RESEARCH ARTICLE

MICROBIOLOGY

Emergent simplicity in microbial community assembly

Joshua E. Goldford[1,2]*, Nanxi Lu[3]*, Djordje Bajić[3], Sylvie Estrela[3], Mikhail Tikhonov[4,5], Alicia Sanchez-Gorostiaga[3], Daniel Segrè[1,6,7], Pankaj Mehta[1,7]†, Alvaro Sanchez[2,3]†

A major unresolved question in microbiome research is whether the complex taxonomic architectures observed in surveys of natural communities can be explained and predicted by fundamental, quantitative principles. Bridging theory and experiment is hampered by the multiplicity of ecological processes that simultaneously affect community assembly in natural ecosystems. We addressed this challenge by monitoring the assembly of hundreds of soil- and plant-derived microbiomes in well-controlled minimal synthetic media. Both the community-level function and the coarse-grained taxonomy of the resulting communities are highly predictable and governed by nutrient availability, despite substantial species variability. By generalizing classical ecological models to include widespread nonspecific cross-feeding, we show that these features are all emergent properties of the assembly of large microbial communities, explaining their ubiquity in natural microbiomes.

Microbial communities play critical roles in a wide range of natural processes, from animal development and host health to biogeochemical cycles (*1*–*3*). Recent advances in DNA sequencing have allowed us to map the composition of these communities with high resolution. This has motivated a surge of interest in understanding the ecological mechanisms that govern microbial community assembly and function (*4*). A quantitative, predictive understanding of microbiome ecology is required to design effective strategies to rationally manipulate microbial communities toward beneficial states.

Surveys of microbiome composition across a wide range of ecological settings, from the ocean to the human body (*2*, *3*), have revealed intriguing empirical patterns in microbiome organization. These widely observed properties include high microbial diversity, the coexistence of multiple closely related species within the same functional group, functional stability despite large species turnover, and different degrees of determinism in the association between nutrient availability and taxonomic composition at different phylogenetic levels (*3*, *5*–*10*). These observations have led to the proposal that common organizational principles exist in microbial community assembly (*6*, *7*). However, the lack of a theory of microbiome assembly is hindering progress toward explaining and interpreting these empirical findings, and it remains unknown which of the functional and structural features exhibited by microbiomes reflect specific local adaptations at the host or microbiome level (*10*) and which are generic properties of complex, self-assembled microbial communities.

Efforts to connect theory and experiments to understand microbiome assembly have typically relied on manipulative bottom-up experiments with a few species (*11*–*13*). Although this approach is useful for providing insights into specific mechanisms of interactions, it is unclear to what extent findings from these studies scale up to predict the generic properties of large microbial communities or the interactions therein. Of note is the ongoing debate about the relative contributions of competition and facilitation (*14*, *15*) and the poorly understood role that high-order interactions play in microbial community assembly (*11*, *16*, *17*). To move beyond empirical observations and connect statistical patterns of microbiome assembly with ecological theory, we need to study the assembly of large numbers of large multispecies microbiomes under highly controlled and well-understood conditions that allow proper comparison between theory and experiment.

Assembly of large microbial communities on a single limiting resource

To meet this challenge, we followed a high-throughput ex situ cultivation protocol to monitor the spontaneous assembly of ecologically stable microbial communities derived from natural habitats in well-controlled environments; we used synthetic (M9) minimal media containing a single externally supplied source of carbon, as well as sources of all of the necessary salts and chemical elements required for microbial life (Fig. 1A). Intact microbiota suspensions were extracted from diverse natural ecosystems, such as various soils and plant leaf surfaces (methods). Suspensions of microbiota from these environments were highly diverse and taxonomically rich (fig. S1), ranging between 110 and 1290 exact sequence variants (ESVs). We first inoculated 12 of these suspensions of microbiota into fresh minimal media with glucose as the only added carbon source and allowed the cultures to grow at 30°C in static broth. We then passaged the mixed cultures in fresh media every 48 hours with a fixed dilution factor of $D = 8 \times 10^{-3}$ for a total of 12 transfers (~84 generations). At the end of each growth cycle, we used 16*S* ribosomal RNA (rRNA) amplicon sequencing to assay the community composition (Fig. 1A and methods). High-resolution sequence denoising allowed us to identify ESVs, which revealed community structure at single-nucleotide resolution (*18*).

Most communities stabilized after ~60 generations, reaching stable population equilibria in nearly all cases (Fig. 1B and fig. S2). For all of the 12 initial ecosystems, we observed large multispecies communities after stabilization that ranged from 4 to 17 ESVs at a sequencing depth of 10,000 reads; further analysis indicated that this is a conservative estimate of the total richness in our communities (figs. S3 and S4 and methods). We confirmed the taxonomic assignments generated from amplicon sequencing by culture-dependent methods, including the isolation and phenotypic characterization of all dominant genera within a representative community (fig. S5).

Convergence of bacterial community structure at the family taxonomic level

High-throughput isolation and stabilization of microbial consortia allowed us to explore the rules governing the assembly of bacterial communities in well-controlled synthetic environments. At the species (ESV) level of taxonomic resolution, the 12 natural communities assembled into highly variable compositions (Fig. 1C). However, when we grouped ESVs by higher taxonomic ranks, we found that all 12 stabilized communities—with very diverse environmental origins—converged into similar family-level community structures dominated by Enterobacteriaceae and Pseudomonadaceae (Fig. 1D). In other words, a similar family-level composition arose in all communities despite their very different starting points. This is further illustrated in fig. S6, where we show that the temporal variability (quantified by the β diversity) in family-level composition is comparable to the variability across independent replicates. The same is not true when we compare taxonomic structure at the subfamily (genus) level.

[1]Graduate Program in Bioinformatics and Biological Design Center, Boston University, Boston, MA 02215, USA. [2]The Rowland Institute at Harvard University, Cambridge, MA 02142, USA. [3]Department of Ecology and Evolutionary Biology, Microbial Sciences Institute, Yale University, New Haven, CT 06511, USA. [4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [5]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA. [6]Departments of Biology and Biomedical Engineering, Boston University, Boston, MA 02215, USA. [7]Department of Physics, Boston University, Boston, MA 02215, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: alvaro.sanchez@yale.edu (A.S.); pankajm@bu.edu (P.M.)

To better understand the origin of the taxonomic variability observed below the family level, we started eight replicate communities from each of the 12 starting microbiome suspensions (inocula) and propagated them in minimal media with glucose, as in the previous experiment. Given that the replicate communities were assembled in identical habitats and were inoculated from the same pool of species, any observed variability in community composition across replicates would suggest that random colonization from the regional pool and microbe-microbe interactions are sufficient to generate alternative species-level community assembly.

For most of the inocula (9 out of 12), replicate communities assembled into alternative stable ESV-level compositions, while still converging to the same family-level attractor described in Fig. 1E (see also fig. S6). One representative example is shown in Fig. 1, F and G; all eight replicates from the same starting inoculum assembled into strongly similar family-level structures, which were quantitatively consistent with those found before (Fig. 1D). However, different replicates contained alternative Pseudomonadaceae ESVs, and the Enterobacteriaceae fraction was constituted by either an ESV from the *Klebsiella* genus or a guild consisting of variable subcompositions of *Enterobacter*, *Raoultella*, and/or *Citrobacter* as the dominant taxa.

For the remaining (3 out of 12) inocula, all replicates exhibited strongly similar population dynamics to each other and equilibrated to similar population structures at all levels of taxonomic resolution (fig. S7). The reproducibility in population dynamics between replicate communities indicates that experimental error is not the main source of variability in community composition. The population bottlenecks introduced by the serial dilutions in fresh media have only a modest effect on the observed variability in population dynamics (fig. S8). However, the dilution factor can influence community assembly through means other than introducing population bottlenecks—for instance, by setting the number of generations in between dilutions and by diluting, to a greater or lesser extent, the environment generated in a previous growth period.

Despite the observed species-level variation in community structure, the existence of family-level attractors suggests that fundamental rules govern community assembly. Recent work on natural communities has consistently found that environmental filtering selects for convergent function across similar habitats, while allowing for taxonomic variability within each functional class (*5*, *6*). In our assembled communities in glucose media, fixed proportions of Enterobacteriaceae and Pseudomonadaceae may have emerged owing to a competitive advantage, given the well-known glucose uptake capabilities of the phosphotransferase system in Enterobacteriaceae and ABC (adenosine triphosphate–binding cassette) transporters in Pseudomonadaceae (*19*). This suggests that the observed family-level attractor may change if we add a different carbon source to our synthetic media.

To determine the effect of the externally provided carbon source on environmental filtering, we repeated the community assembly experiments with eight replicates of all 12 natural communities, this time using one of two alternative single carbon sources—citrate or leucine—instead of glucose. Consistent with previous experiments using glucose minimal media, communities that assembled on citrate or leucine contained large numbers of species: At a sequencing depth of

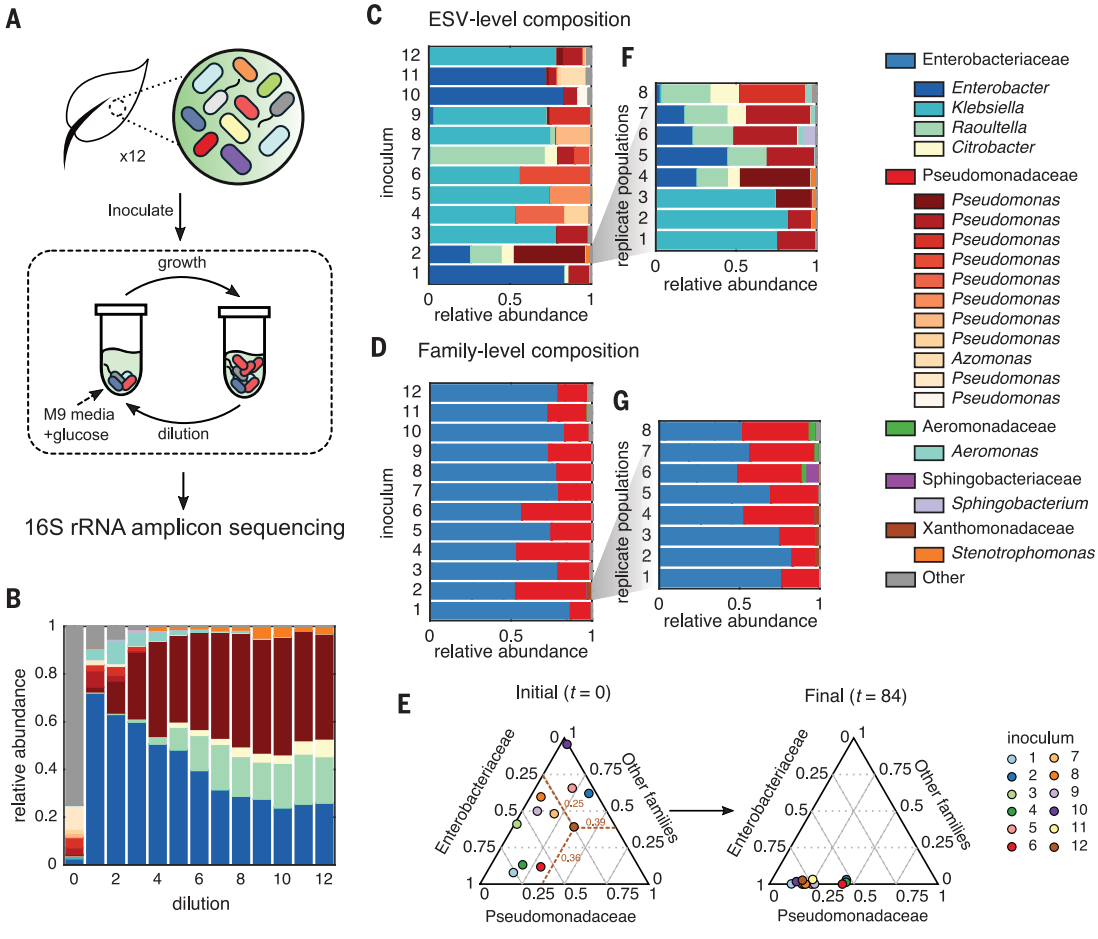

Fig. 1. Top-down assembly of bacterial consortia. (**A**) Experimental scheme: Large ensembles of taxa were obtained from 12 leaf and soil samples and used as inocula in serial dilution cultures containing synthetic media supplemented with glucose as the sole carbon source. After each transfer, 16S rRNA amplicon sequencing was used to assay bacterial community structure. (**B**) Analysis of the structure of a representative community (from inoculum 2) after every dilution cycle (about seven generations) reveals a five-member consortium from the *Enterobacter*, *Raoultella*, *Citrobacter*, *Pseudomonas*, and *Stenotrophomonas* genera. The community composition of all 12 starting inocula after 84 generations is shown at (**C**) the exact sequence variant (ESV) level or (**D**) the family taxonomic level, converging to characteristic fractions of Enterobacteriaceae and Pseudomonadaceae. (**E**) Simplex representation of family-level taxonomy before (t = 0) and after (t = 84) the passaging experiment. (**F** and **G**) Experiments were repeated with eight replicates from a single source (inocula 2). Communities converged to very similar family-level distributions (G) but displayed characteristic variability at the genus and species level (F).

10,000 reads, communities stabilized on leucine contained 6 to 22 ESVs, and communities stabilized on citrate contained 4 to 22 ESVs. As was the case for glucose, replicate communities assembled on citrate and leucine also differed widely in their ESV-level compositions, while converging to carbon source–specific family-level attractors (Fig. 2A and figs. S9 and S10).

Family-level community similarity (Renkonen similarity) was, on average, higher between communities passaged on the same carbon source (median, 0.88) than between communities passaged from the same environmental sample (median, 0.77; one-tailed Kolmogorov-Smirnov test, $P < 10^{-5}$; fig. S11). Communities stabilized on citrate media had a significantly lower fraction of Enterobacteriaceae (Mann-Whitney U test, $P < 10^{-5}$) and were enriched in Flavobacteriaceae (Mann-Whitney U test, $P < 10^{-5}$) relative to communities grown on glucose; communities stabilized on leucine media had no growth of Enterobacteriaceae and were enriched in Comamonadaceae relative to communities grown on glucose (Mann-Whitney U test, $P < 10^{-5}$) or citrate (Mann-Whitney U test, $P < 10^{-5}$).

These results suggest that the supplied source of carbon governs community assembly. To quantify this effect, we used a machine learning approach and trained a support vector machine to predict the identity of the supplied carbon source from the family-level community composition. We obtained a cross-validation accuracy of 97.3% (Fig. 2B and methods). Importantly, we found that considering the tails of the family-level distribution (as opposed to just the two dominant taxa) increased the predictive accuracy (Fig. 2B), which indicates that carbon source–mediated determinism in community assembly extends to the entire family-level distribution, including the rarer members.

Rather than selecting for the most fit single species, our environments select complex communities that contain fixed fractions of multiple coexisting families whose identities are determined by the carbon source in a strong and predictable manner (fig. S11). We hypothesized that taxonomic convergence might reflect selection by functions that are conserved at the family level. Consistent with this idea, we find that the inferred community metagenomes assembled in each type of carbon source exhibit substantial clustering by the supplied carbon source (Fig. 2C) and are enriched in pathways for its metabolism (fig. S11). When we spread the stabilized communities on agarose plates, we routinely found multiple identifiable colony morphologies per plate, showing that multiple taxa within each community are able to grow independently on (and thus compete for) the single supplied carbon source. This suggests that the genes and pathways that confer each community with the ability to metabolize the single supplied resource are distributed among multiple taxa in the community, rather than being present only in the best-competitor species.

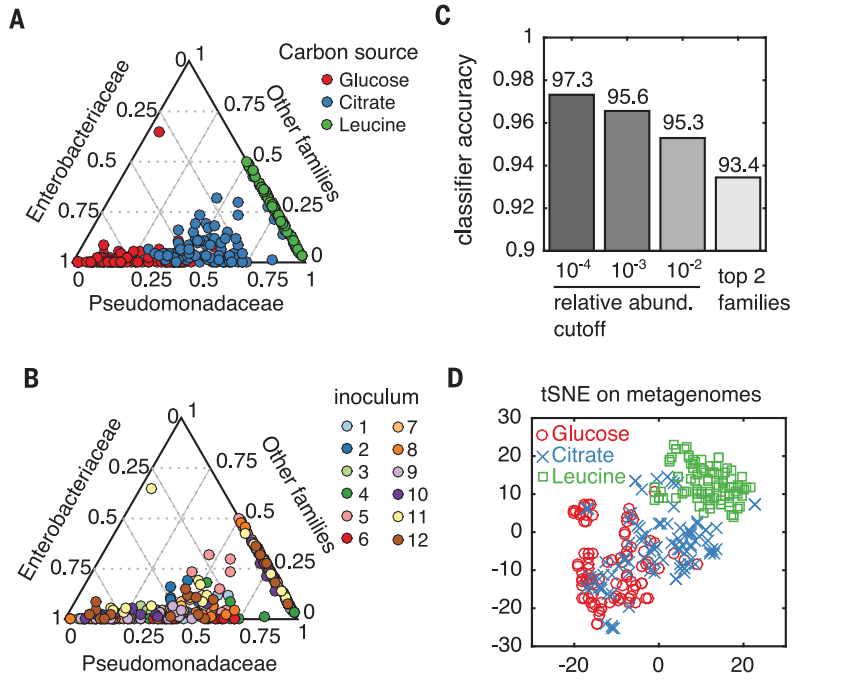

Fig. 2. Family-level and metagenomic attractors are associated with different carbon sources. (**A** and **B**) Family-level community compositions are shown for all replicates across 12 inocula grown on either glucose, citrate, or leucine as the limiting carbon source. Data points are colored by carbon source (A) or initial inoculum (B). (**C**) A support vector machine (methods) was trained to classify the carbon source from the family-level community structure. Low-abundance taxa were filtered using a predefined cutoff (*x* axis) before training and performing 10-fold cross-validation (averaged 10 times). Classification accuracy with only Enterobacteriaceae and Pseudomonadaceae resulted in a model with ~93% accuracy (rightmost bar), while retaining low-abundance taxa (relative abundance cutoff of 10^{-4}) yielded a classification accuracy of ~97% (leftmost bar). (**D**) Metagenomes were inferred using PICRUSt (*40*) and dimensionally reduced using *t*-distributed stochastic neighbor embedding (tSNE), revealing that carbon sources are strongly associated with the predicted functional capacity of each community.

Widespread metabolic facilitation stabilizes competition and promotes coexistence

Classic consumer-resource models indicate that when multiple species compete for a single, externally supplied growth-limiting resource, the only possible outcome is competitive exclusion unless specific circumstances apply (*20*–*25*). However, this scenario does not adequately reflect the case of microbes, whose ability to engineer their own environments both in the laboratory (*26*–*29*) and in nature (*30*, *31*) is well documented. Thus, we hypothesized that the observed coexistence of competitor species in our experiments may be attributed to the generic tendency of microbes to secrete metabolic by-products into the environment, which could then be used by other community members.

To determine the plausibility of niche creation mediated by metabolic by-products, we analyzed one representative glucose community in more depth. We isolated members of the four most abundant genera in this community (*Pseudomonas*, *Raoultella*, *Citrobacter*, and *Enterobacter*), which together represented ~97% of the total population in that community (Fig. 3A). These isolates had different colony morphologies and were also phenotypically distinct (fig. S5). All isolates were able to form colonies in glucose agarose plates, and all grew independently in glucose as the only carbon source, which indicates that each isolate could compete for the single supplied resource. All four species were able to stably coexist with one another when the community was reconstituted from the bottom up by mixing the isolates together (fig. S5). To test the potential for cross-feeding interactions in this community, we grew monocultures of the four isolates for 48 hours in synthetic M9 media containing glucose as the only carbon source (Fig. 3B). At the end of the growth period, the glucose concentration was too low to be detected, indicating that all of the supplied carbon had been consumed and that any carbon present in the media originated from metabolic by-products previously secreted by the cells. To test whether these secretions were enough to support growth of the other species in that community, we filtered the leftover media to remove cells and added it to fresh M9 media as the only source of carbon (Fig. 3B). We found that all isolates were able to grow on every other

isolate's secretions (e.g., Fig. 3C), forming a fully connected facilitation network (Fig. 3D). Growth on the secretions of other community members was strong, often including multiple diauxic shifts (fig. S12), and the amount of growth on secretions was comparable to that on glucose (fig. S13), suggesting that the pool of secreted by-products is diverse and abundant in this representative community.

To find out whether growth on metabolic by-products is common among our communities, we thawed 95 glucose-stabilized communities (seven or eight replicates from each of 12 initial environmental habitats) and grew them again on glucose as the only carbon source for an extra 48-hour cycle. In all 95 communities, glucose was completely exhausted after 24 hours of growth (Fig. 3E), yet most communities continued growing after the glucose had been depleted (Fig. 3E). Moreover, community growth on the secreted by-products was strong: On average, communities produced ~25% as much biomass on the secretions alone as they did over the first 24 hours when glucose was present (Fig. 3F). Propidium iodide staining and phase-contrast imaging of communities at the single-cell level identified low numbers of permeabilized or obviously lysed cells (fig. S14). This supports the hypothesis that metabolic by-product secretion (rather than cell lysis) is the dominant source of the observed cross-feeding. However, lytic events that leave no trace in the form of empty bacterial cell envelopes would not have been detected in our micrographs, so a contribution from cell death to our results cannot be entirely ruled out. Other mechanisms may also operate together with facilitation in specific communities to support high levels of biodiversity (*16*, *24*, *32*–*34*). In experiments where the cultures were well mixed by vigorous shaking, we also found communities containing multiple taxa, indicating that spatial structure is not required for coexistence (fig. S15). In addition, we did not observe effects from temporal competitive niches in our experiments (fig. S16).

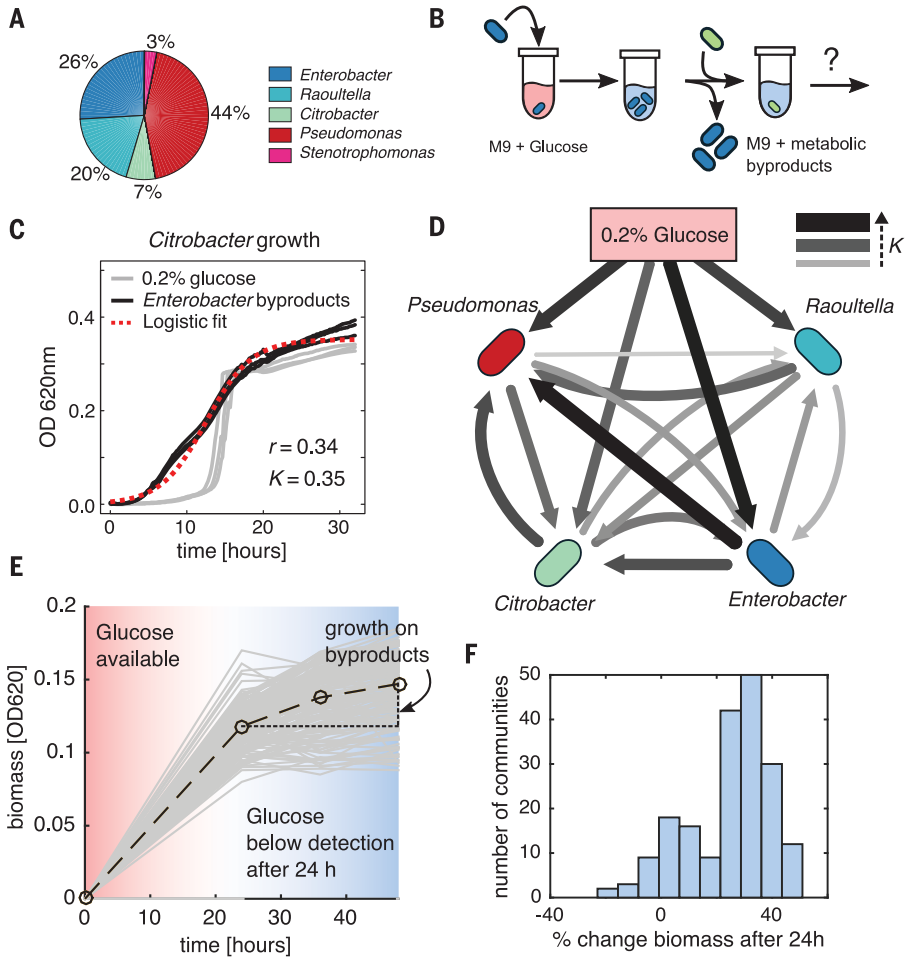

Fig. 3. Nonspecific metabolic facilitation may stabilize competition for the supplied resource. (**A**) Representatives of the four most abundant genera in a representative community (percentages shown in the pie chart) were isolated on M9 minimal glucose medium. (**B**) Experimental setup: Isolates were independently grown in 1X M9 media supplemented with 0.2% glucose for 48 hours, after which cells were filtered out from the suspension. The filtrate was mixed 1:1 with 2X M9 media in the absence of any other carbon sources and used as the growth media for all other isolates (methods). (**C**) Three replicate growth curves of the *Citrobacter* isolate on either M9-glucose media (gray) or the M9-filtrate media from *Enterobacter* monoculture (black). Maximum growth rate (r) and carrying capacity (K) were obtained by fitting to a logistic growth model. (**D**) All isolates were grown on every other isolate's metabolic by-products, and logistic models were used to fit growth curves. We plotted the fitted growth parameters (carrying capacity) as edges on a directed graph. Edge width and color encode the carrying capacity of the target node isolate when grown using the secreted by-products from the source node isolate. Edges from the top node encode the carrying capacity on 0.2% glucose. (**E** and **F**) Growth curves of 95 stabilized communities in M9 glucose media (gray lines) were obtained by measuring the optical density at 620 nm (OD620) at different incubation times. Open circles represent the mean OD620 over all communities at different time points, joined by a dashed line as a guide to the eye. Communities grew on average an additional 25% after glucose had been entirely depleted (~24 hours).

Recent work has suggested that alteration of the pH by bacterial metabolism may also have important growth-limiting effects (*35*, *36*) and can be a driver of microbial community assembly. Our results suggest that although individual isolates can substantially acidify their environment when grown in glucose as monocultures (e.g., the pH drops to 4.85 in *Citrobacter* monocultures and to 5.55 in *Enterobacter* monocultures after 48 hours), our stabilized communities exhibit only modest changes in pH as they grow in glucose minimal media, dropping by less than 1 unit in most communities and stabilizing to pH 6.5 in all cases after 48 hours of growth (fig. S17). In other carbon sources, such as leucine, the pH is even more stable than in glucose (fig. S17). Altogether, our results suggest that acidification by fermentation may be "buffered" by the community relative to the effect observed in monocultures. Although beyond the scope of this work, efforts to elucidate the roles of other mechanisms that may stabilize competition, such as phage predation (*23*) or nontransitive competition networks (*16*), will more fully characterize the landscape of interactions in these microcosms.

A generic consumer-resource model recapitulates experimental observations

Our experiments indicate that competition for a single limiting nutrient may be stabilized by nonspecific metabolic facilitation, leading to coexistence. To test whether this feature alone promotes coexistence, we simulated a community assembly process on a single supplied carbon source, using a version of the classic MacArthur consumer-resource model (*37*), which was modified to include nonspecific cross-feeding interactions. Cross-feeding was modeled through a

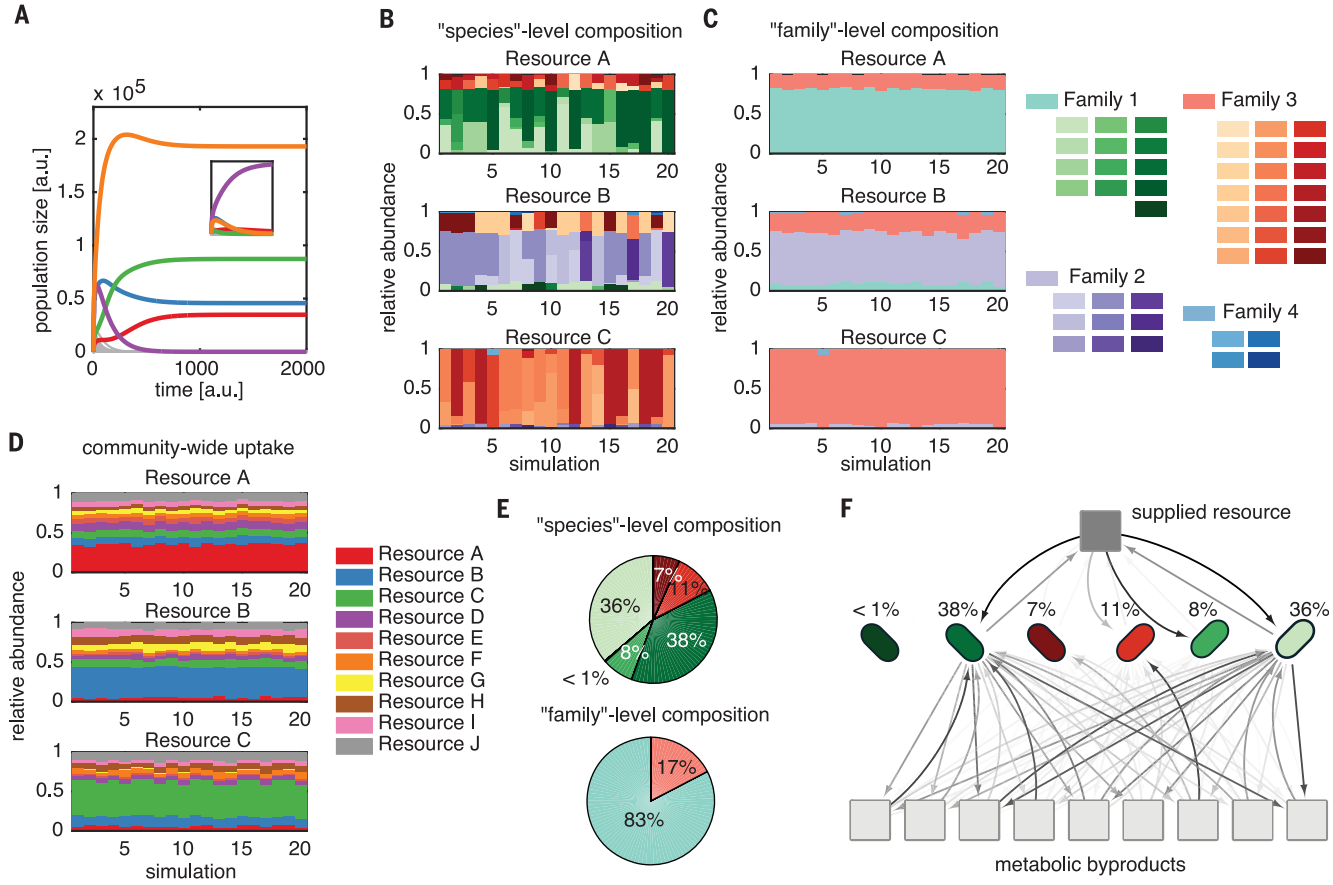

Fig. 4. A simple extension of classic ecological models recapitulates experimental observations. MacArthur's consumer-resource model was extended to include 10 by-product secretions along with consumption of a single primary limiting nutrient (supplementary materials), controlled by secretion coefficient $D_{\beta\alpha}$, which encodes the proportion of the consumed resource α that is transformed to resource β and secreted back into the environment. Consumer coefficients were sampled from four distributions, representing four "families" of similar consumption vectors (fig. S19 and supplementary text). (**A**) Simulations using randomly sampled secretion and uptake rates resulted in coexistence of multiple competitors, whereas setting secretion rates to zero eliminated coexistence (inset). a.u., arbitrary units. (**B** and **C**) Random ecosystems often converged to similar "family"-level structures (C), despite variation in the "species"-level structure (B). The "family"-level attractor changed when a different resource was provided to the same community (lower plots). (**D**) The total resource uptake capacity of the community was computed (supplementary materials) and is, like the family-level structure, highly associated with the supplied resource. (**E**) Communities that formed did not simply consist of single representatives from each family, but often of guilds of several species within each family, similar to what we observed experimentally. (**F**) The topology of the flux distribution shows that surviving species all compete for the primary nutrient, and competition is stabilized by differential consumption of secreted by-products. The darkness of the arrows encodes the magnitude of flux.

stoichiometric matrix that encodes the proportion of a consumed resource that is secreted back into the environment as a metabolic by-product (supplementary materials). Setting this matrix to zero results in no by-products being secreted and recovers the classic results for the consumer-resource model in a minimal environment with one resource: The species with the highest consumption rate of the limiting nutrient competitively excludes all others (Fig. 4A, inset). However, when we drew the stoichiometric matrix from a uniform distribution (while ensuring energy conservation) and initialized simulations with hundreds of "species" (each defined by randomly generated rates of uptake of each resource), coexistence was routinely observed (Fig. 4A). All of the coexisting "species" in this simulation were generalists, capable of growing independently on the single supplied resource and on each other species's secretions.

Our experiments showed that the family-level community composition is strongly influenced by the nature of the limiting nutrient, which may be attributed to the metabolic capabilities associated with each family. We modeled this scenario by developing a procedure that sampled consumer coefficients from four metabolic "families," ensuring that consumers from the same family were metabolically similar (supplementary materials and fig. S18). We randomly sampled a set of 100 consumer vectors (or "species") from four families, then simulated growth for 20 random subsets of 50 species on one of three resources. As in our experimental data (Fig. 2A), simulated communities converged to similar family-level structures (Fig. 4C), despite displaying variation at the species level (Fig. 4B). We confirmed the correspondence between family-level convergence and functional convergence by computing the community-wide metabolic capacity per simulation, resulting in a predicted community-wide resource uptake rate for each resource (supplementary materials). Communities grown on the same resource converged to similar uptake capacities with an enhanced ability to consume the limiting nutrient (Fig. 4D). Importantly, this functional convergence was exhibited even when consumers were drawn from uniform distributions, with no enforced family-level consumer structure, suggesting that the emergence of functional structure at the

community level is a universal feature of consumer-resource models (fig. S19).

We frequently observed that several species belonging to the same metabolic family could coexist at equilibrium. These "guilds" of coexisting consumers from the same family were capable of supporting the stable growth of rare (<1% relative abundance) taxa (Fig. 4E), similar to what we observed in our experimental data (Fig. 1, C and E). Our model suggests that species are stabilized by a dense facilitation network (Fig. 4F), consistent with observations of widespread metabolic facilitation in experiments (Fig. 3D). Thus, we find that simulations of community dynamics with randomly generated metabolisms and resource uptake capabilities capture a wide range of qualitative observations from our experiments and recapitulate previous empirical observations in natural communities (*3*, *10*).

Discussion

In the absence of a theory of microbiome assembly, it is often difficult to determine whether empirically observed features of natural microbiomes are the result of system-specific determinants, such as evolutionary history and past selective pressures at the host level (*10*), or whether they are simply generic emergent properties of large self-assembled communities. Our results show that the generic statistical properties of large consumer-resource ecosystems include large taxonomic diversity even in simple environments, a stable community-level function in spite of species turnover, and a mixture of predictability and variability at different taxonomic depths in how nutrients determine community composition. All of these features are not only observed in our experiments, but also have been reported in systems as diverse as the human gut (*3*, *10*), plant foliages (*6*), and the oceans (*2*, *38*).

Our theoretical results thus provide an explanation for the ubiquity of these empirical findings and suggest that they may reflect universal and generic properties of large self-assembled microbial communities. In spite of their simplicity, consumer-resource models may not only capture many of the generic qualitative features observed in the experiments, but also recapitulate the more subtle aspects, including the existence of temporal blooms in species that eventually go extinct and family-level similarity of communities (fig. S20 and Fig. 4A). However, the models lack biochemical detail and thus do not have the resolution to explain other experimental results such as pH changes, diauxic shifts, or the fact that glucose and citrate communities are more similar to each other than they are to those stabilized in leucine (Fig. 2A).

The theory and simple experimental setup described above also allowed us to identify widespread mechanisms that lead to the assembly of large, stable communities. We find evidence that densely connected cross-feeding networks may stabilize competition within guilds of highly related species that are all strong competitors for the supplied carbon source. Such cross-feeding networks naturally lead to collective rather than pairwise interactions, supporting the hypothesis that higher-order interactions play a critical stabilizing role in complex microbiomes (*16*, *17*). Whether these findings are generic in more complex environments with a larger number of externally supplied resources remains to be elucidated. For instance, the experiments and theory presented in this work indicate that the stabilized microbial communities consist of metabolic generalists, rather than metabolic specialists (*39*), capable of consuming both the supplied resource and metabolic by-products. It is unclear whether these findings are generalizable to microbial communities adapted to static environments where metabolic specialization may confer fitness advantages (*39*). We propose that high-throughput top-down approaches to community assembly that are amenable to direct mathematical modeling are an underused but highly promising avenue to identify generic mechanisms and statistical rules of microbiome assembly, as well as a stepping stone toward developing a quantitative theory of the microbiome.

REFERENCES AND NOTES

1. P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034–1039 (2008).
2. S. Sunagawa *et al*., *Science* **348**, 1261359 (2015).
3. Human Microbiome Project Consortium, *Nature* **486**, 207–214 (2012).
4. E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, D. A. Relman, *Science* **336**, 1255–1262 (2012).
5. P. J. Turnbaugh *et al*., *Nature* **457**, 480–484 (2009).
6. S. Louca *et al*., *Nat. Ecol. Evol.* **1**, 0015 (2016).
7. S. Louca, L. W. Parfrey, M. Doebeli, *Science* **353**, 1272–1277 (2016).
8. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, *Science* **350**, aac9323 (2015).
9. C. Burke, P. Steinberg, D. Rusch, S. Kjelleberg, T. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14288–14293 (2011).
10. L. A. David *et al*., *Nature* **505**, 559–563 (2014).
11. J. Friedman, L. M. Higgins, J. Gore, *Nat. Ecol. Evol.* **1**, 109 (2017).
12. N. M. Vega, J. Gore, *PLOS Biol.* **15**, e2000633 (2017).
13. D. R. Hekstra, S. Leibler, *Cell* **149**, 1164–1173 (2012).
14. K. R. Foster, T. Bell, *Curr. Biol.* **22**, 1845–1850 (2012).
15. K. Z. Coyte, J. Schluter, K. R. Foster, *Science* **350**, 663–666 (2015).
16. J. M. Levine, J. Bascompte, P. B. Adler, S. Allesina, *Nature* **546**, 56–64 (2017).
17. E. Bairey, E. D. Kelsic, R. Kishony, *Nat. Commun.* **7**, 12285 (2016).
18. B. J. Callahan *et al*., *Nat. Methods* **13**, 581–583 (2016).
19. G. Gottschalk, *Bacterial Metabolism* (Springer, 1979).
20. R. Macarthur, R. Levins, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 1207–1210 (1964).
21. F. M. Stewart, B. R. Levin, *Am. Nat.* **107**, 171–198 (1973).
22. D. Tilman, *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
23. F. Rodriguez-Valera *et al*., *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
24. E. D. Kelsic, J. Zhao, K. Vetsigian, R. Kishony, *Nature* **521**, 516–519 (2015).
25. J. Grilli, G. Barabás, M. J. Michalska-Smith, S. Allesina, *Nature* **548**, 210–213 (2017).
26. M. Basan *et al*., *Nature* **528**, 99–104 (2015).
27. N. Paczia *et al*., *Microb. Cell Fact.* **11**, 122 (2012).
28. R. F. Rosenzweig, R. R. Sharp, D. S. Treves, J. Adams, *Genetics* **137**, 903–917 (1994).
29. S. K. Hansen, P. B. Rainey, J. A. J. Haagensen, S. Molin, *Nature* **445**, 533–536 (2007).
30. R. Baran *et al*., *Nat. Commun.* **6**, 8289 (2015).
31. M. S. Datta, E. Sliwerska, J. Gore, M. F. Polz, O. X. Cordero, *Nat. Commun.* **7**, 11965 (2016).
32. P. B. Rainey, M. Travisano, *Nature* **394**, 69–72 (1998).
33. P. Chesson, *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
34. B. R. Levin, *Science* **175**, 1272–1274 (1972).
35. J. Cremer, M. Arnoldini, T. Hwa, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6438–6443 (2017).
36. C. Ratzke, J. Gore, *PLOS Biol.* **16**, e2004248 (2018).
37. R. MacArthur, *Theor. Popul. Biol.* **1**, 1–11 (1970).
38. A. C. Martiny, A. P. K. Tai, D. Veneziano, F. Primeau, S. W. Chisholm, *Environ. Microbiol.* **11**, 823–832 (2009).
39. T. Taillefumier, A. Posfai, Y. Meir, N. S. Wingreen, *eLife* **6**, 1–65 (2017).
40. M. G. I. Langille *et al*., *Nat. Biotechnol.* **31**, 814–821 (2013).

ACKNOWLEDGMENTS

We thank the Goodman laboratory at Yale and the Brucker laboratory at the Rowland Institute for their technical help in the early phases of this project. We also thank M. L. Osborne for technical assistance and members of the Sanchez, Mehta, and Segrè groups for helpful discussions. **Funding:** The funding for this work partly results from a Scialog Program sponsored jointly by the Research Corporation for Science Advancement and the Gordon and Betty Moore Foundation through grants to Yale University and Boston University by the Research Corporation and by the Simons Foundation. This work was also supported by a young investigator award from the Human Frontier Science Program to A.S. (RGY0077/2016) and by NIH NIGMS grant 1R35GM119461 and a Simons Investigator award in the Mathematical Modeling of Living Systems (MMLS) to P.M.; D.S. and J.EG. additionally acknowledge funding from the Defense Advanced Research Projects Agency (purchase request no. HR0011515303, contract no. HR0011-15-C-0091), the U.S. Department of Energy (DE-SC0012627), the NIH (T32GM100842, 5R01DE024468, R01GM121950, and Sub_P30DK036836_P&F), the National Science Foundation (1457695), the Human Frontier Science Program (RGP0020/2016), and the Boston University Interdisciplinary Biomedical Research Office.
Author contributions: J.E.G. designed experiments, collected data, wrote data analysis code, developed the models, ran simulations, analyzed data, and wrote the paper. N.L. generated all sequencing data, designed experiments, collected and analyzed data, and wrote the paper. D.B., S.E., and A.S.-G. designed experiments, collected and analyzed data, and wrote the paper. M.T. provided guidance during data analysis and contributed to the writing of the manuscript. D.S. provided guidance, contributed to the design of the project, and wrote the paper. P.M. developed the consumer-resource model, ran simulations, supervised and contributed to the design of the project, and wrote the paper. A.S. designed the experiments, collected data, supervised the project, and wrote the paper.
Competing interests: The authors declare that no competing interests exist in relation to this manuscript. **Data and materials availability**: Isolates and communities are available upon request. Data analysis and simulation codes are available via GitHub at https://github.com/jgoldford/mcrm. The 16S sequencing data and the metadata file have been deposited in the NCBI SRA database with ID SRP144982 (https://www.ncbi.nlm.nih.gov/sra/SRP144982).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6401/469/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S21
References (*41*–*47*)

25 January 2018; accepted 19 June 2018
10.1126/science.aat1168

www.sciencemag.org/content/361/6401/469/suppl/DC1

Supplementary Materials for

Emergent simplicity in microbial community assembly

Joshua E. Goldford*, Nanxi Lu*, Djordje Bajić, Sylvie Estrela, Mikhail Tikhonov, Alicia Sanchez-Gorostiaga, Daniel Segrè, Pankaj Mehta†, Alvaro Sanchez†

*These authors contributed equally to this work.
†Corresponding author. Email: alvaro.sanchez@yale.edu (A.S.); pankajm@bu.edu (P.M.)

Published 3 August 2018, *Science* **361**, 469 (2018)
DOI: 10.1126/science.aat1168

This PDF file includes:

Materials and Methods

Isolating microbial communities from natural ecosystems

Leaf or soil samples (~1 g) were collected from natural environments using sterile tweezers and placed in 15 mL falcon tubes. In the lab, 10 mL of 5 % NaCl buffer was added to each sample and allowed to incubate for ~48 hours at room temperature. 40% glycerol stock solutions were prepared from aqueous sample suspensions and frozen at -80 °C for storage.

Preparation of 96-well media plates

All media contained 0.07 C-mole/L of carbon source (glucose, citrate or leucine) and was sterile-filtered with a 0.22 μm filter (Millipore). Stock solutions of carbon sources were stored at 4 °C for no more than 1 month. M9 media was prepared from concentrated stocks of M9 salts (without $MgSO_4$ or $CaCl_2$) and stock solutions of $MgSO_4$ and $CaCl_2$. 500 μL cultures containing 450 μL of sample and 50 μL stock carbon source were grown in 96 deep-well plates (VWR). For the first two cell passages, cycloheximide was added to the media at a concentration of 200 μg/mL to inhibit eukaryotic growth.

Passaging microbial populations

Starting inocula were obtained directly from the initial microbiota solution by inoculating 4 μL into 500 μL culture media. For each sample, 4 μL of the culture medium was dispensed into all 60 wells of the fresh media plate. Cultures were allowed to grow for 48 hours at 30 °C in static broth, then each culture was homogenized by pipetting up and down 10 times before passaging. Passaging was performed by taking 4 μL from each culture to use as inocula in 500 μL of fresh media, and cells were allowed to grow again. Cultures were passaged 12 times (~84 generations). Optical density (OD620) was used to measure biomass in cultures after the 48-hour growth cycle. Samples to be sequenced were collected and stored by spinning down in a micro-centrifuge for 10 min at 14,000 RPM at room temperature. Cell pellets were stored at -20 °C.

DNA extraction, library preparation and sequencing

Cell pellets were re-suspended and incubated at 37 °C for 30 min in enzymatic lysis buffer (20 mM Tris-HCl, 2mM sodium EDTA, 1.2% Triton X-100) and 20 mg/mL of lysozyme from chicken egg white (Sigma-Aldrich) to lyse the cell walls of Gram-positive bacteria. Following cell lysis, the DNA extractions were performed following the DNeasy 96 protocol for animal tissues (Qiagen). The clean DNA was eluted in 100 μL elution buffer of 10 mM Tris-HCl, 0.5 mM EDTA at pH 9.0. DNA concentration was quantified using Quan-iT PicoGreen dsDNA Assay Kit (Molecular Probes, Inc.) and normalized to 5 ng/μL for subsequent 16S rRNA sequencing. 16S rRNA amplicon library preparation was conducted using a dual-index paired-end approach developed by Kozich et al (*41*). Briefly, PCR-amplified libraries were prepared using dual-index primers (F515/R806) to generate amplicons spanning the V4 region of the 16S rRNA gene, then pooled and sequenced using the Illumina MiSeq platform. For each sample, a

30-cycle PCR was performed in duplicate in 20 µL reaction volumes using 5 ng of DNA, dual index primers, and AccuPrime Pfx SuperMix (Invitrogen). Thermocycling conditions consisted of a 2-min initial denaturation step at 95 °C, followed by 30 cycles of the following PCR scheme: (a) 20-second denaturation at 95 °C, (b) 15-second annealing at 55 °C, and (c) 5-min extension at 72 °C. PCR was terminated after a 10-min extension step at 72 °C. After pooling amplicons from duplicate reactions, the PCR products were purified and normalized using the SequalPrep PCR cleanup and normalization kit (Invitrogen). Libraries were then pooled and sequenced using Illumina MiSeq v2 reagent kit, which generated 2x250 base pair paired-end reads at the Yale Center for Genome Analysis (YCGA). For shaking control experiments (Fig. S15), library preparation and sequencing was performed at SeqMatic (Fremont, CA). Sequencing and library preparation were identical when compared to the procedure described above, except primers targeted the V3-V4 region of 16S rRNA gene.

16S rRNA sequencing analysis

QIIME 1.9.0 (*42*) was used to demultiplex and remove barcodes, indexes and primers from raw files, producing FASTQ files with for both the forward and reverse reads for each sample. Dada2 version 1.1.6 was used to infer exact sequence variants (ESVs) from each sample (*18*). Briefly, forward and reverse reads were trimmed to 220 and 160 nucleotides, respectively. All other parameters were set to default values. Sequences below 230 or above 242 nucleotides were discarded (indicative or poor merging of paired reads). Chimeric PCR products from two related species (i.e. Bimeras) were removed using the "tableMethod" parameter set to "consensus." A naive Bayes classifier was used to assign taxonomy to Exact Sequence Variants (ESVs) using the SILVA version 123 database (*43*). Metagenome inference was performed using PICRUSt (*40*). ESVs were assigned to OTUs using the greengenes database version 13.5 using the QIIME function *pick_closed_reference_otus.py*, with a 97 % similarity cutoff. Communities were normalized using the *normalize_otus.py* function in PICRUSt, and the metagenomes were estimated using the *estimate_metagenome.py* routine. We note however that imputed metagenomes may be biased by unequal annotation of representative species as well as variability between taxa with similar 16S sequences but different genome composition.

Fermentation assays and isolation of strains

Four bacterial strains from a representative community stabilized in glucose were isolated and identified taxonomically. The community was plated onto 0.5 % agarose Petri-dishes containing M9 supplemented with 0.2% glucose and were allowed to grow for 48 hours at 30 °C. Single colonies were then picked from these plates according to their colony morphologies, re-streaked on fresh agarose plates and grown for another 48 hours at 30 °C. Single colonies from each isolate grown for 48 hours at 30 °C in liquid M9 supplemented with 0.2% glucose were finally stored at -80 °C in 40% glycerol. Isolates were also identified according to their differential ability to ferment the following 16 carbohydrates: adonitol, arabinose, cellobiose, dextrose, dulcitol, fructose, inositol, lactose, mannitol, mannose, melibiose, raffinose, rhamnose, salicin, sucrose, and xylose (Fig S5 A-B). Fermentation ability was assessed using a phenol red broth base with an added carbohydrate at a final concentration of 1% w/v, except for cellobiose (0.25%) due to its low solubility. Each isolate was grown on an agarose plate, and a single

colony was picked and re-suspended into 100 µL 1x PBS. 2 µL of each isolate was inoculated into 50 µL of Phenol red broth + carbon source (in a 384 well-plate, Corning). Spectrophotometric measurements of phenol red (OD450 and OD551) were measured after 0, 12, 16, and 19 hours of incubation. Clustering of O.D. profiles after 19 hours revealed four distinct phenotypic profiles, consistent with morphologies (Fig. S5C). Taxonomic assignments of isolates were verified using full-length 16S rRNA sequencing of DNA extracted from single colonies grown on agarose plates (GENEWIZ), using the online RDP classifier (51).

Reconstitution of isolates from a representative community

To test whether the dominant species isolated from the glucose stabilized communities are able to coexist, we constructed a four-strain community with four strains isolated from one representative community (C2R4). The four isolates belong to four different genera (*Raoultella*, *Enterobacter*, *Pseudomonas*, and *Citrobacter*) and were chosen because they are the most dominant species in the community and display distinctive morphologies, facilitating plate counting. To ensure that the starting densities were similar for all four isolates, single colonies were picked, resuspended into PBS 1x, and the optical densities were normalized to a OD620 of 0.15. The initial inoculum was prepared by mixing the four isolates in 1:1:1:1 ratio. 4 µL of the initial inoculum was transferred to 500 µL fresh media M9 with 0.2% Glucose (3 replicate communities) and cultures were incubated at 30°C (Fig. S5D) . Every 48 hours, 4µL from each replicate community was transferred to 500µL of fresh growth media for a total of 7 transfers (14 days). OD620nm measurements were conducted every 48 hours and the four isolates were enumerated by colony counts on M9+ 0.2% glucose agar plates on Transfer 5 (day 10) and Transfer 7 (day 14). We found that the four isolates were able to stably coexist after 7 transfers (14 days). *Raoultella* was the most abundant strain, followed by *Enterobacter*, and then *Pseudomonas*, and *Citrobacter* (see Fig S5E).

Metabolic facilitation assay and measurement of glucose depletion

To determine whether microbial cross-feeding is a potential mechanism that enables coexistence, four isolates from a single representative community were inoculated in 5 mL of M9 media with 0.2% glucose, then incubated for 48 hours at 30 °C (Fig. 3A). Cells were then separated from the spent media (SM) using the following procedure: cells were centrifuged at 3000 rpm for 10 min, and SM was filter-sterilized and stored at 4 °C. Cells were re-suspended in the same volume of PBS, and washed two times times by centrifugation (3000rpm, 10min). Cells were diluted to an OD620 of 0.24 prior to inoculation. There was no detectable glucose remaining in any SM as measured using the Glucose GO Assay Kit (Sigma), with the exception of the SM from *Pseudomonas*, which was adequately controlled for (see main text). SM was then mixed 1:1 with fresh 2X M9 media with no carbon source. Each isolate was inoculated in each isolate's SM-based M9 in triplicate at 1% v/v in a 384 well plate (Corning). The plate was incubated in a standard plate reader (Thermo 498 Scientific), and OD620 was measured every 10 min at 30 °C.

We sought to determine whether glucose-stabilized communities were able to grow after glucose depletion, which would suggest that biomass accumulation is attributed to consumption of metabolic byproducts. For this, 95 glucose-stabilized communities were inoculated in a 96 deep-well plate from frozen stock in 500 µL of M9+0.2% glucose. Two initial transfers with 48 hours incubation were performed as previously described (30 °C no shaking). The third transfer was performed in duplicate and with a final volume of 600 µL. From these two plates, 100 µL samples were taken at 24, 36, 48 and 56 hours. OD620 was measured, followed by the measurement of glucose using the Glucose GO Assay Kit (Sigma). Glucose concentrations were inferred using linear regression from the standard curve, although no sample at any time showed detectable levels.

Cell death measurements

Samples were obtained at 12-hour intervals to measure the accumulation of biomass and determine the frequency of dead cells. Bacteria stained with the LIVE/DEAD BacLight Bacterial Viability Kit (L-7012, Invitrogen) following manufacturer instructions were spotted on 1% agarose pads. Microscopy was performed on an Eclipse Ti-E microscope (Nikon, Tokyo, Japan), equipped with Perfect Focus System (Nikon), a phase-contrast objective Plan Apochromat 100X/1.40 NA (Nikon), and an ORCA-Flash4.0 V2 Digital CMOS camera (Hamamatsu Photonics, Hamamatsu City, Japan). Red fluorescence of dead cells was recorded with a Texas Red bandpass filter. Images were acquired with MetaMorph software (Molecular Devices, Sunnyvale, CA, USA) and analyzed with Microbe J (52). The images were processed with Adobe Photoshop (CC2015.5). For Fig S14C, we counted between 235-2565 cells.

Low abundant growth with no supplied carbon source

Passaging experiments were performed using M9 synthetic media with no additional carbon sources, which resulted in the stabilization of very low abundancy microbial communities (Fig. S4). Growth was often several orders of magnitude lower than growth on either the primary nutrient (Fig. S4C) or secreted byproducts (Fig. 3E-F), indicating that metabolic consumption of secreted byproducts is more likely to contribute to stabilizing competition than consumption of low levels of latent resources in the deionized water. To determine community richness resulting from growth on the provided resource, we estimated the abundance of 16S amplicon reads deriving from contamination either by cross-well contamination or microbial growth on the low levels of total organic carbon in deionized water (Fig. S4A-B). For each of the 12 initial points, communities were propagated for 84 generations with either with M9 and 0.2% glucose, or M9 and no additional carbon source. We plated communities on 0.5% agarose plates containing M9 minimal media and 0.2% D-glucose to determine the colony forming units (CFU) per ml (Fig. S4C). CFU/ml was used as a proxy for total cell number in the community because of the strong correlation with cell counting using a hemocytometer (Fig. S4D). The relative contribution of CFU for growth on water alone compared to growth on D-glucose was then used as a relative frequency cutoff for each of the 12 initial communities, respectively (Fig. S4E). These values allowed us to estimate lower bounds for community diversity derived from the supplied the carbon source (Fig. S6B).

Measurement of community pH dynamics during a growth cycle

To measure the fluctuations in pH during the 48 hour growth cycle, we thawed communities stabilized and cultured them for an additional 48 hours. We chose a single representative community for each initial inoculum (12) and carbon source (3) used in the paper, resulting in a total of 36 communities. We inoculated these communities from frozen stock into M9+0.2% of the corresponding carbon source (glucose, citrate or leucine) for 48 hours at 30 °C. For each sample, we took 4 μL and re-inoculated the sample into fresh media, and measured the pH after 0, 12, 24, 36, and 48 hours of growth. pH was measured by spotting 4 μL of culture media onto indicator paper (Watman). The pH of the fresh media was measured as a control. he results are shown in Fig. S17B. Media with glucose showed lowest pH of 6.5 at the end of 48 hours of growth. Media with citrate started at pH 6.0 but ended at pH 7.0. Media with leucine stayed stably at above pH 6.5 and finally at pH 7.0. We performed similar experiments with isolates obtained from a representative community grown on M9+0.2% glucose (Fig S5), and found that monocultures acidify the media significantly more than the community (Fig S16A).

Growth of stable consortia on different carbon sources to enrich for potential rare taxa

To more fully characterize the community structure of our microcosms, we shifted communities stabilized on M9+0.2% citrate media to M9+0.2% glutamine media for an additional 42 generations. We obtained eight communities passaged on M9+0.2%citrate for 84 generations, and grew these communities on M9+0.2% glutamine for an additional 42 generations transfers. We sequenced these communities following the protocols described above, and obtained ~25,000 reads per sample. For communities grown on glutamine, we only observed 0-3 additional ESVs per sample.

Statistical tests for Beta diversity differences

The covariates explored in this study are the regional pool of species (initial environmental inocula) and the carbon source supplied in the media. Between samples, we used Renkonen similarity at the family taxonomic level as a measure of beta diversity between communities, which is defined as:

$$D(x, y) = 1 - \frac{1}{2}\sum_i |x_i - y_i|$$

where x_i and y_i are the abundance of taxon i in sample X and sample Y respectively. We computed the family-level Renkonen similarities between all samples and grouped pairwise similarities if pairs were passaged on the same carbon source, or if pairs of samples originated from the same inocula. We used the Renkonen similarity to determine if community similarity was higher between samples from the same time series or from different replicates as genus and family taxonomic rank (Fig S6C). We used the one-tailed Kolmogorov-Smirnov test (MATLAB function *kstest2.m*) to determine if the pairwise similarities grouped by carbon source were on average higher than pairwise similarities grouped by initial inocula (see Fig S11C).

Test of temporal variation and replicate variation

We estimated the variability in commmuntiy composition from different replicates from inoculum 2 (see Fig 1F) and compared this to the variability in community composition between the last three transfers in our passaging experiment. To calculate the variability across replicates, we computed the Renkonen Similarity between each pair of replicates after the last transfer (transfer 12). To calculate the temporal variation within a single replicate, we calculated the Renkonen Similarity within a replicate at transfers 10,11,12. We used only the final three transfers to ensure that the community composition has had enough transfers to stabilize and to ensure that the number of similarity scores used to assess the temporal variation was similar to the number similarity scores used to the assess the replicate variation ($N = 24$ within time-series, and $N = 28$ between time-series). We then assessed if replicate variations at the genus and family level were larger than the temporal variations at the same taxonomical resolution using a standard non-parametric test (in this case the Mann-Whitney U test). The statistical test showed that the replicate variation is significantly larger than the temporal variation at the genus level ($P = 1.1 \times 10^{-5}$) while at the family level this was not the case ($P = 0.0624$).

Prediction of media carbon source from community structure

To quantify the predictive capacity of community structure (both at the taxonomic and functional levels) for the supplied carbon source, we trained and evaluated multi-class support vector machine (SVM) models or random forest classifiers and measured the model accuracy. SVMs were constructed by using the MATLAB function *fitecoc* and evaluated using 10-fold cross validation in Fig. 2C or leave one out cross-validation in Fig. S18. Leave-one-out cross-validation was performed by training the SVM on all samples except one, and predicting the carbon source from the sample left out of the training set. Features used in the in the SVM were either the clr-transformed relative abundances at the family taxonomic level in Fig. 2C or the clr-transformed inferred metagenome composition in Fig. S18. To obtain a list of variable importance scores, we trained a random forest classifier using the same feature set using the *TreeBagger* function in MATLAB with 100 trees and default parameters (Fig S11B).

Supplementary Text

Microbial Consumer Resource Model

The model presented in the paper is a modification of Robert MacArthur's consumer resource model (*33*, *37*, *44*), which models the per-capita growth of species as a function of resource consumption rate. We begin by first re-stating the dynamics of individual species, followed by a modified form of resource dynamics that include environmental modification during bacterial growth.

Let us denote the set of all possible resources by R_α where $\alpha = 1, \dots, M$. Furthermore, let us denote the set of all species by N_i where $i = 1, \dots, S$. Each species is characterized by a resource

utilization matrix $C_{i\alpha}$, which represents the rate at which the species *i* uptakes resource α. Furthermore, there is a resource quality function $\Delta w_{i\alpha}$ which captures the amount of biomass of species *i* produced per unit of resource α uptaken while maintaining energy balance (see below) . Assuming that for each species *i* there exists a minimum maintenance energy required for growth m_i, the per capita growth rate of species *i* is:

$$\frac{1}{N_i}\frac{dN_i}{dt} = \sum_{\alpha} \Delta w_{i\alpha} C_{i\alpha} R_\alpha - m_i$$

This assumes populations die if they cannot achieve minimum growth rate to survive m_i. The principal modification to the MacArthur's consumer resource model is the addition of a stoichiometric matrix that encodes the proportion of consumed resources that are transformed into new resources and secreted back into the environment. A wide variety of bacterial heterotrophs are capable of excreting a large fraction of the carbon input through overflow metabolism even under aerobic conditions (*26*, *27*).

To model the bacterial secretion of metabolic byproducts, let the matrix $D^i{}_{\beta\alpha}$ be a stoichiometric matrix that encodes the number of molecules of resource β secreted by to the environment species *i* per molecule of resource α it uptakes. Thus, the rate of production of resource β by species i is proportional to the sum over all resources of the rate that a species takes up resource α times the stoichiometric parameter $D^i{}_{\beta\alpha}$

$$\sum_{\alpha,i} D^i_{\beta\alpha}\, C_{i\alpha} R_\alpha N_i$$

giving rise to the full dynamical equation for the abundance of resource β.

$$\frac{dR_\beta}{dt} = \frac{K_\beta - R_\beta}{\tau_\beta} - \sum_i C_{i\beta}\, R_\beta N_i + \sum_{\alpha,i} D^i_{\beta\alpha}\, C_{i\alpha} R_\alpha N_i$$

where K_β is the initial resource abundance supplied in fresh media, and τ_β is the replenishing (i.e. transfer) rate during batch culture passaging. Note that we represent the efficiency of resource α with the parameter $\Delta w_{i\alpha} = w_\alpha - \sum_\beta D^i_{\beta\alpha}\, w_\beta$, which ensures that energy is balanced in our model. In slightly more detail, we denote the maximum ATP yield of resource α by w_α. Recall when species *i* consumes resource α it make byproducts β according to the stoichiometric matrix $D^i{}_{\alpha\beta}$. To ensure energy balance, the maximum energy that can be extracted in such a process is the difference between the ATP yield of resource α and the total ATP yield of all the metabolic byproducts. Explicitly, this is just given by $\Delta w_{i\alpha} = w_\alpha - \sum_\beta D^i_{\beta\alpha}\, w_\beta$.

The first term in the resource dynamics equation deliberately chooses the simplest (linear) supply rate of resource β. Alternative, more complex choices for this function are of course also possible, for instance one that would capture the periodic but pulsatile nature of resource addition in our experiments. Likewise, the constant maintenance rate m_i is also the simplest possible functional choice for this parameter. These are the forms originally proposed by MacArthur and colleagues and are the most commonly used in the literature. Therefore, we adopted them for simplicity and to avoid the potential introduction of more complex ecological features, such as temporal niches. As shown in Fig. S21, the main general qualitative results reported in this paper (i.e. the coexistence of many taxa on a single supplied resource, and functional convergence in spite of taxonomic variability across similar habitats) do not change if we choose more complex supply and maintenance functions that reflect more closely our experiments

Ensuring energy conservation

For heterotrophic, aerobic bacteria, energy and carbon sources are often coupled within reduced organic substrates (*19*). Following the laws of thermodynamics, the total energy (or free energy) available from resources supplied in the environment constrains the total energy secreted back into the environment. However, energy (or free energy) is not well defined in our far from equilibrium dynamical equations. This quantity is indirectly associated with the resource quality, w, which is a phenomenological parameter that represents the relative gain in a limiting factor (e.g. carbon or energy) per consumed resource. Our model assumes that the limiting factor is linear in the growth rate, which is expected if species are catabolically-limited, and w_α is the ATP yield for a resource α.

To ensure energy is not created during the metabolism of a resource, we ensure that the secretion matrix, $D^i_{\beta\alpha}$ is constrained by the following relation:

$$\sum_\beta w_\beta D^i_{\beta\alpha} \; < \; w_\alpha$$

Sampling of consumer species according to functional groups

To simulate the scenario where consumers are non-randomly distributed and taxonomically related, we sampled consumer coefficients from a prior distribution where ``families'' of consumers share similar consumption coefficients. In this formulation, consumer coefficients are drawn from Dirichlet distributions, and the Dirichlet concentration parameter encodes the family-level consumption preferences and variability. In our model, sampling from a Dirichlet distribution results in stochastically partitioning a fixed amount of cellular resources dedicated for nutrient uptake (e.g. transporters) into groups, and the concentration parameter fixes the average across these samples.

The family-level consumption properties are represented by two parameters, $\theta_{\alpha,f}$ and Ω_f where $\theta_{\alpha,f}$ is the concentration parameter for resource α belonging to family f, and K_f is the magnitude

of the all concentration parameters, such that: $\sum_\alpha \theta_{\alpha,f} = \Omega_f$. For family f, we wish to construct a family of consumers with a tunable degree of preference for resource $\alpha = f$. Thus we first sample $\theta'_{\alpha=f}$ using the following relation:

$$\theta'_{\alpha=f} \sim \text{Normal}(\mu, \sigma^2),$$

where Normal(μ,σ^2) denotes a Gaussian distribution of mean μ and standard deviation σ. Note that in all simulations μ and σ are chosen to be bounded between 0 and 1. For other concentration parameters we first sample them from a uniform distribution, $\theta'_{\alpha\neq f} \sim \text{Uniform}(0,1)$. The concentration parameters are then normalized using the following formula:

$$\theta_{\alpha\neq f} = (1-\theta_{\alpha=f})\frac{\theta'_{\alpha\neq f}}{\sum_{\gamma\neq f}\theta'_\gamma}$$

Resulting in a set of concentration parameters $\theta_{\alpha,f}$. Note that the parameters μ and σ control how much of a "specialist" a family of consumers will be. For all simulations we choose $\mu = 0.4$ and $\sigma = 0.01$.

We next used the family-specific parameters $\theta_{\alpha,f}$ and Ω_f to compute dirichlet concentration parameters to sample uptake coefficients for individual consumers belonging to family f. We first draw *relative* uptake rates for a "species" from a family of consumers using the following formula:

$$c'_{i,1}, c'_{i,2}, \ldots, c'_{i,M} \sim \text{Dirichlet}(\Omega_f\theta_1, \Omega_f\theta_2, \ldots, \Omega_f\theta_M)$$

where Ω_f controls the total variability with each family. A high Ω_f ensures that ``species" are very similar, where a low $\Omega_f$ results in ``species" that are variable. For our simulations, we chose $\Omega_f = 100$ for all families.

Each sample from a Dirichlet results in a set of consumption coefficients that sum to unity, such that: $\sum_{\alpha=1,..,M} c'_{i\alpha} = 1$. If we used these values directly as uptake coefficients, then we may obtain cases where coexistence is unbounded, recently investigated in detail using similar consumer resource models (*39*, *45*), which arises from a linear constraint on the sum of uptake coefficients. We thus drew a new random value, $T_i \sim \text{Normal}(1,0.01)$, for each "species" i that relaxed this constraint. Consumer coefficients were then computed using the following function:

$$c_{i\alpha} = T_i c'_{i\alpha}$$

Numerical Simulations

Choosing Parameters

For all simulations, we set the number of species to be $N = 100$ and the number of resources to be $M = 10$. The resource qualities, the resource replenishment rates, the maintenance and the growth rate multipliers were set to unity, such that: $w_{i\alpha} = \tau_\alpha = m_i = b_i = 1$ for all species i and resources α. We initialized simulations to model dynamics on a single externally supplied resource γ by setting $K_\alpha = 10^6$ if $\alpha = \gamma$ and 0 otherwise. For all simulations, we assumed that the stoichiometric matrix is species-independent, such that $D^i_{\beta\alpha} = D_{\beta\alpha}$. Stoichiometric matrices were drawn from a uniform distribution, such that:

$$D_{\beta\alpha} \sim \text{uniform}(0, 1/M)$$

Note that by setting the upper bound of $D_{\beta\alpha} \leq 1/M$ and $w_{i\alpha} = 1$, we ensure that energetic constraints are not violated.

Time-courses

In Fig 4, consumer matrices were drawn from Dirichlet distributions (see previous section), while in Fig. S19, consumer matrices were drawn from uniform distributions. Simulations were performed in MATLAB 2015a using ODE solver ode15s. Simulations were performed for at-least 10^4 timesteps, where the vast majority of simulations resulting in reaching stable equilibria in roughly 500 timesteps. Code is available on the following GitHub repository: https://github.com/jgoldford/mcrm.

Metagenomic analysis and comparison with experiment

Based on our experimental results, we expected that the collection of genes in the community (the metagenome) would be associated with the externally-supplied resource (e.g., glucose, citrate, or leucine). To compare to the model, we implicitly assume that the metagenome is associated with the community-wide uptake capability of externally supplied resources. This assumption requires that gene dosage is positively associated with the activity of transporters (*46*).

From experimental data, we estimated the metagenome from 16S rRNA amplicon sequencing data using PICRUSt (*40*). The gene abundance profiles were normalized to sum to unity, and were transformed using the centered log-ratio transform (*47*). Formally, for a composition x, we define the the centered log-ratio transform (clr) as:

$$clr(x) = z = \left[\ln\left(\frac{x_1}{g(x)}\right), \ldots, \ln\left(\frac{x_D}{g(x)}\right)\right]$$

where $g(x) = \sqrt[D]{\prod_i x_i}$ [1], where D represents the length of composition vector x. We then construct a matrix, Z, where z_{ik} represents the clr-transformed abundances for gene i in sample

[1] For all metagenome samples, a small value, $\epsilon = 10^{-20}$ was added to each x_i to prevent $g(x)$from becoming zero.

k. We then used tSNE (t-distributed Stochastic Neighbor Embedding) to reduce the dimensionality of the clr-transfomed metagenome matrix Z as seen in Figure 2c and in the main text. In Fig. S19, the fraction of the metagenome that is dedicated to Leucine degradation (KEGG Module M00036) was computed for each sample, then grouped by the externally-supplied resource *x*-axis), revealing a strong concordance between the presence of a specific limiting nutrient and the community-wide metabolism for that limiting nutrient.

To compare experiments to the model, we first simulated the population dynamics and found the steady state abundance for each species i, N_i^*. We then computed the total uptake of resource α (which we denote as Y_α) as:

$$Y_\alpha = \sum_i C_{i\alpha} N_i^*$$

For each simulation k on a resource γ, we constructed a matrix of community wide uptake rates with matrix elements equal to $Y_{k\gamma}$. The total uptake capacity per simulation was normalized to sum to unity, and was transformed using the clr transform, just like in the case with inferred metagenomic data. Dimensionality reduction was then performed on this matrix using tSNE, and plotted in the Figure S19.

Monod model

Microbes in a community can coexist in an environment with a single limiting resource if strains have a peak fitness at some intermediate concentration of the limiting resource (*21*). We investigated whether this mechanism may be responsible for coexistence by isolating the dominant taxa from a representative community, and measuring the growth rates at various concentrations to estimate parameters used in a Monod growth model. First, isolates were obtained via plating, then grown in minimal M9 salts media supplemented with glucose at concentrations ranging from 0.01 - 0.2 %. For each strain i on glucose concentration S, we fit a curve to the following logistic equation:

$$\frac{1}{N_i}\frac{dN_i}{dt} = r_i(S)\left(1 - \frac{N_i}{K_i(S)}\right)$$

where $r_i(S)$ is the maximum per capita growth rate, and $K_i(S)$ is the carrying capacity of strain i on a carbon source with abundance S. Monod parameters for each species μ_i and κ_i were then fitted using the following function:

$$r_i(S|\mu_i, \kappa_i) = \frac{\mu_i S}{\kappa_i + S}$$

These parameters where then used in the following dynamic growth and substrate equations:

$$\frac{1}{N_i}\frac{dN_i}{dt} = \frac{\mu_i S}{\kappa_i + S} - m_i$$

$$\frac{dS}{dt} = \frac{\alpha_s - S}{\tau} - \sum_i \frac{N_i}{Y_i}\frac{\mu_i S}{\kappa_i + S}$$

where Y_i is the yield coefficient for growth on glucose, $\alpha_s = 0.2$ % is the supply added every time step $\tau = 48$ hours. We set $Y_i = 42$ (in units of O.D. per percent glucose) for each species [2]. We also assume that the maintenance energy is 7.6 mmol ATP per gram cell dry weight (gCDW) per hour which corresponds to a growth rate of approximately 0.02 hour^{-1} [3]. Simulations were performed in MATLAB 2015a, using the ode45 solver, and all fitting to experimental data was done using the *fit.m* function in MATLAB. Fitted Monod curves are plotted in Fig. S16A, and the outcome of a representative simulation are plotted in Fig. S16B. Note that in Fig. S16B, initial conditions were chosen to match experimental relative abundances after the passaging experiment (generation 84). In all simulations, *Raoultella* outcompeted all other strains leading to competitive exclusion.

[2] A yield coefficient of 0.5 g/ g glucose was used for each species (BNID 105318). Assuming that gCDW/cell is roughly 150 fg (BNID: 103894), and 1 O.D. per mL is 8×10^8 cells (BNID: 100985), then $\frac{0.5\ \text{gCDW}}{1\ \text{g glucose}} \times \frac{0.01\ \frac{\text{g glucose}}{\text{mL}}}{\%\ \text{glucose}} \times \frac{1\ \text{cell}}{150 \times 10^{-15}\ \text{gCDW}} \times \frac{1\ \text{O.D.}}{8 \times 10^8\ \frac{\text{cells}}{\text{mL}}} = 42\ \frac{\text{O.D.}}{\%\ \text{glucose}}$

[3] The value of maintenance energy was estimated used *Escherichia coli* measurements on glucose minimal media during exponential growth (BNID:111285). This value was converted into the estimated minimum per capita growth rate per hour using the following dimensional analysis: $\frac{7.6 \times 10^{-3}\ \text{mole ATP}}{1\ \text{gCDW} \times \text{h}} \times \frac{1\ \text{mole glucose}}{36\ \text{mole ATP}} \times \frac{1\ \%\ \text{glucose}}{0.01\ \text{g glucose.}} \times \frac{0.000012\ \text{gCDW}}{1\ \text{O.D.}_{600\ nm}} \times \frac{42\ \text{O.D.}_{600\ nm}}{\%\ \text{glucose}} = 0.0181\ \text{h}^{-1}$

Supplementary Figures

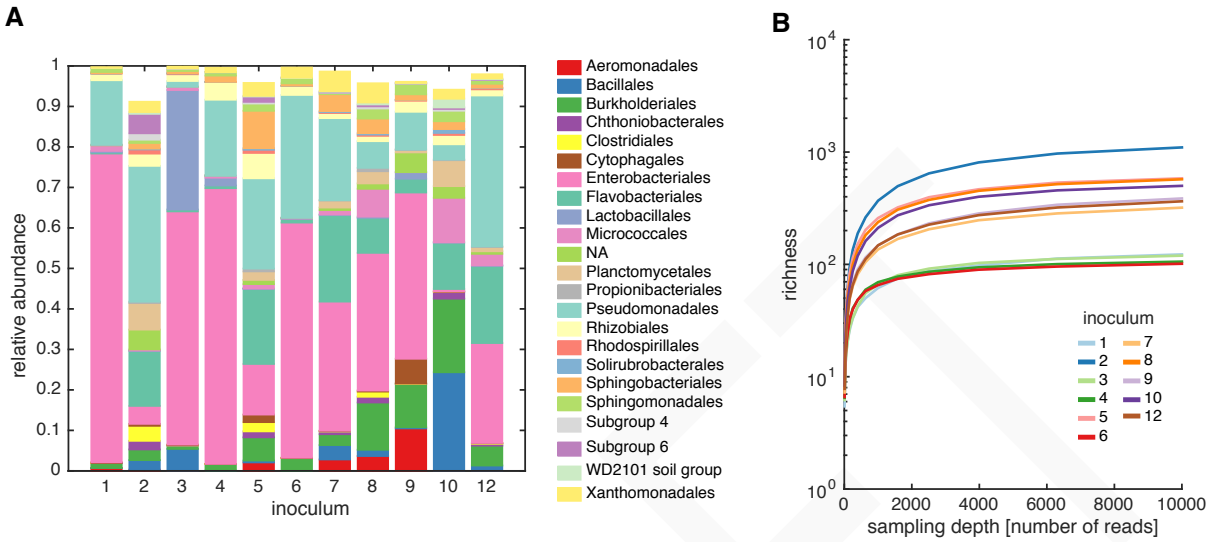

Fig S1: Characterization and diversity of microbiomes isolated from plant and soil samples. (A) 16S sequencing results for 11/12 initial inocula (labeled 1-10, 12 on the *x*-axis). Stacked bar-plots show the community composition at the Order taxonomic level. B) Rarefaction curves for each inoculum community; the average of 100 random samples of a fixed sampling size (*x*-axis) was plotted against the number of unique exact sequence variants (ESV) (*y*-axis). The number of unique 16S sequences spanned an order of magnitude, ranging from 110-1290 exact sequence variants. Note that we were unable to generate amplicon libraries for inoculum 11.

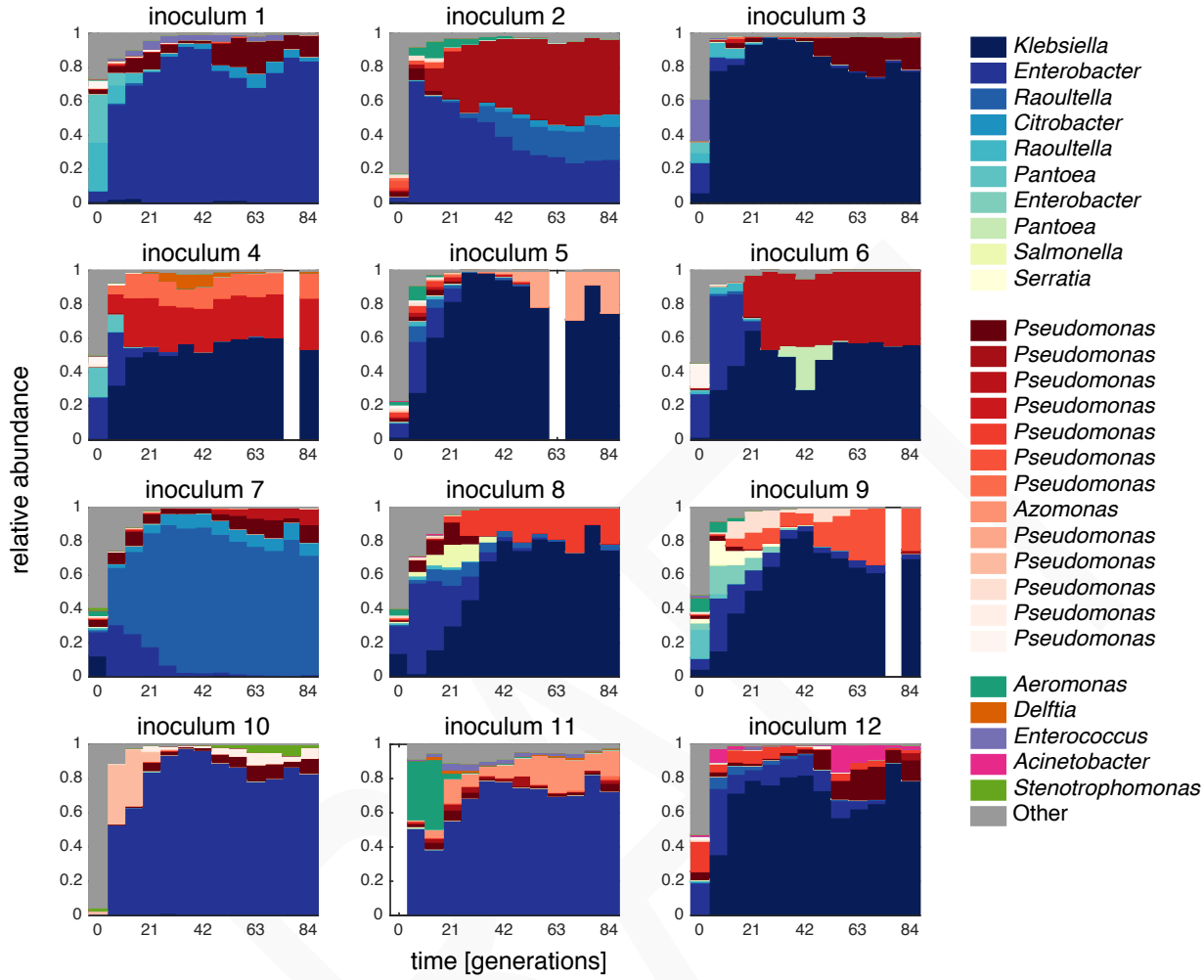

Fig S2: Dynamics of ex-situ community composition over 84 generations in glucose-supplemented media. Communities were transferred into fresh media every 48 hours, allowing approximately seven growth generations per transfer. After each transfer, we determined the community composition using 16S rRNA amplicon sequencing (see methods). The relative abundance of each taxon was plotted as a function of time (generations). All inocula appear to reach stable community structures by the 60th generation.

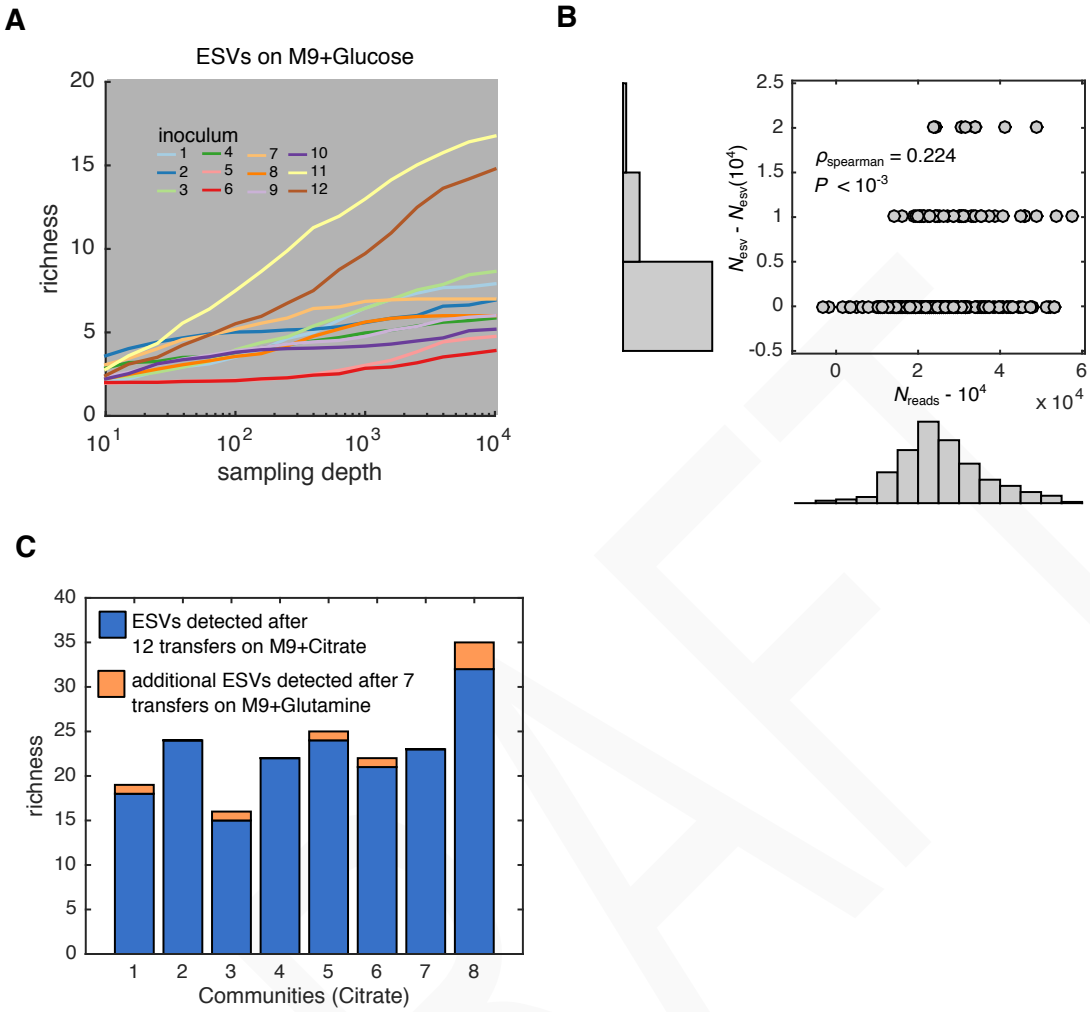

Fig S3: Presence of sparse rare taxa in *ex situ* assembled microbial communities. (A) Rarefaction curves were produced by subsampling a fixed number of reads and computing the number of unique exact sequence variants (ESVs). The plot shows the average over 100 samples at each fixed sampling depth (*x*-axis) for each of the 12 inocula. (B) For each stabilized community, we aimed to estimate the presence of sparse rare taxa on our stabilized communities by measuring the number of additional ESVs detected at sampling depths above 10,000 reads. We plotted the number additional reads above 10,000 (*x*-axis) vs the number of additional ESVs detected at sampling depths above 10,000 reds (*y*-axis). Although there appears to be a positive correlation between additional sampling depth and additional reads, at-most 2 additional ESVs were detected at sampling depths of ~60,000 reads. (C) To further quantify the presence of rare taxa in our samples, we took eight communities stabilized on M9+citrate and passaged them on M9+glutamine for an additional 7 transfers, and sequenced at an average depth of 25,000 reads. The number of ESVs detected in the communities passaged on M9+citrate is plotted as blue bars, and the additional ESVs detected in the communities passaged on M9+glutamine are plotted as orange bars, where between 0-3 additional ESVs were detected when passaged on glutamine.

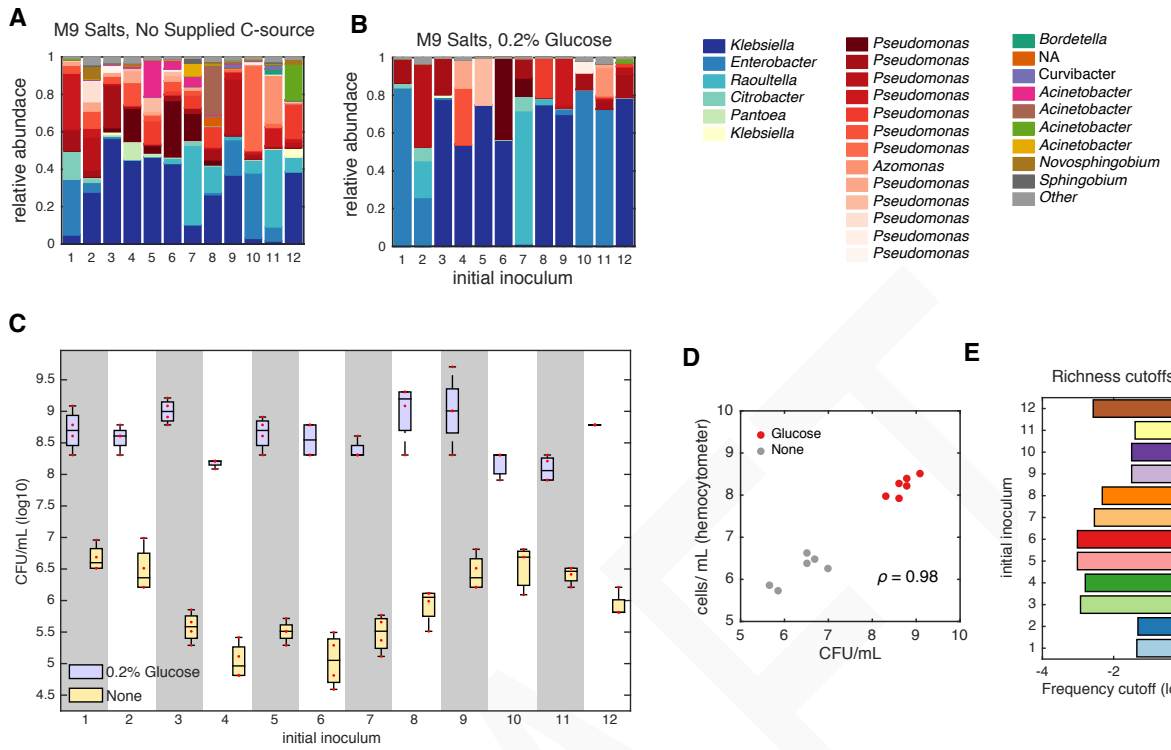

Fig S4: Low levels of bacterial growth with no externally supplied carbon source. (a) We repeated passaging experiments without an externally supplied carbon source, and observed that widespread and diverse communities survive over the course of 84 generations. Communities were similar in structure to glucose supplemented communities (b), but with higher diversity. c) To determine the richness of communities surviving primarily on the externally supplied resource, we plated the communities after 84 generations and counted colony forming units (CFU). We plotted the CFU/mL in replicates of four for each inoculum passaged either on M9 with 0.2% glucose (blue) or on M9 with no supplemented carbon source (yellow). In all cases, populations sizes were orders of magnitude lower when no carbon source was provided compared to population sizes of communities grown on glucose. d) hemocytometer cell counting was performed to verify that CFU accurately recapitulated cell densities. For 12 samples, hemocytometer counting was performed and compared to CFU counts, exhibiting strong positive correlation (Pearson's correlation, $\rho = 0.98$). e) Measurements of absolute population sizes allowed us to define relative abundance cutoffs for communities grown on glucose, ensuring that growth of taxa above the relative abundance cutoff was primarily a consequence of the externally supplied glucose.

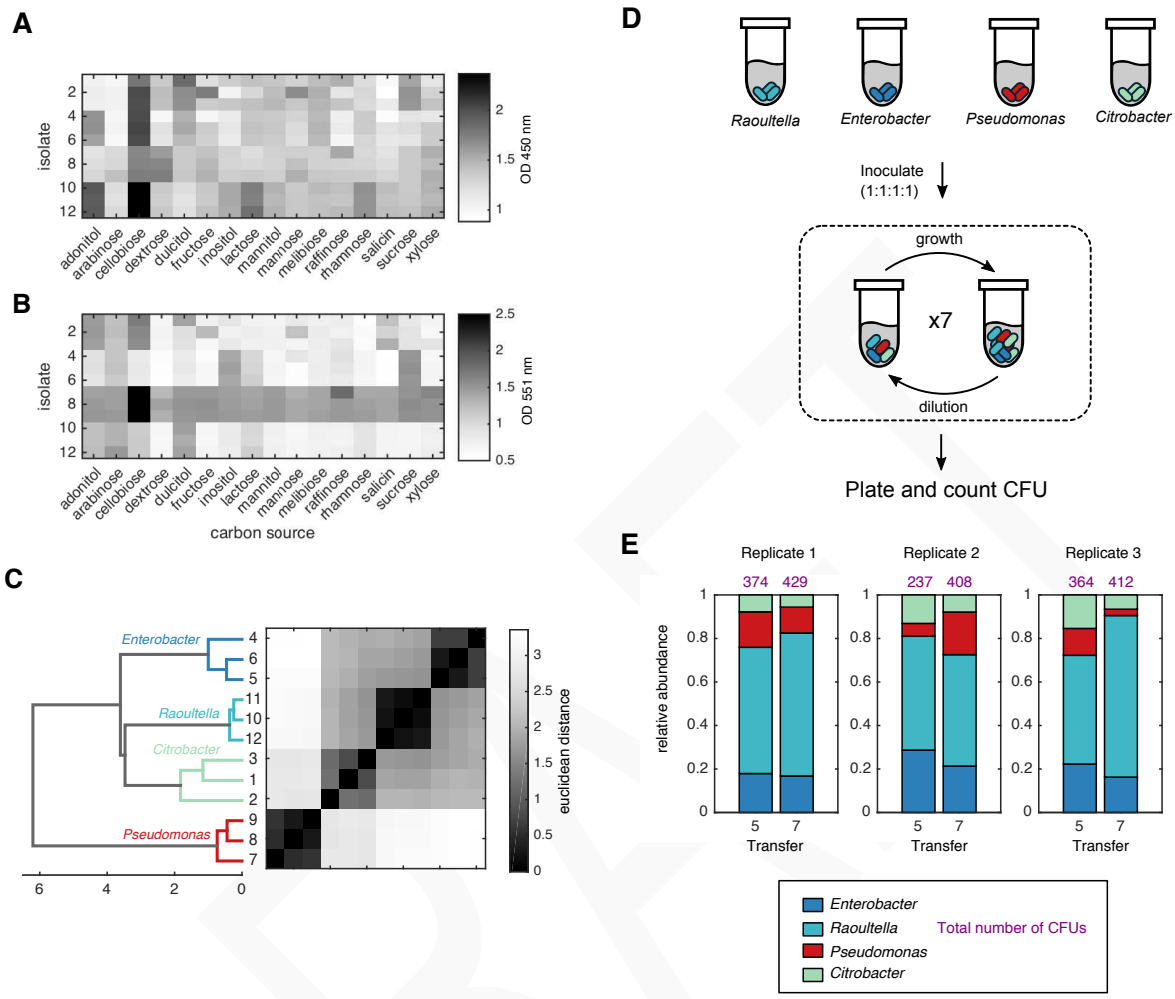

Fig S5: Four strains from a representative community coexist in reconstituted communities. 12 isolates were picked from a representative community from inoculum 2 with 4 distinct morphologies. (A-B) Isolates were grown in phenol red broth with the addition of one of 16 carbon sources. Optical density (OD) was measured at 450 nm and 551 nm after 19 hours to track the degree of acidification from fermentation. (C) The O.D. profiles were hierarchically clustered, revealing 4 clusters of isolates with distinct fermentation profiles, corroborating morphology and sequencing results. These results indicate that the 12 isolates belong to one of four taxa, (D) To see if these four taxa could coexist without the presence of other community members, we inoculated M9+0.2 glucose with equal proportions of each taxa, passaged them for seven dilution cycles and plated the final populations. We counted the colony forming units (CFUs) and distinguished each taxon based on morphology. (E) The relative abundance of three replicates at transfers show that all four taxa coexist after seven transfers.

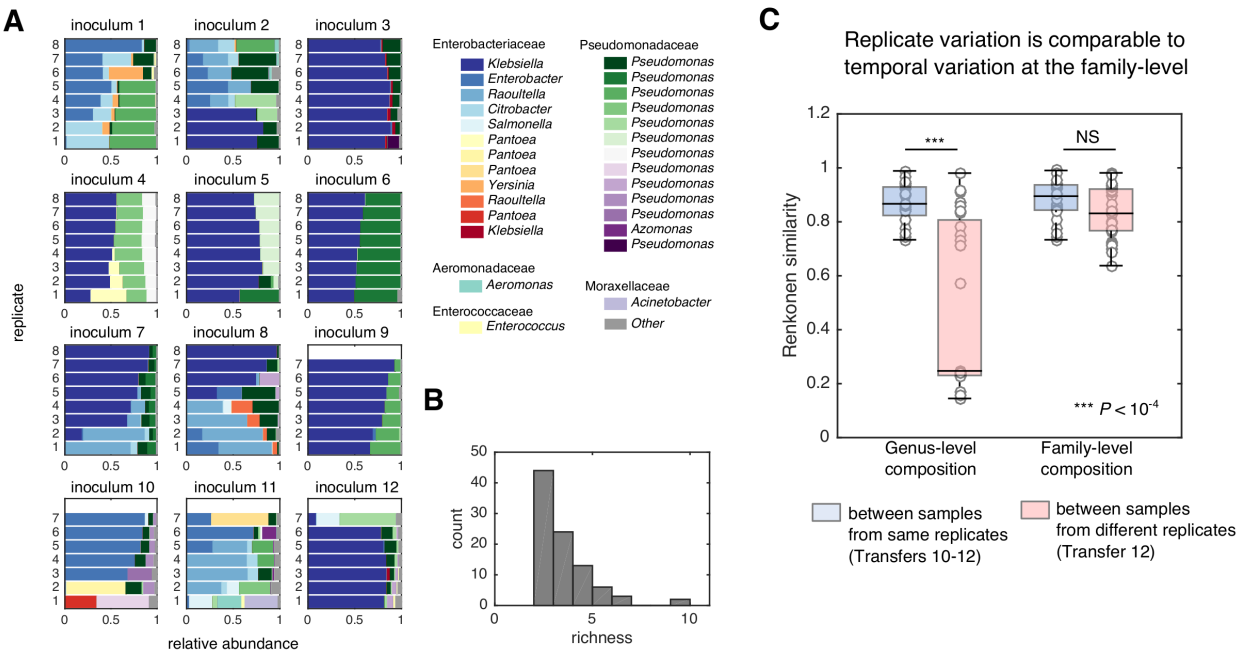

Fig S6: The community structure from the same inocula can be highly variable and the genus level, but similar at the family level. Passaging experiments of microbial communities on M9 + 0.2 % glucose were repeated with up to 8 replicates per inoculum. (A) Each subplot is the relative abundance of the exact sequence variants (ESVs) for all replicates originating from the same inoculum. Note that for each inoculum, fixed points range from multiple (e.g. inoculum 2) to a single attractor (e.g. inoculum 6). (B) The distribution of richness (see Fig. S4) estimates across all communities formed in (A) showed that all large-scale competitive experiments retained at-least 2 sequence variants, and the majority (48/92) retained more than four sequence variants. (C) . To characterize the variability of community structure across different starting replicates at various levels of taxonomic resolution, we computed the Renkonen similarity (at both genus and family-levels) between replicate communities from inocula 2 after 12 transfers. As a comparison, we computed the Renkonen similarity between samples obtained at the end of the last three transfers (transfer 10-12) within the same replicate. The boxplots are distributions of Renkonen similarities between both within replicates (blue) and between replicates (red) at the genus (left) and family (right) taxonomic levels. Communities are significantly less similar at the genus level when comparing between replicates vs. within replicates (Mann-Whitney U-test: $P < 10^{-4}$), while communities are of comparable similarity at the family level when comparing samples from different replicates vs. samples from the same replicate (Mann-Whitney U-test: $P = 0.06$).

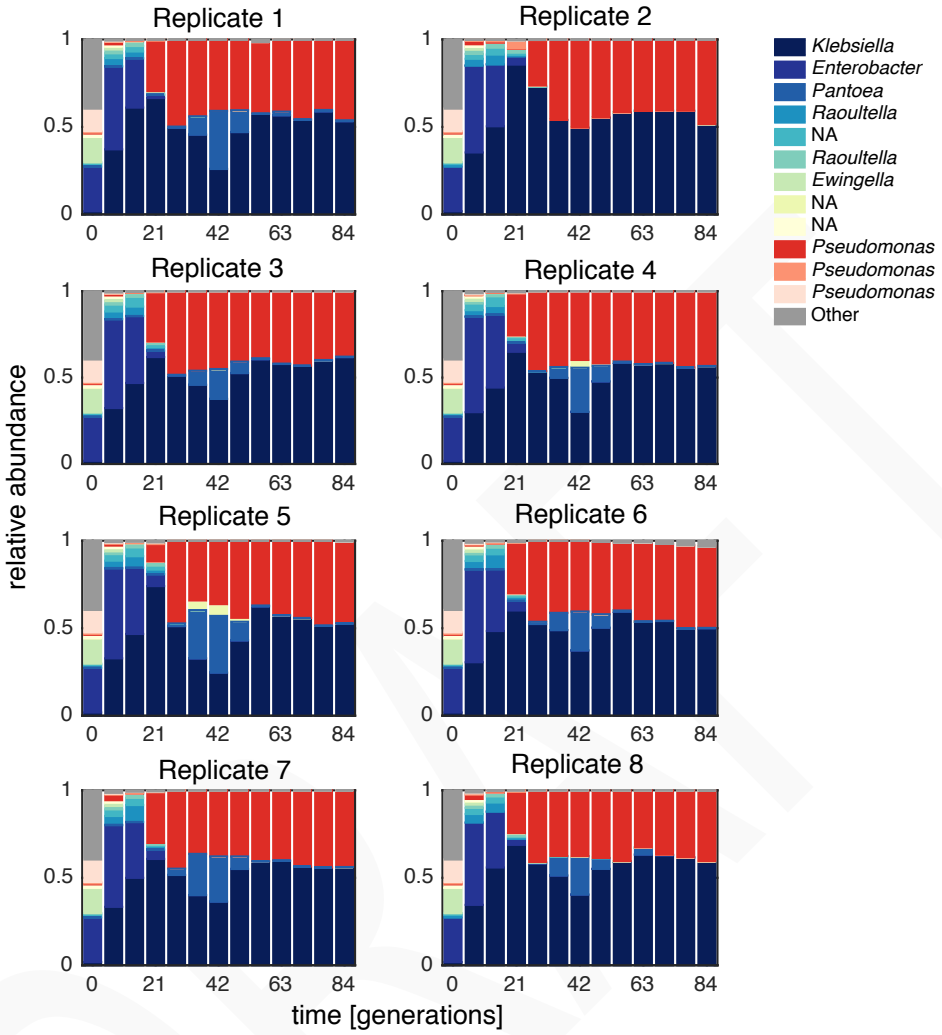

Fig S7: Inoculum 6 exhibits strongly deterministic population dynamics. We performed replicate passaging experiments starting with inoculum 6 and found nearly reproducible population dynamics. Each subplot shows the relative abundance of sequence variants (*y*-axis) during the course of the passaging experiment (*x*-axis). Notably, in 7/8 replicates, a bloom of a *Pantoea* sequence variant occurred at the 42nd generation.

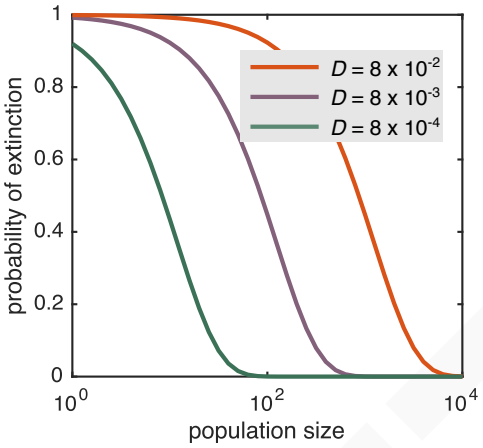

Fig S8: Bottlenecks imposed by dilutions are unlikely to induce extinctions. We calculate the probability of extinction by stochastic sampling as a function of the size of the population for a given species, for the dilution factor we apply in our experiments (D=0.008; purple line) as well as for 10-fold larger (red) and 10-fold smaller (green) dilution factors. We note that all of the ESVs that we detect in our community 16S sequencing have population sizes of at least 10,000.

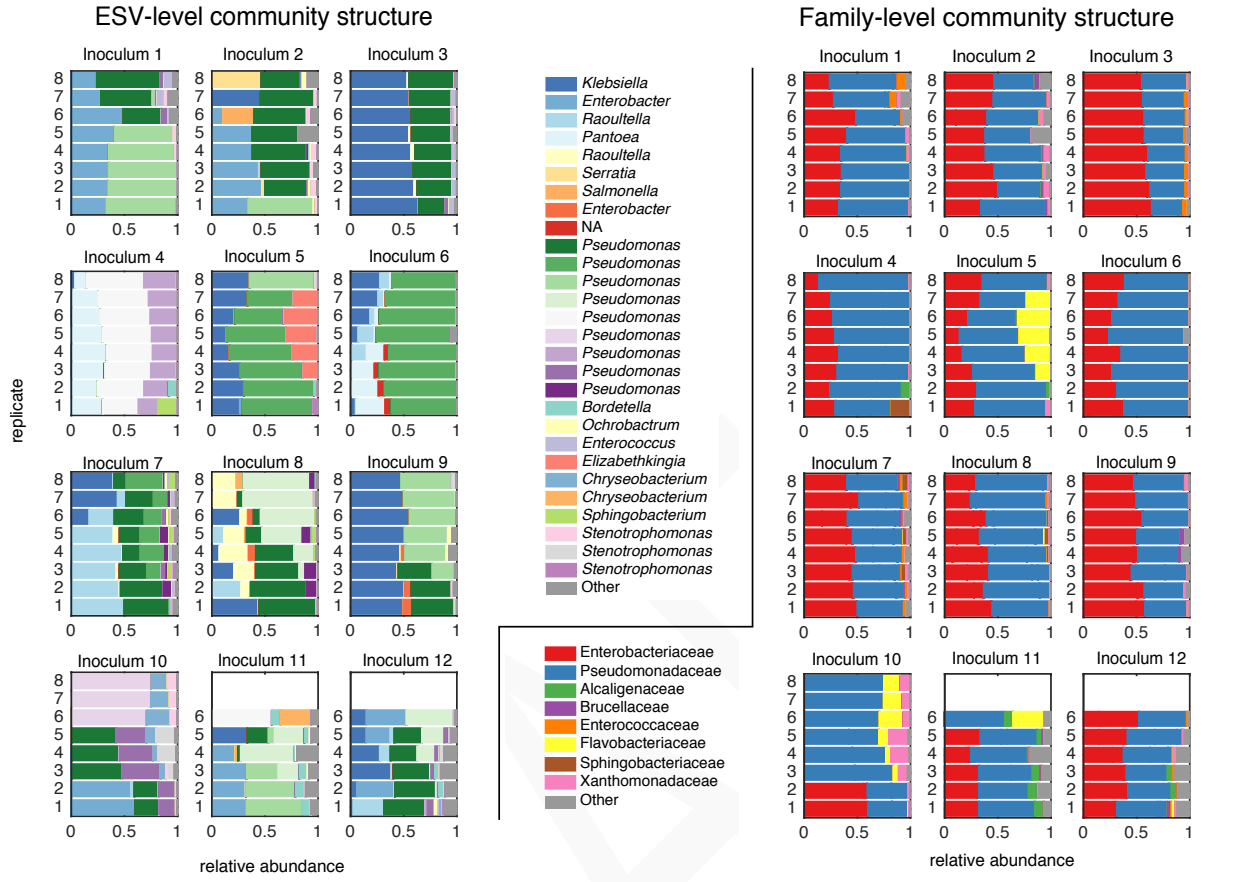

Fig S9: Community structure at ESV and family level on citrate. Passaging experiments of microbial communities on M9 + 0.07 C-mole/L citrate were performed with up to 8 replicates per inoculum, as in the case with glucose.

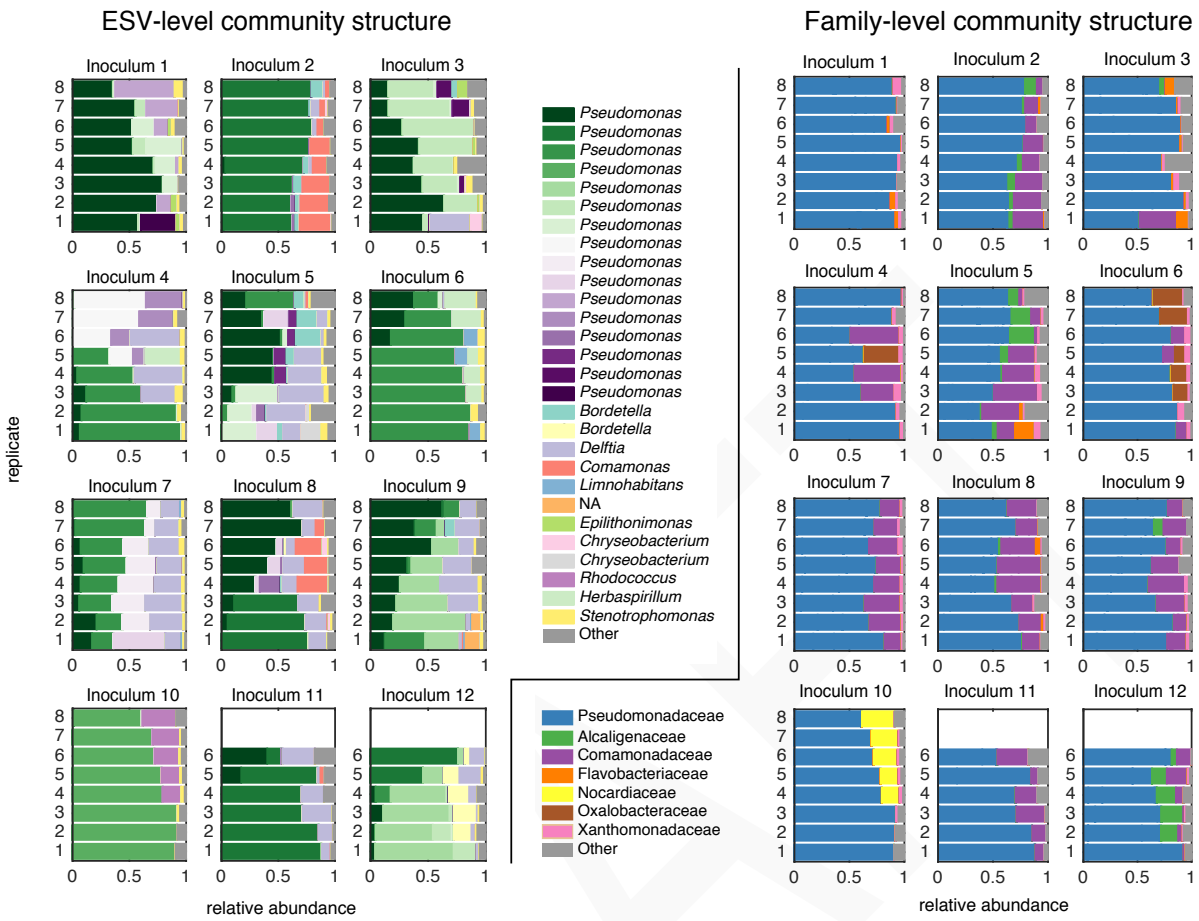

Fig S10: Community structure at ESV and family level on leucine. Passaging experiments of microbial communities on M9 + 0.07 C-mole/L leucine were performed with up to 8 replicates per inoculum.

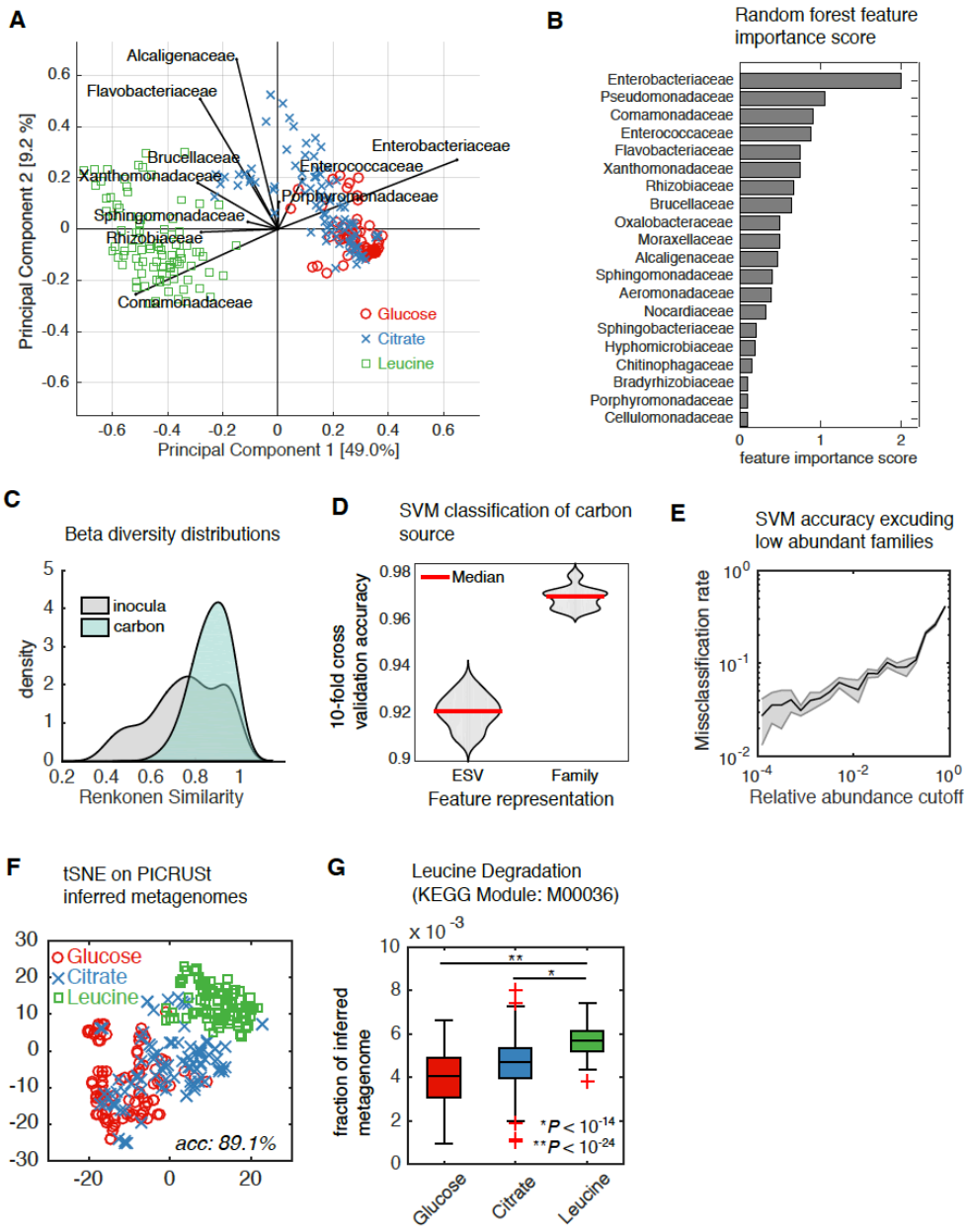

Fig S11: Family-level composition is a strong taxonomic predictor of the externally-supplied carbon source.

(A) The family-level community composition was log-transformed and dimensionally reduced using principal component analysis. Like in Fig 3A, family-level community structure was strongly associated with the carbon source in the media. A biplot was used to show which taxa were correlated with the first two principal components. (B) A random forest classifier was trained to predict carbon source from the family-level community structure, and out-of-bag feature importance scores are reported, confirming that the abundance of Enterobacteriaceae and Pseudomonadaceae are important predictors of carbon source. (C) The distributions of Renkonen similarities between family-level compositions between samples either grown on the same

carbon source (light blue, N=12558) or between samples from the same inocula (grey N=3056) are plotted, revealing that the communities grown on the same carbon source are more similar than communities grown from the same inocula (one-tailed Kolmogorov-Smirnov test; $P < 10^{-5}$). (D) A support vector machine (SVM) classifier was used to train a model to predict the carbon source (glucose, citrate or leucine) from the clr-transformed community structure at the ESV or family level. Models were trained using different coarse-graining descriptions of community structure based on taxonomy (x-axis) and the 10-fold cross-validation accuracy (repeated 10 times) for each model is reported on the y-axis. (E) An SVM was retrained using families above a pre-defined threshold (x-axis), and the misclassification rate (1-accuracy) is reported on the y-axis, revealing that low-abundant families aid in model performance. (F) Metagenome compositions were imputed using PICRUSt (*40*) and embedded in a two-dimensional space using t-distributed stochastic neighbor embedding (tSNE). (G) The summed abundance of genes belonging to the leucine degradation KEGG module (M00036) are plotted for all samples using a boxplot, where samples are grouped by the limiting carbon source (x-axis). Leucine degradation genes are enriched in communities grown on leucine relative to communities grown on citrate (Mann Whitney U-test: $P < 10^{-14}$) or glucose (Mann Whitney U-test: $P < 10^{-24}$).

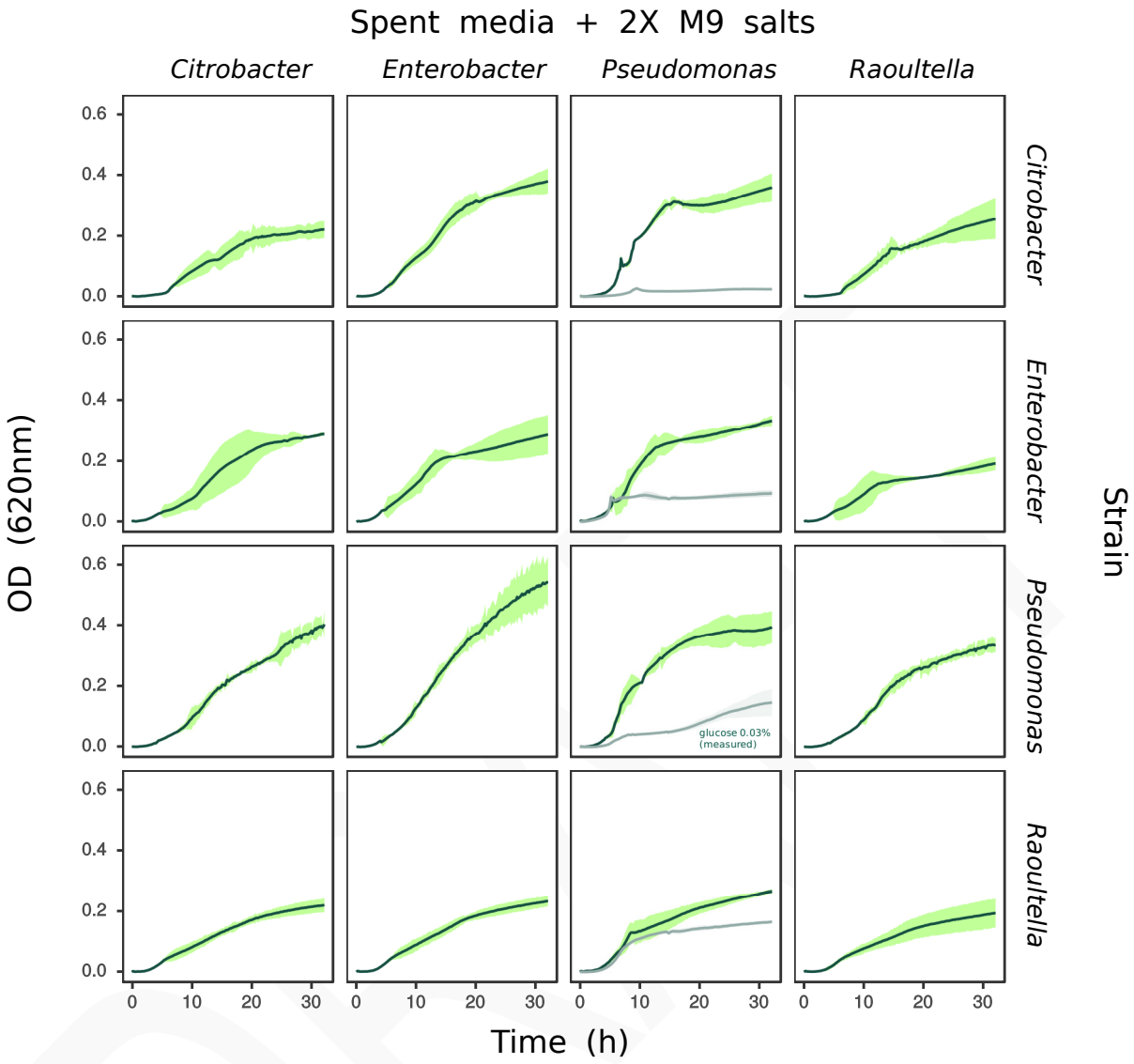

Fig S12: Isolates grown on each other's metabolic byproducts. Isolates were grown for 48 hours and the spent media (SM) was used to synthesize a new growth media (see Methods). Each subplot is the growth curve of one of four strains (vertical axis) on media synthesized from byproducts secreted during monoculture growth of the strain on the horizontal axis. Plots show the average growth across 3 replicates, and shaded regions denote the 95% confidence interval. Note that *Pseudomonas* spent media contained at a residual abundance of glucose (0.03%). Light grey lines show growth on M9+0.03% glucose, which is less than the growth on *Pseudomonas* spent media. This indicates that the growth on the spent media from *Pseudomonas* is not solely explained by the availability of residual glucose.

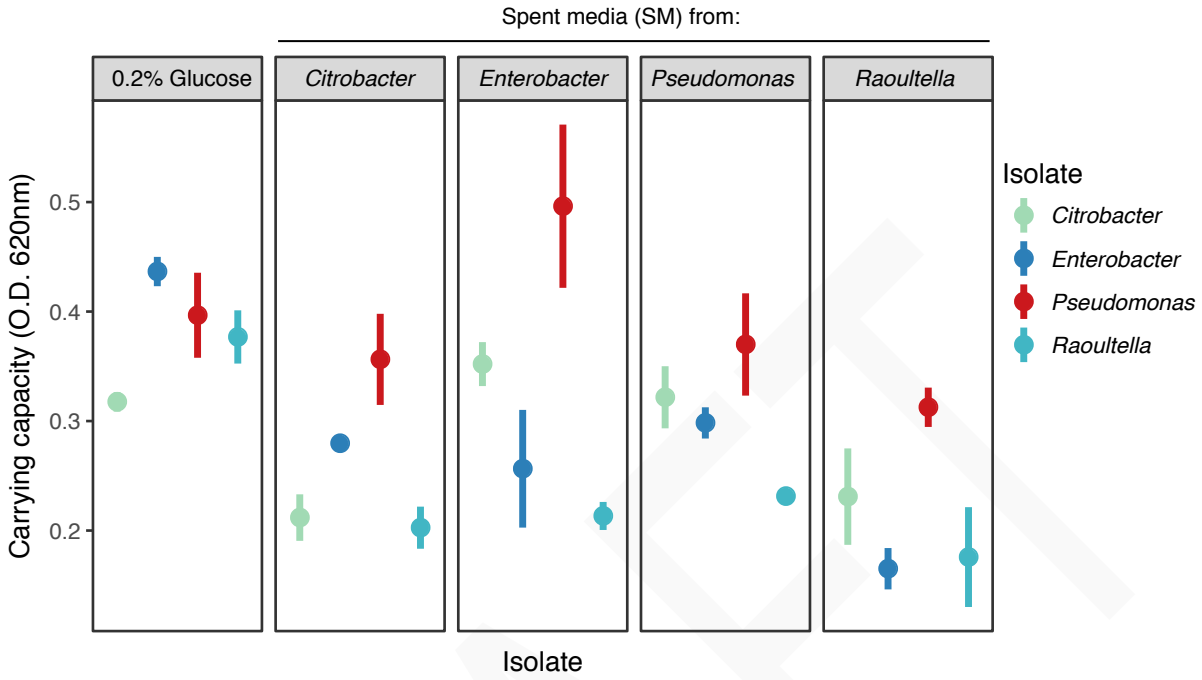

Fig S13: Carrying capacities on secreted byproducts are comparable to growth on glucose. Logistic growth curves were fitted to each growth curve measured in Fig. S12 and the distribution of carrying capacities for each isolate (grouped box-plots) grown on glucose or indicated isolate's spent media is plotted in each subplot.

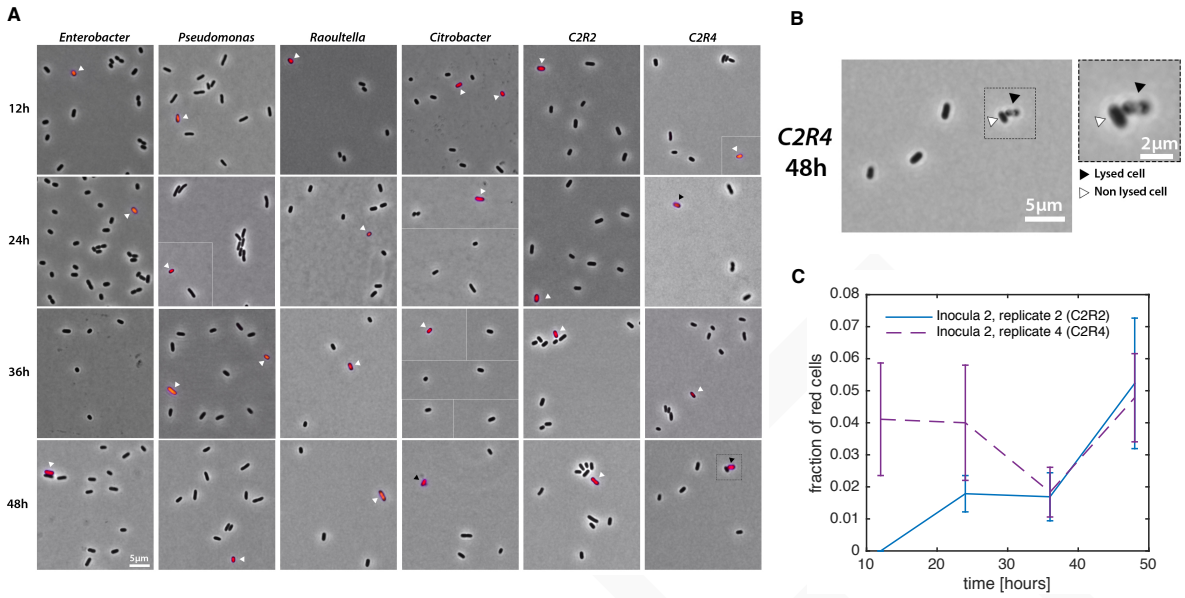

Fig S14: Cell death and lysis are not likely the major source of secreted resources. We used a live/dead cell assay (see methods) to estimate the number of dead cells in two replicate communities (replicate 2 and 4) and monocultures of isolates obtained from replicate 4 from inoculum 2 after 12, 24, 36 and 48 hours of growth on minimal media with glucose. (A) Images show representative dead cells (red fluorescence) for all samples. White triangles appear next to cells that have not lysed, while black triangles appear next to cells that have lysed (lysis is shown in adjacent insets). (B) An example of lysed cell (black triangle) and non-lysed cell (white triangle). (C) The fraction of cells that stained red is on the *y*-axis, which is a proxy for cell death. Error estimates were generated by using the measured binomial sampling variance. It is worth noting that in Fig. 3F in the main text, the average increase in biomass is approximately 5-fold greater than the proportion of estimated dead cells, suggesting that consumption of dead cellular material is not sufficient to explain results presented in Fig. 3F.

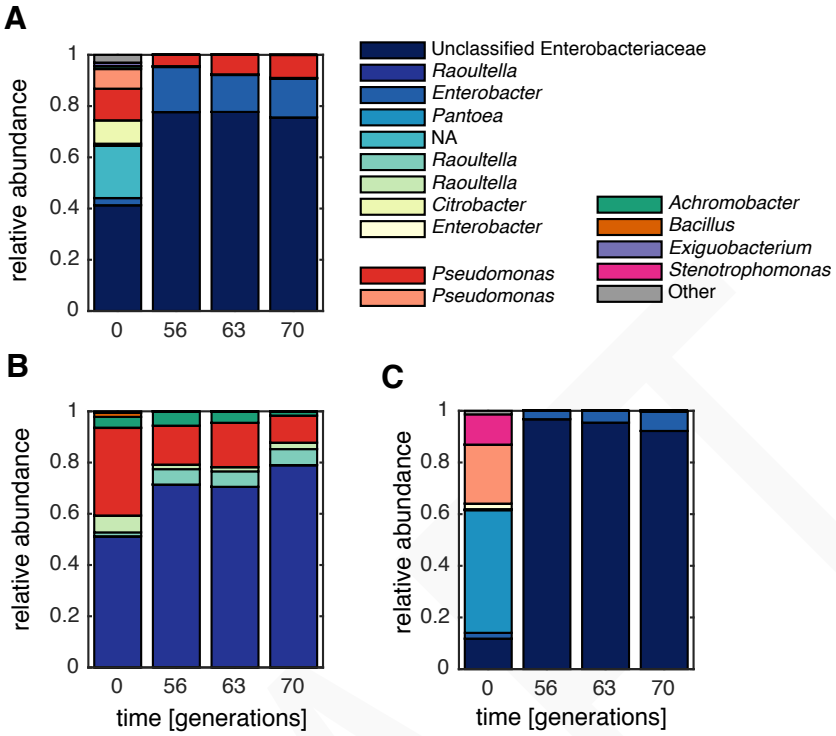

Fig S15: Repeating the experiment with vigorous shaking does not result in massive loss of coexistence. Spatial structure in our 96-well plate format could also allow for coexistence of microbial species (*33*). Thus, experiments were repeated for three separate inocula passaged on media with M9+0.2% glucose, but while vigorously shaking cultures at 200 RPM. In all cases, no single strain outcompeted all other strains, suggesting that coexistence is stable even without potential spatial heterogeneity.

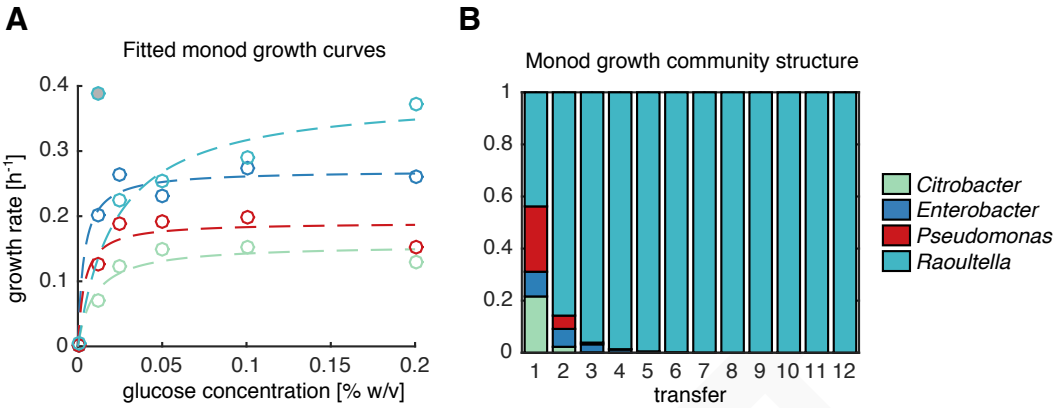

Fig S16: Effect of resource abundance on the growth rates of individual species. A potential mechanism for coexistence among microbes in an environment with a single limiting nutrient is each species has maximal fitness at least one intermediate level of the limiting nutrient (*21*). Thus, isolates from a representative community were grown at various concentrations of glucose (subplot (A), *x*-axis), and the initial growth rate was measured (See Monod model section in the Supporting Information), and fitted to a Monod growth model. *Raoultella* displayed unusually high growth rates at low glucose concentrations. In (A), we removed this outlier (grey dot) at very low resource abundances. We used the Monod parameters to simulate a batch culture passaging experiment (B), and found that *Raoultella* competitively excludes all other species *in silico*. If the outlier observed at low growth rates is retained, *Raoultella* still competitively excludes all other species. Together, these results indicate that there is no supporting evidence of resource abundance-dependent fitness effects that lead to coexistence amongst these strains.

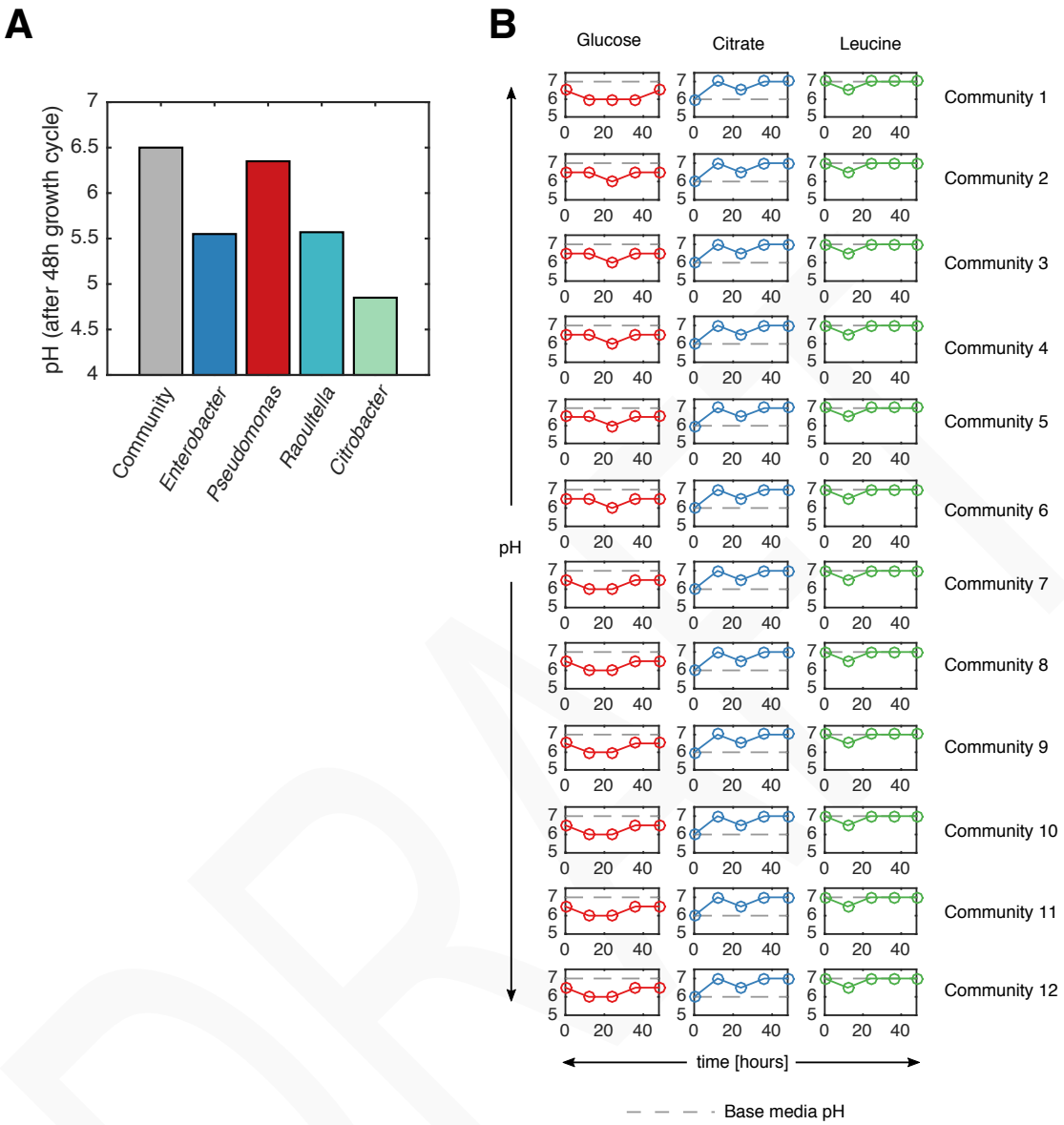

Fig S17: Communities buffer pH fluctuations during growth: (A) pH was measured after the 48 hour growth cycle in a representative community (grey bar), and compared to the pH of the media after growth of individual isolates (colored bars). Monocultures lowered the pH more than the community. (B) To determine whether communities buffered pH fluctuations generically, we thawed stable communities from each inocula passaged on either glucose, citrate and leucine, and passaged them for one additional growth cycle and measured the pH at 12 hour intervals (colored lines). Interestingly, pH dropped initially for communities grown on glucose or leucine, but increased during the the last phase of growth. Only for communities grown on citrate did the pH change by ~ 1 pH unit, suggesting that there are not major pH fluctuations during the growth of these microbial communities.

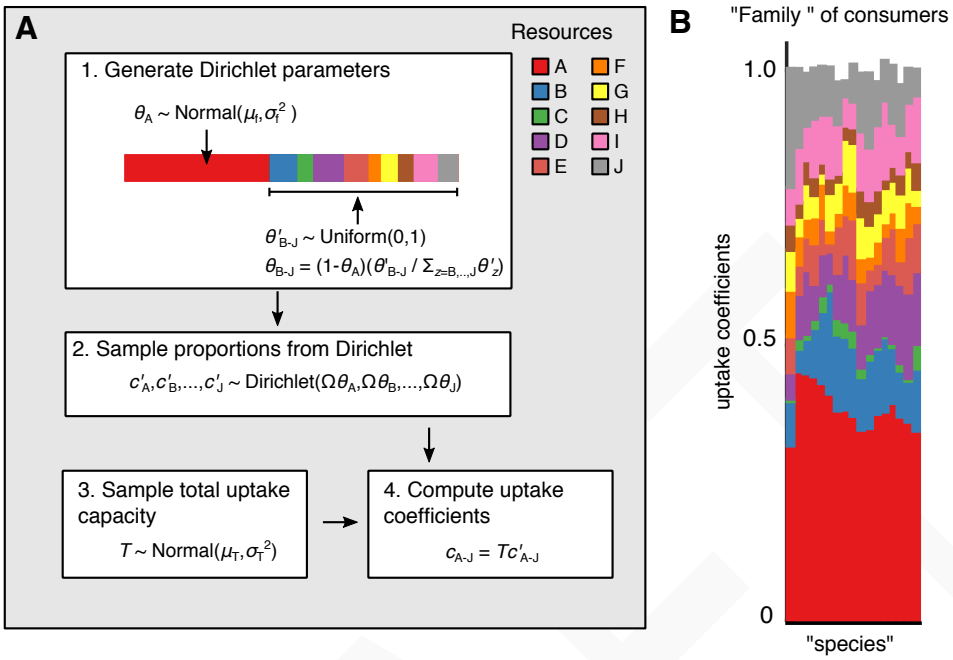

Fig S18: Generation of "families" of consumers in consumer resource models. (A) A flow diagram describing the processes of generating families of consumers in consumer resource models. (1) First we define a set of parameters for a Dirichlet distribution specifying the proportion the consumer's total uptake rate taken by each resource (θ_{resource}), where we each family has a preferred resource (red). (2) We then sample uptake proportions for each resource α, c'_α from the Dirichlet distribution, and multiply these values with a species dependent total uptake capacity (Step 3, T) to obtain the consumption rate of resource α for each consumer. (B) A stacked bar plot showing the uptake coefficients (consumption rates, $c_{i\alpha}$) for each sampled consumer and resource. Although each species has different uptake rates for each resource, each consumer has a high uptake coefficient for resource A.

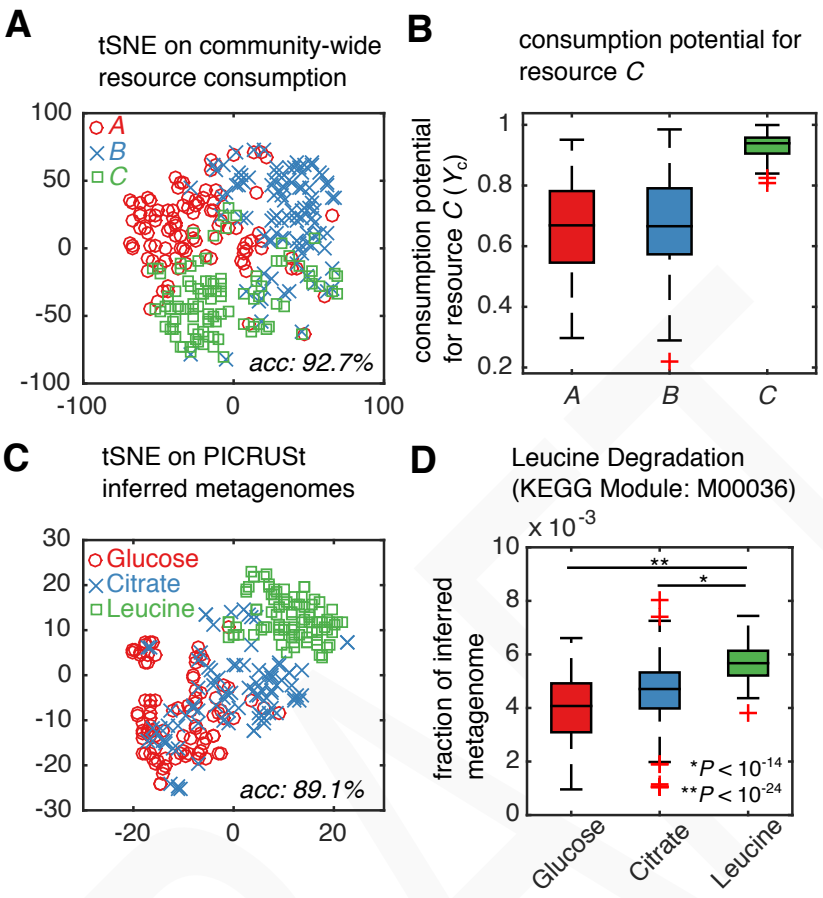

Fig S19: Functional clustering is observed in both consumer resource models and experiments. (A) Simulations of the microbial consumer resource model (see SI text) was performed by randomly sampling consumer and stoichiometric matrices from uniform distributions, then supplying one of three resources in the environment (denoted as A, B and C here), and the communities' capacity to consume each resource was computed (Supplemental information). t-distributed stochastic neighbor embedding (tSNE) was used to reduce dimensionality of the resource uptake vectors and plotted in 2-D, which revealed clustering of uptake capacity based on the identity of the resource in the environment. (B) The distribution of community-wide uptake capacity for resource C when grown on three different resources (*x*-axis). Note that even in the presence of stabilizing mechanisms like cross-feeding, the dominant signal is the capacity to uptake the primary nutrient. (C) predictions from the model are compared to experiment, where we performed dimensionality reduction on inferred metagenomes. We then computed the relative abundance of genes used for leucine degradation (D), showing that communities grown in leucine are enriched genes involved in leucine degradation relative to communities grown in citrate (Mann Whitney U-test: $P < 10^{-14}$) or glucose (Mann Whitney U-test: $P < 10^{-24}$). Note that in (A) and (C), SVMs were trained to predict the carbon source from either the community-wide uptake rates (in A) or the metagenome (in C), and the leave-one out cross-validation accuracy is reported in the lower right corner.

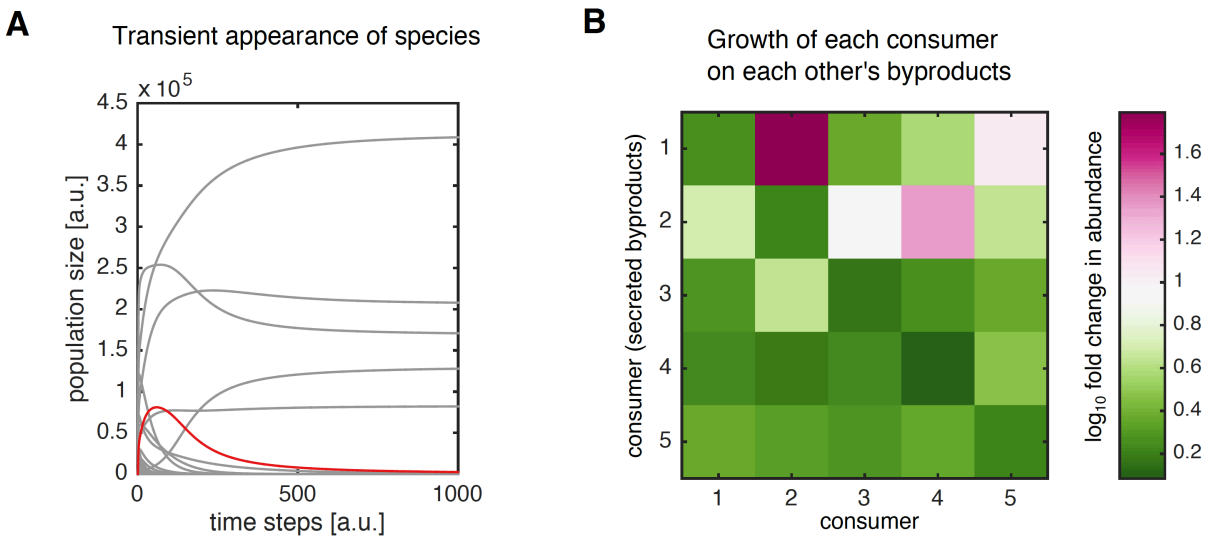

Fig S20: Microbial consumer resource model can recapitulate key experimental findings: (A) Numerical simulations of the microbial consumer resource model (MCRM) often display trajectories of species that grow to large densities before going extinct at steady state (red line), similar to experimental results found in Fig. S7. (B) We performed simulations modeling the experiment performed in Fig. 3, where individual consumers produced byproducts that were used as substrates for the growth of other strains. For this simulation, a stable community was obtained using random sampling of consumer and secretion rates resulting in a 5 "species" community. For each "species", we simulated batch culture growth by not resupplying resources, and obtaining the secreted byproducts after 48 time steps. We used these byproducts as the input resources for simulations of batch culture for each isolate. The fold change in population size for each consumer (*x*-axis) growth on the byproducts of each consumer (*y*-axis) is presented as a heat-map. Notice that all values are above 0, indicating that each "species" grew on the byproducts of others.

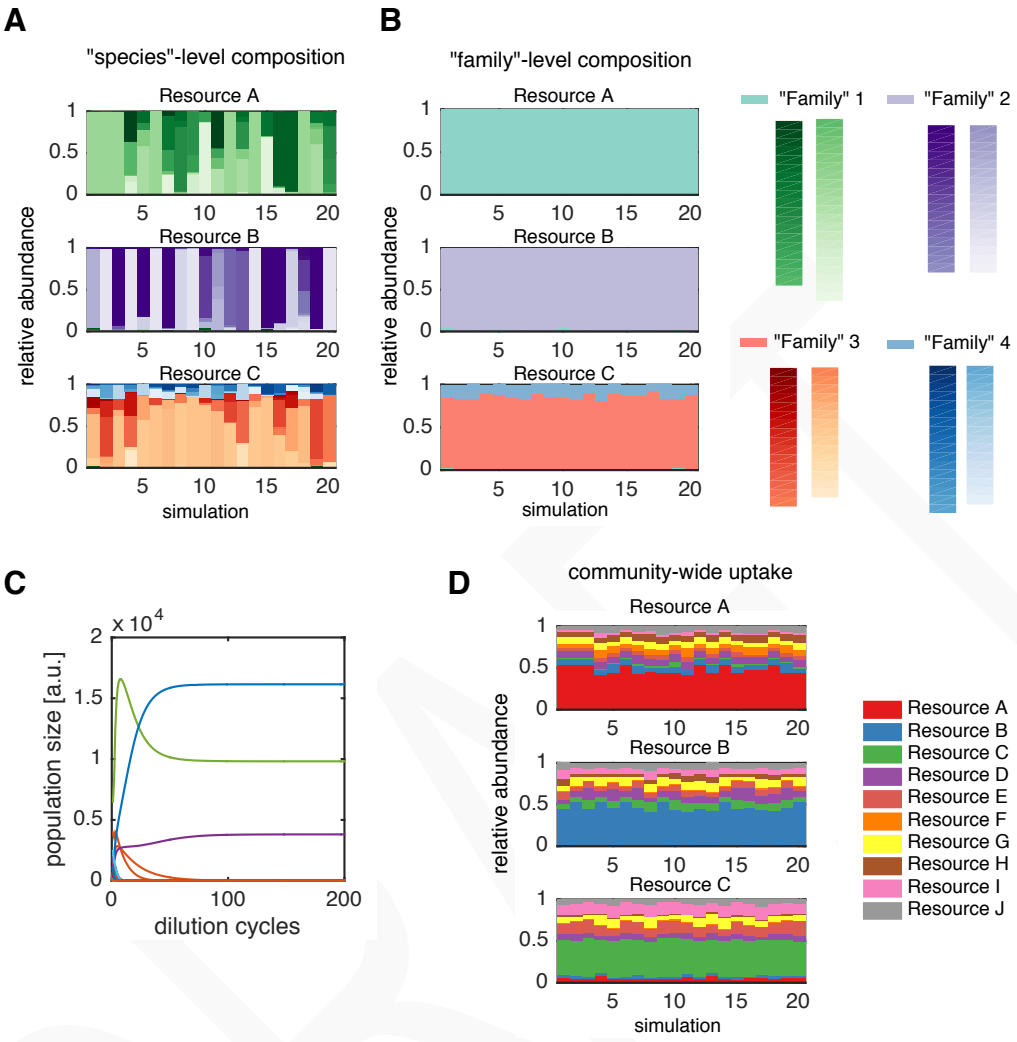

Fig S21: Major qualitative features of the model are unaffected by an oscillating resource supply. To determine whether the qualitative features of the MCRM are affected by the functional form of the resource supply term, we simulated growth dynamics in batch culture with dilutions after each growth cycle (48 time steps). We rescaled both the uptake coefficients and secretion parameters to ensure resources were not immediately consumed during the growth cycle (see SI text). All other parameters were the same as those presented in Figure 4 of the main text. Like the simulations presented in the main text using continuous resource dynamics, simulations with non-continuous resource dynamics resulted in (A) species variability, (B) family-level convergence, (C) coexistence and (D) functional convergence just as in the chemostat simulations presented in main text.

References and Notes

1. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008). doi:10.1126/science.1153213 Medline

2. S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, Tara Oceans coordinators, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, P. Bork, Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015). doi:10.1126/science.1261359 Medline

3. Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012). doi:10.1038/nature11234 Medline

4. E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. M. Bohannan, D. A. Relman, The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012). doi:10.1126/science.1224203 Medline

5. P. J. Turnbaugh, M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, J. I. Gordon, A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009). doi:10.1038/nature07540 Medline

6. S. Louca, S. M. S. Jacques, A. P. F. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, M. Doebeli, High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* **1**, 0015 (2016). doi:10.1038/s41559-016-0015 Medline

7. S. Louca, L. W. Parfrey, M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016). doi:10.1126/science.aaf4507 Medline

8. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A phylogenetic perspective. *Science* **350**, aac9323 (2015). doi:10.1126/science.aac9323 Medline

9. C. Burke, P. Steinberg, D. Rusch, S. Kjelleberg, T. Thomas, Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14288–14293 (2011). doi:10.1073/pnas.1101591108 Medline

10. L. A. David, C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma, M. A. Fischbach, S. B. Biddinger, R. J. Dutton, P. J. Turnbaugh, Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014). doi:10.1038/nature12820 Medline

11. J. Friedman, L. M. Higgins, J. Gore, Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 109 (2017). doi:10.1038/s41559-017-0109 Medline

12. N. M. Vega, J. Gore, Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. *PLOS Biol.* **15**, e2000633 (2017). doi:10.1371/journal.pbio.2000633 Medline

13. D. R. Hekstra, S. Leibler, Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* **149**, 1164–1173 (2012). doi:10.1016/j.cell.2012.03.040 Medline

14. K. R. Foster, T. Bell, Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* **22**, 1845–1850 (2012). doi:10.1016/j.cub.2012.08.005 Medline

15. K. Z. Coyte, J. Schluter, K. R. Foster, The ecology of the microbiome: Networks, competition, and stability. *Science* **350**, 663–666 (2015). doi:10.1126/science.aad2602 Medline

16. J. M. Levine, J. Bascompte, P. B. Adler, S. Allesina, Beyond pairwise mechanisms of species coexistence in complex communities. *Nature* **546**, 56–64 (2017). doi:10.1038/nature22898 Medline

17. E. Bairey, E. D. Kelsic, R. Kishony, High-order species interactions shape ecosystem diversity. *Nat. Commun.* **7**, 12285 (2016). doi:10.1038/ncomms12285 Medline

18. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016). Medline

19. G. Gottschalk, *Bacterial Metabolism* (Springer, 1979).

20. R. Macarthur, R. Levins, Competition, habitat selection, and character displacement in a patch environment. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 1207–1210 (1964). doi:10.1073/pnas.51.6.1207 Medline

21. F. M. Stewart, B. R. Levin, Partitioning of Resources and the Outcome of Interspecific Competition: A Model and Some General Considerations. *Am. Nat.* **107**, 171–198 (1973). doi:10.1086/282825

22. D. Tilman, *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).

23. F. Rodriguez-Valera, A.-B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pasić, T. F. Thingstad, F. Rohwer, A. Mira, Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009). doi:10.1038/nrmicro2235 Medline

24. E. D. Kelsic, J. Zhao, K. Vetsigian, R. Kishony, Counteraction of antibiotic production and degradation stabilizes microbial communities. *Nature* **521**, 516–519 (2015). doi:10.1038/nature14485 Medline

25. J. Grilli, G. Barabás, M. J. Michalska-Smith, S. Allesina, Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210–213 (2017). Medline

26. M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson, T. Hwa, Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature* **528**, 99–104 (2015). doi:10.1038/nature15765 Medline

27. N. Paczia, A. Nilgen, T. Lehmann, J. Gätgens, W. Wiechert, S. Noack, Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microb. Cell Fact.* **11**, 122 (2012). doi:10.1186/1475-2859-11-122 Medline

28. R. F. Rosenzweig, R. R. Sharp, D. S. Treves, J. Adams, Microbial evolution in a simple unstructured environment: Genetic differentiation in Escherichia coli. *Genetics* **137**, 903–917 (1994). Medline

29. S. K. Hansen, P. B. Rainey, J. A. J. Haagensen, S. Molin, Evolution of species interactions in a biofilm community. *Nature* **445**, 533–536 (2007). doi:10.1038/nature05514 Medline

30. R. Baran, E. L. Brodie, J. Mayberry-Lewis, E. Hummel, U. N. Da Rocha, R. Chakraborty, B. P. Bowen, U. Karaoz, H. Cadillo-Quiroz, F. Garcia-Pichel, T. R. Northen, Exometabolite niche partitioning among sympatric soil bacteria. *Nat. Commun.* **6**, 8289 (2015). doi:10.1038/ncomms9289 Medline

31. M. S. Datta, E. Sliwerska, J. Gore, M. F. Polz, O. X. Cordero, Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat. Commun.* **7**, 11965 (2016). doi:10.1038/ncomms11965 Medline

32. P. B. Rainey, M. Travisano, Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998). doi:10.1038/27900 Medline

33. P. Chesson, Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000). doi:10.1146/annurev.ecolsys.31.1.343

34. B. R. Levin, Coexistence of two asexual strains on a single resource. *Science* **175**, 1272–1274 (1972). doi:10.1126/science.175.4027.1272 Medline

35. J. Cremer, M. Arnoldini, T. Hwa, Effect of water flow and chemical environment on microbiota growth and composition in the human colon. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6438–6443 (2017). doi:10.1073/pnas.1619598114 Medline

36. C. Ratzke, J. Gore, Modifying and reacting to the environmental pH can drive bacterial interactions. *PLOS Biol.* **16**, e2004248 (2018). doi:10.1371/journal.pbio.2004248 Medline

37. R. MacArthur, Species packing and competitive equilibrium for many species. *Theor. Popul. Biol.* **1**, 1–11 (1970). doi:10.1016/0040-5809(70)90039-0 Medline

38. A. C. Martiny, A. P. K. Tai, D. Veneziano, F. Primeau, S. W. Chisholm, Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ. Microbiol.* **11**, 823–832 (2009). doi:10.1111/j.1462-2920.2008.01803.x Medline

39. T. Taillefumier, A. Posfai, Y. Meir, N. S. Wingreen, Microbial consortia at steady supply. *eLife* **6**, 1–65 (2017). doi:10.7554/eLife.22644 Medline

40. M. G. I. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko, C. Huttenhower,

Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013). doi:10.1038/nbt.2676 Medline

41. J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, P. D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013). doi:10.1128/AEM.01043-13 Medline

42. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010). doi:10.1038/nmeth.f.303 Medline

43. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013). doi:10.1093/nar/gks1219 Medline

44. P. Chesson, MacArthur's consumer-resource model. *Theor. Popul. Biol.* **37**, 26–38 (1990). doi:10.1016/0040-5809(90)90025-Q

45. A. Posfai, T. Taillefumier, N. S. Wingreen, Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Phys. Rev. Lett.* **118**, 028103 (2017). doi:10.1103/PhysRevLett.118.028103 Medline

46. J.-H. Hehemann, P. Arevalo, M. S. Datta, X. Yu, C. H. Corzett, A. Henschel, S. P. Preheim, S. Timberlake, E. J. Alm, M. F. Polz, Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* **7**, 12860 (2016). doi:10.1038/ncomms12860 Medline

47. J. Aitchison, *The Statistical Analysis of Compositional Data* (Chapman & Hall, 1986).