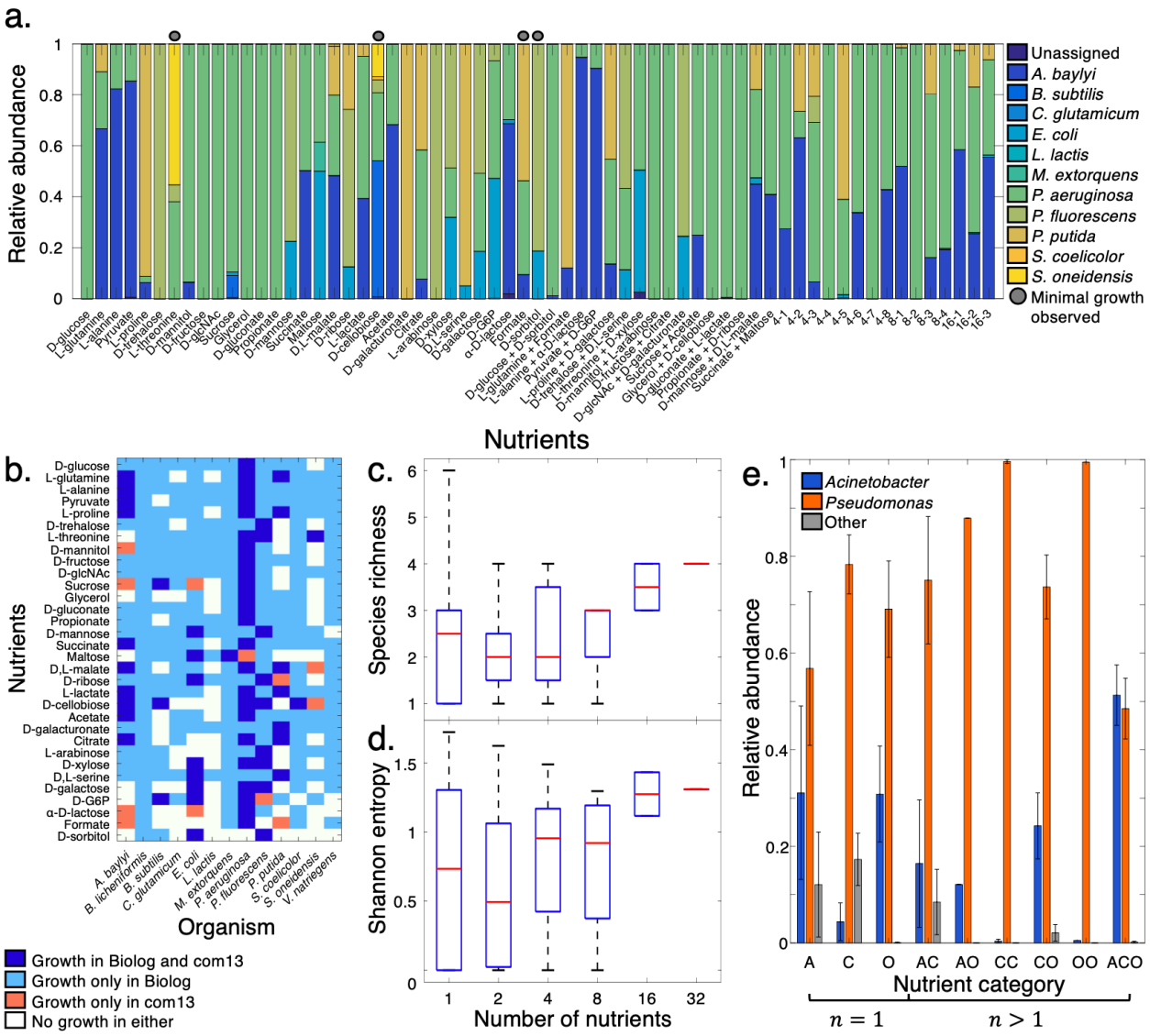


Figure 3. Endpoint taxonomic composition and species-specific growth profiles of 13-species community in 63 unique environmental conditions containing combinations of up to 32 carbon sources (com13). **(a).** Species relative abundances. Environmental compositions for conditions with more than two carbon sources are abbreviated, and are made up of conditions containing fewer carbon sources (e.g. condition 4-1 contains the carbon sources in the first two 2-carbon source conditions, 8-1 contains those in the first 4-carbon source conditions, etc.). Complete environmental compositions are outlined in Supplementary Table 3. Gray markers above selected conditions indicate low community growth. Initial compositions, as well as environment-specific distributions across all replicates, are found in Supplementary Figure 17a. **(b).** Species-specific differences in growth capabilities between monoculture (Biolog assay, Supplementary Figure 2a) and single-carbon source community contexts. **(c,d).** Community species richness **(c)** and Shannon entropy **(d)** against the number of provided nutrients. **(e).** Relative abundances of organism groupings (*Acinetobacter*, *Pseudomonas*, and 'Other') according to nutrient type (A: amino acid, C: carbohydrate, O: organic acid). Single nutrient conditions ($n = 1$) contain only one nutrient type, while multiple-nutrient conditions ($n > 1$) contain at least one of the nutrient types shown. For example, the L-glutamine condition would be categorized under 'A' for amino acid, while the L-glutamine + formate condition would be categorized as 'AO' as it contains both an amino acid and an organic acid. Nutrient-specific type designations are provided in Supplementary Table 2. Error bars indicate standard deviation.

In addition to analyzing global diversity patterns, our data allowed us to quantify possible nonlinearities in how two nutrient sets jointly affect ecosystem properties relative to expectations based on their individual effects. Such deviations, akin to the concept of epistasis in genetics[35], can be defined using simple scores for species richness and Shannon entropy (Methods). These scores, termed E_S and E_H, quantify the differences between the observed S or H values of the community on a particular environment AB and the expected value based on the corresponding lower-complexity conditions A and B. We found that the distributions for both scores were unimodal and highly skewed toward negative values, with the vast majority (90.3% for E_S and 83.9% for E_H) being less than or equal to zero (Supplementary Figure 21a,b). These distributions indicate that combinations of nutrient sets often yield communities that are less diverse than what would be expected by solely considering individual nutrient sets. Although these differences appeared to be less pronounced between environments with higher nutrient complexity (Supplementary Figure 21c,d), we found that species richness and Shannon entropy of lower-complexity environments were poorly indicative of these properties in more complex conditions.

Mapping nutrient identity to community diversity

The general trends described above do not provide biological insight into the specific role that individual carbon sources could play in fostering diversity. We therefore asked whether specific instances of competition and coexistence observed in our dataset could be associated with specific nutrient combinations. We began by performing a hierarchical clustering of all 63 nutrient combinations in com13 based on the mean organism abundances in each condition, which allowed us to identify patterns between environments spanning our entire dataset. In addition to revealing specific nutrient combinations that conferred similar structures (Supplementary Figure 22), we noticed that three distinct organism groupings emerged from this analysis, highlighting potentially important functional relationships. We termed these groups *Acinetobacter* (containing *A. baylyi*), *Pseudomonas* (containing *P. aeruginosa*, *P. fluorescens*, and *P. putida*, and "Other" (containing the remaining nine organisms), and quantified their relative abundances based on the nutrient types that their communities were grown on: amino acids, carbohydrates, and organic acids (Figure 3e). We noticed that, in communities supplied only with one carbohydrate, *Pseudomonas* organisms grew to high average relative abundances (78%) – an effect that was amplified in communities supplied with multiple carbohydrates. In communities cultured on either one amino acid or one organic acid, *Acinetobacter* was able to more effectively grow in the presence of *Pseudomonas* organisms, with relative abundances averaging 42.7% and 34.8%, respectively. Although *Acinetobacter* did not reach as high relative abundances in conditions with two different nutrient types, the most balanced community compositions were observed when all three nutrient types were combined.

Although non-*Acinetobacter* or *Pseudomonas* organisms were less uniformly prevalent, certain environments yielded more diverse taxonomic compositions. For example, the community cultured on D-cellobiose was composed of six different organisms with *B. subtilis* reaching the highest abundance (53.4%). Despite our observation that an organism's performance in monoculture is rarely correlated with its persistence in a community (Supplementary Figure 19), the ability of *B. subtilis* to efficiently metabolize D-cellobiose (Supplementary Figure 2a) may provide clues to its high abundance in this case. Specifically, *B. subtilis* has been documented to convert cellobiose to L-lactate[47,48]. Our initial phenotypic assays showed that *P. aeruginosa* and *S. oneidensis*, the next most prevalent organisms in this community, can utilize L-lactate, indicating the possibility of cross-feeding. Despite this apparent cooperation, *B. subtilis*, along with all other community members, was overtaken by *P. aeruginosa* when D-cellobiose was

combined with another carbon source, glycerol. The presence of this additional nutrient therefore removed the competitive advantage of *B. subtilis*, exemplifying the nonintuitive effects the addition of nutrients can have on the community composition.

Given these examples of how community composition can change with the addition of a new carbon source, we sought to determine in a more comprehensive and systematic way if individual nutrients were more broadly associated with taxonomic diversity. To do this, we generated a simple linear regression model that relates the presence of each carbon source to the Shannon entropy exhibited by the communities. We calculated regression coefficients for each nutrient, allowing us to estimate the contribution of each carbon source to taxonomic balance independent of the number of nutrients (Supplementary Table 6). We estimated D-galactose to be the most highly associated with greater community evenness, with a regression coefficient of 0.76. Indeed, the environment containing only D-galactose yielded a community with a relatively even composition of *E. coli*, *P. aeruginosa*, and *P. fluorescens*, while the community grown in the two-carbon-source condition with D-galactose also contained three organisms (*A. baylyi*, *P. aeruginosa*, and *P. putida*). Despite these relatively balanced communities at lower complexities, the presence of a nutrient significantly associated with lower Shannon entropy like L-arabinose can overpower the effects of nutrients like D-galactose. This effect is most clearly observed in the 8-carbon source condition containing both these nutrients, which resulted in the complete dominance of *P. aeruginosa*. Therefore, nutrients that disproportionately favor the growth of a single organism can eclipse the effects of those that promote greater taxonomic evenness in some environments.

Community diversity depends on the distribution of usable resources

The above results indicate that, surprisingly, increased environmental complexity does not necessarily result in increased ecosystem diversity. While specific taxonomic groupings and nutrient combinations seem to explain some individual observations, we hypothesized that a much broader and general principle may need to be invoked in order to explain the overall patterns we identified. Specifically, we sought to determine whether some properties of the nutrient use capabilities of the organisms in our communities could be accounting for the low degrees of taxonomic diversity we observed. We therefore first aimed at implementing detailed mathematical models to explain the taxonomic compositions of com3 and com4, which were strikingly low in biodiversity and most often dominated by the organisms with the broadest nutrient utilization capabilities (Supplementary Figure 23). We applied our consumer resource model to these consortia by explicitly parametrizing the nutrient use capabilities of the organisms with species-specific growth data (Supplementary Figure 7b). Moreover, we integrated the potential for metabolic cross-feeding of secreted metabolites[27,34] informed by flux-balance predictions of metabolic turnover (Methods, Figure 4a). This parametrization enabled us to make quantitative estimates of community growth trajectories and metabolic exchange (Supplementary Figure 24), and accurately predicted the dominance of *P. aeruginosa* – which had the broadest set of usable carbon sources – across most conditions in com4. However, our model could not fully explain the dominance of *S. oneidensis* in com3 or the increases in yield we observed experimentally (Supplementary Figure 25).

Despite some inaccuracies, our experimentally-parametrized model accurately recapitulated the low levels of taxonomic diversity we observed *in vitro*. We thus asked whether, beyond the specific details of our small consortia, this effect could be primarily explained by a community's distribution

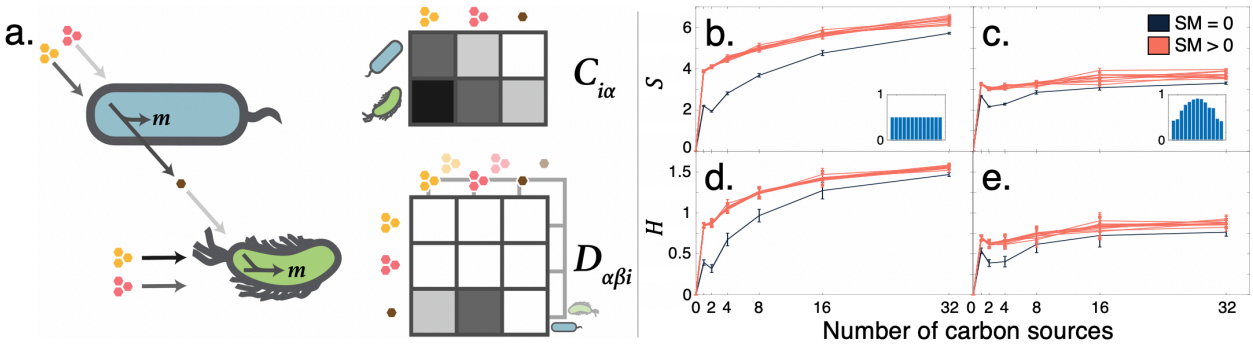


Figure 4. Consumer resource modeling framework and predictions. **(a).** Schematic of consumer resource model. Nutrients (pink and yellow) are utilized by organisms according to the stoichiometric matrix C and converted into biomass m. In this example, the bottom green organism is able to utilize both nutrients more efficiently than the top blue one (as denoted in the shades in the C matrix and of the nutrient-organism arrows. The blue organism converts both the pink and yellow nutrients into a brown metabolite according to the species-specific stoichiometric matrix D. This secreted metabolite cannot be consumed by the blue organism in this example, but it can be utilized by the green organism as indicated by the C matrix and the arrows. **(b-e).** Distributions of species richness S **(b,c)** and Shannon entropy H **(d,e)** for 13-species communities with uniform **(b,d)** and experimentally-derived **(c,e)** nutrient utilization preferences. Each line represents the trajectory of species richness or Shannon entropy relative to the number of carbon sources supplied for a certain number (0-10) of secreted metabolites (SM) made available. Insets represent species-specific nutrient utilization probabilities. Error bars are s.e.m.

of the number of usable resources. To do this, we generated simulated 4-species communities whose constituent organisms had either uniform or uneven nutrient use capabilities determined via random sampling (Methods). These models showed that the presence of a generalist organism robustly decreased community diversity, in a way that strongly depended on the presence (but not necessarily quantity) of available secreted metabolites (Supplementary Figure 26a-d). We hypothesized that this effect could also be responsible for the limited diversity observed experimentally in com13, given that its constituent organisms displayed an uneven distribution of nutrient use capabilities. We thus used our consumer resource model to simulate the growth of two sets of 13-species communities: one set in which each organism had the same probability of consuming a provided nutrient, and one in which this probability was based on the experimentally-measured proportion of nutrients consumed by each com13 organism in our initial phenotypic assay (Supplementary Figure 2a).

We found that, while species richness and Shannon entropy were predicted to markedly increase for communities with uniform nutrient use distributions (reaching a maximum of ~7 coexisting species), they remained relatively flat in communities with experimentally-derived nutrient use distributions (reaching a maximum of ~3 coexisting species, a value very similar to experimental observations (Figure 3c)) (Figure 4b-e). These results suggest that reduced taxonomic diversity in complex environments could be the outcome of competition in communities with uneven metabolic capabilities, which can be slightly modulated by the availability of secreted byproducts (Supplementary Figure 26e-h).

Discussion

Given the chemical and taxonomic complexity of natural ecosystems, it can be challenging to disentangle the effect of multiple substrates on microbial community phenotypes. By

systematically assembling synthetic microcosms, we found that greater environmental complexity could lead to increased average yields in small communities, and – transiently – in larger ones. Except for a few specific nutrient combinations, however, greater environmental complexity did not lead to significantly higher degrees of average taxonomic diversity. We found evidence that this effect can be caused by uneven distributions of metabolic capabilities across a community's constituent organisms, providing further support that appropriate mathematical models can be crucial to understanding ecological patterns across environments[34]. Future work in synthetic ecology could specifically address the feasibility of generating stable, highly diverse engineered communities by combining organisms with similar distributions of metabolic capabilities. Furthermore, given the nonlinear effects of combining different sets of carbon sources, we envisage that the design of synthetic communities with defined properties will require careful mappings of nutrient compositions to community phenotypes.

Our observation that broader taxonomic classes display coherent dynamics in an environment-dependent manner (Figure 3e) is consistent with previous observations of natural communities in single carbon sources[27]. However, additional attributes not explicitly monitored or modeled in our study could have important effects on yield and diversity. These attributes, including environmental pH[49,50], spatial structure[51,52], and different resource concentrations[53,54], could be explicitly addressed in future extensions. Given these limitations, however, it remains to be seen whether our results will hold in natural environments and more complex communities. Interestingly, although our most complex consortia were still very small compared to natural microbiota, their species abundance distributions mirrored those of aquatic and terrestrial ecosystems (Supplementary Figure 18)[41–43,55–57], suggesting that even small, engineered consortia such as ours tend to converge to naturally-occurring compositional structures.

Overall, our study suggests that even in well-defined communities and environmental conditions, the dependence of diversity on environmental complexity is not intuitive, making it difficult to assume general relationships for large natural microbiomes. Nonetheless, our results suggest that it may be possible to distill general principles or phenomenological laws that could help understand how microbial communities behave under molecularly complex environments. While our increase of environmental complexity at constant total carbon does not mimic any real disturbance in microbial ecosystems, it suggests that chemical diversity may have nonlinear, but quantifiable consequences on ecosystem diversity and yield, alongside the effects of changes in total nutrients. We also envisage that phenomenological relationships similar to the ones reported here could inform multiscale ecosystem and biogeochemical cycle models[12,58,59], e.g. by parametrizing microbial community properties as a function of environmental complexity.

Methods

Selection and initial metabolic profiling of organisms

In order to maximize the chance of obtaining communities with diverse taxonomic profiles from different environmental compositions, the organisms selected were drawn from a number of bacterial taxa known to employ varying metabolic strategies. In addition, given the growing relevance of synthetic microbial communities to industrial and biotechnological applications[60–63], we chose to employ bacterial species that have previously been used as model organisms and have well-characterized metabolic capabilities. This criterion, paired with the availability of flux-balance models associated with a majority of these organisms, allows us to explore the metabolic

mechanisms observed in our various experimental conditions with higher confidence. These selection principles resulted in a set of 15 candidate bacterial organisms spanning three bacterial phyla (Actinobacteria, Firmicutes, and Proteobacteria) (Supplementary Table 1, Supplementary Figure 1a).

A microtiter plate-based phenotypic assay was used to assess the metabolic capabilities of each of the 15 candidate organisms. Each organism, stored in glycerol at -80˚C, was initially grown in 3mL of Miller's LB broth (Sigma-Aldrich, St. Louis, MO) for 18 hours with shaking at 300 rpm at each organism's recommended culturing temperature (Supplementary Table 1). To maximize oxygenation of the cultures and prevent biofilm formation, culture tubes were angled at 45˚ during this initial growth phase. Candidate organism *Streptococcus thermophilus* was found to have produced too little biomass in this time period and was grown for an additional 8 hours. Each culture was then separately washed three times by centrifuging at 8000 rpm for 2 minutes, removing the supernatant, suspending the pellet in 1mL of M9 minimal medium with no carbon source, and vortexing or triturating to homogenize. The cultures were then diluted to OD600 0.5 ± 0.1 as read by a microplate reader (BioTek Instruments, Winooski, VT) and distributed into each well of three PM1 Phenotype MicroArray Plates (Biolog Inc., Hayward, CA) per organism at final OD600 of 0.05. The carbon sources in the PM1 plates (Supplementary Table 2, Supplementary Figure 1b) were resuspended in 150 μl of M9 minimal media prepared from autoclaved M9 salts (BD, Franklin Lakes, NJ) and filter-sterilized $MgSO_4$ and $CaCl_2$ prior to inoculation. The cultures in each PM1 plate were incubated at each organism's recommended culturing temperature with shaking at 300rpm for 48 hours. After this growing period, the OD600 of each culture was measured by a microplate reader to quantify growth. To account for evaporation in the outer wells of the plates, which could yield in inflated OD readings, three 'evaporation control' plates with no carbon source were inoculated with bacteria at a final OD600 of 0.05 and incubated at 30˚C for 48 hours. The averaged OD600 readings of these plates were subtracted from the readings of the bacterial growth plates to correct for evaporation. A one-tailed t-test was performed using these corrected OD600 values to determine significance of growth above the value of the negative controls ($p < 0.05$). These final growth yields for the 15 candidate organisms are reported in Supplementary Figure 2a, and aggregated analyses of the growth profiles of the organisms are reported in Supplementary Figure 2b-d.

After this initial metabolic profiling, *Streptococcus thermophilus* was not included in any of the subsequent experiments as it displayed too low of a growth rate in the initial overnight growth phase and grew very minimally (no more than OD600 0.2) in fewer than 20% of the carbon sources in the PM1 plate. After inclusion in an initial mixed-culture experiment (com14, Supplementary Table 5), *Salmonella enterica* was also removed from future experiments due to its high levels of growth on all but one of the PM1 plate carbon sources. Its exclusion, meant to prevent its complete dominance in the subsequent mixed-culture experiments, resulted in a final set of 13 bacterial organisms.

For experiments involving a subset of the 13 organisms, the organisms were chosen to ensure they could be differentiated via agar plating. In the 3-species community experiment involving *E. coli*, *M. extorquens*, and *S. oneidensis* in combinations of 5 nutrients (com3a), the organisms were selected based on their easily differentiable colony morphologies (Supplementary Figure 27). In the second 3- and 4-species community experiments (*B. subtilis*, *M. extorquens*, *P. aeruginosa*, and *S. oneidensis* (com3 and com4)), selection was informed by differentiable colony morphology and additional metabolic criteria based on generalist-specialist relationships. Absolute growth yield on the Biolog PM1 plates was also considered, with the goal of including

both high- and low- yielding organisms. Therefore, *P. aeruginosa* (high-yield generalist), *B. subtilis* (low-yield specialist), *M. extorquens* (low-yield generalist), and *S. oneidensis* (high-yield specialist) were selected.

Selection of nutrients and assembly of combinatorial media

The nutrients used in all experiments were selected from the 95 carbon sources contained in the Biolog PM1 Phenotype MicroArray Plate. This plate contains a variety of molecule types such as mono- and disaccharides, sugar alcohols, organic acids, amino acids, and more complex polymers (Supplementary Table 2, Supplementary Figure 1b). Using the metabolic profiling experiments for each individual organism as a basis (Supplementary Figure 2a), different criteria were established to choose the carbon sources used in each experiment depending on the desired complexity of the environment. An overarching criterion was that each experiment contain at least one sugar, one organic acid, and one amino acid to increase the possibility of synergistic interactions between nutrients and nutrient use orthogonality between the organisms.

For the communities grown in 5 carbon sources (i.e. com3a and com13a in D-Glucose, pyruvate, D-glcNAc, L-proline, and L-threonine), the following criteria were applied: D-glucose was selected as it resulted in the highest yield of each of the individual organisms, pyruvate was an organic acid with relatively high yields, D-glcNAc was a more complex sugar that resulted in varying individual growth yields, and L-proline and L-threonine were amino acids that resulted in generally high and low individual species yields, respectively. Communities were grown in all combinations of these five carbon sources (5 conditions of 1 carbon source, 10 of 2, 10 of 3, 5 of 4, and 1 of 5) for a total of 31 unique environmental compositions (Supplementary Table 4).

The carbon sources for the 32-nutrient experiments were selected based on the following criteria: nutrients in which generalists individually displayed low levels of growth but favored at least one specialist (3 nutrients), nutrients that resulted in high-variance in growth yields across species (5 nutrients), and nutrients that resulted in low-variance in growth yields across species (7 nutrients). These criteria were meant to increase the probability of observing more taxonomically diverse communities. The remaining 21 nutrients were selected based on the total species-specific yields they conferred (Supplementary Figure 2a), with higher-yielding nutrients being prioritized. Communities were grown in selected combinations of these 32 carbon sources (32 conditions of 1 carbon source, 16 of 2, 8 of 4, 4 of 8, 2 of 16, and 1 of 32) for a total of 63 unique environmental compositions (Supplementary Table 3). The selected combinations were chosen based on the Biolog biomass yields of the 13 organisms under each carbon source, with the lowest-yielding nutrient (D-sorbitol) being paired with the highest (D-glucose) followed by the second-lowest and second-highest, etc.

Combinatorial media conditions were assembled by resuspending each carbon source in distilled water to stock concentrations of 1.25 mol C/L and filter sterilizing using 0.2 μm membrane filter units (Nalgene, Rochester, NY). A liquid handling system (Eppendorf, Hamburg, Germany) was used to distribute the individual carbon source stocks in the appropriate combinations in 96-well plates. The combinatorial nutrient stocks were then sterilized with filter plates (Pall Corporation, Port Washington, NY) via centrifugation at 4000rpm for 2 minutes. These were then combined with M9 minimal medium (containing M9 salts, $MgSO_4$, $CaCl_2$, and no carbon) and filter-sterilized water to final working concentrations of 50 mM C in 96 deep-well plates (USA Scientific, Ocala, FL) for a total volume of 300 μl. This working concentration was selected such that all organisms would not grow beyond the linear range of OD600 for biomass measurements.

Culturing and quantification of microbial communities in combinatorial media

All communities were assembled using a bottom-up approach, with each organism initially grown separately and diluted to the same starting concentrations before being combined in mixed cultures. For all experiments, each organism was inoculated from a glycerol stock stored at -80˚C into 3 mL of LB broth and incubated at 30˚C with shaking at 300 rpm for 18 hours. The culture tubes were angled at 45˚ to prevent biofilm formation and encourage oxygenation of the cultures. The overnight cultures were then washed three times by centrifuging at 8000 rpm for 2 minutes, resuspending in M9 medium without carbon, and vortexing and triturating if necessary to homogenize. The individual cultures were then diluted to OD600 of 0.5 ± 0.05, combined at equal proportions, and inoculated in biological triplicate into the combinatorial media plates at final concentrations of OD600 0.05 in 300 μl. Each plate additionally contained three control wells: one uninoculated well with 50 mM C of D-glucose to control for contamination, and two inoculated wells with no nutrient to assess the decay of the initial inocula. The communities were grown at 30˚C with shaking at 300 rpm for periods of 24 or 48 hours before each passage. At each passaging step, the cultures were triturated 10 times to ensure the communities were homogenized and 10 μl were transferred to 290 μl of fresh media for the subsequent growth period. Yields of the cultures were quantified by transferring 150 μl to clear 96-well plates (Corning, Corning, NY) and reading absorbance (OD600) using a microplate reader (BioTek-Synergy HTX). Biomass quantities are reported as the difference between the raw OD600 readings of each sample and the mean OD600 value of the negative control wells. A summary of the organisms, nutrients, and culturing conditions for each experiment is found in Supplementary Table 5.

Communities to be sequenced were centrifuged at 4000 rpm for 10 minutes and the supernatant was removed. Cell pellets were stored at -20˚C until DNA collection was performed using a 96-well genomic DNA purification kit (Invitrogen). To harvest the DNA, each cell pellet was resuspended in 180μl lysis buffer containing 25 mM Tris-HCl, 2.5 mM EDTA, 1% Triton X-100, and 20 mg/ml Lysozyme (Sigma-Aldrich). The samples were mixed by vortexing and incubated at 37˚C for 30 minutes, after which 20mg/ml of RNase A (Invitrogen, Carlsbad, CA) and 20 mg/ml of Proteinase K (Invitrogen) with PureLink Pro 96 Genomic Lysis/Binding Buffer (Invitrogen) were added. The samples were mixed by vortexing and centrifuged after each reagent was added. The samples were incubated at 55˚C for 30 minutes, after which 200 μl of 100% ethanol (Sigma-Aldrich) were added. DNA from the resulting lysates was purified using a vacuum manifold according to the purification kit protocol (Invitrogen). Purified DNA was normalized to 15 ng/μl using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA). Library preparation was performed based on a paired-end approach developed by Preheim *et al.*[64], which targets the V4 region of the 16s rRNA gene with the primers U515F (5'-GTGCCAGCMGCCGCGGTAA) and E786R (5'-GGACTACHVGGGTWTCTAAT). Libraries were sequenced using an Illumina MiSeq platform at either the MIT BioMicroCenter, Cambridge, MA (com13a) or at QuintaraBio, Boston, MA (com13). QIIME2[65] was used to demultiplex raw files and produce FASTQ files for forward and reverse reads for each sample. DADA2[66] was used to infer sequence variants from each sample and a naïve Bayes classifier was used to assign taxonomic identities using a custom reference database with a 95% confidence cutoff.

Communities to be assayed by agar plating were diluted by a factor of 10^4 and spread on LB agar plates using autoclaved glass beads. Plates were prepared by autoclaving and distributing 18 mL of LB agar (Sigma-Aldrich) into petri dishes using a peristaltic pump (TriTech, Los Angeles, CA). Inoculated plates were incubated at 30˚C and imaged after 72 hours using a flatbed scanner for

colony counting. Colony counts for com3 and com4 were adjusted based on a standard dilution of the community members at equal concentrations measured by OD.

Significance between growth yields under differing environments was determined using a one-sided two-sample t-test with significance cutoffs of 0.05, 0.01, and 0.001. For hierarchical clustering analysis of communities, Spearman correlation coefficients were computed either for pairs of environments or pairs of organisms based on normalized vectors of species abundances. Hierarchical clustering was performed on the correlation coefficients using the 'clustergram' function in MATLAB, which calculated distances between clusters using the UPGMA method based on Euclidean distance.

Computation of epistasis scores for species richness and Shannon entropy

In order to quantify how taxonomic diversity and balance could change in incrementally more complex environments, we first established definitions of expected values of species richness S and Shannon entropy H based on their values in lower-complexity conditions. Let a combined set of nutrients AB be defined as the union of nutrient sets A and B. For nutrient sets A, B, and AB, the vectors of species abundances in each set are defined as V_A, V_B, and V_{AB}, respectively. The species richness values S for each set are therefore simply the number of positive species abundance values in each vector. Based on the organisms that survived in sets A and B, we establish the naïve assumption that all those organisms will also survive in set AB. We therefore define $V_{expected,AB}$ as $V_A \cup V_B$, and $S_{expected,AB}$ as the number of organisms contained in this set. Our epistasis score for species richness $E_{S,AB}$ becomes:

$$E_{S,AB} = S_{AB} - S_{expected,AB}.$$

We use the same definition of expected species abundances to calculate Shannon entropy H. We first calculate the observed H for nutrient set AB as $H_{AB} = -\sum V_{AB} \log_2 V_{AB}$, and the expected H for nutrient set AB based on A and B as $H_{expected,AB} = -\sum V_{expected,AB} \log_2 V_{expected,AB}$. The epistasis score for Shannon entropy $E_{H,AB}$ is therefore:

$$E_{H,AB} = H_{AB} - H_{expected,AB}.$$

Flux balance modeling

To estimate the number of secreted and absorbed metabolites in com4, we used experimentally-validated genome-scale models for each of the four organisms[67–70]. Genome-scale models are mathematical representations of the known metabolic capacities of individual organisms[71]. They are constrained by known maximum metabolic fluxes v_{max} through internal and transport reactions, as well as by reaction stoichiometric constraints represented by a matrix S. Flux balance analysis (FBA), a mathematical optimization technique, can then be applied to the models in order to define the metabolic fluxes v within the organism's network that will maximize a particular objective, such as growth[72]. This technique allows us to interrogate the growth rate of organisms under specific environmental conditions, as well as rates of nutrient consumption and metabolite secretion.

Our application of FBA uses the COBRA toolbox[73] and is largely based off of an implementation used in a previous study[31]. Here, we employed FBA to simulate the growth of the four organisms in com4 in the 63 combinatorial medium conditions we tested experimentally. We first defined an *in silico* M9 minimal medium consisting of the various inorganic molecules present in the *in vitro*

minimal medium (Supplementary Table 7). These molecules were provided to the genome-scale models at nonlimiting availabilities by setting the corresponding maximum flux bounds v_{max} to 1000 mmol/gDW/hr. Depending on the environmental condition, we supplied each *in silico* organism with the appropriate carbon sources by setting the corresponding maximum flux bounds v_{max} to 10 mmol/gDW/hr. We then applied FBA by maximizing the growth rate and minimizing the sum of the fluxes in the network. This latter step was employed in order to more closely model proteome usage and minimize metabolite cycling throughout the network[74]. The optimization problem applied is therefore:

$$\begin{gathered}\min|v|,\\ \text{s.t.:}\\ S \cdot v = 0,\\ v_{min} \leq v \leq v_{max}.\end{gathered}$$

If an *in silico* organism grew on a given environmental condition, we recorded which organic metabolites were predicted to be taken up and secreted. These are summarized for all environments in Supplementary Table 8.

Consumer resource modeling

We employed a dynamical modeling framework to simulate the yields of arbitrary communities in increasingly complex environments, as well as the relative abundances of com4. The model is an adaptation of Robert MacArthur's consumer resource model[75–77] and a subsequent modification by Marsland *et al.*[34], which simulates the abundances of organisms over time as a function of resource availability, metabolic preferences, and exchange of secreted metabolites.

We define the individual species abundances as N_i for $i = 1, \ldots, S$, and the resource abundances as R_α for $\alpha = 1, \ldots M$. The key variable in calculating the abundances N_i is a stoichiometric resource utilization matrix $C_{i\alpha}$, which defines the uptake rate per unit concentration of resource α by organism i (Figure 4a). To calculate the growth of each organism on each resource, this matrix is multiplied by a Monod consumption term $R_\alpha/(k_{i,a} - R_\alpha)$ that simulates concentration-dependent resource depletion. Each consumed resource type α with abundance R_α is therefore consumed by organism i at a rate $C_{i\alpha}\, R_\alpha/(k_{i,a} - R_\alpha)$. These resources α are then transformed into other resources β by the organisms via a species-specific normalized stoichiometric matrix $D_{\alpha\beta i}$. A fraction l of the resultant metabolic flux is returned to the environment to be made available to other organisms, while the rest is used for growth. In addition to these resource consumption terms, the species abundances are also defined by (i) a species-specific conversion factor from energy uptake to growth rate g_i, (ii) a scaling term representing the energy content of each resource w_α , (iii) a quantity representing the minimal energy uptake required for maintenance of each species m_i, and (iv) a dilution rate d. These terms are further defined in Supplementary Table 7. Taken together, the species abundances N_i over time are defined by:

$$\frac{dN_i}{dt} = g_i N_i \left[\sum_\alpha w_\alpha (1 - l_\alpha) C_{i\alpha} \frac{R_\alpha}{k_{i,a} - R_\alpha} - m_i \right] - dN_i \tag{1}$$

The resource abundances R_α are first defined by a continuous external supply rate κ_α in addition to the constant dilution rate d (Supplementary Table 9). Each resource is then consumed, in a

manner dependent on the matrix $C_{i\alpha}$, and converted to other nutrients based on the stoichiometric matrix $D_{\alpha\beta i}$:

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - dR_\alpha - \sum_i C_{i\alpha} N_i \frac{R_\alpha}{k_{i,a} - R_\alpha} + \sum_{i,\beta} D_{\alpha\beta i} \frac{w_\beta}{w_\alpha} l_\beta C_{i\beta} N_i \frac{R_\beta}{k_{i,\beta} - R_\beta} \quad (2)$$

We selected the parameters for our equations based on experimental observations and quantities obtained from the literature. The dilution rate d was based on the 48-hour dilution timeframe we used in our experiments, in which 10 μl of culture was passaged into a total of 300 μl of fresh, uninoculated media. The resource input rate κ_α was defined similarly for all nutrients, equaling the initial concentration of each nutrient divided by a 48-hour dilution timeframe. Kinetic growth curves for com14 (Supplementary Table 5, Supplementary Figure 11a) were used to estimate the orders of magnitude for the remaining parameters, based on the community reaching an average of approximately 2.4×10^8 CFU/mL (OD600 0.3) within 20 hours in 50 mM C of D-glucose. The conversion factor from energy uptake to growth rate g_i, as well as the energy content of each resource w_α, were set to 1 and 1×10^8, respectively, in order to approximate this magnitude of growth. The Monod resource consumption half-velocity constant was set to 3000 g/mL for all nutrients in order to approximate the experimentally-observed growth timeframe. Lastly, the minimum energy requirements for all organisms m_i were informed by the community yields at steady state and the leakage fraction l_α was set to 0.8 based on community simulations in Marsland *et al.*[34]. These quantities are summarized in Supplementary Table 9.

In our initial CRM simulations, in which we generated a null model for community yield under increasing environmental complexity (Supplementary Figure 5), we simulated the growth phenotypes of a community containing $S = 10$ different arbitrary organisms. These organisms were defined by randomly-populated nutrient preference matrices C, in which each organism had a 25% chance of being able to utilize a particular nutrient. In this way, the organisms had comparatively even, though not necessarily intersecting, nutrient utilization profiles. Each community was simulated 50 times in each environment, so that the nutrient preference matrices could be randomly repopulated. This process allowed us to more effectively sample the large space of possible nutrient utilization matrices and obtain a clearer indication of how mean community yields changed in response to increasing environmental complexity. The environments were generated from a set of $M = 256$ arbitrary nutrients, which were combined in a scheme similar to that of com3, com4, and com13: 256 conditions with one nutrient, 128 conditions with two nutrients, and so on, up until one condition with all 256 nutrients. In total, this resulted in 10,240 unique *in silico* growth experiments. The metabolite conversion matrix D was set to zero, in order to simulate growth exclusively on the provided nutrients. The initial species and resource abundances were set to 6×10^6 CFU/mL and 1.5 g/mL, respectively, to approximate the initial OD600 of 0.05 and the initial nutrient concentration of 50 mM C of glucose used in our experiments. We then simulated the growth of these communities over the course of 288 hours (based on com3, com4, and com13 culturing timescale) with a timestep of 0.01 hours and quantified their yields at the endpoints.

For our simulations of species dynamics in com3 and com4, we parametrized the nutrient preference matrix C based on the growth yields of each individual organism on the 32 individual carbon sources (Supplementary Figure 7b). To determine the fraction of each nutrient α that could be converted to a secreted metabolite β, we used flux-balance models for each of the four relevant organisms[67–70]. We calculated the secretion fluxes of metabolic byproducts from the

organisms under each of the 32 individual nutrients using the technique described above, and used the ratio of secretion to intake fluxes to populate the D matrix. This matrix was then normalized across each primary nutrient to ensure conservation of mass. Nutrient uptake efficiencies for these secreted metabolites were then assigned according to the same monoculture growth data used to define preferences for the primary nutrients. As with our initial arbitrary communities, initial species abundances were set to 6×10^6 CFU/mL and initial nutrient abundances were set to 1.5 g/mL. We then simulated the growth and potential metabolic exchange of all four community members in coculture over the course of 288 hours with a timestep of 0.01 hours in the 63 environmental conditions for com3 and com4 (Supplementary Table 3).

To simulate the growth of randomly-defined communities containing generalist and specialist organisms, we again relied on random sampling to populate the resource preference matrix C. For our 4-species communities, we used a more stringent generalist-specialist designation based on how many nutrients were able to be consumed by the top organisms in our Biolog assay: an organism was classed as a generalist if it was able to grow on more than 90% of the nutrients (e.g. *P. aeruginosa*), and a specialist if it was able to grow on fewer than 50% of the nutrients (e.g. *B. subtilis*). Therefore, if an organism was a generalist it had a 90% chance of consuming any given nutrient, versus a 50% chance if it was a specialist. For our first 13-species community, each organism had a probability of 0.5 of utilizing a particular nutrient. In our second 13-species community, this probability was defined by the proportion of nutrients each com13 organism was able to utilize in our Biolog assay. For the nutrient conversion matrix D, we established a fixed probability of a particular organism i converting a nutrient α into a metabolite β of 25%. We carried out 50 random simulations of these communities, randomly repopulating the C and D matrices each time to adequately sample the possible space of nutrient preferences and conversions. These simulations were also run with the same parameters, initial conditions, and timeframe as those for com4.

Acknowledgements

We thank Melisa L. Osborne, Christopher P. Mancuso, Jennifer M. Bhatnagar, and Ilija Dukovski for experimental advice and technical assistance. We are also grateful to Sean P. Mullen for access to the robotic liquid-handling system and to James E. Fifer for relevant experimental training. Additionally, we thank past and present members of the Segrè group for helpful discussions on the research and the manuscript, particularly David B. Bernstein, Joshua E. Goldford, Demetrius DiMucci, Elena Forchielli, and Devlin Moyer. A.R.P. is supported by a Howard Hughes Medical Institute Gilliam Fellowship and a National Academies of Sciences, Engineering, and Medicine Ford Foundation Predoctoral Fellowship. We gratefully acknowledge support from the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research through the Microbial Community Analysis and Functional Evaluation in Soils SFA Program (m-CAFEs) under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory, as well as the National Institutes of Health (grants 5R01DE024468, R01GM121950), the National Science Foundation (grants 1457695 and NSFOCE-BSF 1635070), the Human Frontiers Science Program (grant RGP0020/2016), and the Boston University Interdisciplinary Biomedical Research Office.

Author contributions

A.R.P and D.S. designed the research. A.R.P. designed and performed experiments, collected data, wrote data analysis code, developed the models, and ran and analyzed simulations. A.R.P. and D.S. wrote the manuscript. Both authors read and approved the final manuscript.

Declaration of interests

The authors declare that no competing interests exist in relation to this manuscript.

References

1. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
2. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12115–20 (2006).
3. Tecon, R. & Or, D. Biophysical processes supporting the diversity of microbial life in soil. *FEMS Microbiol. Rev.* **41**, 599–623 (2017).
4. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science (80-.).* **320**, 1034–1039 (2008).
6. Preusser, S., Marhan, S., Poll, C. & Kandeler, E. Microbial community response to changes in substrate availability and habitat conditions in a reciprocal subsoil transfer experiment. *Soil Biol. Biochem.* **105**, 138–152 (2017).
7. Yang *et al.* Role of Dietary Nutrients in the Modulation of Gut Microbiota: A Narrative Review. *Nutrients* **12**, 381 (2020).
8. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
9. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12**, 69 (2014).
10. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).
11. Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
12. Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nat. Rev. Microbiol.* **18**, 35–46 (2020).
13. Moran, M. A. The global ocean microbiome. *Science (80-.).* **350**, aac8455–aac8455 (2015).
14. Sonnenburg, E. D. & Sonnenburg, J. L. Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metab.* **20**, 779–786 (2014).
15. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
16. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl. Acad. Sci.* **107**, 14691–14696 (2010).
17. Bolnick, D. I. *et al.* Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). *Ecol. Lett.* **17**, 979–987 (2014).
18. Li, H. *et al.* Diet Diversity Is Associated with Beta but not Alpha Diversity of Pika Gut Microbiota. *Front. Microbiol.* **7**, (2016).
19. Ringø, E. *et al.* Effect of dietary components on the gut microbiota of aquatic animals. A never-ending story? *Aquac. Nutr.* **22**, 219–282 (2016).
20. Lindemann, S. R. *et al.* Engineering microbial consortia for controllable outputs. *ISME J.* **10**, 2077–2084 (2016).
21. Klitgord, N. & Segrè, D. Environments that Induce Synthetic Microbial Ecosystems. *PLoS Comput. Biol.* **6**, e1001002 (2010).
22. Lawson, C. E. *et al.* Common principles and best practices for engineering microbiomes. *Nat. Rev. Microbiol.* **17**, 725–741 (2019).

23. Hardin, G. The competitive exclusion principle. *Science* **131**, 1292–7 (1960).
24. Tilman, D. *Resource Competition and Community Structure*. (Princeton University Press, 1982).
25. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).
26. Vila, J. C. C., Liu, Y.-Y. & Sanchez, A. Dissimilarity–Overlap analysis of replicate enrichment communities. *ISME J.* (2020) doi:10.1038/s41396-020-0702-7.
27. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science (80-.).* **361**, 469–474 (2018).
28. Estrela, S. *et al.* Environmentally Mediated Social Dilemmas. *Trends Ecol. Evol.* **34**, 6–18 (2019).
29. Harcombe, W. Novel cooperation experimentally evolved between species. *Evolution (N. Y).* **64**, 2166–72 (2010).
30. San Roman, M. & Wagner, A. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *PLOS Comput. Biol.* **14**, e1006340 (2018).
31. Pacheco, A. R., Moel, M. & Segrè, D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* **10**, 103 (2019).
32. Replansky, T. & Bell, G. The relationship between environmental complexity, species diversity and productivity in a natural reconstructed yeast community. *Oikos* **118**, 233–239 (2009).
33. Fonte, E. S. *et al.* The Combination of Different Carbon Sources Enhances Bacterial Growth Efficiency in Aquatic Ecosystems. *Microb. Ecol.* **66**, 871–878 (2013).
34. Marsland, R., Cui, W. & Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Sci. Rep.* **10**, 1–17 (2020).
35. Segrè, D., DeLuna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
36. Dykhuizen, D. & Davies, M. An Experimental Model: Bacterial Specialists and Generalists Competing in Chemostats. *Ecology* **61**, 1213–1227 (1980).
37. Bauchop, T. & Elsden, S. R. The Growth of Micro-organisms in Relation to their Energy Supply. *Microbiology* **23**, 457–469 (1960).
38. Monod, J. The Growth of Bacterial Cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
39. Wang, X., Xia, K., Yang, X. & Tang, C. Growth strategy of microbes on mixed carbon sources. *Nat. Commun.* **10**, 1–7 (2019).
40. Muscarella, M. E., Boot, C. M., Broeckling, C. D. & Lennon, J. T. Resource heterogeneity structures aquatic bacterial communities. *ISME J.* **13**, 2183–2195 (2019).
41. Eguíluz, V. M. *et al.* Scaling of species distribution explains the vast potential marine prokaryote diversity. *Sci. Rep.* **9**, 18710 (2019).
42. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5970–5 (2016).
43. Hoffmann, K. H. *et al.* Power law rank-abundance models for marine phage communities. *FEMS Microbiol. Lett.* **273**, 224–8 (2007).
44. Redondo-Nieto, M. *et al.* Genome Sequence of the Biocontrol Strain Pseudomonas fluorescens F113. *J. Bacteriol.* **194**, 1273 (2012).
45. Liu, Y., Rainey, P. B. & Zhang, X.-X. Molecular mechanisms of xylose utilization by *P seudomonas fluorescens*: overlapping genetic responses to xylose, xylulose, ribose and mannitol. *Mol. Microbiol.* **98**, 553–570 (2015).
46. Baumann, P. & Baumann, L. Catabolism of d-fructose and d-ribose by Pseudomonas doudoroffii. *Arch. Microbiol.* **105**, 225–240 (1975).
47. S, R.-G., C, H.-B., E, M., G, G. & A, M. Homolactic Fermentation From Glucose and Cellobiose Using Bacillus Subtilis. *Microb. Cell Fact.* **8**, (2009).

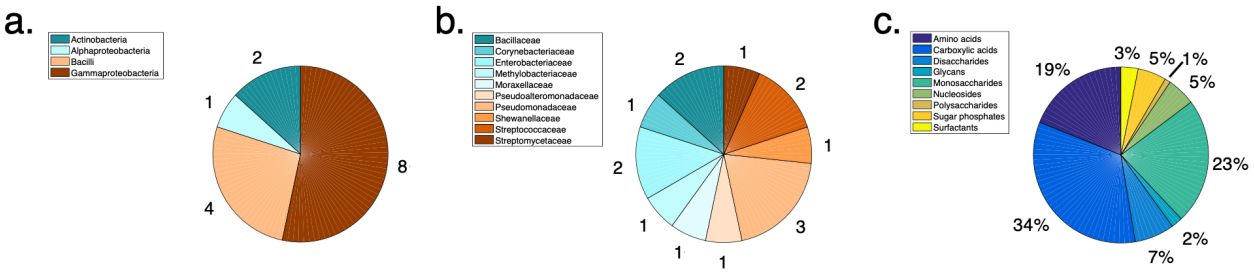
48. Kim, H., Lee, S. J. & Shin, K.-S. Characterization of new oligosaccharide converted from cellobiose by novel strain of Bacillus subtilis. *Food Sci. Biotechnol.* **27**, 37 (2018).
49. Jin, Q. & Kirk, M. F. pH as a Primary Control in Environmental Microbiology: 1. Thermodynamic Perspective. *Front. Environ. Sci.* **6**, 21 (2018).
50. Ratzke, C., Denk, J. & Gore, J. Ecological suicide in microbes. *Nat. Ecol. Evol.* **2**, 867–872 (2018).
51. Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* **7**, 1104–1115 (2014).
52. Kelsic, E. D., Zhao, J., Vetsigian, K. & Kishony, R. Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nature* **521**, 516–519 (2015).
53. Ratzke, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nat. Ecol. Evol.* 1–8 (2020) doi:10.1038/s41559-020-1099-4.
54. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25 (2010).
55. Acinas, S. G. *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551–4 (2004).
56. Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**, 4399–406 (2001).
57. Schloss, P. D. & Handelsman, J. Toward a Census of Bacteria in Soil. *PLoS Comput. Biol.* **2**, e92 (2006).
58. Anthony, M. A., Crowther, T. W., Maynard, D. S., van den Hoogen, J. & Averill, C. Distinct Assembly Processes and Microbial Communities Constrain Soil Organic Carbon Formation. *One Earth* **2**, 349–360 (2020).
59. Ebrahimi, A., Schwartzman, J. & Cordero, O. X. Cooperation and spatial self-organization determine rate and efficiency of particulate organic matter degradation in marine bacteria. *Proc. Natl. Acad. Sci.* **116**, 23309–23316 (2019).
60. Zomorrodi, A. R. & Segrè, D. Synthetic Ecology of Microbes: Mathematical Models and Applications. *J. Mol. Biol.* **428**, 837–861 (2016).
61. Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N. & Silver, P. A. Better together: engineering and application of microbial symbioses. *Curr. Opin. Biotechnol.* **36**, 40–49 (2015).
62. Bernstein, H. C. & Carlson, R. P. Microbial Consortia Engineering for Cellular Factories: in vitro to in silico systems. *Comput. Struct. Biotechnol. J.* **3**, e201210017 (2012).
63. Shong, J., Jimenez Diaz, M. R. & Collins, C. H. Towards synthetic microbial consortia for bioprocessing. *Curr. Opin. Biotechnol.* **23**, 798–802 (2012).
64. Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A. & Alm, E. J. Distribution-Based Clustering: Using Ecology To Refine the Operational Taxonomic Unit. *Appl. Environ. Microbiol.* **79**, 6593–6603 (2013).
65. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
66. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
67. Henry, C. S., Zinner, J. F., Cohoon, M. P. & Stevens, R. L. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biol.* **10**, R69 (2009).
68. Peyraud, R. *et al.* Genome-scale reconstruction and system level investigation of the metabolic network of Methylobacterium extorquens AM1. *BMC Syst. Biol.* **5**, (2011).
69. Oberhardt, M. A., Puchałka, J., Fryer, K. E., Martins dos Santos, V. A. P. & Papin, J. A.

Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **190**, 2790–803 (2008).
70. Pinchuk, G. E. *et al.* Constraint-Based Model of Shewanella oneidensis MR-1 Metabolism: A Tool for Data Analysis and Hypothesis Generation. *PLoS Comput. Biol.* **6**, e1000822 (2010).
71. Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120 (2014).
72. Orth, J. D., Thiele, I. & Palsson, B. Ø. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
73. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
74. Holzhütter, H.-G. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.* **271**, 2905–2922 (2004).
75. MacArthur, R. Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970).
76. Chesson, P. MacArthur's consumer-resource model. *Theor. Popul. Biol.* **37**, 26–38 (1990).
77. Chesson, P. Mechanisms of Maintenance of Species Diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2003).

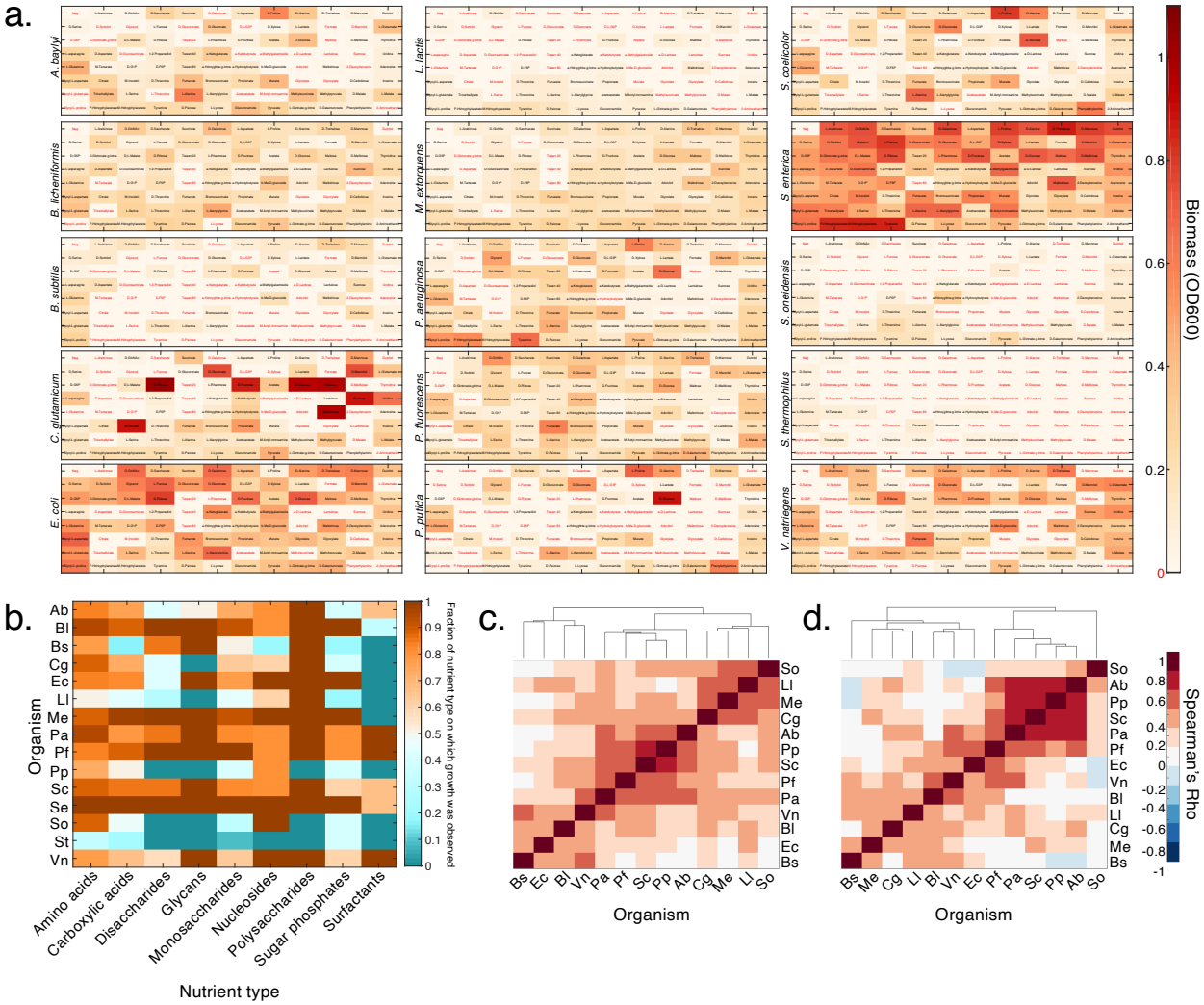
Supplementary Information

The effects of environmental complexity on microbial community yield and structure

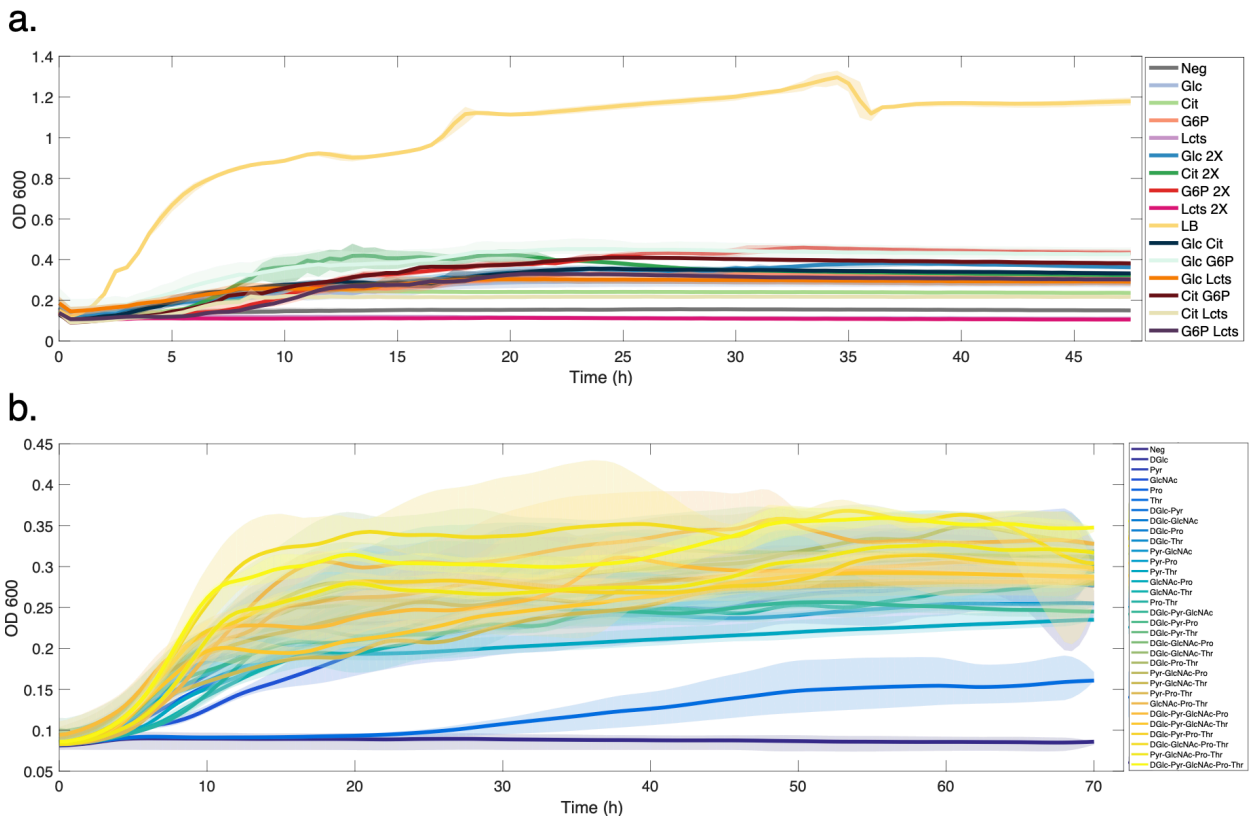
Alan R. Pacheco and Daniel Segrè



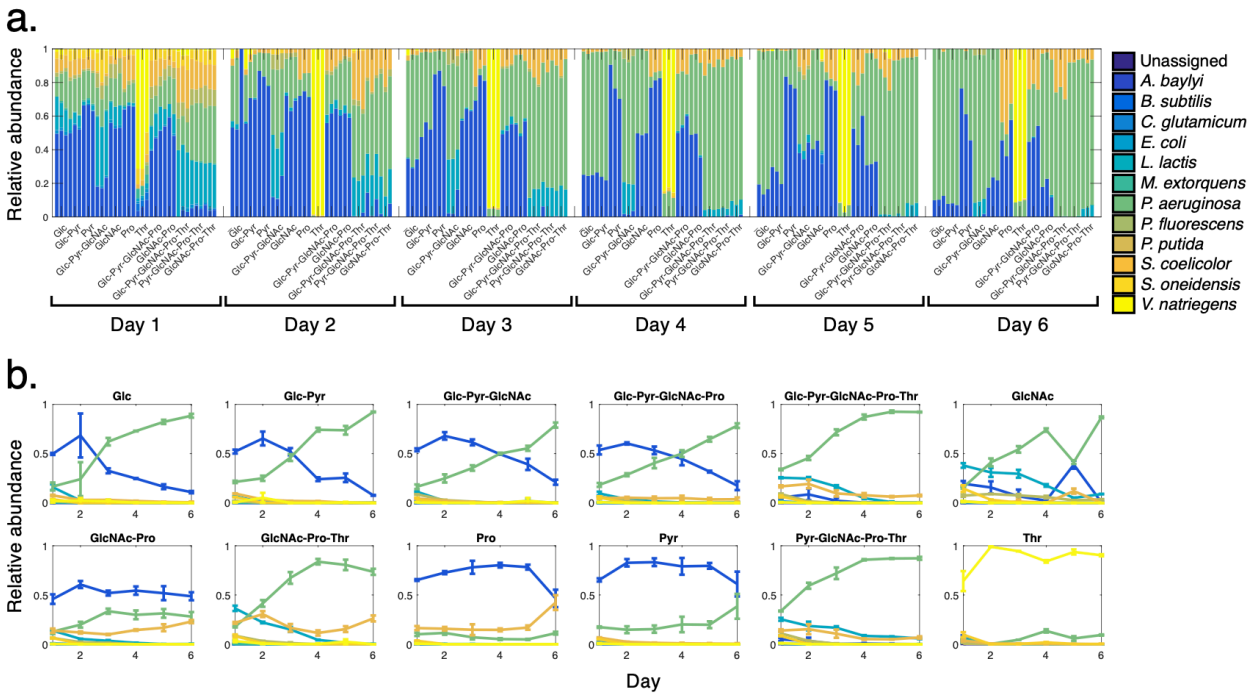
Supplementary Figure 1. Organisms and carbon sources used in experiments. **(a-b).** Phylum- **(a)** and family-level **(b)** groupings of 15 bacterial organisms. **(c).** Nutrient categories for 95 carbon sources.



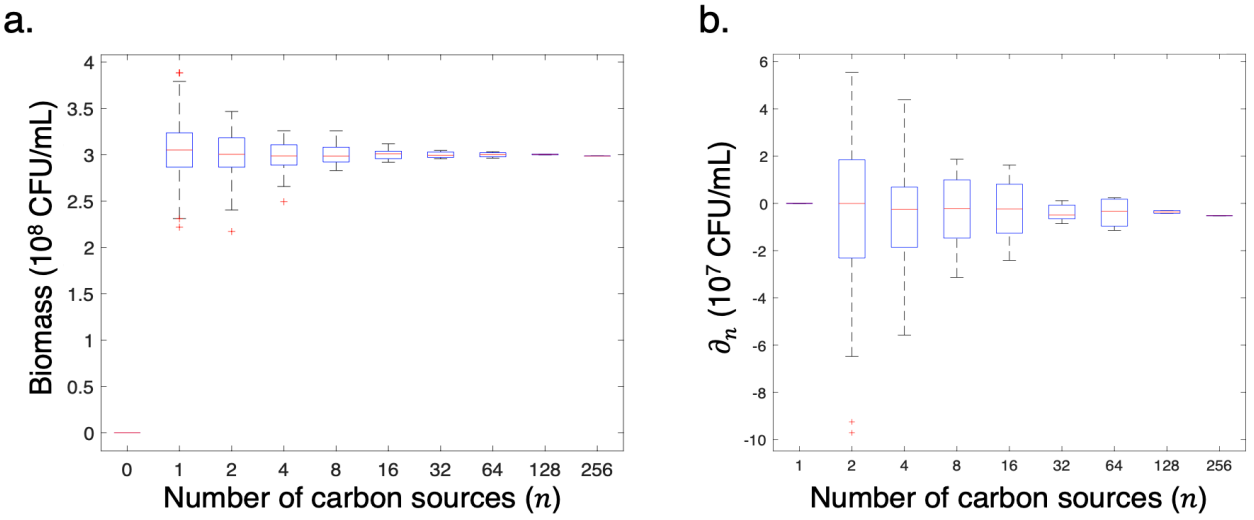
Supplementary Figure 2. Results of Biolog phenotypic assay. **(a).** Average (3-replicate) species-specific growth profiles on Biolog carbon sources in M9 minimal medium after 48 hours. Raw OD600 values were corrected for liquid evaporation and a significance cutoff was applied to determine growth above the levels of the negative controls (Methods). Nutrients on which growth was not observed are marked in red. **(b).** Growth capabilities of all 15 organisms on 95 Biolog nutrients. Values displayed are the fraction of a specific nutrient type on which an organism displayed growth. **(c, d).** Hierarchical clustering of 13 selected organisms based on Spearman correlations of growth profiles on all 95 **(c)** and on 32 selected nutrients **(d)**. Organisms are abbreviated as: Ab: *A. baylyi*, Bl: *B. licheniformis*, Bs: *B. subtilis*, Cg: *C. glutamicum*, Ec: *E. coli*, Ll: *L. lactis*, Me: *M. extorquens*, Pa: *P. aeruginosa*, Pf: *P. fluorescens*, Pp: *P. putida*, Sc: *S. coelicolor*, So: *S. oneidensis*, Vn: *V. natriegens*.



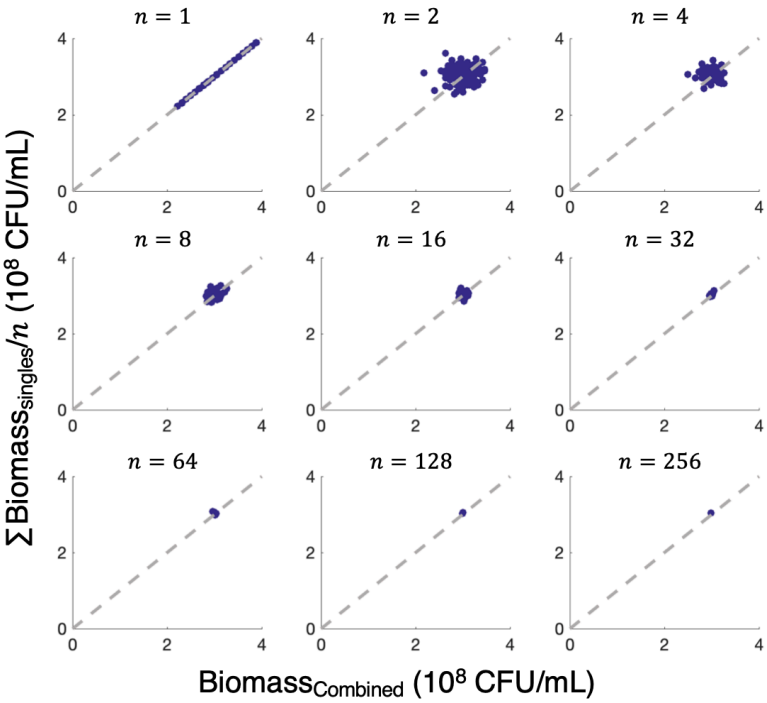
Supplementary Figure 3. Growth curves for multispecies communities in combinatorial environments. **(a).** Growth trajectories of 14-species communities (com14, Supplementary Table 5) on combinations and different concentrations (25 mM C and 50 mM C) of D-glucose (Glc), citrate (Cit), glucose-6-phosphate (G6P), and a-D-lactose (Lcts). Double-carbon source conditions contain 50 mM C of total carbon source, for 25 mM C of each individual nutrient. **(b).** Growth trajectories of 13-species community (com13a) on equimolar concentrations (50 mM C) of five carbon sources: D-glucose (DGlc), pyruvate (Pyr), GlcNAc (GlcNAc), L-proline (Pro), and L-threonine (Thr). Shaded regions in both plots indicate standard deviations across three replicates per condition.



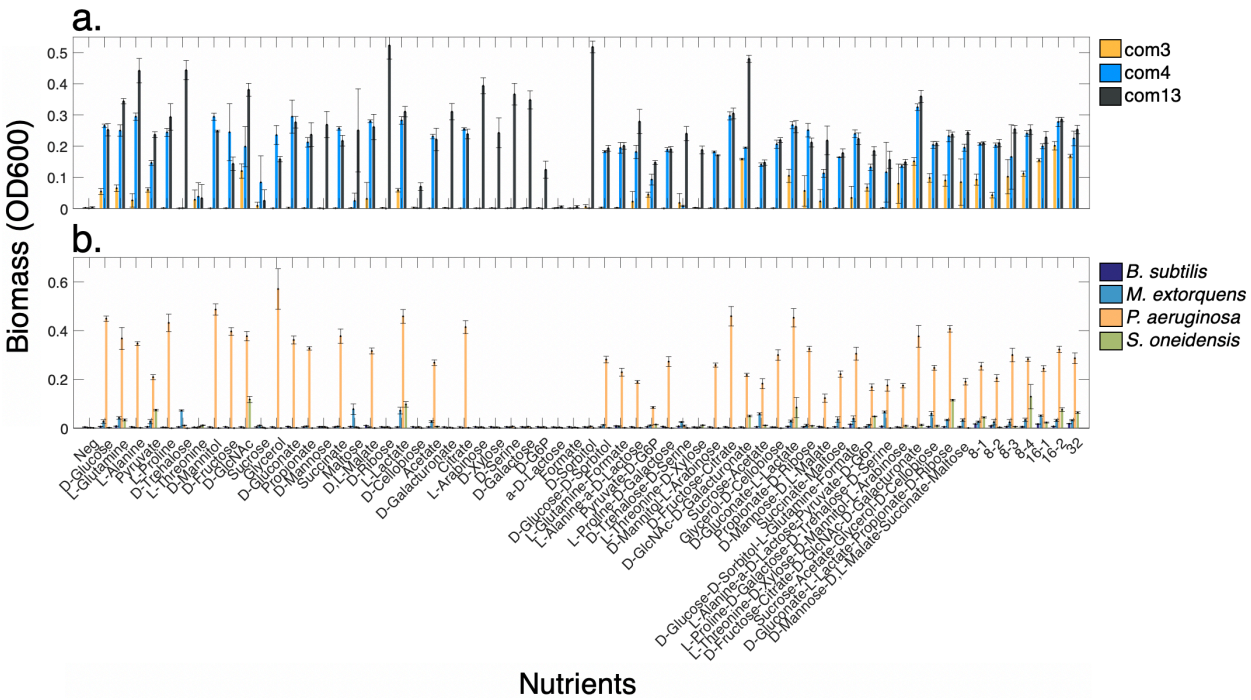
Supplementary Figure 4. Taxonomic data for 13-species community grown on five carbon sources (com13a, Supplementary Table 5): D-glucose (DGlc), pyruvate (Pyr), GlcNAc (GlcNAc), L-proline (Pro), and L-threonine (Thr). **(a).** Relative abundance plots of all replicates over time. **(b).** Species relative abundance trajectories over time. Error bars indicate standard deviation.



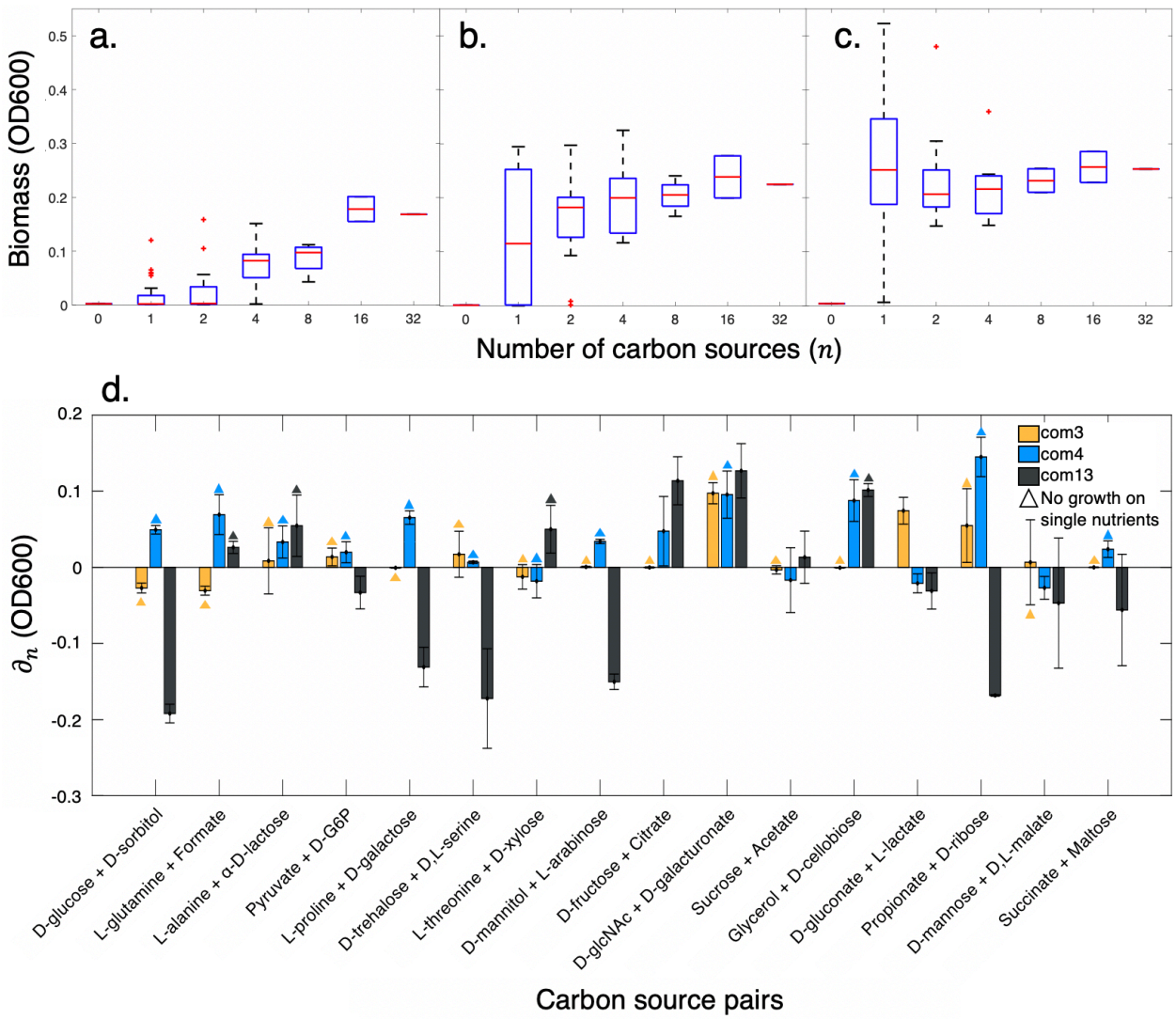
Supplementary Figure 5. Consumer-resource model-predicted yields for simulated 10-member communities. **(a).** Final biomass values after simulated 288 hours of community growth (corresponding to the full experimental timescale of com3, com4, and com13) on different combinations of nutrients (Methods). Here, final community yields converge to a median based on the total nutrient concentration (approximately 3×10^8 CFU/mL). **(b).** Differences ∂_n between community biomass yields on n carbon sources and the scaled sum of yields on the corresponding single carbon sources. No significant changes in mean ∂_n was observed relative to single-nutrient conditions. Here, as with all subsequent boxplots, the red central mark indicates the median, the top and bottom box edges indicate the 25th and 75th percentiles, respectively, the whiskers extend to the most extreme points not considered outliers, and the red '+' symbols indicate outliers plotted individually.



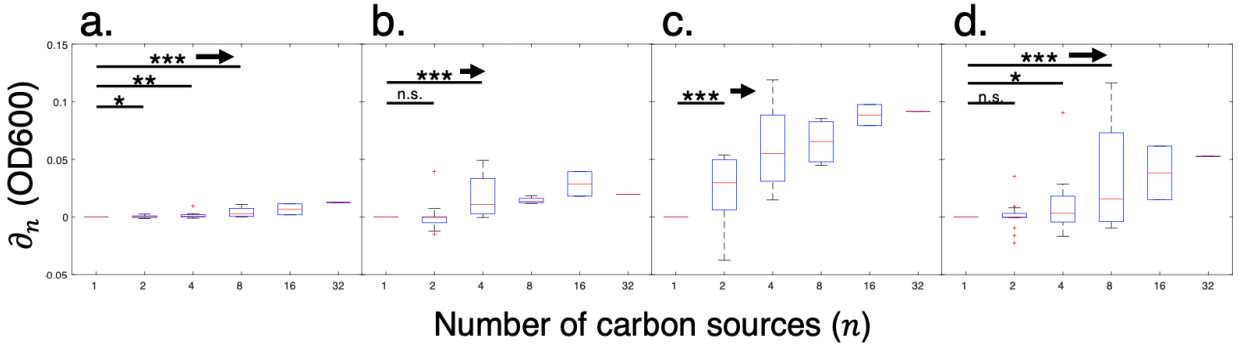
Supplementary Figure 6. Expanded representation of consumer resource model-predicted biomass yields for simulated 10-member communities. For each data point, its position on the x-axis indicates the final community biomass on a combination of nutrients n and its position on the y-axis indicates the scaled sum final community biomass on the corresponding individual nutrients.



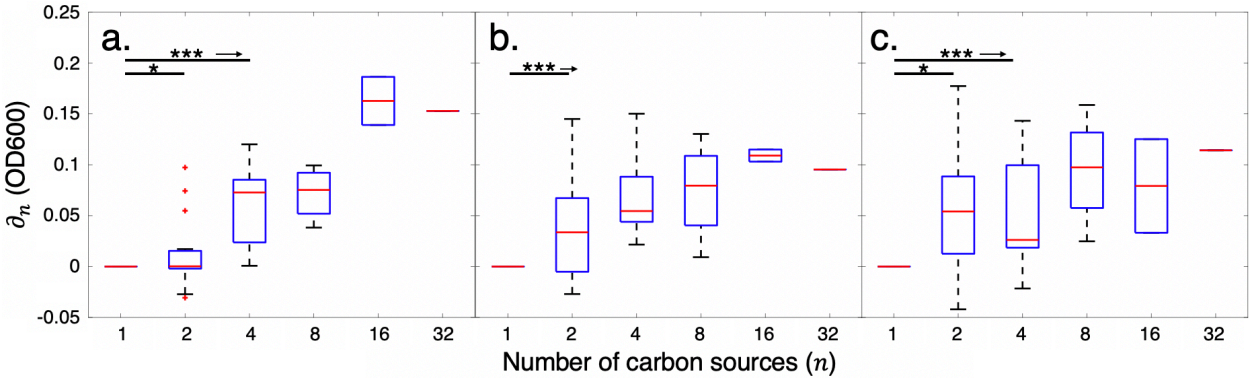
Supplementary Figure 7. Endpoint nutrient combination-specific growth yields for multispecies communities com3, com4, and com13 **(a)** (Supplementary Table 5) and single organisms **(b)** on combinations of up to 32 carbon sources. Error bars indicate standard deviation.



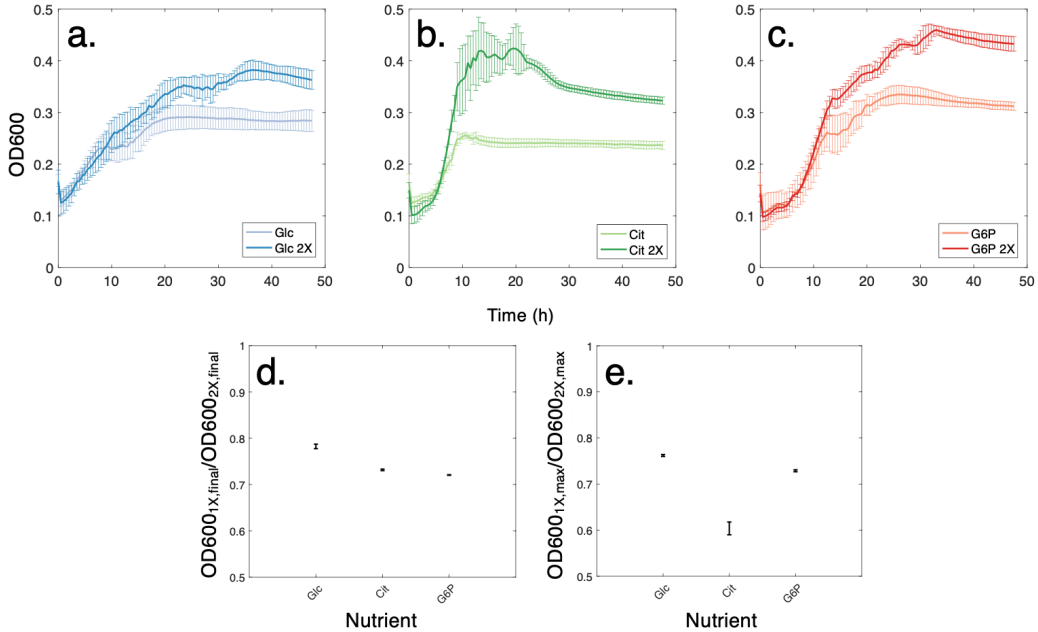
Supplementary Figure 8. Endpoint aggregated growth yields for multispecies communities com3 **(a)**, com4 **(b)**, and com13 **(c)** relative to number of carbon sources. **(d)** Differences ∂_n for com3, com4, and com13 between community yields on pairs of carbon sources and the scaled sum of community yields on the corresponding single carbon sources. Error bars indicate standard deviation.



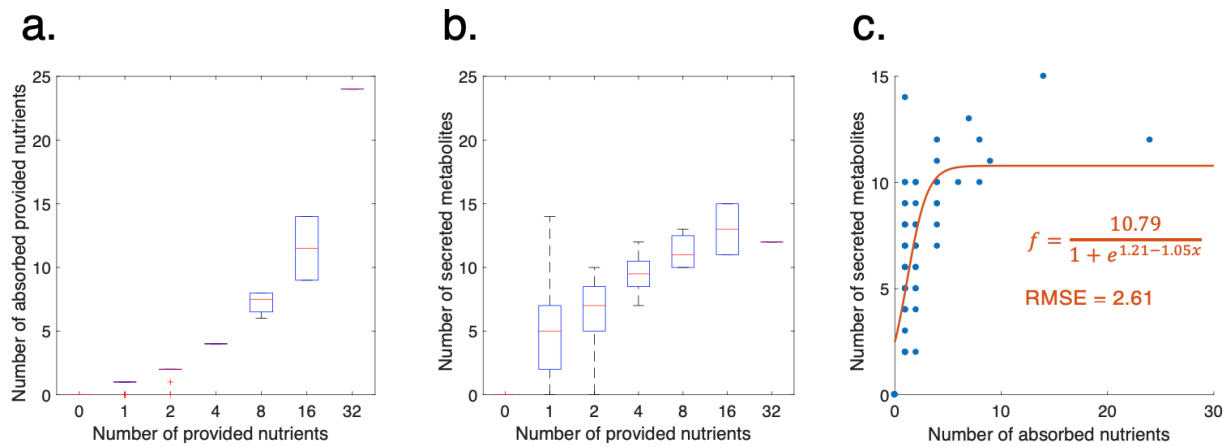
Supplementary Figure 9. Differences ∂_n between single organism growth yields on n carbon sources and the scaled sum of yields on the corresponding single carbon sources for **(a)** *B. subtilis*, **(b)** *M. extorquens*, **(c)** *P. aeruginosa*, and **(d)** *S. oneidensis*.



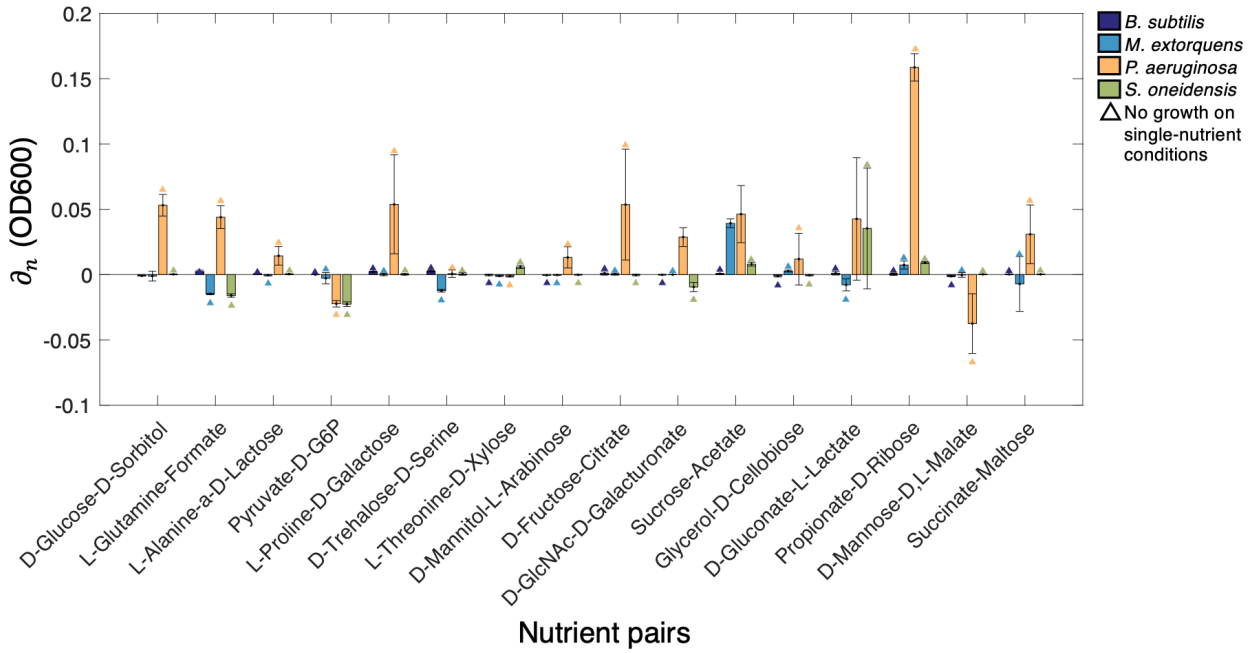
Supplementary Figure 10. Differences ∂_n between community growth yields on n carbon sources and the scaled sum of community yields on the corresponding single carbon sources for **(a)** com3, **(b)** com4, and **(c)** com13 after the first 48 hours of growth.



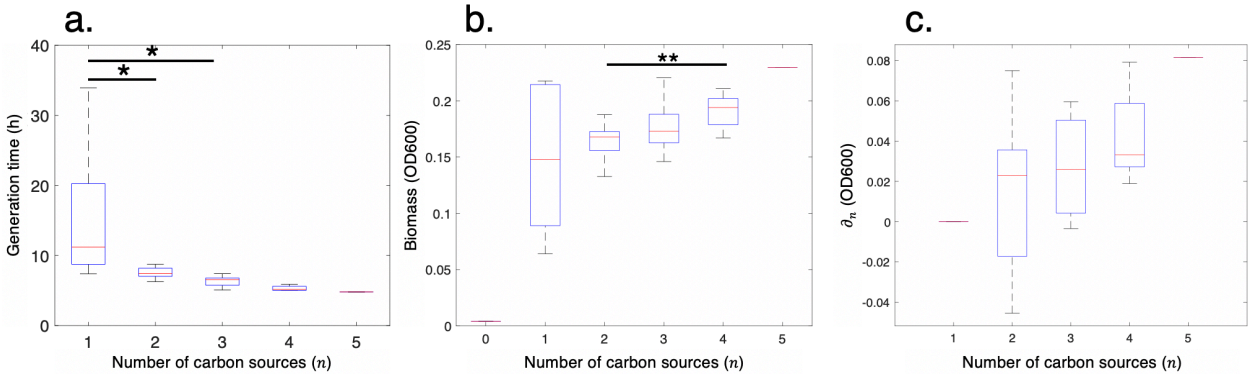
Supplementary Figure 11. Analysis of community growth yields in single (25 mM C) and double (50 mM C) concentrations of nutrients. **(a-c).** Growth trajectories for 14-species community (com14) in two concentrations of D-glucose (Glc, **a**), citrate (Cit, **b**), and glucose-6-phosphate (G6P, **c**). **(d).** Ratio of community biomass values between single and double nutrient concentrations at endpoint **(d)** and for maximum OD values **(e)**. Error bars indicate standard deviation.



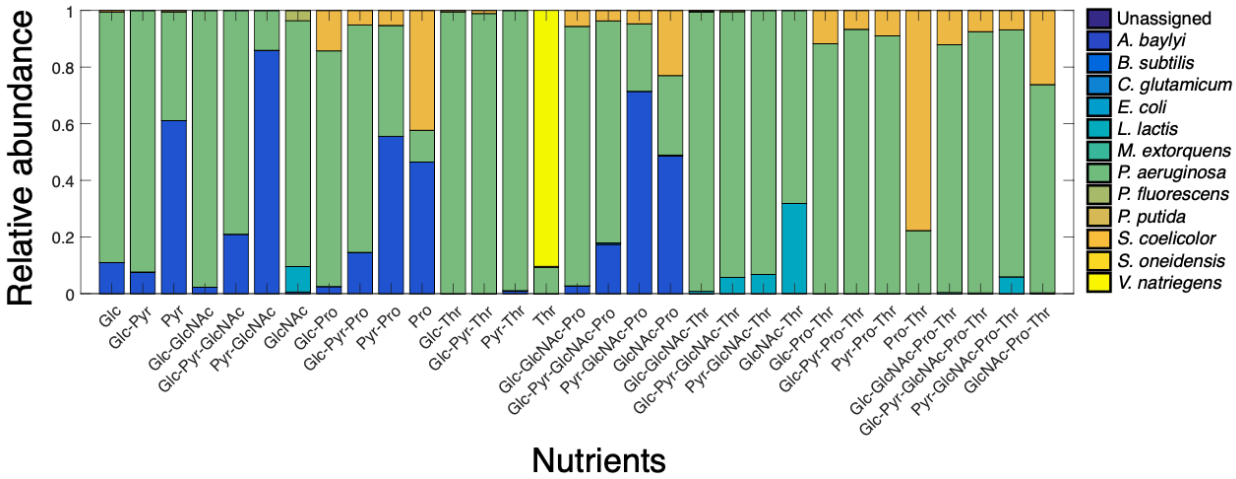
Supplementary Figure 12. Flux balance analysis-predicted number of absorbed and secreted metabolites for 4-species community made up of *B. subtilis*, *M. extorquens*, *P. aeruginosa*, and *S. oneidensis* in combinations of up to 32 carbon sources. (a). Number of nutrients taken up by any of the four organisms vs. number of provided nutrients. (b). Number of secreted metabolites secreted by any of the four organisms vs. the number of provided nutrients. (c). Relationship between number of secreted metabolites and absorbed nutrients. We found that a logistic function provided the best fit for this data, indicating a sharp rise followed by a plateau in the number of unique secreted metabolites.



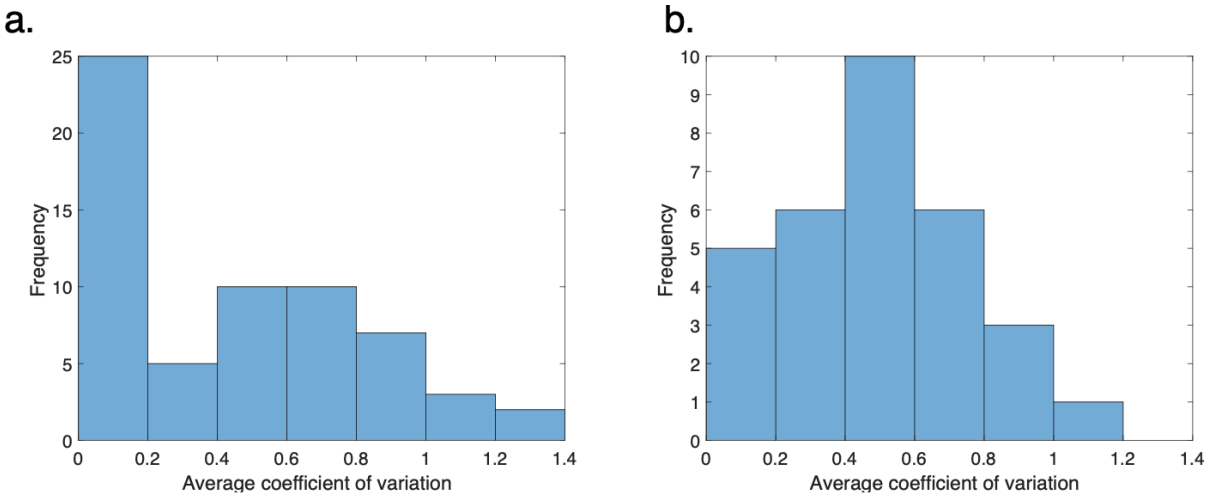
Supplementary Figure 13. Differences ∂_n for four single organisms between yields on pairs of carbon sources and the scaled sum of yields on the corresponding single carbon sources. For *M. extorquens*, for example, very minimal growth was observed on sucrose and acetate alone (OD600 0.01 ± 0.001 and 0.03 ± 0.004, respectively), whereas higher growth was observed on the combination of the two (OD600 0.06 ± 0.004), resulting in a mean ∂_n value of 0.02. In addition, *P. aeruginosa* displayed an increase on D-gluconate and L-lactate (OD600 0.36 ± 0.016 on D-gluconate, 0.46 ± 0.028 on L-lactate, and 0.45 ± 0.038 on the combination of the two), yielding a mean ∂_n value of 0.04. Error bars indicate standard deviation.



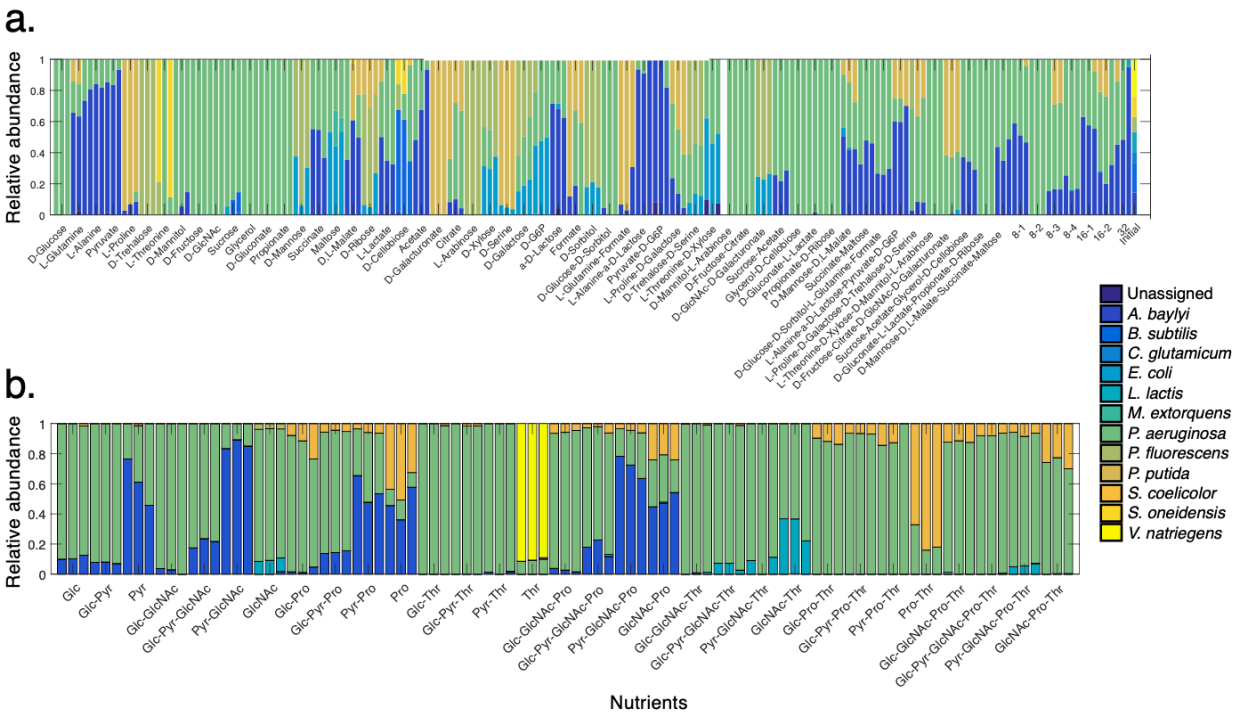
Supplementary Figure 14. Growth phenotype for 13-species community com13a. **(a).** Generation time vs. number of nutrients for 13-species community grown on five carbon sources (com13a, Supplementary Table 5). Generation time was calculated by obtaining the maximum slope in biomass (Supplementary Figure 3b) using a moving window encompassing 5 hours. **(b-c).** Aggregated growth yields **(b)** and differences ∂_n between single organism growth yields on n carbon sources and the scaled sum of yields on the corresponding single carbon sources **(c)**.



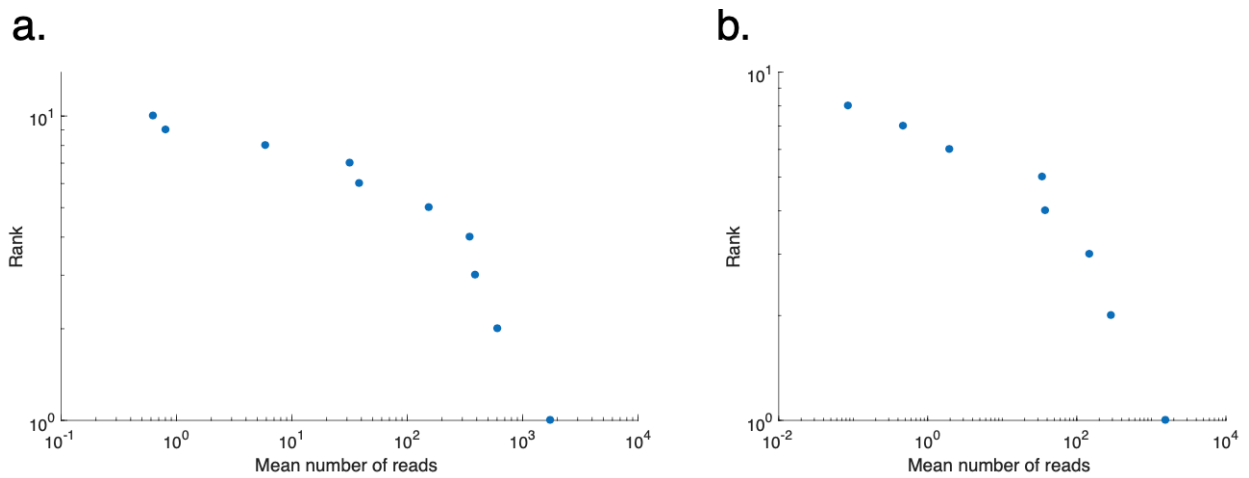
Supplementary Figure 15. Mean relative abundances (averaged over 3 replicates) for 13-species community grown on five carbon sources (com13a): D-glucose (DGlc), pyruvate (Pyr), GlcNAc (GlcNAc), L-proline (Pro), and L-threonine (Thr).



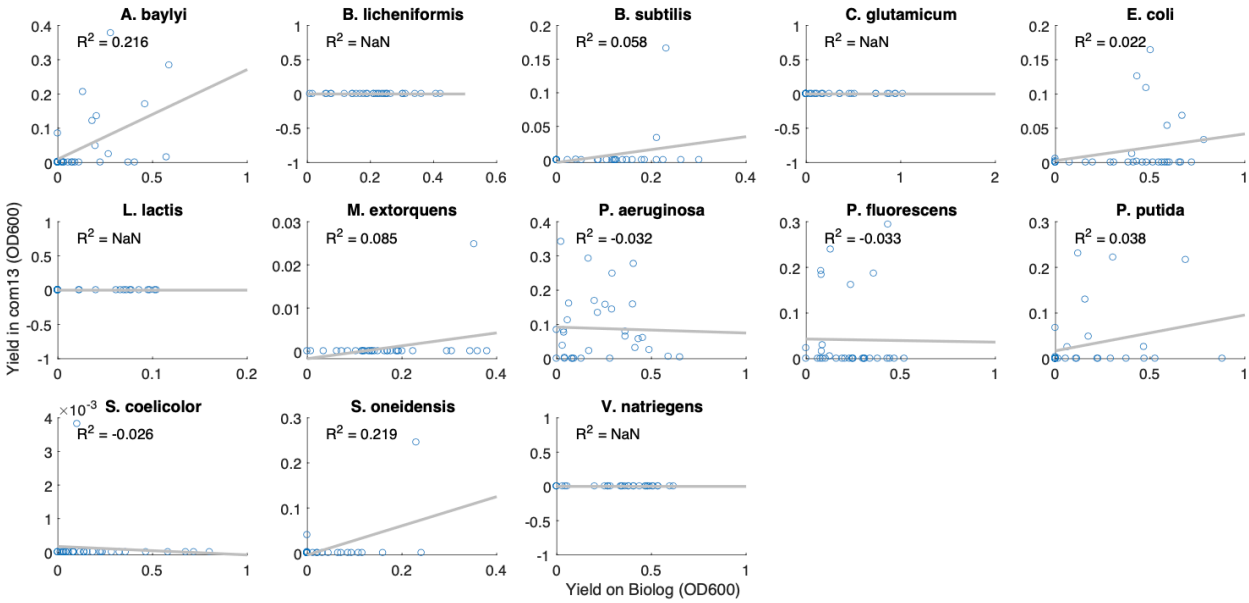
Supplementary Figure 16. Distributions of inter-replicate coefficients of variation for 13-species community grown on 32 carbon sources (com13, **a**) and 13-species community grown on 5 carbon sources (com13a, **b**).



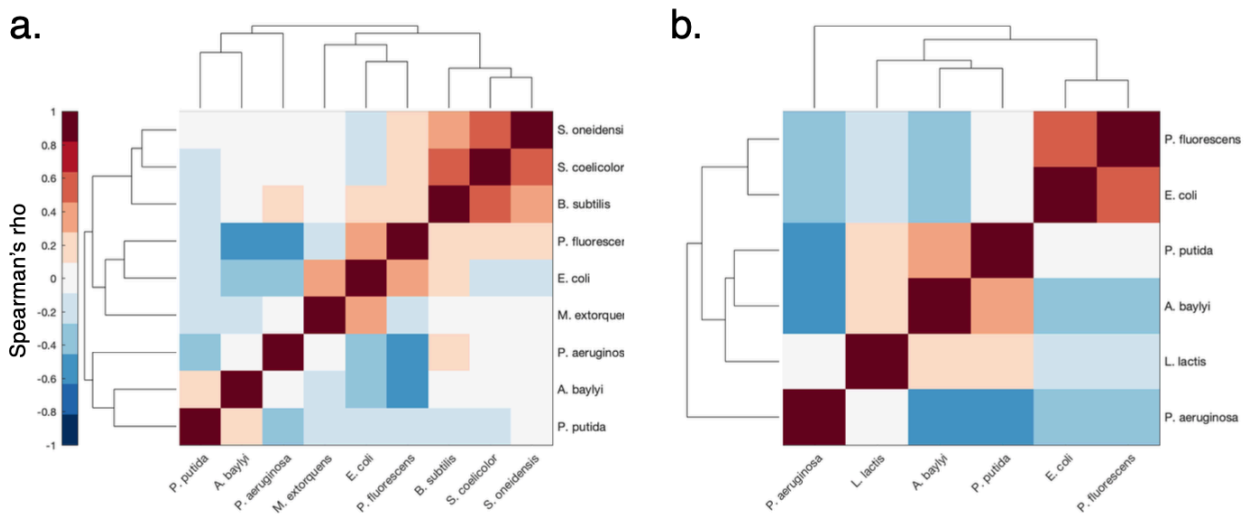
Supplementary Figure 17. Endpoint taxonomic distributions over all replicates for 13-species community grown on 32 carbon sources (com13, **a**) and 13-species community grown on 5 carbon sources (com13a, **b**). We encountered general consistency in community composition between comparable conditions of the two 13-species experiments. The D-glucose, pyruvate, and D-GlcNAc conditions had the same dominant organisms in both experiments (*P. aeruginosa*, *A. baylyi*, and *P. aeruginosa*, respectively), and the L-proline condition was composed of *A. baylyi*, *P. aeruginosa*, and a third organism in both experiments with only the identity of the third organism being different (*P. putida* in com13 and *S. coelicolor* in com13a). Nonetheless, in com13, L- threonine resulted in a dominance of *S. oneidensis* in two of the replicates and of *P. aeruginosa* in one replicate, while com13a resulted in dominance of *V. natriegens* in all three replicates. However, com13 grew very minimally in L-threonine (OD600 0.03 ± 0.04) in comparison to com13a (0.15 ± 0.02), in addition to having a very high inter-replicate coefficient of variation for this condition (1.36), suggesting that it is an outlier.



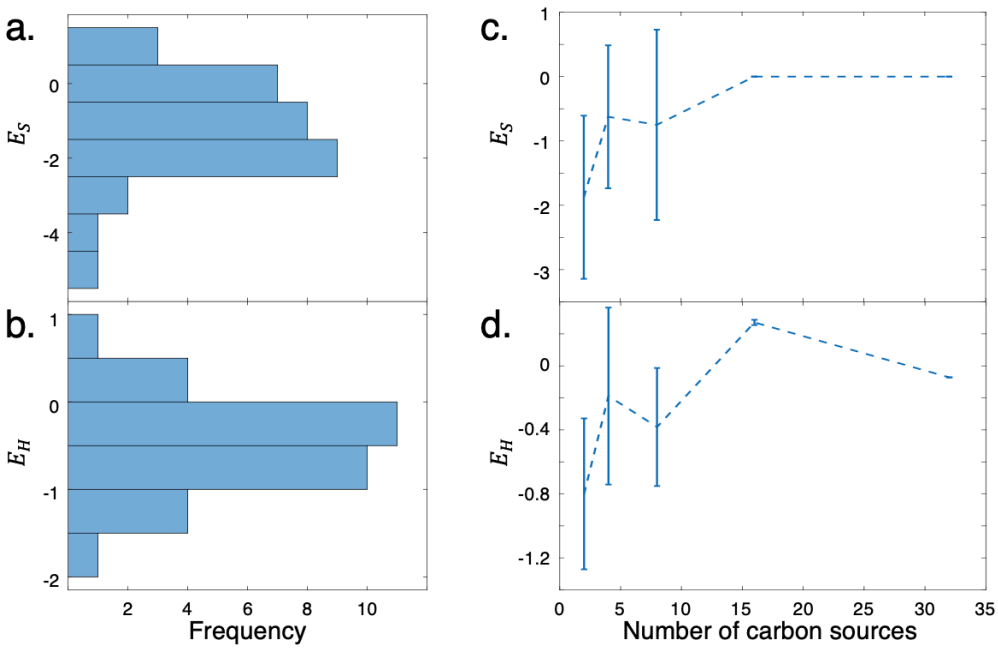
Supplementary Figure 18. Rank-abundance plots for 13-species community grown on 32 carbon sources (com13, **a**) and 13-species community grown on 5 carbon sources (com13a, **b**). These relationships displayed decay patterns separated at characteristic scales of approximately 10^2 reads, resembling double-power law relationships previously observed in a variety of natural ecosystems orders of magnitude more complex than our model communities[1–3]. Despite difference in scale, this rank-abundance relationship suggests fundamental structural similarities in community composition across experimental systems. Moreover, this similarity extends to the scaling and prevalence of very low-abundance taxa, indicating the abundances of these community members are accurately represented within our populations.



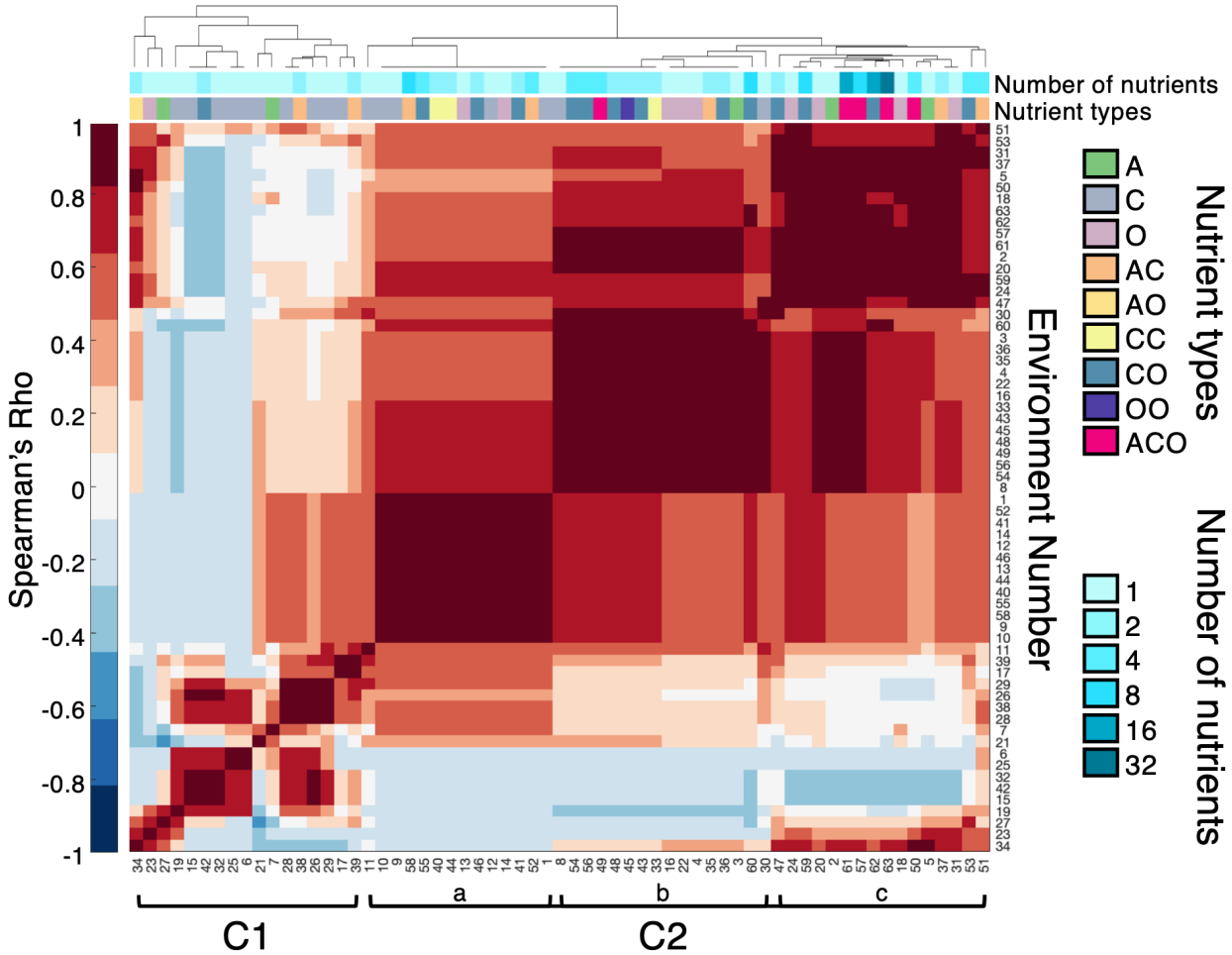
Supplementary Figure 19. Organism-specific yields in 13-species community grown on 32 carbon sources (com13) vs. in monoculture, with best-fit lines calculated using ordinary least squares regression.



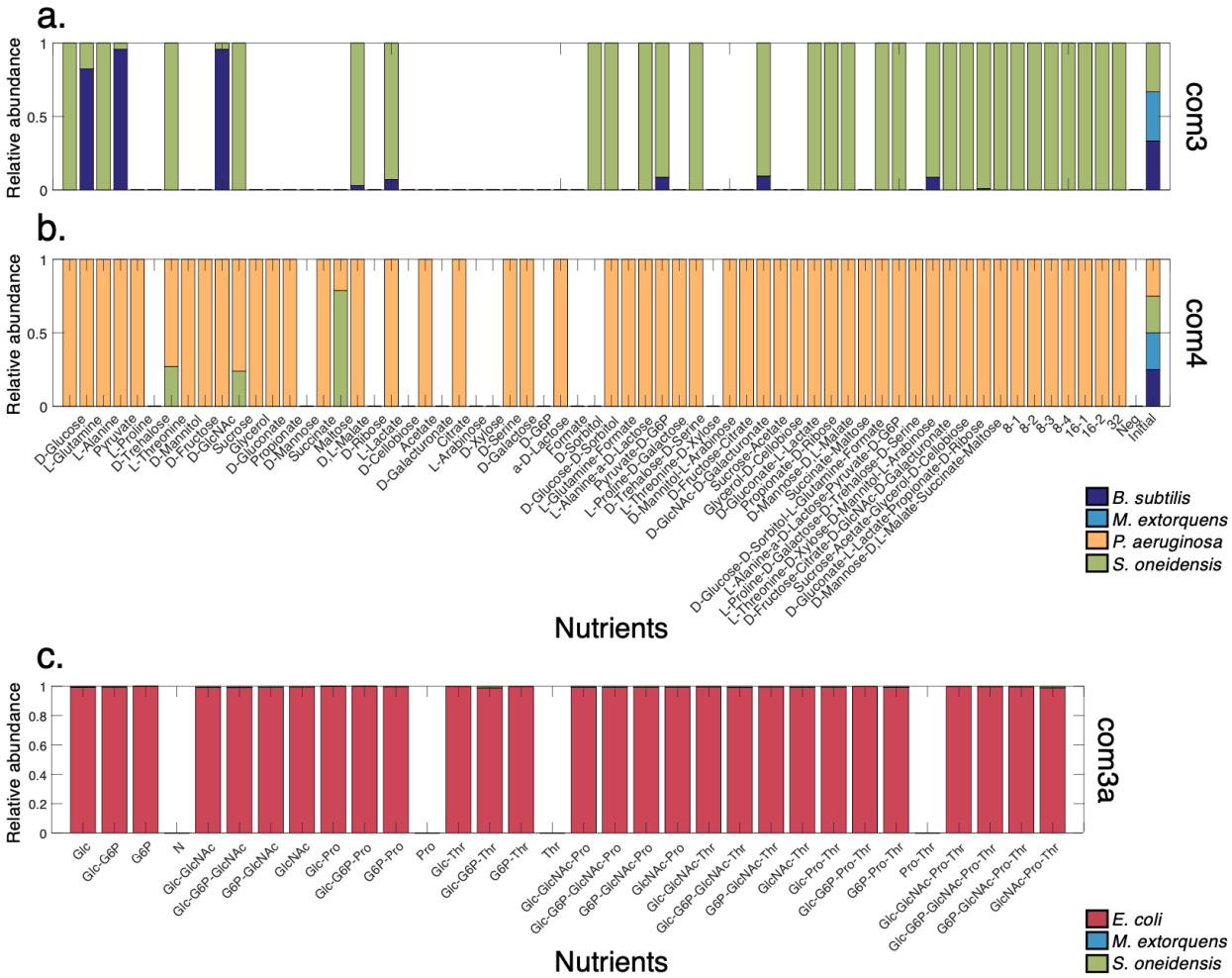
Supplementary Figure 20. Hierarchical clustering of Spearman correlations between organism relative abundances across com13 environments (Supplementary Table 3). **(a).** Clustering of species-species correlations across single-nutrient conditions. We found that the overall structure of the species-species clusters in mixed cultures was dramatically different from that of the monocultures (Supplementary Figure 2d), with much lower degrees of interspecies similarities. In com13, higher degrees of similarity were observed between *B. subtilis*, *S. coelicolor*, and *S. oneidensis*, likely due to their co-occurrence in D-cellobiose. This similarity was not observed in the monoculture data, in which *B. subtilis* correlated more strongly with organisms that did not remain in com13. The species-species clustering of com13 also highlighted the profound dissimilarities between *P. fluorescens* and both *A baylyi* and *P. aeruginosa*. These anticorrelations contrast with the monoculture data, as *P. fluorescens* had displayed relatively high degrees of similarity with both organisms. This difference further clarifies the competitive effects observed between *P. fluorescens* and these two organisms in a community setting, which leads to exclusion of specific organisms despite their ability to utilize the provided nutrients. **(b)** Clustering of species-species correlations across multiple-nutrient conditions. We observed that the similarities between *E. coli* and *P. fluorescens* were more pronounced than in the single-nutrient conditions, reflecting the ability of these two organisms to coexist in across different environments (e.g. D-trehalose + D-serine and D-glcNAc + D-galacturonate). Despite also having the ability to coexist in more complex environments, the correlations between *A. baylyi* and *P. aeruginosa* decreased in the multiple-nutrient conditions. In fact, *P. aeruginosa* was found to be the most dissimilar organism in these environments, likely due to its ability to exclude all other organisms in many conditions.



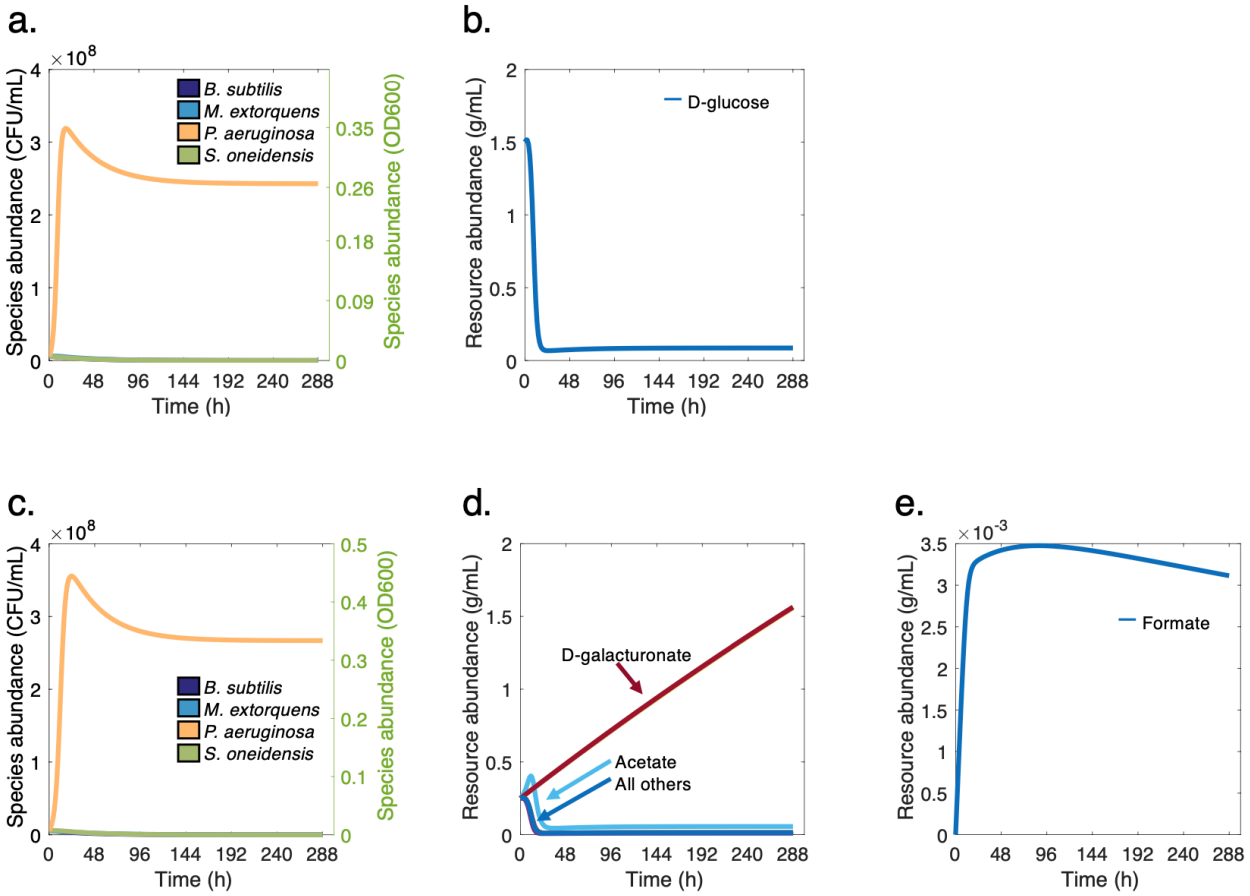
Supplementary Figure 21. Distributions of species richness and Shannon entropy epistasis scores for com13. **(a,b)**. Distributions of epistasis scores for species richness E_S **(b)** and Shannon entropy E_H **(b)**. **(c,d)**. Epistasis scores for species richness E_S **(c)** and Shannon entropy E_H **(d)**, grouped by number of provided nutrients. For each specific pair of nutrient combinations A and B, expected species richness and Shannon entropy values are computed for the union of the two combinations AB. This expected value is then subtracted from the observed values calculated for AB, resulting in epistasis scores (Methods). Negative epistasis scores indicate that the species richness and Shannon entropy of the communities on a given nutrient combination were lower than the values expected from less complex nutrient compositions. Error bars indicate standard deviation.



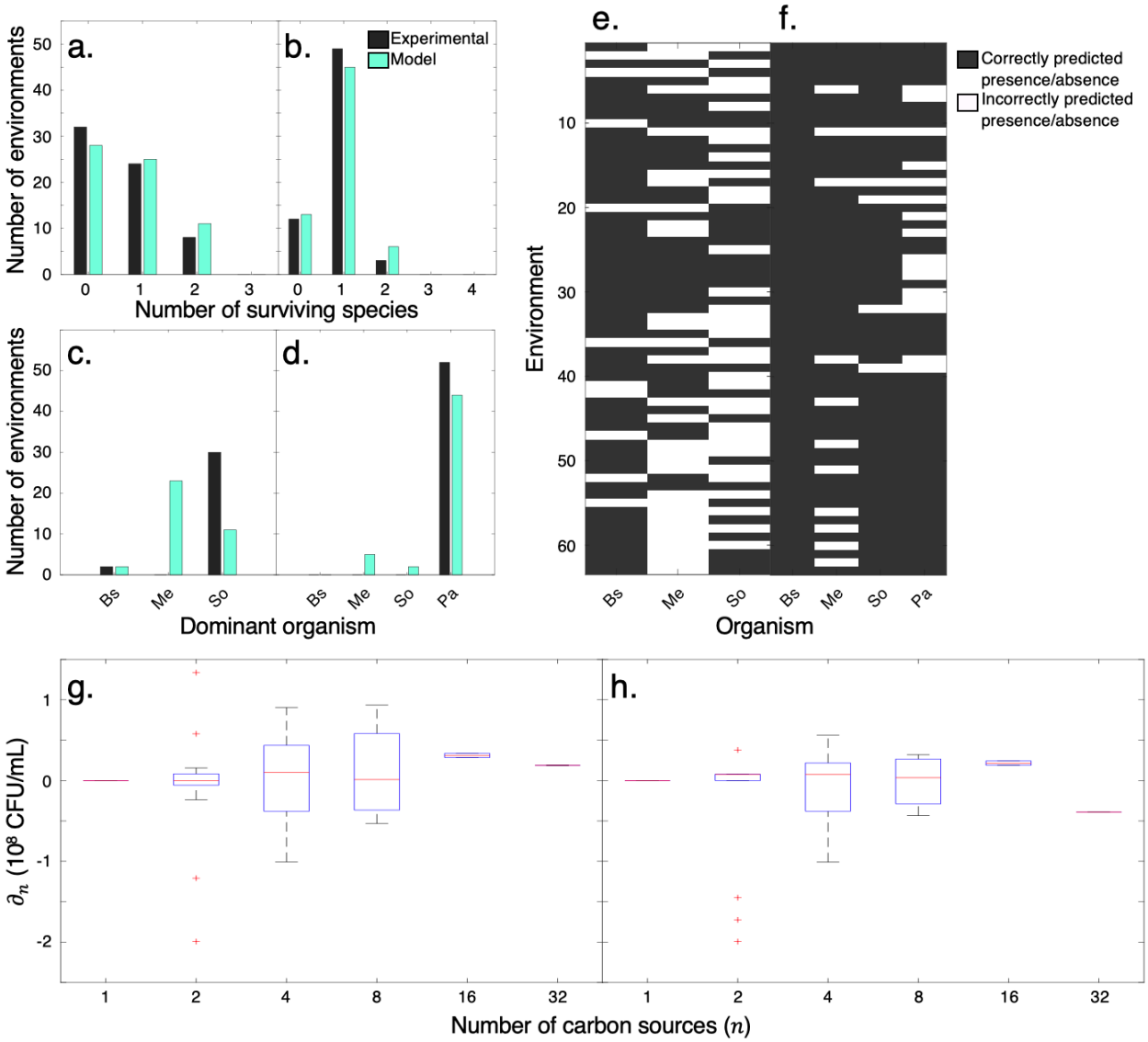
Supplementary Figure 22. Hierarchical clustering of 63 nutrient combinations according to Spearman correlations between community taxonomic distributions. Clusters C1 and C2a-c are designated according to higher-level branches. Compositions of the 63 environments are provided in Supplementary Table 3. C1 contains 17 conditions and C2 contains the remaining 46. Conditions that were clustered closely in C1 included D-mannose and D-glcNAc + D-galacturonate (conditions 15 and 42), which displayed almost equal distributions of *E. coli* and *P. fluorescens*, as well as conditions 26, 28, 29, and 38 (D-xylose, D-galactose, D-G6P, and D-trehalose + D,L-serine), which displayed distributions of *E. coli*, *P. aeruginosa*, and *P. fluorescens*. Subcluster C2a contains a variety of nutrient combinations ranging from one to 8 carbon sources. Despite this variability, these environments all resulted in communities that were dominated by *P. aeruginosa*. Subcluster C2b was mainly represented by environments containing organic acids, which yielded communities composed of *A. baylyi* and *P. aeruginosa*. Subcluster C2c contained the most environmentally-complex conditions (with both 16-carbon source and the 32-carbon source condition) and displayed some of the most pronounced differences when compared to those in other clusters. In particular, conditions 15, 32, and 42 (D-mannose, D-sorbitol, and D-glcNAc + D-galacturonate) displayed the most dissimilar community compositions compared to subcluster C2c.



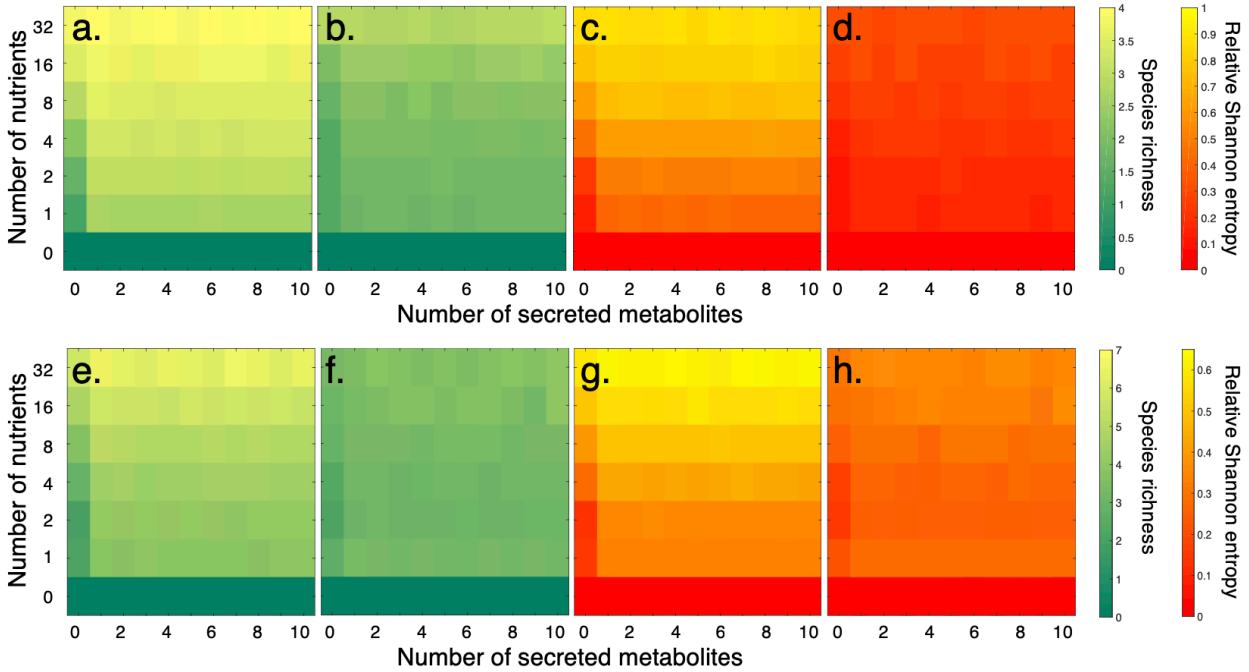
Supplementary Figure 23. Endpoint species relative abundances for multispecies communities com3 **(a)** and com4 **(b)**, and com3a **(c)**. *P. aeruginosa* was most often dominant in com4, outcompeting the other community members in 52 cases. In addition to the dominance of *P. aeruginosa* in com4, we noticed that *S. oneidensis* was dramatically overrepresented in our com3 experiment. This distribution was striking, as we expected *M. extorquens* to be most often dominant given its wider breadth of nutrient utilization capabilities in monoculture (Supplementary Figure 2a). **(c)**. An additional experiment containing *E. coli*, *M. extorquens*, and *S. oneidensis* (com3a, Supplementary Table 5) similarly highlighted the ability of a single organisms to overtake small communities. Relative abundances for com3 and com4 are adjusted based on a calibration of CFU counts to equal OD600 values. Unabbreviated environmental compositions for com3 and com4 are outlined in Supplementary Table 3. Unique conditions for com3a are: D-glucose (DGlc), pyruvate (Pyr), GlcNAc (GlcNAc), L-proline (Pro), L-threonine (Thr), and a no-carbon negative control (N). Missing bars indicate no growth in the specified condition.



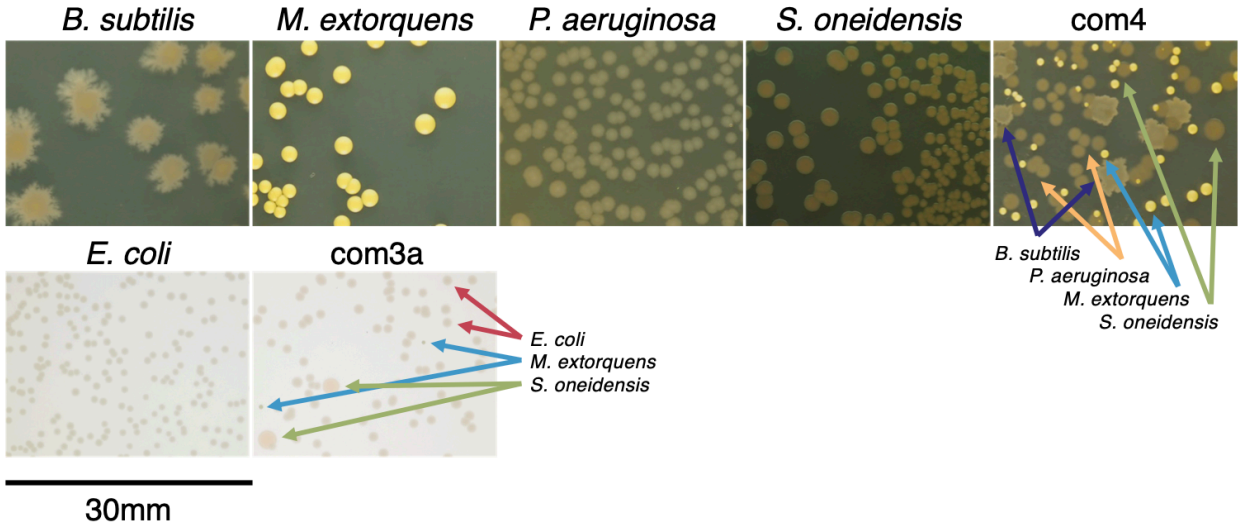
Supplementary Figure 24. Example consumer resource model-predicted growth phenotypes for com4 grown on D-glucose **(a)** and D-fructose, citrate, D-glcNAc, D-galacturonate, sucrose, acetate, glycerol, and D-cellobiose **(b)**, (conditions 1 and 59, respectively, Supplementary Table 3). Community compositions and yields recapitulate those observed experimentally (Supplementary Figure 7a, Supplementary Figure 22a) and reveal environment-specific nutrient utilization patterns. Specifically, the community grown on the second condition shown is predicted to consume all of the provided nutrients except D-galacturonate, which accumulates in the medium. Moreover, the community is predicted to first rapidly secrete then stop producing formate, which becomes diluted over time.



Supplementary Figure 25. Consumer resource model predictions compared to com3 and com4 experiments. **(a,b).** Frequency of number of environments by number of surviving species for com3 **(a)** and com4 **(b)** as observed experimentally (gray) and predicted by the consumer resource model (light blue). **(c,d).** Frequency of environments in which each organism was predicted to be dominant in com3 **(c)** and com4 **(d)**, as observed experimentally (gray) and predicted by the consumer resource model (light blue). Organisms are abbreviated as follows: Bs: *B. subtilis*, Me: *M. extorquens*, So: *S. oneidensis*, and Pa: *P. aeruginosa*. **(e,f).** Environment-by-environment comparison of model-predicted and experimentally-observed presence/absence of individual organisms in com3 **(e)** and com4 **(f)**. Species-specific accuracies are: 84.1% for Bs, 52.4% for Me, and 47.6% for So in com3, and 100% for Bs, 82.5% for Me, 92.1% for So, and 74.6% for Pa in com4. **(g,h).** Yield differences ∂_n between community yields on n carbon sources and the scaled sum of community yields on the corresponding single carbon sources for the *in silico* **(a)** com3, **(b)** com4 simulations. No statistically-significant increasing trend in yield differences was identified for either simulated community, despite the presence of secreted metabolites. Though secreted metabolites are predicted to be present across many simulations, (Supplementary Table 8) it may be that their concentrations are too low to produce marked changes in community yields at higher levels of environmental complexity, contrasting with our experimental results (Figure 2).



Supplementary Figure 26. Distributions of average model-predicted species richness and Shannon entropy for 4- and 13-species communities. **(a-d).** Average species richness **(a,b)** and relative Shannon entropy **(c,d)** for randomly-parametrized four-species communities containing either one **(b,d)** or no generalists **(a,c)**. **(e-h).** Average species richness **(e,f)** and relative Shannon entropy **(g,h)** for 13-species communities with either flat **(e,g)** or experimentally-derived **(f,h)** resource utilization capabilities. Relative Shannon entropy is computed by dividing the model-predicted Shannon entropy for a given simulation by the maximum theoretical Shannon entropy for a 4- or 13-species community.



Supplementary Figure 27. Differentiable colony morphologies for organisms assayed by agar plating.

Supplementary Table 1. Complete list of organisms used in experiments.

Organism name	Strain	ATCC	Temp. (˚C)	Flux Balance Model
Acinetobacter baylyi	ADP1	33305	37	[4]
Bacillus licheniformis	46	14580	37	
Bacillus subtilis	168	23857	37	[5]
Corynebacterium glutamicum	534	13032	37	
Escherichia coli	MG1655	25922	37	[6]
Lactococcus lactis	MG1363	19257	37	[7]
Methylobacterium extorquens	AM1	43645	30	[8]
Pseudomonas aeruginosa	PA01	9027	37	[9]
Pseudomonas fluorescens	28/5	13525	26	
Pseudomonas putida	KT2440	47054	26	[10]
Streptomyces coelicolor	A3(2)	23899	28	[11]
Shewanella oneidensis	MR-1	700550	30	[12]
Salmonella enterica	LT2	27106	37	[13]
Streptococcus thermophilus	LMG18311	BAA-250	37	
Vibrio natriegens	111	14048	26	

Supplementary Table 2. List and ontology of 95 carbon sources in Biolog PM1 plate.

Nutrient name	Class 1	Class 2	Class 3
1,2-Propanediol	Sugar alcohols	Monosaccharides	Carbohydrates
2-Aminoethanol	Sugar alcohols	Monosaccharides	Carbohydrates
2-Deoxyadenosine	Ribonucleosides	Nucleosides	Nucleic acids
a-D-Lactose	Disaccharides	Disaccharides	Carbohydrates
a-Hydroxyglutarate-g-lactone	Esters	Carboxylic acids	Organic acids
a-Hydroxybutyrate	2-Oxocarboxylic acids	Carboxylic acids	Organic acids
a-Ketobutyrate	2-Oxocarboxylic acids	Carboxylic acids	Organic acids
a-Ketoglutarate	2-Oxocarboxylic acids	Carboxylic acids	Organic acids
a-Methylgalactoside	Carbohydrate derivatives	Glycans	Carbohydrates
Acetate	Monocarboxylic acids	Carboxylic acids	Organic acids
Acetoacetate	3-Oxocarboxylic acids	Carboxylic acids	Organic acids
Adenosine	Ribonucleosides	Nucleosides	Nucleic acids
Adonitol	Sugar alcohols	Monosaccharides	Carbohydrates
b-Me-D-glucoside	Carbohydrate derivatives	Glycans	Carbohydrates
Bromosuccinate	Monocarboxylic acids	Carboxylic acids	Organic acids
Citrate	Tricarboxylic acids	Carboxylic acids	Organic acids
D-Alanine	Other amino acids	Amino acids	Peptides
D-Aspartate	Other amino acids	Amino acids	Peptides
D-Cellobiose	Disaccharides	Disaccharides	Carbohydrates
D-F6P	Sugar 6 phosphates	Sugar phosphates	Carbohydrates
D-Fructose	Ketoses	Monosaccharides	Carbohydrates
D-G1P	Sugar 1 phosphates	Sugar phosphates	Carbohydrates
D-G6P	Sugar 6 phosphates	Sugar phosphates	Carbohydrates
D-Galactose	Aldoses	Monosaccharides	Carbohydrates
D-Galacturonate	Uronic acids	Carboxylic acids	Organic acids
D-GlcNAc	Amino sugars	Monosaccharides	Carbohydrates
D-Galactonate-g-lactone	Esters	Carboxylic acids	Organic acids
D-Gluconate	Sugar acids	Monosaccharides	Carbohydrates
D-Glucosaminate	Aldonic acids	Carboxylic acids	Carbohydrates
D-Glucose	Aldoses	Monosaccharides	Carbohydrates
D-Glucuronate	Dicarboxylic acids	Carboxylic acids	Organic acids
D-Malate	Hydroxycarboxylic acids	Carboxylic acids	Organic acids
D-Mannitol	Sugar alcohols	Monosaccharides	Carbohydrates
D-Mannose	Aldoses	Monosaccharides	Carbohydrates
D-Melibiose	Disaccharides	Disaccharides	Carbohydrates

D-Psicose	Ketoses	Monosaccharides	Carbohydrates
D-Ribose	Aldoses	Monosaccharides	Carbohydrates
D-Saccharate	Dicarboxylic acids	Carboxylic acids	Organic acids
D-Serine	Other amino acids	Amino acids	Peptides
D-Sorbitol	Sugar alcohols	Monosaccharides	Carbohydrates
D-Threonine	Other amino acids	Amino acids	Peptides
D-Trehalose	Disaccharides	Disaccharides	Carbohydrates
D-Xylose	Aldoses	Monosaccharides	Carbohydrates
D,L-G3P	Sugar 3 phosphates	Sugar phosphates	Carbohydrates
D,L-Malate	Hydroxycarboxylic acids	Carboxylic acids	Organic acids
Dulcitol	Sugar alcohols	Monosaccharides	Carbohydrates
Formate	Monocarboxylic acids	Carboxylic acids	Organic acids
Fumarate	Dicarboxylic acids	Carboxylic acids	Organic acids
Glucuronamide	Hexoses	Monosaccharides	Carbohydrates
Glycerol	Sugar 3 phosphates	Sugar phosphates	Carbohydrates
Glycolate	Monocarboxylic acids	Carboxylic acids	Organic acids
Glycyl-L-aspartate	Dipeptides	Amino acids	Peptides
Glycyl-L-glutamate	Dipeptides	Amino acids	Peptides
Glycyl-L-proline	Dipeptides	Amino acids	Peptides
Glyoxylate	2-Oxocarboxylic acids	Carboxylic acids	Organic acids
Inosine	Ribonucleosides	Nucleosides	Nucleic acids
L-Alanine	Common amino acids	Amino acids	Peptides
L-Alanylglycine	Dipeptides	Amino acids	Peptides
L-Arabinose	Aldoses	Monosaccharides	Carbohydrates
L-Asparagine	Common amino acids	Amino acids	Peptides
L-Aspartate	Common amino acids	Amino acids	Peptides
L-Fucose	Deoxy sugars	Monosaccharides	Carbohydrates
L-Galactonate-g-lactone	Esters	Carboxylic acids	Organic acids
L-Glutamate	Common amino acids	Amino acids	Peptides
L-Glutamine	Common amino acids	Amino acids	Peptides
L-Lactate	Hydroxycarboxylic acids	Carboxylic acids	Organic acids
L-Lyxose	Aldoses	Monosaccharides	Carbohydrates
L-Malate	Hydroxycarboxylic acids	Carboxylic acids	Organic acids
L-Proline	Common amino acids	Amino acids	Peptides
L-Rhamnose	Deoxy sugars	Monosaccharides	Carbohydrates
L-Serine	Common amino acids	Amino acids	Peptides
L-Threonine	Common amino acids	Amino acids	Peptides
Lactulose	Disaccharides	Disaccharides	Carbohydrates
M-Acetyl-mannosamine	Hexosamines	Monosaccharides	Carbohydrates

M-Hydroxyphenylacetate	Monocarboxylic acids	Carboxylic acids	Organic acids
M-Inositol	Sugar alcohols	Monosaccharides	Carbohydrates
M-Tartarate	Hydroxycarboxylic acids	Carboxylic acids	Organic acids
Maltose	Disaccharides	Disaccharides	Carbohydrates
Maltotriose	Polysaccharides	Polysaccharides	Carbohydrates
Methylpyruvate	Esters	Carboxylic acids	Organic acids
Methylsuccinate	Dicarboxylic acids	Carboxylic acids	Organic acids
Mucate	Aldaric acids	Carboxylic acids	Organic acids
P-Hydroxyphenylacetate	Monocarboxylic acids	Carboxylic acids	Organic acids
Phenylethylamine	Amines	Amino acids	Peptides
Propionate	Monocarboxylic acids	Carboxylic acids	Organic acids
Pyruvate	2-Oxocarboxylic acids	Carboxylic acids	Organic acids
Succinate	Dicarboxylic acids	Carboxylic acids	Organic acids
Sucrose	Disaccharides	Disaccharides	Carbohydrates
Thymidine	Deoxyribonucleosides	Nucleosides	Nucleic acids
Tricarballylate	Tricarboxylic acids	Carboxylic acids	Organic acids
Tween 20	Polysorbates	Surfactants	Surfactants
Tween 40	Polysorbates	Surfactants	Surfactants
Tween 80	Polysorbates	Surfactants	Surfactants
Tyramine	Amines	Amino acids	Peptides
Uridine	Ribonucleosides	Nucleosides	Nucleic acids

Supplementary Table 3. Nutrient pairings in 32-carbon source experiments (com3, com4, com13). Each individual nutrient combination is numbered and highlighted with alternating background colors, for a total of 63 unique conditions containing varying numbers of carbon sources (nCS).

nCS = 1, 32 envs.	nCS = 2, 16 envs.	nCS = 4, 8 envs.	nCS = 8, 4 envs.	nCS = 16, 2 envs.	nCS = 32, 1 env.
1 D-Glucose	33 D-Glucose	D-Glucose	D-Glucose	D-Glucose	D-Glucose
2 L-Glutamine	D-Sorbitol	49 D-Sorbitol	D-Sorbitol	D-Sorbitol	D-Sorbitol
3 L-Alanine	34 L-Glutamine	L-Glutamine	L-Glutamine	L-Glutamine	L-Glutamine
4 Pyruvate	Formate	Formate	57 Formate	Formate	Formate
5 L-Proline	35 L-Alanine	L-Alanine	L-Alanine	L-Alanine	L-Alanine
6 D-Trehalose	a-D-Lactose	50 a-D-Lactose	a-D-Lactose	a-D-Lactose	a-D-Lactose
7 L-Threonine	36 Pyruvate	Pyruvate	Pyruvate	Pyruvate	Pyruvate
8 D-Mannitol	D-G6P	D-G6P	D-G6P	61 D-G6P	D-G6P
9 D-Fructose	37 L-Proline	L-Proline	L-Proline	L-Proline	L-Proline
10 D-GlcNAc	D-Galactose	51 D-Galactose	D-Galactose	D-Galactose	D-Galactose
11 Sucrose	38 D-Trehalose	D-Trehalose	D-Trehalose	D-Trehalose	D-Trehalose
12 Glycerol	D,L-Serine	D,L-Serine	58 D,L-Serine	D,L-Serine	D,L-Serine
13 D-Gluconate	39 L-Threonine	L-Threonine	L-Threonine	L-Threonine	L-Threonine
14 Propionate	D-Xylose	52 D-Xylose	D-Xylose	D-Xylose	D-Xylose
15 D-Mannose	40 D-Mannitol	D-Mannitol	D-Mannitol	D-Mannitol	D-Mannitol
16 Succinate	L-Arabinose	L-Arabinose	L-Arabinose	L-Arabinose	L-Arabinose
17 Maltose	41 D-Fructose	D-Fructose	D-Fructose	D-Fructose	63 D-Fructose
18 D,L-Malate	Citrate	Citrate	Citrate	Citrate	Citrate
19 D-Ribose	42 D-GlcNAc	53 D-GlcNAc	D-GlcNAc	D-GlcNAc	D-GlcNAc
20 L-Lactate	D-Galacturonate	D-Galacturonate	59 D-Galacturonate	D-Galacturonate	D-Galacturonate
21 D-Cellobiose	43 Sucrose	Sucrose	Sucrose	Sucrose	Sucrose
22 Acetate	Acetate	Acetate	Acetate	Acetate	Acetate
23 D-Galacturonate	44 Glycerol	54 Glycerol	Glycerol	Glycerol	Glycerol
24 Citrate	D-Cellobiose	D-Cellobiose	D-Cellobiose	62 D-Cellobiose	D-Cellobiose
25 L-Arabinose	45 D-Gluconate	D-Gluconate	D-Gluconate	D-Gluconate	D-Gluconate
26 D-Xylose	L-Lactate	55 L-Lactate	L-Lactate	L-Lactate	L-Lactate
27 D,L-Serine	46 Propionate	Propionate	Propionate	Propionate	Propionate
28 D-Galactose	D-Ribose	D-Ribose	D-Ribose	D-Ribose	D-Ribose
29 D-G6P	47 D-Mannose	D-Mannose	60 D-Mannose	D-Mannose	D-Mannose
30 a-D-Lactose	D,L-Malate	56 D,L-Malate	D,L-Malate	D,L-Malate	D,L-Malate
31 Formate	48 Succinate	Succinate	Succinate	Succinate	Succinate
32 D-Sorbitol	Maltose	Maltose	Maltose	Maltose	Maltose

Supplementary Table 4. Nutrient pairings in 5-carbon source experiments (com3a, com13a). Each background color indicates one environmental composition, for a total of 31 unique conditions containing varying numbers of carbon sources (nCS).

nCS = 1, 5 envs.	nCS = 2, 10 envs.	nCS = 3, 10 envs.	nCS = 4, 5 envs.	nCS = 5, 1 env
D-Glucose	D-Glucose	D-Glucose	D-Glucose	D-Glucose
Pyruvate	Pyruvate	Pyruvate	Pyruvate	Pyruvate
D-GlcNAc	D-Glucose	D-GlcNAc	D-GlcNAc	D-GlcNAc
L-Proline	D-GlcNAc	D-Glucose	L-Proline	L-Proline
L-Threonine	D-Glucose	Pyruvate	D-Glucose	L-Threonine
	L-Proline	L-Proline	Pyruvate	
	D-Glucose	D-Glucose	D-GlcNAc	
	L-Threonine	Pyruvate	L-Threonine	
	Pyruvate	L-Threonine	Pyruvate	
	D-GlcNAc	D-Glucose	D-GlcNAc	
	Pyruvate	D-GlcNAc	L-Proline	
	L-Proline	L-Proline	L-Threonine	
	Pyruvate	D-Glucose	D-Glucose	
	L-Threonine	D-GlcNAc	Pyruvate	
	D-GlcNAc	L-Threonine	L-Proline	
	L-Proline	D-Glucose	L-Threonine	
	D-GlcNAc	L-Proline	D-Glucose	
	L-Threonine	L-Threonine	D-GlcNAc	
	L-Proline	Pyruvate	L-Proline	
	L-Threonine	D-GlcNAc	L-Threonine	
		L-Proline		
		Pyruvate		
		D-GlcNAc		
		L-Threonine		
		D-GlcNAc		
		L-Proline		
		L-Threonine		

Supplementary Table 5. Descriptions of combinatorial nutrient experiments. All experiments were carried out at 30˚C and provided 50 mM C per well (com14 also grown at 25 mM C).

Name	Number of organisms	Organisms	Number of nutrients	Number of passages	Passaging frequency
com3	3	*B. subtilis*, *M. extorquens*, *S. oneidensis*	32 (Supplementary Table 3)	6	48h
com3a	3	*E. coli*, *M. extorquens*, *S. oneidensis*	5 (Supplementary Table 4)	0 (24h batch)	0
com4	4	com3 + *P. aeruginosa*	32 (Supplementary Table 3)	6	48h
com13	13	com4 + *A. baylyi*, *B. licheniformis*, *C. glutamicum*, *E. coli*, *L. lactis*, *P. fluorescens*, *P. putida*, *S. coelicolor*, *V. natriegens*	32 (Supplementary Table 3)	6	48h
com13a	13	com13	5 (Supplementary Table 4)	6	24h
com14	14	com13 + *S. enterica*	4 (Supplementary Figure 3a)	0 (24h batch)	0

Supplementary Table 6. Regression coefficients for individual carbon sources significantly positively or negatively associated with higher Shannon entropy in com13.

Nutrient	Coefficient	P-value
L-Glutamine	0.619	2.78E-02
D,L-Malate	0.858	2.40E-03
D-Ribose	0.795	4.90E-03
D-Cellobiose	1.049	2.00E-04
Citrate	1.049	2.00E-04
D-Xylose	0.673	1.69E-02
D-Galactose	0.752	7.70E-03
D-Glucose	-0.826	3.50E-03
D-Fructose	-0.951	8.00E-04
Glycerol	-1.285	0.00E+00
D-Gluconate	-0.872	2.10E-03
Propionate	-1.205	0.00E+00
L-Arabinose	-1.105	1.00E-04

Supplementary Table 7. Minimal medium composition for flux-balance modeling.

Metabolite	BIGG metabolite name[14]
Calcium	ca2[e]
Chloride	cl[e]
Cobalt	cobalt2[e]
Copper^{2+}	cu2[e]
Iron^{2+}	fe2[e]
Iron^{3+}	fe3[e]
Hydrogen	h2[e]
Water	h2o[e]
Potassium	k[e]
Magnesium	mg2[e]
Manganese	mn2[e]
Molybdate	mobd[e]
Sodium	na1[e]
Ammonium	nh4[e]
Nickel^{2+}	ni2[e]
Nitrate	no3[e]
Oxygen	o2[e]
Phosphate	pi[e]
Sulfate	so4[e]
Zinc^{2+}	zn2[e]

Supplementary Table 8. List of unique organic molecules predicted to be secreted across all flux-balance simulations.

Metabolite	BIGG metabolite name[14]
Acetate	ac[e]
D-alanine	ala-D[e]
Allantoin	alltn[e]
Citrate	cit[e]
Formate	for[e]
Formamide	frmd[e]
Fumarate	fum[e]
Glycolaldehyde	gcald[e]
D-gluconate	glcn-D[e]
Glycolate	glyclt[e]
L-lactate	lac-L[e]
D-malate	mal-D[e]
R-pantothenate	pnto-R[e]
Riboflavin	ribflv[e]
Spermidine	spmd[e]
Succinate	succ[e]
Trehalose	tre[e]
Urea	urea[e]

Supplementary Table 9. Descriptions and quantities for consumer resource model state variables and parameters. Resource supply and dilution rates are defined by 48-hour experimental timescale.

Variable/parameter	Description	Units	Quantity
N_i	Abundance of organism i	CFU/mL	variable
R_α	Abundance of nutrient α	g/mL	variable
$C_{i\alpha}$	Uptake rate per unit concentration of resource α by organism i	mL/hr	variable
$D_{\alpha\beta i}$	Proportion of nutrient α converted to nutrient β by organism i	unitless	variable
l_α	Leakage fraction for resource α	unitless	0.25 [15]
g_i	Conversion factor from energy uptake to growth rate for organism i	1/energy	1 [15]
w_α	Energy content of resource α	energy/g	1×10^8
$k_{i,a}$	Half velocity constant for resource uptake	g/mL	3000
m_i	Minimal energy uptake for maintenance of each species i	energy/hr	0.05
κ_α	External supply of resource α	g/mL/hr	$\frac{1.5g/mL}{48h}$
d	Dilution rate	1/hr	$\left(\frac{10\mu l}{300\mu l}\right) \times 48h^{-1}$

References

1. Eguíluz, V. M. *et al.* Scaling of species distribution explains the vast potential marine prokaryote diversity. *Sci. Rep.* **9**, 18710 (2019).
2. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5970–5 (2016).
3. Hoffmann, K. H. *et al.* Power law rank-abundance models for marine phage communities. *FEMS Microbiol. Lett.* **273**, 224–8 (2007).
4. Durot, M. *et al.* Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst. Biol.* **2**, 85 (2008).
5. Henry, C. S., Zinner, J. F., Cohoon, M. P. & Stevens, R. L. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biol.* **10**, R69 (2009).
6. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011).
7. Flahaut, N. A. L. *et al.* Genome-scale metabolic model for Lactococcus lactis MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.* **97**, 8729–8739 (2013).
8. Peyraud, R. *et al.* Genome-scale reconstruction and system level investigation of the metabolic network of Methylobacterium extorquens AM1. *BMC Syst. Biol.* **5**, (2011).
9. Oberhardt, M. A., Puchałka, J., Fryer, K. E., Martins dos Santos, V. A. P. & Papin, J. A. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J. Bacteriol.* **190**, 2790–803 (2008).
10. Nogales, J., Palsson, B. Ø. & Thiele, I. A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* **2**, 79 (2008).
11. Kim, M. *et al.* Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in Streptomyces coelicolor A3(2). *Biotechnol. J.* **9**, 1185–1194 (2014).
12. Pinchuk, G. E. *et al.* Constraint-Based Model of Shewanella oneidensis MR-1 Metabolism: A Tool for Data Analysis and Hypothesis Generation. *PLoS Comput. Biol.* **6**, e1000822 (2010).
13. Thiele, I. *et al.* A community effort towards a knowledge-base and mathematical model of the human pathogen Salmonella Typhimurium LT2. *BMC Syst. Biol.* **5**, 8 (2011).
14. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522 (2016).
15. Marsland, R., Cui, W. & Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Sci. Rep.* **10**, 1–17 (2020).