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,†], and **Alvaro Sanchez**[2,3,†]

[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.

Abstract

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 16S 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.

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 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.

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).

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 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

1. Falkowski PG, Fenchel T, Delong EF, Science 320, 1034–1039 (2008). [PubMed: 18497287]

2. Sunagawa S et al., Science 348, 1261359 (2015). [PubMed: 25999513]

3. Human Microbiome Project Consortium, Nature 486, 207–214 (2012). [PubMed: 22699609]

4. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA, Science 336, 1255–1262 (2012). [PubMed: 22674335]

5. Turnbaugh PJ et al., Nature 457, 480–484 (2009). [PubMed: 19043404]

6. Louca S et al., Nat. Ecol. Evol 1, 0015 (2016).

7. Louca S, Parfrey LW, Doebeli M, Science 353, 1272–1277 (2016). [PubMed: 27634532]

8. Martiny JBH, Jones SE, Lennon JT, Martiny AC, Science 350, aac9323 (2015). [PubMed: 26542581]

9. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T, Proc. Natl. Acad. Sci. U.S.A 108, 14288–14293 (2011). [PubMed: 21825123]

10. David LA et al., Nature 505, 559–563 (2014). [PubMed: 24336217]

11. Friedman J, Higgins LM, Gore J, Nat. Ecol. Evol 1, 109 (2017). [PubMed: 28812687]

12. Vega NM, Gore J, PLOS Biol. 15, e2000633 (2017). [PubMed: 28257456]

13. Hekstra DR, Leibler S, Cell 149, 1164–1173 (2012). [PubMed: 22632978]

14. Foster KR, Bell T, Curr. Biol 22, 1845–1850 (2012). [PubMed: 22959348]

15. Coyte KZ, Schluter J, Foster KR, Science 350, 663–666 (2015). [PubMed: 26542567]

16. Levine JM, Bascompte J, Adler PB, Allesina S, Nature 546, 56–64 (2017). [PubMed: 28569813]

17. Bairey E, Kelsic ED, Kishony R, Nat. Commun 7, 12285 (2016). [PubMed: 27481625]

18. Callahan BJ et al., Nat Methods 13, 581–583 (2016). [PubMed: 27214047]

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

20. Macarthur R, Levins R, Proc. Natl. Acad. Sci. U.S.A 51, 1207–1210 (1964). [PubMed: 14215645]

21. Stewart FM, Levin BR, Am. Nat 107, 171–198 (1973).

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

23. Rodriguez-Valera F et al., Nat. Rev. Microbiol 7, 828–836 (2009). [PubMed: 19834481]

24. Kelsic ED, Zhao J, Vetsigian K, Kishony R, Nature 521, 516–519 (2015). [PubMed: 25992546]

25. Grilli J, Barabás G, Michalska-Smith MJ, Allesina S, Nature 548, 210–213 (2017). [PubMed: 28746307]

26. Basan M et al., Nature 528, 99–104 (2015). [PubMed: 26632588]

27. Paczia N et al., Microb. Cell Fact 11, 122 (2012). [PubMed: 22963408]

28. Rosenzweig RF, Sharp RR, Treves DS, Adams J, Genetics 137, 903–917 (1994). [PubMed: 7982572]

29. Hansen SK, Rainey PB, Haagensen JAJ, Molin S, Nature 445, 533–536 (2007). [PubMed: 17268468]

30. Baran R et al., Nat. Commun 6, 8289 (2015). [PubMed: 26392107]

31. Datta MS, Sliwerska E, Gore J, Polz MF, Cordero OX, Nat. Commun 7, 11965 (2016). [PubMed: 27311813]

32. Rainey PB, Travisano M, Nature 394, 69–72 (1998). [PubMed: 9665128]

33. Chesson P, Annu. Rev. Ecol. Syst 31, 343–366 (2000).

34. Levin BR, Science 175, 1272–1274 (1972). [PubMed: 4551427]

35. Cremer J, Arnoldini M, Hwa T, Proc. Natl. Acad. Sci. U.S.A 114, 6438–6443 (2017). [PubMed: 28588144]

36. Ratzke C, Gore J, PLOS Biol. 16, e2004248 (2018). [PubMed: 29538378]

37. MacArthur R, Theor. Popul. Biol 1, 1–11 (1970). [PubMed: 5527624]

38. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW, Environ. Microbiol 11, 823–832 (2009). [PubMed: 19021692]

39. Taillefumier T, Posfai A, Meir Y, Wingreen NS, eLife 6,1–65 (2017).

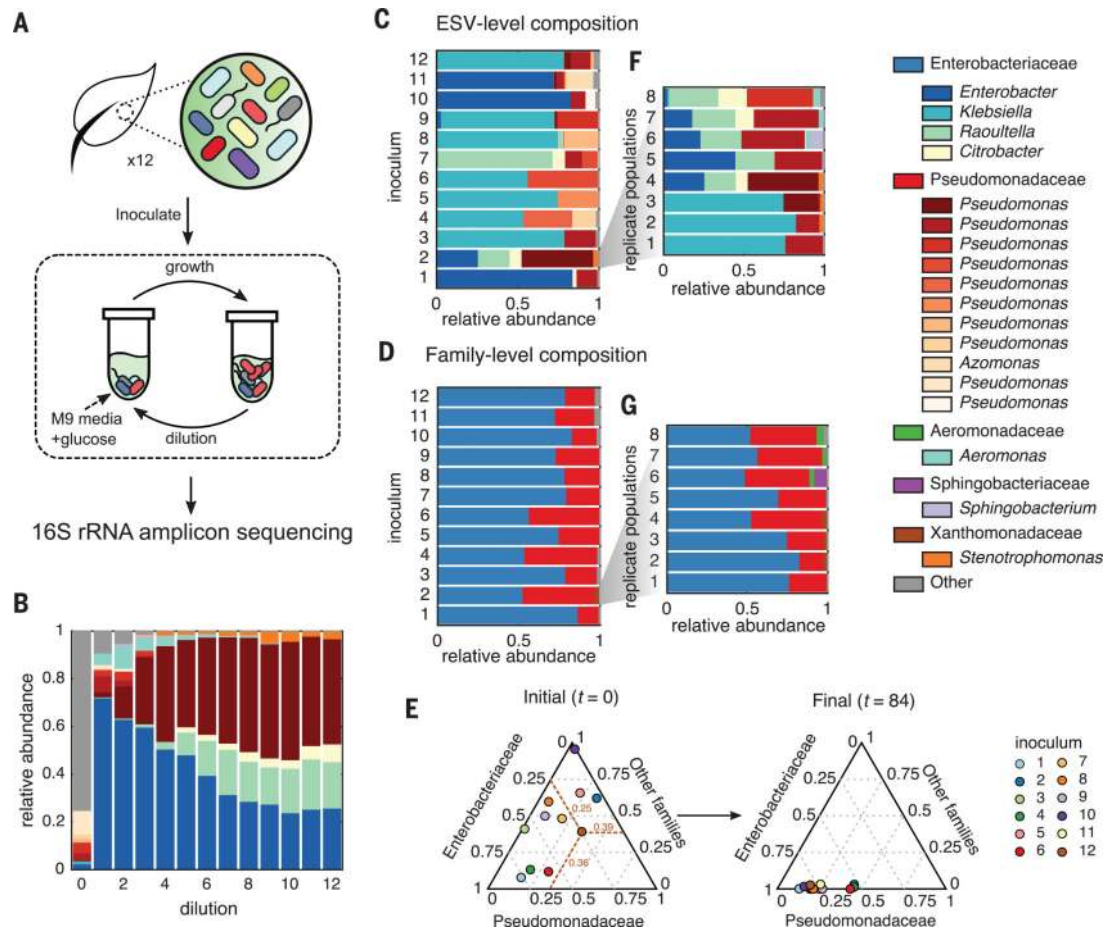40. Langille MGI et al., Nat. Biotechnol 31, 814–821 (2013). [PubMed: 23975157]

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, 16*S* 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).
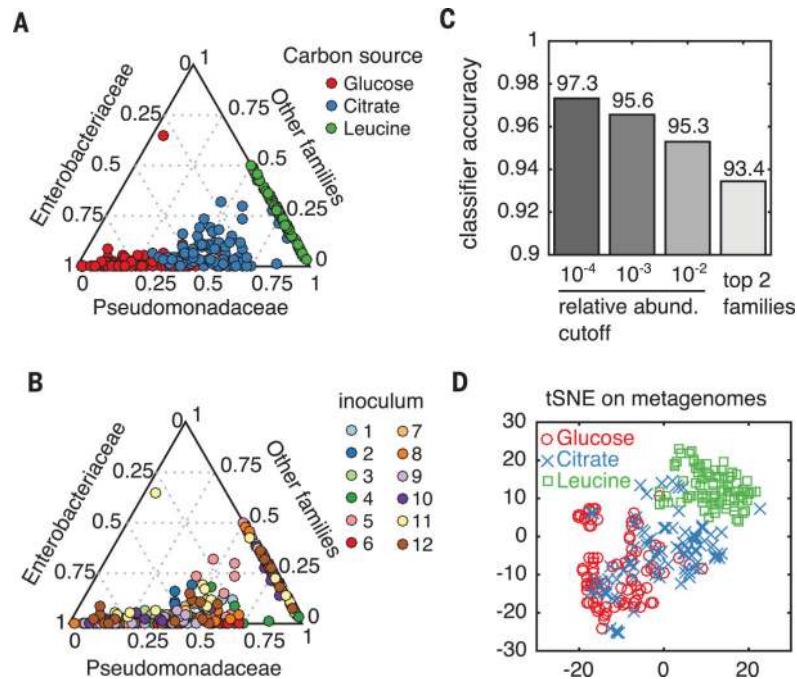
Fig. 2. Family-level and metagenomic attractors are associated with different carbon sources.
(**A** and **B**) Family-level community compositions are shown for all replicates across 12 inocula grown on either glucose, citrate, or leucine as the limiting carbon source. Data points are colored by carbon 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.
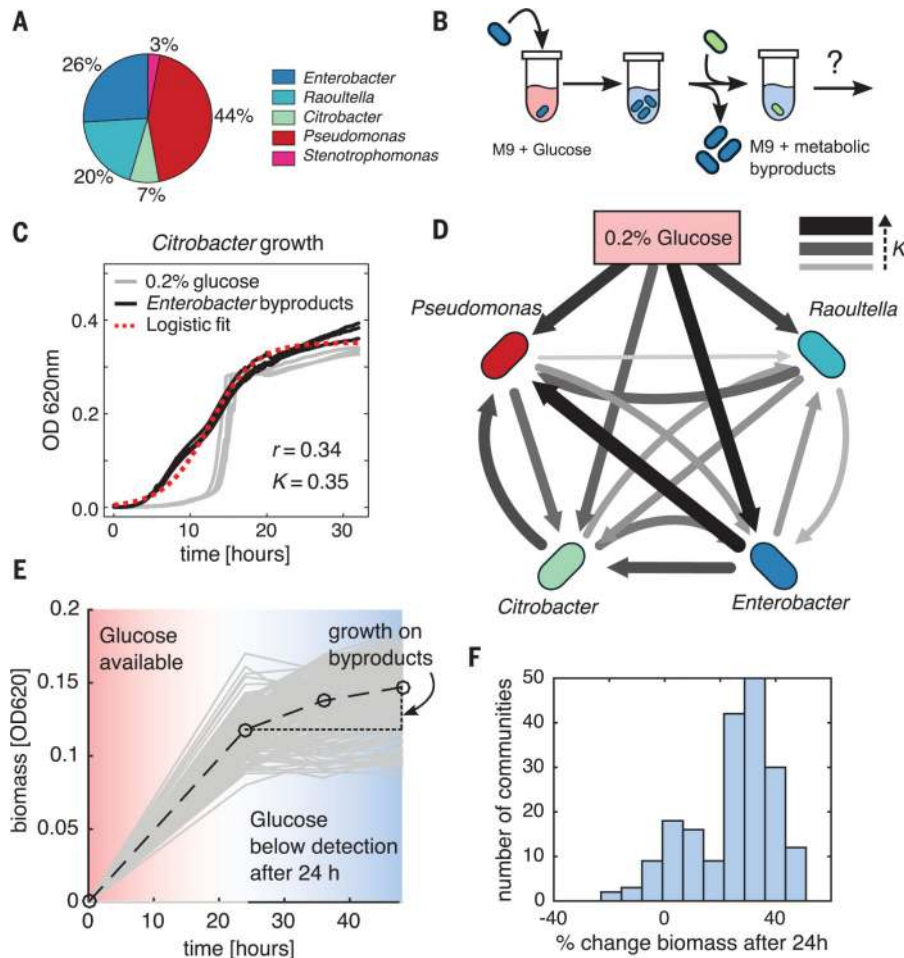
Fig. 3. Nonspecific metabolic facilitation may stabilize competition for the supplied resource.
(**A**) Representatives of the four most abundant genera in a representative community
(percentages shown in the pie chart) were isolated on M9 minimal glucose medium. (**B**)
Experimental setup: Isolates 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).
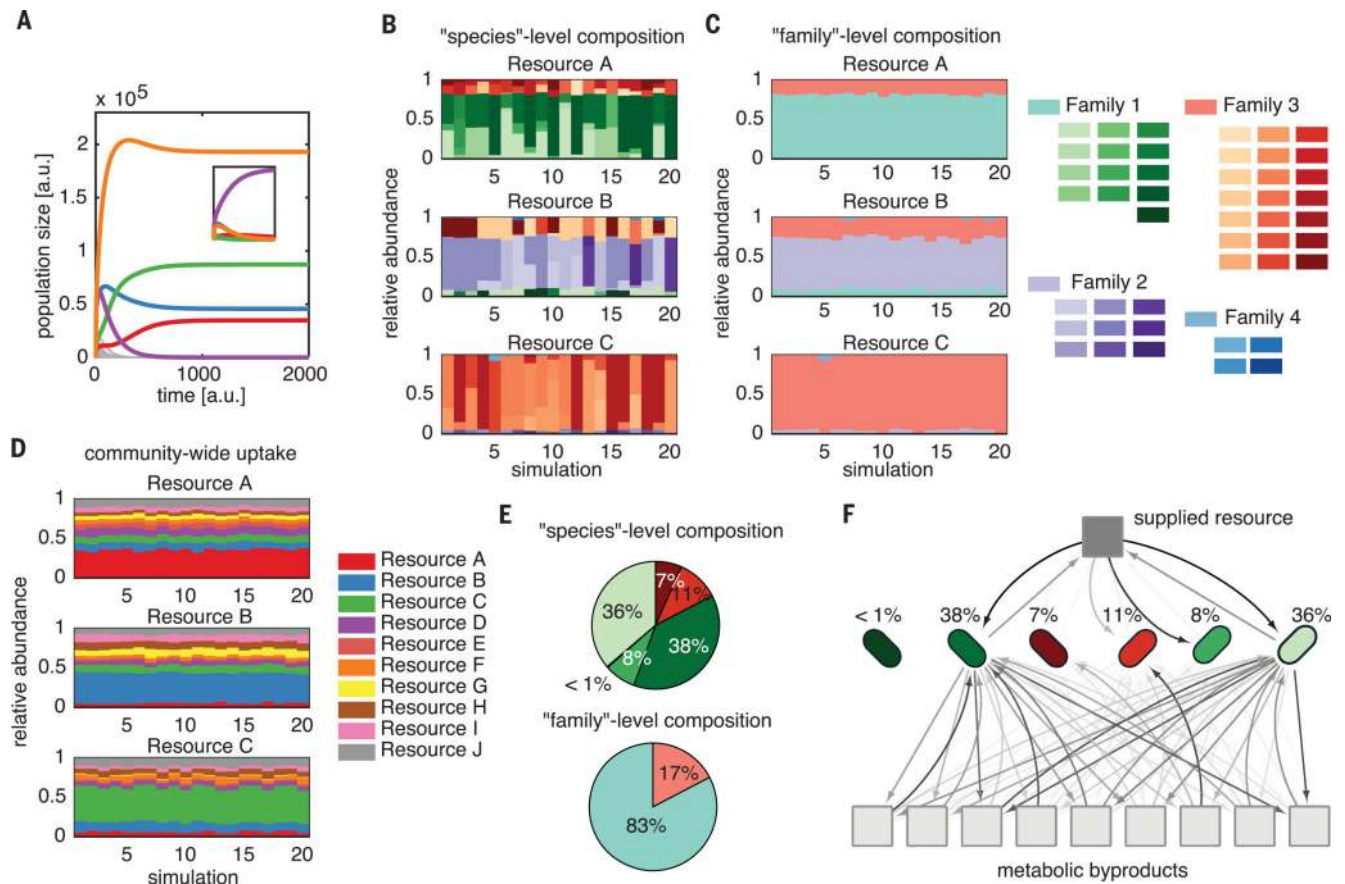
Fig. 4. A simple extension of classic ecological models recapitulates experimental observations.
MacArthur's consumer-resource model was extended to include 10 by-product secretions
along with consumption of a single primary limiting nutrient (supplementary materials),
controlled by secretion coefficient $D_{\beta\alpha}$, 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.