Available energy fluxes drive a phase transition in the diversity, stability, and functional structure of microbial communities

Robert Marsland III, Wenping Cui, Joshua Goldford,* Alvaro Sanchez,† Kirill Korolev, and Pankaj Mehta
Department of Physics, Boston University, 590 Commonwealth Ave., Boston, MA 02215
(Dated: May 31, 2018)

A fundamental goal of microbial ecology is to understand what determines the diversity, stability, and structure of microbial ecosystems. The microbial context poses special conceptual challenges because of the strong mutual influences between the microbes and their chemical environment through the consumption and production of metabolites. By analyzing a generalized consumer resource model that explicitly includes cross-feeding, stochastic colonization, and thermodynamics, we show that complex microbial communities generically exhibit a sharp transition as a function of available energy fluxes from a "resource-limited" regime where community structure and stability is shaped by energetic and metabolic considerations to a diverse regime where the dominant force shaping microbial communities is competition. These two regimes have distinct species abundance patterns, different functional profiles, and respond differently to environmental perturbations. Our model reproduces large-scale ecological patterns observed across multiple experimental settings such as nestedness and differential beta diversity patterns along energy gradients. We discuss the experimental implications of our results and possible connections with disorder-induced phase transitions in statistical physics.

Microbial communities inhabit every corner of our planet, from our own nutrient-rich guts to the remote depths of the ocean floor. Different environments harbor very different levels of microbial diversity: e.g., in some samples of non-saline water at mild temperature and pH, nearly 3,000 coexisting types of bacteria can be detected, whereas at ambient temperatures warmer than 40° C, most cataloged samples contain fewer than 100 distinct variants [43]. The functional structure of these communities is also highly variable, with functional traits often reflecting the environment in which the communities are found [23, 43]. A central goal of microbial community ecology is to understand how these variations in diversity, stability and functional structure [47] arise from an interplay of environmental factors such as energy and resource availability [13, 30] and ecological processes such as competition [9, 17, 28, 31] and stochastic colonization [8, 24, 26, 46].

This endeavor is complicated by the fact that microbes dramatically modify their abiotic environments through consumption and secretion of organic and inorganic compounds. This happens on a global scale, as in the Great Oxidation Event two billion years ago [4, 38], and also on smaller scales relevant to agriculture, industry and medicine. In this sense, every microbe is an "ecosystem engineer" [25]. Metabolic modeling and experiments suggests that metabolically-mediated syntrophic interactions should be a generic feature of microbial ecology [19, 22, 50] and that complex microbial communities can self-organize even in constant environments with no spatial structure or predation [16, 19]. For these reasons, there has been significant interest in developing new models for community assembly suited to the microbial setting [7, 20, 21, 42, 45].

Here, we present a statistical physics-inspired consumer resource model for microbial community assembly that builds upon the simple model introduced in [19] and explicitly includes energetic fluxes, stochastic colonization, syntrophy, and resource competition. We focus on modeling complex communities with many species and metabolites. By necessity, any mathematical model of such a large, diverse ecosystem will contain thousands of parameters that are hard to measure. To circumvent this problem, we take a statistical physics approach where all consumer preferences and metabolic parameters are drawn from random distributions.

This approach to modeling complex systems has its root in the pioneering work of Wigner on the spectrum of heavy nuclei [48] and was adapted by May to ecological settings [33]. Recently, there has been a renewed interest in using these ideas to understand complex systems in both many-body physics (reviewed in [10]) and community assembly [1, 3, 5, 6, 12, 15, 18, 19, 26, 45]. The key insight underlying this approach is that generic and reproducible large-scale patterns observed across multiple settings likely reflect *typical* properties, rather than fine tuned features of any particular realization or community. Consistent with this idea, it was recently shown that a generalized consumer resource model with random parameters can reproduce many of the patterns observed in experiments where natural communities were grown in synthetic minimal environments [19].

In this paper, we ask how varying the energy flux into an ecosystem and the amount of cross-feeding affects microbial community assembly. We find that metabolic communities generically exhibit a phase transition between a low diversity and a high diversity regime. These two phases have qualitatively different functional structures, respond differently to environmental perturba-

* Bioinformatics Program, Boston University
† Department of Ecology and Evolutionary Biology, Yale University

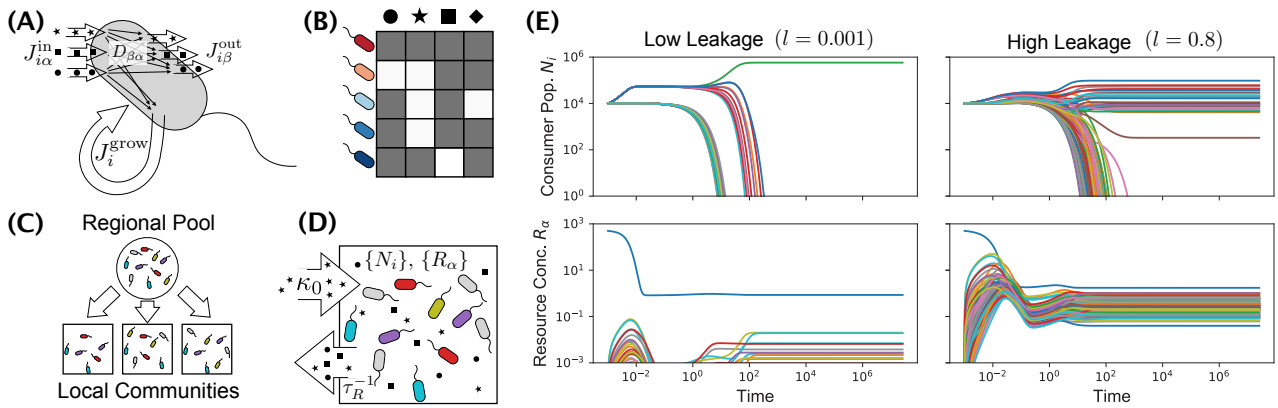

FIG. 1. **Microbial communities engineer complex chemical environments using a single energy source.** (A) Schematic of microbe-mediated energy fluxes in the Thermodynamic Microbial Consumer Resource Model. Each cell of species $i\,(=1,2,\ldots S)$ supplies itself with energy through import of resources, generating an incoming energy flux $J^{\rm in}_{i\alpha}$ for each resource type $\alpha\,(=1,2,\ldots M)$. A fraction l_α of this energy leaks back into the environment in the form of metabolic byproducts, with each byproduct type carrying an outgoing energy flux $J^{\rm out}_{i\beta} = \sum_\alpha l_\alpha D_{\beta\alpha} J^{\rm in}_{i\alpha}$. The remaining energy, $J^{\rm grow}_i$, is used to replicate the cell. (B) Each species is defined by a vector of consumer preferences that encode its capacity for harvesting energy from each resource type. These vectors comprise a consumer matrix $c_{i\alpha}$. (C) A regional pool of species is randomly generated, and communities are initialized with random subsets of these species to simulate stochastic colonization. (D) Each community is supplied with a constant flux κ_0 of a single resource type ($\alpha = 0$), and all resources are continuously diluted at a fixed rate τ_R^{-1}. (E) Consumer populations N_i and resource concentrations R_α as a function of time for two realizations of this model, with low ($l = 0.001$) and high ($l = 0.8$) levels of uniform metabolic leakage (see Appendix for parameters).

tions, and give rise to distinct large-scale biodiversity patterns. We show our model predictions are consistent with data from the Tara Oceans database [41] and the Earth Microbiome Project [43].

I. THERMODYNAMIC MICROBIAL CONSUMER RESOURCE MODEL

The starting point for our analysis is a new model that adapts MacArthur's Consumer Resource Model [31] to the microbial context by including energetics, stochastic colonization, and the exchange and consumption of metabolites. We consider the population dynamics of S species of consumers (e.g., microbes) competing for M types of resources. We are interested in large, diverse ecosystems where $S, M \gg 1$. A schematic summarizing our model is shown in Figure 1.

To set up a thermodynamically consistent model , we keep track of energy fluxes (denoted by J) and mass fluxes (denoted by ν) in the system. The rate at which an individual of species i imports energy via resource α depends on the resource concentration R_α as well as on the consumer's vector of resource preferences $c_{i\alpha}$ through the expression:

$$J^{\rm in}_{i\alpha} = w_\alpha \nu^{\rm in}_{i\alpha} = w_\alpha \sigma(c_{i\alpha} R_\alpha), \tag{1}$$

where $\sigma(x)$ encodes the functional response and w_α is the energy density of resource α. In the main text, we focus on Type-I responses where $\sigma(x) = x$, but most of our results also hold for the case where $\sigma(x)$ is a Monod function. In the microbial context the consumer preferences $c_{i\alpha}$ can be interpreted as expression levels of transporters for each of the resources.

We model leakage and secretion by letting a fraction l_α of this imported energy return to the environment, so that the power available to the cell for sustaining growth is equal to

$$J^{\rm grow}_i = \sum_\alpha (1 - l_\alpha) J^{\rm in}_{i\alpha}. \tag{2}$$

By keeping l_α between 0 and 1, we guarantee that the community does not spontaneously generate usable energy in violation of the Second Law of Thermodynamics. We assume that a fixed quantity m_i of power per cell is required for maintenance of species i, and that the per-capita growth rate is proportional to the remaining energy flux, with proportionality constant g_i. Under these assumptions, the time-evolution of the population size N_i of species i can be modeled using the equation

$$\frac{dN_i}{dt} = g_i N_i \left[J^{\rm grow}_i - m_i \right]. \tag{3}$$

The leaked energy flux $J^{\rm out}_i = \sum_\alpha l_\alpha J^{\rm in}_{i\alpha}$ from each cell of species i is partitioned among the M possible resource types via the biochemical pathways operating within the cell. We assume that all species share a similar core metabolism, encoded in a matrix $D_{\beta\alpha}$. Each element of $D_{\beta\alpha}$ specifies the fraction of leaked energy from resource α that is released in the form of resource β. Note

that by definition, $\sum_\beta D_{\beta\alpha} = 1$. Thus, in our model the resources that are excreted into the environment are intimately coupled to the resources a cell is consuming. The outgoing energy flux contained in metabolite β is given by

$$J^{\text{out}}_{i\beta} = w_\beta \nu^{\text{out}}_{i\beta} = \sum_\alpha D_{\beta\alpha} l_\alpha J^{\text{in}}_\alpha. \tag{4}$$

The dynamics of the resource concentrations depend on the incoming and outgoing mass fluxes $\nu^{\text{in}}_{i\alpha}$ and $\nu^{\text{out}}_{i\alpha}$, which are related to the energy fluxes via the energy densities w_α. In terms of these quantities, we have

$$\frac{dR_\alpha}{dt} = h_\alpha + \sum_j N_j(\nu^{\text{out}}_{j\alpha} - \nu^{\text{in}}_{j\alpha}), \tag{5}$$

with h_α encoding the dynamics of externally supplied resources. In this manuscript, we focus on the case where the microbial communities are grown in a chemostat with a single externally supplied resource $\alpha = 0$ (Figure 1). In this case, the resource dynamics can be described by choosing $h_\alpha = \kappa_\alpha - \tau_R^{-1} R_\alpha$, with all the κ_α set to zero except for κ_0. These equations for N_i and R_α, along with the expressions for $J^{\text{in}}_{i\alpha}$ and $J^{\text{out}}_{i\alpha}$, completely specify the ecological dynamics of the model.

This model has been implemented in a freely available open-source Python package "Community Simulator" (Marsland et al. in preparation). The package can be downloaded from https://github.com/robertvsiii/community-simulator, which also contains a Jupyter notebook with scripts for reproducing the figures.

II. RESULTS

To assess the typical community structure and resource pool stability for ecosystems obeying Equations (1)-(5), we randomly generated a binary $S \times M$ consumer preference matrix $c_{i\alpha}$ with $S = 200$ species and $M = 100$ resources, along with a sparse random chemistry $D_{\alpha\beta}$. We chose $c_{i\alpha}$ so that each species consumed an average of 10 kinds of resource out of the 100 possible. The full sampling procedure is detailed in the Supporting Information, which also contains results with $c_{i\alpha}$ drawn from a Gaussian or Gamma distribution. We set the mean m_i equal to 1, with standard deviation 0.1, and set all the w_α equal to 1. Finally, we made all the leakage fractions identical, with $l_\alpha = l$ for all α. To assess the amount of variability in the results, we initialized 10 different communities by seeding each one with a random subset of 100 species from the full 200-species pool. This simulates the stochastic colonization frequently observed in microbial ecosystems, where the community composition can randomly vary depending on the set of microbes this particular local environment happened to be exposed to [35]. Figure 1 shows typical dynamical trajectories in the presence of high ($l = 0.8$) and extremely low leakage ($l = 0.001$).

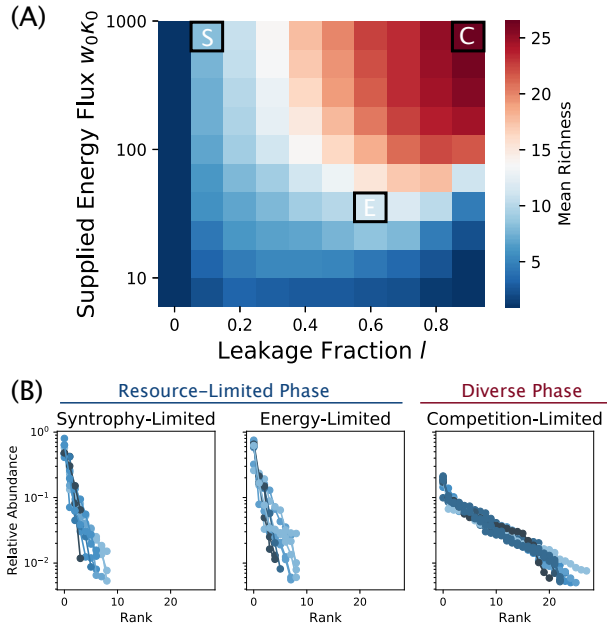

FIG. 2. **Steady-state richness as a function of metabolic leakage l and externally supplied energy flux $w_0\kappa_0$.** We generated 200 species, initialized 10 communities of 100 species each from this pool, and ran the dynamics to steady state under different combinations of $w_0\kappa_0$ and l (see main text and Appendix for parameters). (A) Heat map summarizing all simulations, colored by the average number of surviving species per steady-state community ("Richness"). Slices through the heat map are plotted in Figure 11 of the Appendix. (B) Community compositions are displayed as rank-abundance curves for one example from each of the three regimes (colored by community richness): syntrophy-limited (S), energy-limited (E) and competition-limited (C). The lines are assigned different shades for clarity. The first two regimes are parts of the same resource-limited phase, manifesting similar statistical properties. The plots are truncated at a relative abundance of 0.5%; see Appendix for full data.

Available energy fluxes drive a phase transition between a "Resource-Limited" and "Diverse" Phase

Our numerical simulations indicate that there is an abrupt and sharp transition between two qualitatively different community structures as we vary the externally supplied energy flux $w_0\kappa_0$ into the ecosystem at sufficiently high levels of leakage/syntrophy l. This transition is reminiscent of phase transitions commonly seen in disordered systems in physics (see Discussion and Appendix). One clear signature of this transition is the sudden change in diversity, as illustrated in Figure 2. At low energy flux or leakage, the diversity is severely limited by resource availability. It is helpful to divide this resource-limited phase into two regimes, based on the detailed mechanism of diversity limitation: a "syntrophy-limited" regime at low l where diversity is limited by the number of available metabolic niches, and an "energy-limited" regime at low $w_0\kappa_0$ where diversity is limited by

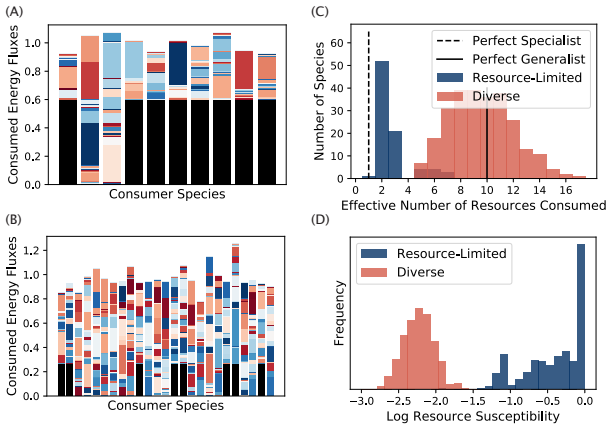

FIG. 3. **Structure and stability of resource dynamics depend on ecological phase**. (A) Consumed energy fluxes $(1-l)J_{i\alpha}^{\text{in}}$ for each of the ten surviving species in the steady state of a community in the resource-limited phase ($l = 0.6$, $w_0\kappa_0 = 28$). The black portion of the bar is the flux $(1-l)J_{i0}^{\text{in}}$ due to the externally supplied resource, and the colored bars represent the contributions of the other resources. Since these communities have reached the steady state, Equation (3) implies that the total height of each bar equals the maintenance cost m_i of the corresponding consumer species. (B) Same as previous panel, but for an example of the diverse phase ($l = 0.9$, $w_0\kappa_0 = 1000$). (C) Simpson diversity M_i^{eff} of steady-state flux vector $J_{i\alpha}^{\text{in}}$ for each species in all 10 communities in the two phases (same parameters as panels A and B). (D) Logarithm of susceptibility $\log_{10} \partial\bar{R}_\alpha/\partial\kappa_\alpha$ of community-supplied resources ($\alpha \neq 0$) to addition of an externally supplied flux κ_α (same parameters as panels A and B).

the amount of energy supplied to the ecosystem. At high leakage, increasing $w_0\kappa_0$ causes a sudden increase in diversity at a certain threshold value. The diversity of this phase is limited only by the level of inter-species competition within the regional species pool, in accordance with classical niche-packing theory [31] as we will discuss below.

The resource-limited and diverse phase have distinct functional structures

To better understand these two phases, we examined the functional traits of members of typical communities in each phase. In the resource-limited phase, many surviving species derive most of their energy directly from the externally supplied resource (Figure 3A). In the diverse phase, by contrast, only a minority of the steady-state community members can consume this resource at all, and even these species receive most of their energy from a diverse array of metabolic byproducts (Figure 3B). We quantified this observation using the Simpson Diversity M_i^{eff} of the incoming resource flux vectors $J_{i\alpha}^{\text{in}}$, which measures the effective number of resources consumed by each species, and is closely related to the inverse participation ratio in statistical physics. The Simpson Diversity is defined by

$$M_i^{\text{eff}} = \left[\sum_\alpha \left(\frac{J_{i\alpha}^{\text{in}}}{J_i^{\text{in}}}\right)^2\right]^{-1}, \tag{6}$$

where $J_i^{\text{in}} = \sum_\alpha J_{i\alpha}^{\text{in}}$ is the total incoming energy flux for each cell of this species. M_i^{eff} approaches 1 for species that obtain the bulk of their energy from a single resource type and approaches M when all resource types are consumed equally. In the resource-limited phase, the distribution of these values is sharply peaked between 1 and 2. In the diverse phase, the peak is located around 10, which is the average number of resources with high transporter expression in our binary sampling scheme for $c_{i\alpha}$. This shows that most community members in the diverse phase utilize multiple energy sources, with the incoming flux spread evenly over all resource types they are capable of consuming.

Responses to resource perturbations differ in two phases

Another important property of microbial ecosystems is how they respond to environmental perturbations. Previous theoretical studies have shown that sufficiently diverse communities can "pin" the resource concentrations in their local environment to fixed values, which are independent of the magnitude of externally supplied fluxes [37, 42, 44]. In these studies, resource pinning occurs only when the community saturates the diversity bound imposed by the principal of competitive exclusion, i.e. when the number of coexisting species is at least as large as the number of resource types. Such maximally diverse communities typically require fine-tuning of the resource utilization profiles or imposition of universal efficiency tradeoffs in cellular metabolism.

In our stochastically assembled communities, the diversity is always much lower than the number of resource types, so we hypothesized that the resource concentrations should not be pinned. To test this idea, we measured the response of the steady-state concentrations $\bar{R}_\alpha$ to changes in external supply rates κ_α, in terms of the "resource susceptibilities" $\partial\bar{R}_\alpha/\partial\kappa_\alpha$ plotted in Figure 3D [1]. Our hypothesis was valid in the resource-limited phase, where many resource susceptibilities are comparable to the susceptibility in the empty chemostat $\partial\bar{R}_\alpha/\partial\kappa_\alpha = \tau_R = 1$. But in the diverse phase, we were surprised to find that the susceptibilities are 100 times smaller than this maximum value. This suggests that resource pinning may be a generic phenomenon, observable in real ecosystems when the energy supply is sufficiently large.

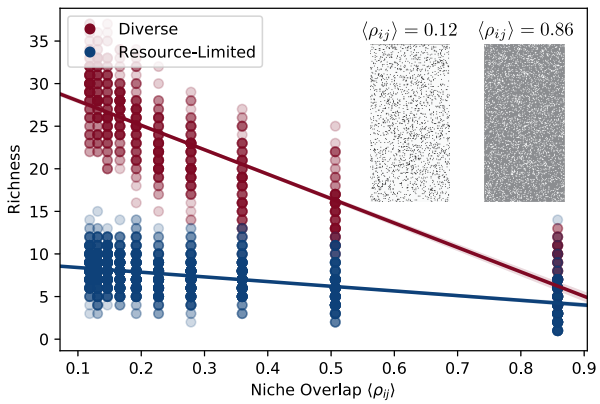

FIG. 4. **Richness of diverse phase depends on generalized niche-overlap**. We took the values of supplied energy flux $w_0\kappa_0$ and leakage fraction l from the three examples highlighted in Figure 2, and varied the average niche overlap $\langle\rho_{ij}\rangle$ between members of the metacommunity. For each $w_0\kappa_0, l$ combination and each value of $\langle\rho_{ij}\rangle$, we generated 10 pools of 200 species, initialized 10 communities of 100 species each from this pool, and ran the dynamics to steady state. The steady-state richness of each community is plotted against the niche overlap. Points are colored by their phase, and solid lines are linear regressions. Inset: $c_{i\alpha}$ matrices that define the regional pool for two different levels of overlap, with dark squares representing high consumption coefficients.

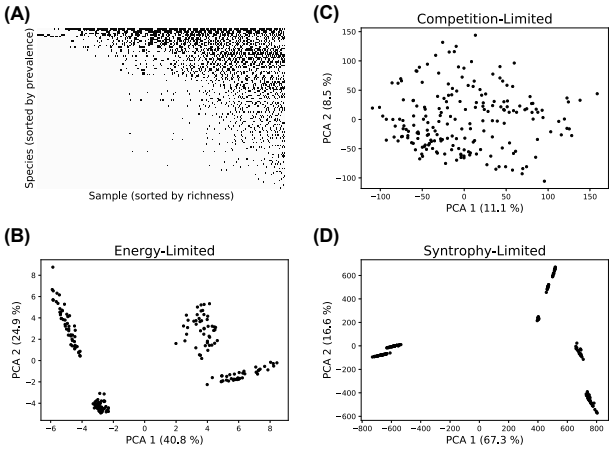

FIG. 5. **Resource-limited phase features community-level environmental filtering**. (A) Presence (black) or absence (white) of all species in all 1,000 communities from the original simulations of Figure 2. (B, C, D) We initialized 200 new communities for each of the three examples highlighted in Figure 2A, by randomly choosing sets of 100 species from the regional pool. Each panel shows the projection of final community compositions $\{N_i\}$ onto the first two principal components of the set of compositions.

Competition between microbes limits richness in diverse phase

In the diverse phase, the number of coexisting species ("richness") is not limited by energy availability or by access to secreted metabolites, but is still much less than the maximal value of $M = 100$ set by the competitive exclusion principle [28]. We hypothesized that the diversity in this phase is limited by the intensity of competition between members of the regional species pool. This can be quantified in terms of the niche overlap, which measures the similarity between consumer preferences, with a bigger niche overlap resulting in greater competition between members of the ecosystem [9, 32]. Figure 4 shows how the richness varies as a function of the average niche overlap $\langle\rho_{ij}\rangle$ in the regional pool, defined by

$$\langle\rho_{ij}\rangle \equiv \left\langle \frac{\sum_\alpha c_{i\alpha}c_{j\alpha}}{\sqrt{\sum_\alpha c_{i\alpha}^2 \sum_\alpha c_{j\alpha}^2}} \right\rangle = \frac{\langle c_{i\alpha}\rangle^2}{\sqrt{\langle c_{i\alpha}^2\rangle\langle c_{j\alpha}^2\rangle}}. \qquad (7)$$

In the diverse phase the mean richness decreases approximately linearly with increasing overlap. The richness of the resource-limited phase, on the other hand, has only a very weak dependence on the niche overlap. These results suggest that competition for resources is the primary driver of community assembly in the diverse phase. Importantly, competition limits the number of coexisting species well below the upper bound imposed by the competitive exclusion principle.

Nestedness and other large scale beta-diversity patterns

To test our model against existing data from surveys of natural communities, we asked about the consequences of the above observations for large-scale patterns of biodiversity. We began by re-examining the simulations in Figure 2. Since all species in the simulations come from the same regional pool, we asked which species are present in each of final communities across all energy fluxes and leakage rates. We found that when we sorted species by prevalence (rows in Fig. 5A) and samples by richness (columns in Fig. 5A) , the community composition generically exhibited a nested structure – less diverse communities had a subset of the species prevalent in the more diverse communities [29, 36]. This suggests that nested structures may generically emerge in community assembly through the interplay of stochastic colonization, competition, and environmental filtering.

Next, we asked if we large-scale beta-diversity patterns could be used to distinguish the resource-limited and diverse phases.We initialized 200 new communities with 100 randomly chosen members from the full regional species pool and simulated these communities to steady state in both the resource-limited and diverse phases (see Appendix for details). This sub-sampling of the full regional species pools mimics the effect of stochastic colonization, where a different random subset of species seeds each community. To better understand beta-diversity

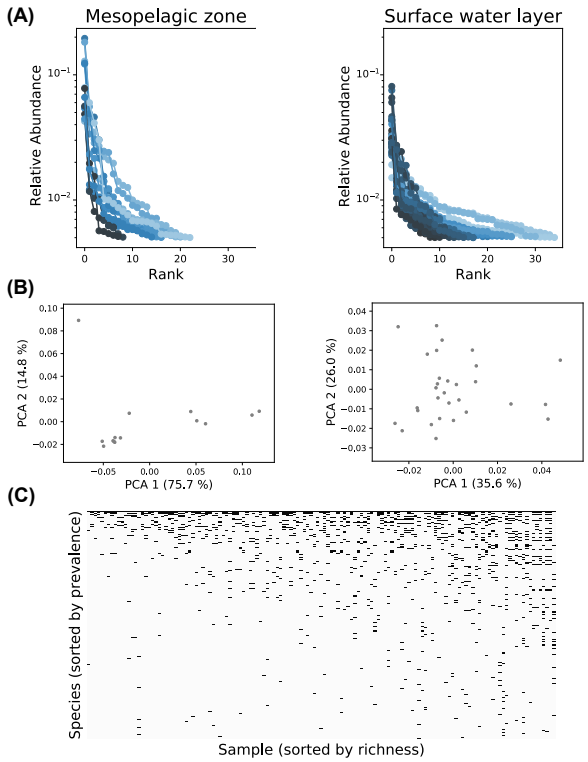

FIG. 6. **Ecological phases and nestedness in microbiome data.** (A) Rank-abundance plots for 16S OTU's above 0.5% relative abundance for tropical samples from the Tara Oceans database [41]. (See Appendix for full curves). "Surface water layer" samples were collected at a depth of 5 meters, while the deep-sea "Mesopelagic Zone" extends from 200 to 1,000 meters, with significantly less access to solar energy. (B) Projections of all community compositions from each environment on their first two principle components. (C) Presence (black) or absence (white) of each OTU above 0.5% relative abundance across all Tara Oceans samples.

signatures in the two phases, we performed a Principal Component Analysis (PCA) on community composition and projected the data onto the first two principal components, as shown in Figure 5B-D. In the resource-limited phase, the communities form distinct clusters that are distinguished by different highly abundant species. This suggests that harsh environments only allow a few species from the regional pool to rise to dominance, and that these dominant species induce clustering of communities. Such "enterotype"-like behavior is a common feature observed in many microbial settings [2] . In contrast, the diverse phase exhibited neither well-defined clusters nor dominant, highly abundant species.

Comparison to microbial datasets

The preceding results suggest that the resource-limited and diverse phases can be distinguished using beta-diversity patterns. To test this prediction, we looked for a publicly available dataset containing a well-defined gradient of energy supply. The Tara Oceans survey collected samples from a range of depths across the world's oceans, and the resulting gradient of solar energy flux provided the desired context (http://ocean-microbiome.embl.de/companion.html)[41]. We analyzed the 16S OTU composition of tropical ocean communities for all 30 sea-surface samples, where solar energy is plentiful, and all 13 samples from the dark deep-sea Mesopelagic Zone where energy fluxes are limited. We focused on species that reached 0.5% relative abundance in at least one of the samples, since the population dynamics of the vast numbers of trace species are likely to depend on environmental heterogeneity and other effects not included in our model. The sea surface data resembles our diverse phase, with relatively uniform biomass distribution over species and a continuum of possible community compositions. In contrast, the Mesopelagic Zone is similar to our resource-limited phase: the dominance of the most abundant species is much more pronounced, and the compositions cluster into four discrete types. While these results are consistent with our model predictions, more samples will have to be collected and analyzed in order to make a definitive comparison.

As mentioned above, our model also gives a natural explanation for the nestedness in the Earth Microbiome Project [43], suggesting that it may be a natural byproduct of complex microbial communities shaped by competition, environmental heterogeneity, and stochastic colonization. To test how generic this feature is, we plotted presence/absence community compositions of all samples from the Tara Oceans dataset, sorting samples by richness and OTU's ("species") by prevalence. To obtain a sufficiently large range of richnesses, we imposed a 0.5% relative abundance threshold for an OTU to count as "present". The resulting pattern in Figure 6 is qualitatively similar to our simulations (Fig. 5D), and to the phylum-level data of the Earth Microbiome Project [43].

III. DISCUSSION

Advances in sequencing technology have opened new horizons for the study of microbial ecology, generating massive amounts of data on the composition of both natural and synthetic communities. But the complexity of these systems make it difficult to extract robust general principles suitable for guiding medical and industrial applications. Numerical simulations provide a powerful tool for addressing this problem. By rapidly iterating numerical experiments under a variety of modeling choices with random parameters, one can identify robust patterns and use the resulting insights to guide targeted experiments.

Following this strategy, we developed a thermodynamic consumer resource model that explicitly includes energetic fluxes and metabolically-mediated cross-feeding and competition. Using this model, we identified two qualitatively distinct phases in these simulations as we

varied the amount of energy supplied to ecosystem and the fraction of energy leaked back into the environment: a low diversity "resource-limited" phase and a "diverse" phase. The structure of the resource-limited phase is strongly constrained by species- and community-level environmental filtering. Each community is dominated by a handful of species, making the community properties sensitive to the idiosyncratic characteristics of these species and susceptible to environmental fluctuations. In the diverse phase, communities exhibit more universal features because they substantially engineer their environments. In particular, the concentrations of resources at steady state are narrowly distributed and roughly independent of the type and amount of externally supplied resource. More over, each species draws its energy roughly equally from all resources rather than subsisting on the externally supplied resource as in the resource-limited phase.

Our model complements other recent efforts at understanding microbial community ecology. Taillefumier *et al.* proposed a similar model of metabolite exchange, and focused on the case where the number of resource types M is equal to 3 [42]. In this case, repeated invasion attempts from a large regional species pool produced optimal combinations of metabolic strategies. Goyal *et al.* examined the opposite limit, with $M = 5,000$, but allowed each species to consume only one type of resource [21]. This generated communities with a tree-like metabolic structure, where each species depends directly on another species to generate its unique food source. In our model, the large number of resource types ($M = 100$ in the current study) makes spontaneous optimization extremely unlikely. In addition, the consumption preferences in the model generically overlap, giving rise to trophic networks qualitatively different from the specialist communities of Goyal *et al.*. The absence of highly specialized metabolic structure in our model makes it especially well-suited for interpreting patterns in large-scale sequence-based datasets such as the Earth Microbiome Project [43].

Our model predictions can also be directly tested using experiments with natural communities in synthetic laboratory environments [11, 19]. Our model predicts that beta diversity patterns can be significantly altered by increasing the ecosystem's energy supply to induce a transition from the resource-limited to the diverse phase. In the experimental set-up in [11], this can be done by directly adding chitinase enzymes to the sludge reactor to increase the degradation of chitin-based organic particles on which the ocean-derived microbial communities exist.

In this work we have largely confined ourselves to studying steady-state properties of well-mixed microbial communities. Complex microbial communities often exhibit complex temporal dynamics with well-defined successions [11, 14, 49]. It will be interesting to explore these dynamical phenomena using our model. It is also well established that spatial structure can give rise to new ecological phenomena [27, 34] and an important area of future work will be to better explore the role of space in microbial community assembly.

Finally, the phase transition observed here is likely closely related to disorder-induced transitions in statistical physics. We suspect that the transition observed in our numerical simulations is closely related to a replica symmetry breaking transition [5] but much more analytic work needs to be done in order to see if this is indeed the case. More generally, this adds to a series of works suggesting that ecological communities fall into distinct ecological phases that are qualitatively different from each other and that communities can undergo phase transitions in response to environmental and ecological variables [3, 5, 6, 12, 15, 26, 39, 40, 45]. These ecological phases often display distinct biodiversity patterns that emerge from the interplay of stochastic processes and competition and may explain the origins of some of the reproducible large-scale patterns seen in sequencing data across multiple environments and communities.

IV. ACKNOWLEDGMENTS

The funding for this work partly results from a Scialog Program sponsored jointly by Research Corporation for Science Advancement and the Gordon and Betty Moore Foundation. This work was also supported by NIH NIGMS grant 1R35GM119461, and by Simons Investigator in the Mathematical Modeling of Living Systems (MMLS) awards to PM and KK.

[1] Madhu Advani, Guy Bunin, and Pankaj Mehta. Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model. *Journal of Statistical Mechanics*, page 033406, 2018.

[2] Manimozhiyan Arumugam, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R Mende, Gabriel R Fernandes, Julien Tap, Thomas Bruls, Jean-Michel Batto, et al. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174, 2011.

[3] Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 2018.

[4] A. Bekker, H. D. Holland, P.-L. Wang, D. Rumble III, H. J. Stein, J. L. Hannah, L. L. Coetzee, and N. J. Beukes. Dating the rise of atmospheric oxygen. *Nature*, 427:117, 2004.

[5] Giulio Biroli, Guy Bunin, and C. Cammarota. Marginally stable equilibria in critical ecosystems. *arXiv*,

1710.03606, 2017.

[6] G. Bunin. Ecological communities with Lotka-Volterra dynamics. *Physical Review E*, 95:042414, 2017.

[7] Stacey Butler and James O'Dwyer. Stability criteria for complex microbial communities. *bioRxiv*, 2018.

[8] Jonathan M Chase. Community assembly: when should history matter? *Oecologia*, 136(4):489, 2003.

[9] Peter Chesson. Macarthur's consumer-resource model. *Theoretical Population Biology*, 37(1):26, 1990.

[10] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65:239, 2016.

[11] Manoshi S. Datta, Elzbieta Sliwerska, Jeff Gore, Martin F. Polz, and Otto X. Cordero. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications*, 7:11965, 2016.

[12] Benjamin Dickens, Charles K. Fisher, and Pankaj Mehta. Analytically tractable model for community ecology with many species. *Physical Review E*, 94:022423, 2016.

[13] Mallory Embree, Joanne K Liu, Mahmoud M Al-Bassam, and Karsten Zengler. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proceedings of the National Academy of Sciences*, 112(50):15450, 2015.

[14] Tim N. Enke, Gabriel E. Leventhal, Matthew Metzger, José T. Saavedra, and Otto X. Cordero. Micro-scale ecology regulates particulate organic matter turnover in model marine microbial communities. *bioRxiv*, 2018.

[15] Charles K. Fisher and Pankaj Mehta. The transition between the niche and neutral regimes in ecology. *PNAS*, 111:13111, 2014.

[16] Jonathan Friedman, Logan M. Higgins, and Jeff Gore. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology and Evolution*, 1:0109, 2017.

[17] G. F. Gause and A. A. Witt. Behavior of mixed populations and the problem of natural selection. *The American Naturalist*, 69:596, 1935.

[18] Theo Gibbs, Jacopo Grilli, Tim Rogers, and Stefano Allesina. The effect of population abundances on the stability of large random ecosystems. *arXiv*, 1708.08837, 2017.

[19] Joshua E. Goldford, Nanxi Lu, Djordje Bajic, Sylvie Estrela, Mikhail Tikhonov, Alicia Sanchez-Gorostiaga, Daniel Segre, Pankaj Mehta, and Alvaro Sanchez. Emergent simplicity in microbial community assembly. *bioRxiv*, 2017.

[20] Benjamin H. Good, Stephen Martis, and Oskar Hallatschek. Directional selection limits ecological diversi cation and promotes ecological tinkering during the competition for substitutable resources. *bioRxiv*, 2018.

[21] Akshit Goyal and Sergei Maslov. Diversity, stability, and reproducibility in stochastically assembled microbial ecosystems. *Physical Review Letters*, 2018.

[22] William R. Harcombe, William J. Riehl, Ilija Dukovski, Brian R. Granger, Alex Betts, Alex H. Lang, Gracia Bonilla, Amrita Kar, Nicholas Leiby, Pankaj Mehta, Christopher J. Marx, and Daniel Segrè. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports*, 7:1104, 2014.

[23] Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H Badger, Asif T Chinwalla, Heather H Creasy, Ashlee M Earl, Michael G FitzGerald, Robert S Fulton, et al. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207, 2012.

[24] Patricio Jeraldo, Maksim Sipos, Nicholas Chia, Jennifer M Brulc, A Singh Dhillon, Michael E Konkel, Charles L Larson, Karen E Nelson, Ani Qu, Lawrence B Schook, F Yang, Bryan A White, and Nigel Goldenfeld. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proceedings of the National Academy of Sciences*, 109(25):9692, 2012.

[25] Clive G. Jones, John H. Lawton, and Shachak Moshe. Organisms as ecosystem engineers. *Oikos*, 69:373, 1994.

[26] David A. Kessler and Nadav M. Shnerb. Generalized model of island biodiversity. *Physical Review E*, 91:042705, 2015.

[27] Kirill S Korolev, Mikkel Avlund, Oskar Hallatschek, and David R Nelson. Genetic demixing and evolution in linear stepping stone models. *Reviews of modern physics*, 82(2):1691, 2010.

[28] Simon A Levin. Community equilibria and stability, and an extension of the competitive exclusion principle. *The American Naturalist*, 104(939):413, 1970.

[29] Mark V. Lomolino. Investigating causality of nestedness of insular communities: Selective immigrations or extinctions? *Journal of Biogeography*, 23:699, 1996.

[30] Michel Loreau. Consumers as maximizers of matter and energy flow in ecosystems. *The American Naturalist*, 145(1):22, 1995.

[31] Robert MacArthur. Species Packing and Competitive Equilibrium for Many Species. *Theoretical Population Biology*, 1:1, 1970.

[32] Robert MacArthur and Richard Levins. The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*, 101:377, 1967.

[33] Robert M. May. *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, N.J., 2001.

[34] Rajita Menon and Kirill S Korolev. Public good diffusion limits microbial mutualism. *Physical review letters*, 114(16):168102, 2015.

[35] Benjamin Obadia, Z.T. Güvener, Vivian Zhang, Javier A. Ceja-Navarro, Eoin L. Brodie, William W. Ja, and William B. Ludington. Probabilistic Invasion Underlies Natural Gut Microbiome Stability. *Current Biology*, 27:1999, 2017.

[36] Bruce D Patterson and Wirt Atmar. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society*, 28(1-2):65–82, 1986.

[37] Anna Posfai, Thibaud Taillefumier, and Ned S. Wingreen. Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Physical Review Letters*, 118:028103, 2017.

[38] Bettina E. Schirrmeister, Muriel Gugger, and Philip C. J. Donoghue. Cyanobacteria and the great oxidation event: Evidence from genes and fossils. *Palaeontology*, 58:769, 2015.

[39] A Carla Staver, Sally Archibald, and Simon Levin. Tree cover in sub-saharan africa: rainfall and fire constrain forest and savanna as alternative stable states. *Ecology*, 92(5):1063, 2011.

[40] A Carla Staver, Sally Archibald, and Simon A Levin. The global extent and determinants of savanna and forest as

alternative biome states. *Science*, 334(6053):230, 2011.

[41] Shinichi Sunagawa, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, Georg Zeller, Daniel R. Mende, Adriana Alberti, Francisco M. Cornejo-Castillo, Paul I. Costea, Corinne Cruaud, Francesco d'Ovidio, Stefan Engelen, Isabel Ferrera, Josep M. Gasol, Lionel Guidi, Falk Hildebrand, Florian Kokoszka, Cyrille Lepoivre, Gipsi Lima-Mendez, Julie Poulain, Bonnie T. Poulos, Marta Royo-Llonch, Hugo Sarmento, Sara Vieira-Silva, Céline Dimier, Marc Picheral, Sarah Searson, Stefanie Kandels-Lewis, Chris Bowler, Colomban de Vargas, Gabriel Gorsky, Nigel Grimsley, Pascal Hingamp, Daniele Iudicone, Olivier Jaillon, Fabrice Not, Hiroyuki Ogata, Stephane Pesant, Sabrina Speich, Lars Stemmann, Matthew B. Sullivan, Jean Weissenbach, Patrick Wincker, Eric Karsenti, Jeroen Raes, Silvia G. Acinas, and Peer Bork. Structure and function of the global ocean microbiome. *Science*, 348:1261359, 2015.

[42] Thibaud Taillefumier, Anna Posfai, Yigal Meir, and Ned S. Wingreen. Microbial consortia at steady supply. *eLife*, 6:e22644, 2017.

[43] Luke R. Thompson, Jon G. Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J. Locey, Robert J. Prill, Anupriya Tripathi, Sean M. Gibbons, Gail Ackermann, Jose A. Navas-Molina, Stefan Janssen, Evguenia Kopylova, Yoshiki Vázquez-Baeza, Antonio González, James T. Morton, Siavash Mirarab, Zhenjiang Zech Xu, Lingjing Jiang, Mohamed F. Haroon, Jad Kanbar, Qiyun Zhu, Se Jin Song, Tomasz Kosciolek, Nicholas A. Bokulich, Joshua Lefler, Colin J. Brislawn, Gregory Humphrey, Sarah M. Owens, Jarrad Hampton-Marcell, Donna Berg-Lyons, Valerie McKenzie, Noah Fierer, Jed A. Fuhrman, Aaron Clauset, Rick L. Stevens, Ashley Shade, Katherine S. Pollard, Kelly D. Goodwin, Janet K. Jansson, Jack A. Gilbert, Rob Knight, and The Earth Microbiome Project Consortium. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551:457, 2017.

[44] Mikhail Tikhonov and Remi Monasson. Collective phase in resource competition in a highly diverse ecosystem. *Physical Review Letters*, 118:048103, 2017.

[45] Mikhail Tikhonov and Remi Monasson. Innovation Rather than Improvement: A Solvable High-Dimensional Model Highlights the Limitations of Scalar Fitness. *Journal of Statistical Physics*, 2018.

[46] Nicole M. Vega and Jeff Gore. Stochastic assembly produces heterogeneous communities in the *Caenorhabditis elegans* intestine. *PLoS Biol.*, 15:e2000633, 2017.

[47] Stefanie Widder, Rosalind J Allen, Thomas Pfeiffer, Thomas P Curtis, Carsten Wiuf, William T Sloan, Otto X Cordero, Sam P Brown, Babak Momeni, Wenying Shou, Helen Kettle, Harry J Flint, Andreas F Haas, Béatrice Laroche, Jan-Ulrich Kreft, Tobias Großkopf, Jef Huisman, Andrew Free, Cristian Picioreanu, Christopher Quince, Isaac Klapper, Simon Labarthe, Barth F. Smets, Harris Wang, and Orkun S Soyer. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal*, 10(11):2557, 2016.

[48] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics (ser. 2)*, 62:548, 1955.

[49] Benjamin E. Wolfe and Rachel J. Dutton. Fermented foods as experimentally tractable microbial ecosystems. *Cell*, 161:49, 2015.

[50] Ali R. Zomorrodi and Daniel Segrè. Synthetic ecology of microbes: Mathematical models and applications. *J. Mol. Biol.*, 428:837, 2016.

V. APPENDIX

In this appendix, we provide a full explanation of the model and its numerical implementation, as well as additional data illustrating the robustness of the qualitative results. All data and code for generating figures can be found at https://github.com/robertvsiii/community-simulator, in the 'phase-transition' folder.

A. Model details

1. *Generalities*

We begin by defining an *energy flux* into a cell $J^{\rm in}$, an energy flux that is used for growth $J^{\rm growth}$, and an outgoing energy flux due to byproduct secretion $J^{\rm out}$. Energy conservation requires

$$J^{\rm in} = J^{\rm growth} + J^{\rm out} \tag{8}$$

for any reasonable metabolic model. Now consider a model with M resources R_β with $\beta = 1 \dots M$ each with "energy" or quality w_β. It will be useful to divide the input and output energy fluxes that are consumed/secreted in metabolite β by $J^{\rm in}_\beta$ and $J^{\rm out}_\beta$ respectively. We define the fraction $f^{\rm out}_\beta$ of the *output energy* secreted as resource β by

$$J^{\rm out}_\beta \equiv f^{\rm out}_\beta J^{\rm out}. \tag{9}$$

We can define corresponding mass fluxes by

$$\nu^{\rm out}_\beta \equiv J^{\rm out}_\beta / w_\beta \tag{10}$$

and

$$\nu^{\rm in}_\beta \equiv J^{\rm in}_\beta / w_\beta \tag{11}$$

In general, all these fluxes depend on the species under consideration and will carry an extra roman index i indicating the species.

We will assume that growth is proportional to the energy flux $J^{\rm growth}$ with some proportionality constant. Thus, we can model the growth of species i by

$$\frac{dN_i}{dt} = g_i N_i (J^{\rm growth}_i - m_i), \tag{12}$$

where g_i converts energy fluxes to growth and m_i is some minimum maintenance energy for species i.

We can model the resource dynamics by functions of the form

$$\frac{dR_\alpha}{dt} = h_\alpha(R_\alpha) - \sum_j N_j \nu^{\text{in}}_{j\alpha} + \sum_j N_j \nu^{\text{out}}_{j\alpha}, \tag{13}$$

where the function h_α describes the resource dynamics in the absence of other consumers. We can consider two kinds of dynamics: renewable and non-renewable. For renewable, we take a linearized form of the dynamics:

$$h^{\text{renew}}_\alpha(R_\alpha) = \kappa_\alpha - \tau_\alpha^{-1} R_\alpha \tag{14}$$

while for non-renewable we take a logistic form for the dynamics

$$h^{\text{non-renew}}_\alpha(R_\alpha) = r_\alpha R_\alpha (K_\alpha - R_\alpha). \tag{15}$$

In the present study, we only consider renewable dynamics.

These equations specify the general dynamics of all the models we consider. Metabolism is encoded in the relationship between input, output, and growth fluxes.

2. *Adding structured metabolism*

We will now specify the form of the input fluxes ν^{in}_β and the output partitioning f^{out}_β. This involves specifying how an input resource is turned into an metabolic byproducts. To try to capture metabolic structure, we will divide the M resources into T classes (e.g. sugars, amino acids, etc.), each with M_A resources where $A = 1, \dots T$ and $\sum_A M_A = M$. We will be interested in capturing coarse metabolic structure (i.e. metabolizing sugars outputs carboxylic acids, etc). We will limit ourselves to considering strictly substitutable resources.

In all consumer resource models, we assume that

$$\nu^{\text{in}}_{i\beta} = \sigma(c_{i\beta} R_\beta) \tag{16}$$

where σ encodes the response function of consumer i for resource α. We will work with three kinds of response functions: Type-I, linear response functions where

$$\sigma_I(x) = x, \tag{17}$$

a Type-II saturating Monod function,

$$\sigma_{II}(x) = \frac{x}{1 + \frac{x}{K}} \tag{18}$$

and a Type-III Hill or sigmoid-like function

$$\sigma_{III}(x) = \frac{x^n}{1 + \left(\frac{x}{K}\right)^n}, \tag{19}$$

where $n > 1$.

In all the simulations of this paper, we assume that resources independently contribute to the growth rate. We define a leakage fraction $0 \le l_\alpha \le 1$ for resource α such that

$$J^{\text{out}}_\alpha = l_\alpha J^{\text{in}}_\alpha. \tag{20}$$

A direct consequence of energy conservation (Equation (8)) is that

$$J^{\text{growth}}_i = \sum_\alpha (1 - l_\alpha) J^{\text{in}}_{i\alpha} = \sum_\alpha (1 - l_\alpha) w_\alpha \sigma(c_{i\alpha} R_\alpha) \tag{21}$$

All that is left is now specifying a structure of the probability of producing a byproduct β when consuming α. Let us denote by $D_{\beta\alpha}$ the fraction of the output energy that is contained in metabolite β when a cell consumes α. Note that by definition $\sum_\beta D_{\beta\alpha} = 1$. These $D_{\beta\alpha}$ and l_α uniquely specify the metabolic model for independent resources and we can write all fluxes in terms of these quantities.

The total energy output in metabolite β is thus

$$J^{\text{out}}_{i\beta} = \sum_\alpha D_{\beta\alpha} l_\alpha J^{\text{in}}_{i\alpha} = \sum_\alpha D_{\beta\alpha} l_\alpha w_\alpha \sigma(c_{i\alpha} R_\alpha). \tag{22}$$

This also yields

$$\nu^{\text{out}}_{i\beta} = \sum_\alpha D_{\beta\alpha} l_\alpha \frac{w_\alpha}{w_\beta} \sigma(c_{i\alpha} R_\alpha) \tag{23}$$

We are now in position to write down the full dynamics in terms of these parameters

$$\begin{aligned} \frac{dN_i}{dt} &= g_i N_i \left[\sum_\alpha (1 - l_\alpha) w_\alpha \sigma(c_{i\alpha} R_\alpha) - m_i \right] \\ \frac{dR_\alpha}{dt} &= h_\alpha(R_\alpha) - \sum_j N_j \sigma(c_{j\alpha} R_\alpha) \\ &\quad + \sum_{j\beta} N_j \sigma(c_{j\beta} R_\beta) \left[D_{\alpha\beta} \frac{w_\beta}{w_\alpha} l_\beta \right] \end{aligned} \tag{24}$$

Notice that when σ is Type-I (linear) and $l_\alpha = 0$ for all α (no leakage or byproducts), this reduces to MacArthur's original model.

3. *Choosing consumer preferences*

We will now choose consumer preferences as follows. We assume that each specialist family has a preference for one resource class A (where $A = 1 \dots F$) with $0 \le F \le T$. The index of the preferred class A will also serve as the name of the family. We will also consider generalists that have no preferences. Let us denote the consumer coefficients for family A by $c^A_{i\alpha}$. Let us also denote the consumer coefficients of the generalist family by $c^{\text{gen}}_{i\alpha}$. We will consider three kinds of models: one where the coefficients are drawn from Gaussian distributions, another where they are drawn from Gamma distributions (which ensure positivity of the coefficients), and finally a discrete, binary preference model.

a. Gaussian consumer preferences The Gaussian model allows a continuous gradation of transporter expression levels. We assume that the variance is fixed to so that for all coefficients for all families

$$\langle(\delta c_{i\alpha}^A)^2\rangle = \langle(\delta c_{i\alpha}^{\text{gen}})^2\rangle = \sigma_c^2. \tag{25}$$

In the generalist family, the mean is also the same for all resources, and is given by

$$\langle c_{i\alpha}^{\text{gen}}\rangle = \frac{\mu_c}{M}. \tag{26}$$

The specialist families sample from a distribution with a larger mean for resources in their preferred class:

$$\langle c_{i\alpha}^A\rangle = \begin{cases} \frac{\mu_c}{M}\left[1+\frac{M-M_A}{M_A}q_A\right], & \text{if } \alpha\in A \\ \frac{\mu_c}{M}(1-q_A), & \text{otherwise,} \end{cases} \tag{27}$$

where M_A is the number of resources in class A, and q_A controls how much more species from family A prefer resources from class A.

b. Gamma consumer preferences We will also consider the case where consumer preferences are drawn from Gamma distributions, which guarantee that all coefficients will be positive. We choose the parameters of the Gamma distribution such that the variance remains fixed for all coefficients for all families:

$$\langle(\delta c_{i\alpha}^A)^2\rangle = \langle(\delta c_{i\alpha}^{\text{gen}})^2\rangle = \sigma_c^2, \tag{28}$$

while the mean takes the same values as in the Gaussian case.

c. Choosing binary preferences In the binary model, there are only two possible expression levels for each transporter: a low level c_0 and a high level $c_0 + c_1$. The elements of $c_{i\alpha}^A$ are given by

$$c_{i\alpha}^A = c_0 + c_1 X_{i\alpha}, \tag{29}$$

where $X_{i\alpha}$ is a binary random variable that equals 1 with probability

$$p_{i\alpha}^A = \begin{cases} \frac{\mu_c}{Mc_1}\left[1+\frac{M-M_A}{M_A}q_A\right], & \text{if } \alpha\in A \\ \frac{\mu_c}{Mc_1}(1-q_A), & \text{otherwise} \end{cases} \tag{30}$$

for the specialist families, and

$$p_{i\alpha}^{\text{gen}} = \frac{\mu_c}{Mc_1} \tag{31}$$

for the generalists.

4. Constructing the metabolic matrix

We choose the metabolic matrix $D_{\alpha\beta}$ according to a three-tiered secretion model. The first tier contains a preferred class of byproducts, such as carboyxlic acids for fermentative and respiro-fermentative bacteria, which includes M_c members. The second contains byproducts of the same class as the input resource. For example, this could be attributed to the partial oxidation of sugars into sugar alcohols, or the antiporter behavior of various amino acid transporters. The third tier includes everything else. We encode this structure in $D_{\alpha\beta}$ by sampling each column of the matrix from a Dirichlet distribution with concentration parameters $d_{\alpha\beta}$ that depend on the byproduct tier, so that on average a fraction f_c of the secreted flux goes to the first tier, while a fraction f_s goes to the second tier, and $f_o = 1 - f_c - f_s$ to the third:

$$d_{\alpha\beta} = \begin{cases} d_0\frac{f_c}{M_c}, & \text{if } \alpha = c \\ d_0\frac{f_s}{M_{A(\beta)}}, & \text{if } \alpha\neq c \text{ and } A(\alpha)=A(\beta) \\ d_0\frac{f_o}{M-M_{A(\beta)-M_c}}, & \text{otherwise.} \end{cases} \tag{32}$$

The parameter d_0 controls the randomness of the partitioning, ranging from the maximum value where all the weight is put on a single resource for $d_0 = 1$, to deterministic partitioning as $d_0 \to \infty$.

B. Simulation and data analysis

1. The Community Simulator

We implemented the above modeling framework in a Python package called "community-simulator," which can be downloaded and installed from https://github.com/robertvsiii/community-simulator. Once this package is installed, the Jupyter notebook included in the 'phase-transition' folder of the repository can be used to regenerate all the figures.

Community Simulator is designed to run dynamics on multiple communities in parallel, inspired by the parallel experiments commonly performed with 96-well plates. The central object of the package is a `Community` class, whose instances are initialized by specifying the initial population sizes and resource concentrations for each parallel "well," along with the functions and parameters that define the population dynamics. This class contains two core methods, also based on this experimental setup. `Propagate(T)` sends each community to a separate CPU (for however many CPU's are available), runs the given population dynamics for a time T using the SciPy function `odeint`, and updates the population sizes and resource concentrations in each well to the time-evolved values. `Passage(f)` initializes a fresh set of wells by adding a fraction $f_{\mu\nu}$ of the contents of each old well ν to each new well μ. (Fresh media can also be added at this point, but this feature was not relevant for the current work). The resulting values of N_i are converted from arbitrary concentration units to actual population sizes using an optional scale factor, and then integer population sizes are obtained by multinomial sampling based on these values.

The Community Simulator package also contains a set of scripts for generating models and randomly sampling parameters. `MakeConsumerDynamics` and `MakeResourceDynamics` from the `usertools` module take a dictionary of assumptions concerning the response type, metabolic regulation and resource replenishment (as described above), and generate the corresponding functions for dN_i/dt and dR_α/dt. The function `MakeMatrices`, from the same module, samples the consumer matrix $c_{i\alpha}$ and the metabolic matrix $D_{\alpha\beta}$ according to the scheme described in the previous section.

2. *Simulation Details*

For this paper, we generated a binary consumer matrix with $c_0 = 0.01$, $c_1 = 1$ and $\mu_c = 10$, and a metabolic matrix with $d_0 = 0.2$. This matrix defined a regional pool of $S = 200$ species, consuming $M = 100$ possible resource types. We used $F = 4$ families and $T = 4$ resource class, but set $q = 0$ and $f_s = f_c = 0.25$ so that class membership had no functional consequence, but merely served as a neutral label. We set $w_\alpha = g_i = 1$ for all i and α, and set the l_α for all resources equal to each other. For the multinomial sampling described above, we worked in units where $N_i = 1$ corresponds to a population of 10^6 cells. We generated dynamics with Type-I response, no regulation, and a "renewable" resource replenishment model. Only resource type 0 was supplied externally, with flux κ_0, and all the other κ_α's were set to zero.

To simulate stochastic colonization, we initialized each of 10 wells with 100 randomly chosen species from the regional pool, with a population size of 10^6 cells per species per well. We propagated each well under Equations (24) for a time $\Delta t = 11,500$, which is much longer than the maximum time required to relax to the steady state for any of the parameter regimes sampled. We used the `Passage` method with $f_{\mu\nu} = \delta_{\mu\nu}$ to periodically eliminate species whose populations became too small. For the large steady-state population sizes we consider here ($\sim 10^4 - 10^9$, see Figure 9 below), the multinomial sampling results in near-deterministic elimination of species whose populations are heading for extinction. We passaged after every 5 time units of propagation from the beginning of the simulation up to time $t = 500$, then every 100 time units until time $t = 1,500$, and finally every 1,000 time units up to the final time $t = 11,500$.

The timeseries shown in Figure 1E was generated under these assumptions, with $w_0\kappa_0 = 500$.

We propagated these 10 initial states using this procedure for 100 different combinations of externally supplied energy flux $w_0\kappa_0$ and leakage fraction l, with 10 $w_0\kappa_0$ values evenly spaced on a logarithmic scale from 10 to 100, and 10 l values evenly spaced from 0 to 0.9. Figure 2 of the main text plots the mean richness over the 10 parallel wells for each combination of $w_0\kappa_0$ and l. The richness is defined as the number of species with non-zero abundance at the end of the simulation.

We focused on one example from each of the three regimes for further analysis:

- **Syntrophy-Limited**: $w_0\kappa_0 = 1000$, $\langle l_\alpha \rangle = 0.1$
- **Energy-Limited**: $w_0\kappa_0 = 28$, $\langle l_\alpha \rangle = 0.6$
- **Competition-Limited**: $w_0\kappa_0 = 1000$, $\langle l_\alpha \rangle = 0.9$.

The rank-abundance plots in Figure 2 of the main text show the population sizes in all 10 wells from each of these examples, after normalizing them by the total biomass $\sum_i N_i$. The plots were truncated at a relative abundance of 0.5% for clarity. Rank-abundance plots for these same three examples in absolute units with no truncation can be found in Figure 9 below.

3. *Susceptibilities*

One important property of an ecosystem is its sensitivity to changes in environmental conditions. Figure 3 of the main text quantifies this sensitivity in terms of a set of susceptibilities, defined by

$$\chi_{\alpha\beta} \equiv \frac{\partial \bar{R}_\alpha}{\partial \kappa_\beta} \tag{33}$$

$$\eta_{i\beta} \equiv \frac{\partial \bar{N}_i}{\partial \kappa_\beta} \tag{34}$$

where $\bar{N}_i$, $\bar{R}_\alpha$ are the steady-state consumer populations and resource concentrations, respectively.

For the case of renewable resource dynamics and Type-I growth, setting Equations (24) equal to zero and differentiating with respect to κ_β yields:

$$0 = \sum_\alpha (1 - l_\alpha) w_\alpha c_{i\alpha} \chi_{\alpha\beta} \tag{35}$$

$$-\tau_\alpha^{-1}\delta_{\alpha\beta} = \sum_\gamma \left(\sum_j c_{j\gamma} N_j \left[D_{\alpha\gamma} \frac{w_\gamma}{w_\alpha} l_\gamma - \delta_{\gamma\alpha} \right] - \tau_\alpha^{-1}\delta_{\gamma\alpha} \right) \times \chi_{\gamma\beta} + \sum_{j\gamma} c_{j\gamma} \left[D_{\alpha\gamma} \frac{w_\gamma}{w_\alpha} l_\gamma - \delta_{\gamma\alpha} \right] R_\gamma \eta_{j\beta} \tag{36}$$

For each value of β, this system of linear equations can be solved for $\chi_{\gamma\beta}$ and $\eta_{j\beta}$ by simply inverting a matrix (once the terms corresponding to extinct species have been removed).

The histograms of Figure 3D in the main text contain the diagonal elements $\chi_{\alpha\alpha}$ for all resources except for the one supplied externally ($\alpha = 0$), which might be expected to behave somewhat differently. The $\chi_{\alpha\alpha}$ values from all 10 parallel communities are included in the histogram. We generated one histogram for the competition-limited regime, and one for the energy-limited regime, using the parameters listed at the end of Section V B 2 above.

4. Niche Overlap

To find out what controls the diversity of the diverse phase, we varied the niche overlap, which controls the intensity of competition within the regional species pool. We did this by holding μ_c fixed, and varying c_1 from its original value of 1 down to a minimum value of 0.12. For each value of c_1, we generated 10 $c_{i\alpha}$ matrices, which each defined a regional pool of 200 species. We then repeated the procedure of Section V B 2 above for each of these regional pools: initializing 10 wells with 100 species and running them to the steady state with the same sequence of propagation and passage steps. The final richness of each community is plotted in Figure 4 of the main text as a translucent point, such that more common richness values are darker. We included all three examples defined at the end of Section V B 2 in the plot, and colored both examples from the resource-limited phase blue, while the competition-limited regime (diverse phase) was colored red.

5. Beta Diversity

To examine the beta diversity patterns in each regime, we initialized 200 wells with 100 randomly chosen species from the regional pool of 200 species, and propagated them to steady state following Section V B 2 under the three different choices of $w_0\kappa_0$ and l listed at the end of that section. To visualize the variation among these communities, we used the Python package scikit-learn to compute the first two principle components of the set of composition vectors in each regime. We then projected the compositions onto the plane spanned by these vectors, and generated a scatter plot of the results. We also computed the percentage of the total variance accounted for by each of these two principal components, and indicated the value in parentheses on each axis.

C. Data Format

The output of all the simulations was saved to a set of Microsoft Excel spreadsheets, which can be easily imported into Python for analysis using the Pandas package. Each simulation generated four files: final consumer populations (‘`Consumers`’), final resource compositions (‘`Resources`’), a metadata summary (‘`Parameters`’), and initial conditions (‘`Initial_State`’). The $c_{i\alpha}$ and $D_{\alpha\beta}$ matrices as well as the m_i and w_α were pickled into a binary file (‘`Realization`’). The file names also include the date on which the data was generated, and a task ID when multiple files were generated on the same day.

The first column in the consumer and resource tables is the index of the simulation run. The second and third columns of the consumer file are the family ID and species ID, respectively. In the resource file, these columns contain the class ID and resource ID. The remaining columns contain the populations/concentrations for each well. The consumer populations are in units of 10^6 cells.

All the parameters that change between runs are included in the metadata file (‘`Parameters`’). The first column of this file is the simulation run index, corresponding to the index in the consumer and resource files.

The initial conditions file contains the initial population sizes for each of the wells, which were the same for all runs within a simulation.

D. Robustness of qualitative results

In this section, we test the robustness of our qualitative results by modifying the modeling assumptions in five ways. We have given each way a descriptive name, which can be used to look up the raw data files from the supplemental data folder using the `file_list.csv` table:

- `main_dataset` is the data from the main text
- `type_II` uses a Type II functional response, with $K = 20$.
- `dense_metabolism` has a dense metabolic matrix with $d_0 = 0.001$.
- `randomness` adds (quenched) random variation to w_α and l_α, with standard deviations 0.1 and 0.03, respectively.
- `Gaussian_sampling` samples the $c_{i\alpha}$'s from Gaussian distributions, with the same mean 0.11 and standard deviation 0.3 as the binary matrix used in the main text.
- `Gamma_sampling` samples the $c_{i\alpha}$'s from Gamma distributions, with the same mean and variance.

The following sections describe each of these choices in more detail. Figures 8, 9, 12 and 13 show the key plots from the main text along with the new versions generated under all these modified assumptions. Figures 10 and 11 display another diversity measure not discussed in the main text: the Simpson Diversity (S. D.). This is defined analogously to the “effective number of resources consumed” presented in Equation (6) of the main text:

$$\text{S.D.} = \left[\sum_i \left(\frac{N_i}{N}\right)^2\right]^{-1} \tag{37}$$

where $N \equiv \sum_i N_i$. As discussed in the main text in connection with resource fluxes, this quantity approaches 1 when there is one large $N_i \approx N$ and all the other populations are very small. It approaches the number of species (i.e., the richness) as the biomass distribution becomes more uniform.

1. Type-II Growth

We chose the Monod parameter $K = 20$ in the Type-II growth simulations in order to ensure that at least one species would survive in the steady state in all simulations. The maximum possible incoming energy flux in the Type-II model is equal to $0.1K$ when $w_\alpha = 1$ and $l = 0.9$, and this must exceed $m_i \approx 1$ for a species to survive. $K = 20$ provides a maximum flux of 2 in this case.

2. Metabolic Matrices

The metabolic matrices $D_{\beta\alpha}$ are plotted in Figure 7 for `main_dataset` and `dense_metabolism` (all other simulations use the same metabolic parameters as `main_dataset`). We see that $d_0 = 0.2$ leads to a very sparse matrix, with only a few secreted byproducts per input resource, while the secretion fractions for $d_0 = 0.001$ are much more uniform.

3. Randomness in w_α and l_α

To relax the assumption of all the w_α's and l_α's being equal, we sampled these two vectors from Gaussian distributions. We chose the standard deviations of the distributions to be small enough that both quantities would almost always be positive, and l_α would remain less than 1.

4. Gaussian and Gamma Sampling

Sampling consumer preferences from the continuous Gaussian and Gamma distributions makes the differential equations much stiffer than in the binary case. To ensure stable operation of the integrator, we "passaged" the cells every 0.1 time units. Each call of the "passage" method zeros out small negative values of resource concentration or consumer population that arise because of numerical error, in addition to setting small consumer populations to zero. This high frequency of passaging made the simulation more computationally intensive, so we only propagated these simulations for 300 time units. We confirmed that the norm of the vector of between per-capita growth rates $\sqrt{\sum_i [(1/N_i)(dN_i/dt)]^2}$ was much less than 1 in all simulations to verify convergence.

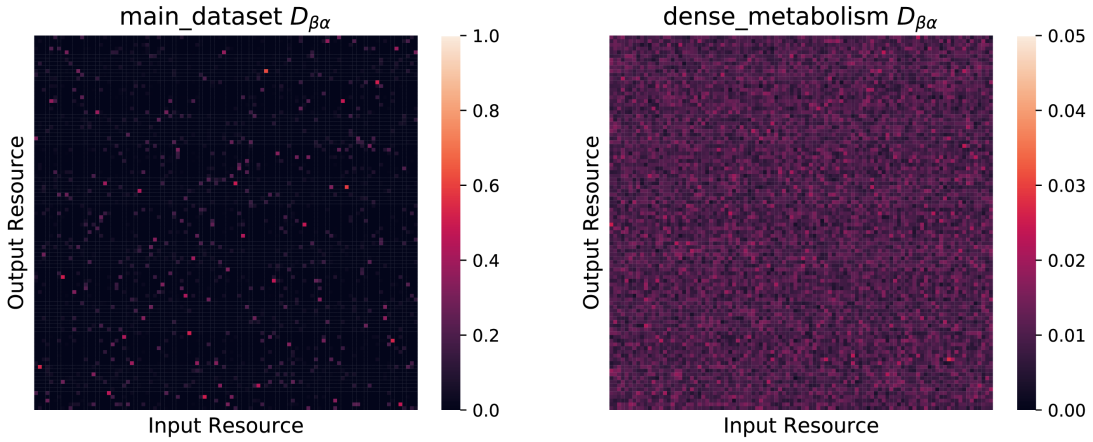

FIG. 7. **"Sparse" and "dense" metabolic matrices** $D_{\beta\alpha}$ Each pixel in the two panels represents an element of the metabolic matrix $D_{\beta\alpha}$. Panels are labeled by the name of the dataset they came from.

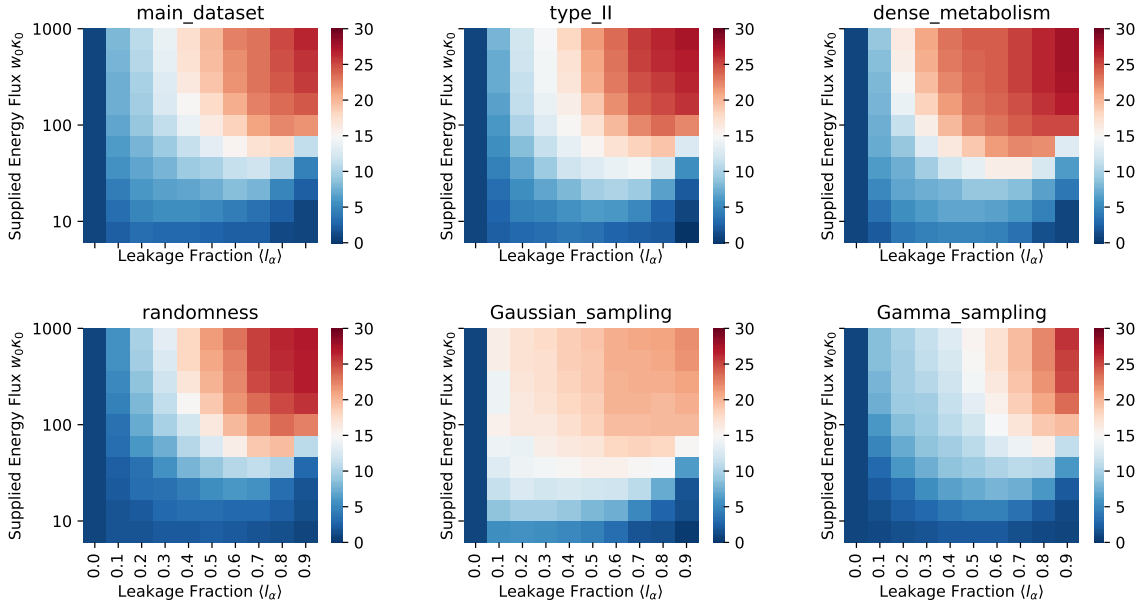

FIG. 8. **Richness vs.** $w_0\kappa_0$ **and** $\langle l_\alpha \rangle$ We generated 200 species, initialized 10 communities of 100 species each from this pool, and ran the dynamics to steady state under different combinations of $w_0\kappa_0$ and $\langle l_\alpha \rangle$, for each of the six model choices listed. The color of each square indicates the mean number of non-extinct species at the end of the simulation, over all 10 communities at each combination of $w_0\kappa_0$ and $\langle l_\alpha \rangle$.

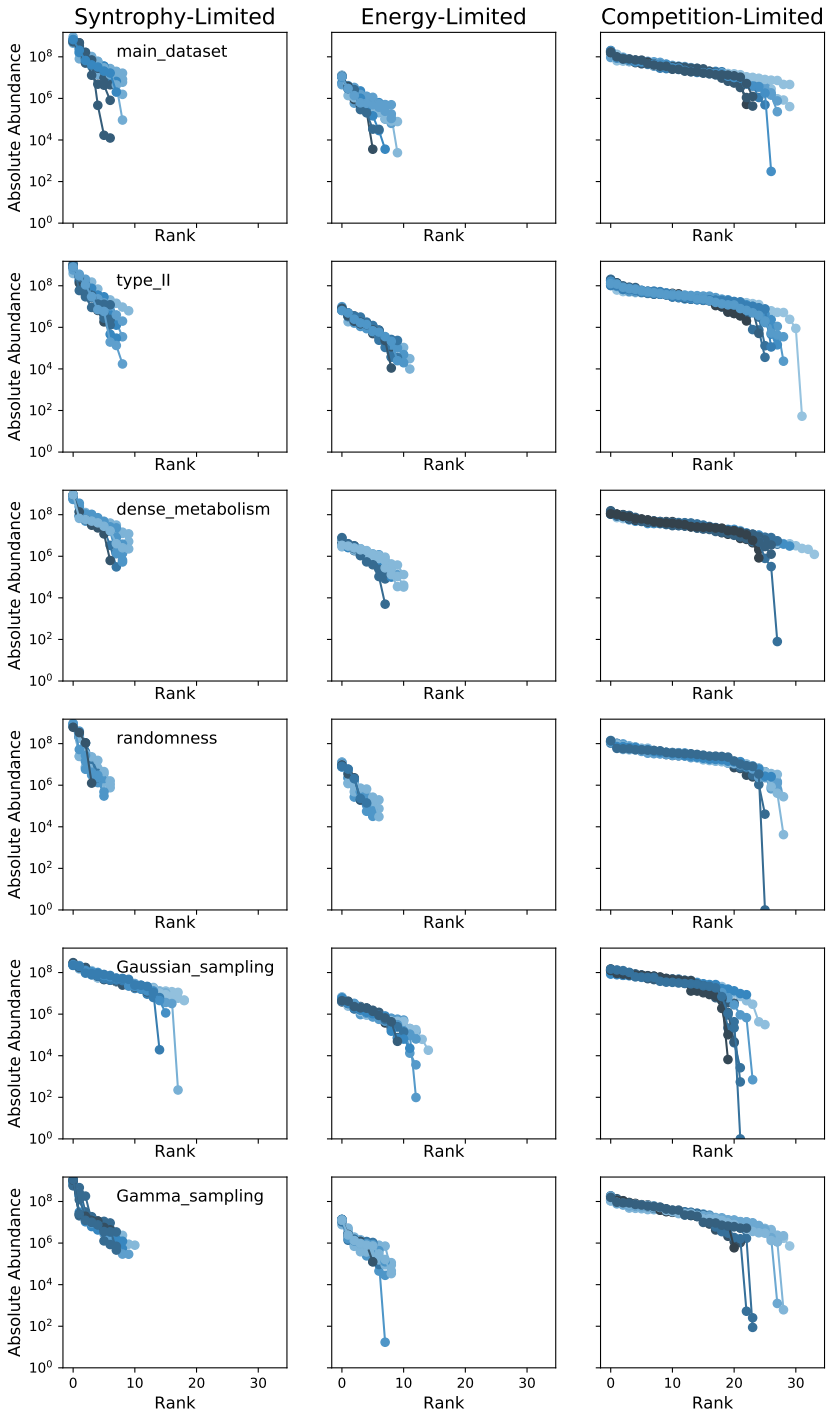

FIG. 9. **Rank-abundance curves in the three regimes** The abundance is plotted in absolute number of cells, as opposed to the relative abundance of the main text, and includes all species, with no truncation. The lower limit of the vertical axis is set to 1 cell, which is the smallest possible population size once the Passage method described in Section V B 2 has generated integer population values. One example from each regime has been chosen as defined at the end of Section V B 2.

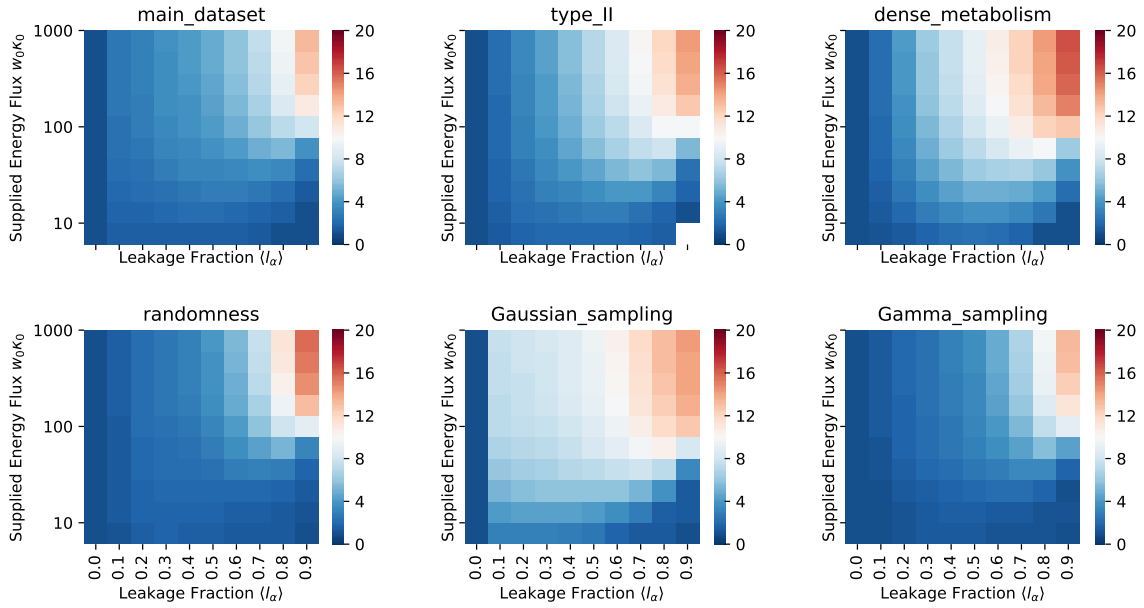

FIG. 10. **Simpson Diversity vs.** $w_0\kappa_0$ **and** $\langle l_\alpha \rangle$. Simpson Diversity was computed according to Equation (37), using the same data as Figure 8.

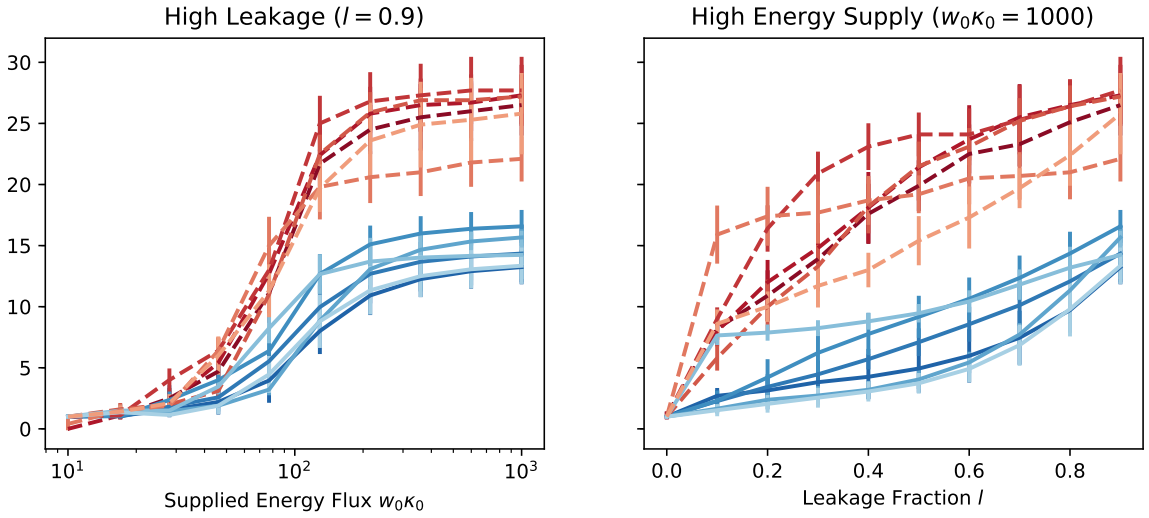

FIG. 11. **Richness (blue solid) and Simpson Diversity (red dotted) for cuts through the heat map** All six modeling choices are plotted, going from dark to light in the order listed in the text. Error bars are plus and minus one standard deviation, where the standard deviation is computed over the 10 parallel communities at each set of parameter values.

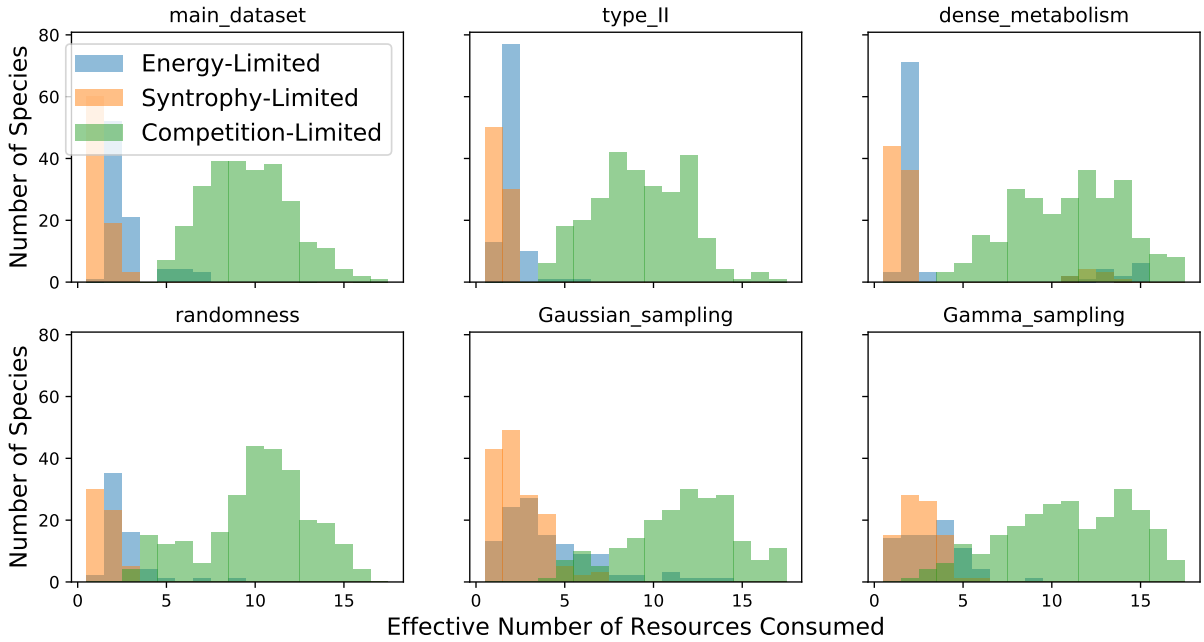

FIG. 12. **Effective number of resources consumed** The effective number of resources consumed M_i^{eff} is computed as described in the main text for the same three examples. The y axis indicates the total number of species falling into each bin from the combination of all 10 parallel communities in each example.

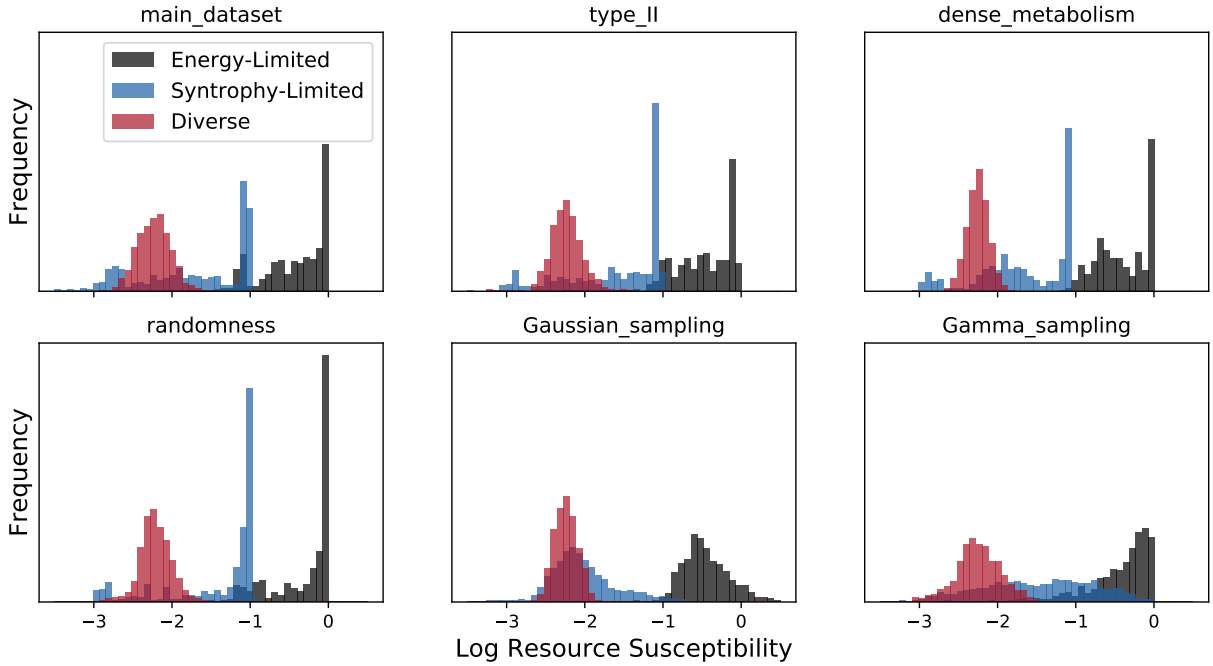

FIG. 13. **Resource susceptibility** Histograms of the logarithm $\log_{10} \partial \bar{R}_\alpha / \partial \kappa_\alpha$ are plotted for the same three examples. The y axis indicates the total number of species falling into each bin from the combination of all resources for all 10 parallel communities in each example, excepting the externally supplied resource $\alpha = 0$.