Macroecological laws in experimental microbial communities

William R. Shoemaker[1,*], Álvaro Sánchez[2], and Jacopo Grilli[1]

1 Quantitative Life Sciences, The Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, 34151, Italy.
2 Department of Microbial Biotechnology, Campus de Cantoblanco, CNB-CSIC, Madrid, Spain.

*** Contact:** `williamrshoemaker@gmail.com`

Abstract

Ecology has historically benefited by characterizing statistical patterns of biodiversity within and across communities. This approach, encompassed by the discipline of macroecology, has achieved considerable success in microbial ecology in recent years. Macroecological approaches have identified universal patterns of diversity and abundance that can be captured by effective models that do not include explicit interactions between community members. Experimentation has simultaneously played a crucial role in the development of our understanding of the ecology of microbes, as the advent of highly replicated time-series has allowed researchers to investigate how ecological forces govern community dynamics. However, there remains a gap between microbial experiments performed in the laboratory and macroecological patterns documented in natural systems, as we do not know if and how experimental manipulations produce macroecological effects. Here, we work to bridge the gap between the experimental manipulation of communities and their macroecological consequences. Using high-replication time-series of experimental microbial communities, we demonstrate that macroecological laws observed in nature can be readily recapitulated in a laboratory setting and unified under the Stochastic Logistic Model of growth (SLM). We found that demographic manipulations and their effect on community-level variation can alter empirical patterns in a manner that diverges from predictions obtained from the SLM. By incorporating experimental details (e.g., number of migrants), we were able to restore the predictive capacity of the SLM by linking demographic manipulations with macroecological effects. Finally, we demonstrate the extent that experimental manipulations are capable of altering macroecological patterns under the SLM, establishing a demarcation between macroecological outcomes we can and cannot observe in a laboratory setting.

Introduction

Microbial communities inhabit virtually every environment on Earth. Through their ubiquity, abundance, and diversity, microorganisms regulate the biogeochemical processes that sustain vast amounts of life. Furthermore, microbial communities play a crucial role in maintaining the health of many macroscopic forms of life, including humans, who have harnessed microbial communities to promote their own well-being [1, 2]. Given their environmental, medical, and economic importance, it is necessary to develop a quantitative theory of ecology that allows researchers to explain,

maintain, and alter the properties of microbial communities. A challenge of this scale is daunting, and the complexity of microbial communities has promoted the engagement of researchers from various disciplines with distinct approaches. Of these approaches, there are two that have substantially contributed towards our quantitative understanding of microbial communities in recent years: macroecology and experimental ecology.

Historically, the field of ecology has achieved considerable success by characterizing generalized statistical patterns of biodiversity, an approach known as macroecology [3–8]. The macroecological approach is, fundamentally, statistical in nature, allowing for quantitative predictions to be made about the typical features of ecological communities without having to specify microscopic ecological forces. The unencumbered nature of this approach has allowed researchers to successfully characterize a diverse array of microbial ecological patterns [9–17] and spurred the development of mathematical models of microbial ecology grounded in statistical physics [18–26]. Through this approach, disparate patterns were recently unified by the observation that the typical microbial community follows three macroecological laws: 1) the abundance of a given community member across sites follows a gamma distribution, 2) the mean abundance of a given community member is not independent of its variance (i.e., Taylor's Law [27, 28]), and 3) the mean abundance of a community member across sites follows a lognormal distribution [29]. These three universal laws can be captured by an intuitive mathematical model of density-dependent growth with environmental noise, the stochastic logistic model (SLM) of growth [29–31]. Building on its utility, the SLM has been successfully extended to quantitatively capture additional empirical microbial macroecological patterns. Examples that explicitly use the SLM include attempts to capture measures of ecological distances and dissimilarities between communities [32], alternative stable-states [33], patterns of richness and diversity at coarse-grained taxonomic and phylogenetic scales [34], and dynamics within and across human hosts at the sub-species level (i.e., strains) [17, 35]. The results of these studies demonstrate that a minimal mathematical model of ecological dynamics can capture a broad assemblage of microbial macroecological patterns.

Alongside the study of macroecological patterns in observational data, experimentation has played a crucial role in the documentation and manipulation of ecological patterns [36, 37]. The advent of 16S rRNA amplicon sequencing permits current researchers to advance past experimental efforts by investigating the ecological dynamics of a large number of microbial communities in a laboratory setting. Through this controlled approach, researchers have begun to determine the extent that the assembly of microbial communities is reproducible and predicable [38–43]. Surprisingly, seemingly simple environments have the capacity to form highly dissimilar microbial communities, even in systems harboring a lone exchangeable resource [44]. Results such as this have led researchers to investigate the degree that the state of a community is susceptible to experimentally imposed ecological forces. A prominent example is the migration of individuals between communities, an ecological force that is amenable to experimental manipulation [45, 46] and can alter the heterogeneity of communities across replicates [44, 47, 48]. However, the effect of migration on the macroecological properties of microbial communities, as well as the overall macroecology of experimental communities, remains to be determined.

Macroecology and experimental ecology as distinct approaches have revealed substantial insight into the fundamental nature of microbial communities, though they are not without their own drawbacks. It is often difficult to identify microscopic mechanistic causes for the macroecological patterns we frequently observe (e.g., the slope of Taylor's Law) [49]. Contrastingly, experimental ecology can inordinately focus on system-specific details and observations that may not be generalizable, a hindrance towards constructing an overall theory of microbial ecology [50, 51]. Providing

mechanistic explanations of universal empirical macroecological patterns alongside their characterization and prediction requires both approaches and is arguably a central goal of microbial ecology. Furthermore, there remains the need to develop and incorporate ecological theory to the study of variation in experimental microbial communities as well as the extent that extrinsic and intrinsic ecological forces contribute to said variation. By applying and extending macroecological theory to highly replicated temporally resolved experiments, we can begin to approach a quantitative theory of microbial ecology.

Here, we attempted to bridge the gap between the experimental manipulation of ecological forces and their resulting macroecological effects. Using high-replication time-series of experimental microbial communities, we first demonstrated that the macroecological laws observed in nature can be recapitulated in a laboratory setting. By connecting these patterns to the predictions of the SLM, we identified a reasonable null model of microbial macroecology which can be used to predict treatment-specific effects. Specifically, we focused on migration treatments that correspond to mainland-island and fully-connected metacommunity models [52, 53]. We examined the ways in which the ecological force of migration can be manipulated, their resulting macroecological outcomes, and in what degree these outcomes can be captured by the SLM. Using these results, we identified when and how the SLM succeeded at predicting the effects of experimental manipulation and where it fell short. By leveraging high-throughput ecological experiments alongside robust statistical patterns, we can strengthen the predictive and quantitative elements of microbial macroecological theory.

Materials and methods

Experimental data

Experimental data was obtained from a previous study where a large number of replicate ecological communities originating from a single progenitor soil sample were propagated in controlled laboratory conditions under regional or global migration regimes [44]. A given replicate community was initiated by inoculating 4 μL (low inoculum) or 40 μL (high inoculum) of the source community into 500 μL of M9 minimal media with 0.2% glucose into a well of a 96 deep-well plate (VWR) at 30°C under static conditions. Transfers were performed every 48 hours, 18 times in total for all replicate communities with a dilution rate $D_{\text{transfer}} = 0.0084$ (Table 1).

The details of the transfer procedure varied to manipulate the effects of different forms of migration. For the regional migration treatment, a 4 μL aliquot of the soil supernatant was added to the 4 μL aliquot from the previous transfer. For the global migration treatment, 4 μL aliquots of all replicate populations were pooled, vortexed, and diluted 10,000-fold. A 4 μL aliquot of the diluted solution was added to the 4 μL aliquot from the previous transfer for each replicate community. At the end of each 48 hr. transfer period samples from each replicate community were mixed with 40% glycerol and cryopreserved at -80°C.

DNA extraction was performed using a QIAGEN DNeasy 96 Blood and Tissue kit. Library preparation for 16S rRNA amplicon sequencing of the V4 region was performed as previously described [41] and PCR products were purified and normalized using the SequalPrep PCR kit (Invitrogen). Sequencing was performed on an Illumina MiSeq (2x250 bp paired-end) and raw reads were processed for demultiplexing and barcode, index, and primer removal using QIIME v1.9 [54]. The number of communities sequenced at each transfer for each treatment can be found in Table S1. In this study we reprocessed all raw FASTQ data from the original study to obtain Amplicon Sequence Variants (ASVs) using DADA2 [55]. We processed our data using the pooled

option inference option so that ASVs with an abundance of one in a given site (i.e., singletons) could be inferred, allowing us to examine the entirety of the empirical sampling distribution. The attractor status of a given replicate community was assigned as previously described [44]. Additional details can be found in the original manuscript [44].

Testing the macroecological predictions of the stochastic logistic model

The SLM is a stochastic differential equation that describes the logistic growth of a population with a time-dependent growth rate. In this context, the SLM for the relative abundance of the ith ASV (x_i) is defined as the following Langevin equation

$$\frac{\partial x_i}{\partial t} = \underbrace{\frac{x_i}{\tau_i}\left(1 - \frac{x_i}{K_i}\right)}_{\text{Self-limiting growth}} + \underbrace{\sqrt{\frac{\sigma_i}{\tau_i}} x_i \cdot \eta(t)}_{\text{Environmental noise}} \tag{1}$$

where $\frac{1}{\tau_i}$, K_i, and σ_i represent the intrinsic growth rate, carrying capacity, and the coefficient of variation of growth rate fluctuations.

Environmental noise is modeled as the product of a linear frequency term (as opposed to demographic noise, which would be modeled as $\sqrt{x_i}$), the compound parameter $\sqrt{\frac{\sigma_i}{\tau_i}}$, and a Brownian noise term $\eta(t)$ that introduces stochasticity. The expected value of $\eta(t)$ is $\langle \eta(t) \rangle = 0$ and the dependence of $\eta(t')$ at time t' on an earlier time t is defined as $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$ [56]. Under this standard definition if the noise term is shifted in time it has zero correlation with itself.

The SLM has recently been shown to explain three empirical macroecological laws that hold across dissimilar environments as follows: 1) predicting a gamma distributed abundance fluctuation distribution, 2) interpreting the power law relationship between mean relative abundance and the variance of relative abundance (i.e., Taylor's Law) in terms of a community member's carrying capacity, and 3) interpreting the lognormally distributed mean relative abundance distribution in terms of underlying carrying capacities [29]. To provide necessary context, we briefly summarize these three laws and their connection to the SLM.

The Abundance Fluctuation Distribution

The stationary form of the Abundance Fluctuation Distribution (AFD) P^* predicted by the SLM can be derived from the stationary state of the partial differential equation obtained from the Fokker-Planck equation corresponding to Eq. 1 (interpreted as an Itô SDE) [56]

$$P^*(x_i) = \frac{1}{\Gamma\left(2\sigma_i^{-1} - 1\right)} \left(\frac{2}{K_i \sigma_i}\right)^{2\sigma_i^{-1} - 1} \exp\left[-\frac{2}{K_i \sigma_i} x_i\right] x_i^{2\sigma_i^{-1} - 2} \tag{2}$$

which is the predicted form of the AFD under the SLM. This distribution is a gamma distribution with the following expected mean and squared coefficient of variation

$$\langle x_i \rangle = K_i \left(1 - \frac{\sigma_i}{2}\right) \tag{3a}$$

$$\frac{\langle x_i^2 \rangle - \langle x_i \rangle^2}{\langle x_i \rangle^2} = \frac{\sigma_i}{2 - \sigma_i} \tag{3b}$$

Defining empirical estimates of $\langle x_i \rangle$ and $\frac{\langle x_i^2 \rangle - \langle x_i \rangle^2}{\langle x_i \rangle^2}$ as $\bar{x}_i$ and β_i^{-1}, we obtain a form of the gamma that can be used to generate predictions

$$P^*(x_i|\bar{x}_i, \beta_i) = \frac{1}{\Gamma(\beta_i)} \left(\frac{\beta_i}{\bar{x}_i}\right)^{\beta_i} \exp\left[-x_i \frac{\beta_i}{\bar{x}_i}\right] x_i^{\beta_i - 1} \tag{4}$$

By using the Poisson limit of a binomial sampling process, the sampling form of the AFD has been derived [29]. The derivation allows us to calculate the probability of obtaining n reads out of a total sampling depth of N for the ith ASV as

$$P(n_i|\bar{x}_i, \beta_i, N) = \frac{\Gamma(\beta_i + n_i)}{n_i! \Gamma(\beta_i)} \left(\frac{\bar{x}_i N}{\beta_i + \bar{x}_i N}\right)^{n_i} \left(\frac{\beta_i}{\beta_i + \bar{x}_i N}\right)^{\beta_i} \tag{5}$$

By setting $n = 0$ we obtain the expression for the probability of not detecting an ASV in a site

$$P(0|\bar{x}_i, \beta_i, N) = \left(1 + \frac{\bar{x}_i N}{\beta_i}\right)^{-\beta_i} \tag{6}$$

from which we can derive the expectation of the fraction of M sites where an ASV is found (i.e., its occupancy).

$$\langle o_i \rangle = 1 - \frac{1}{M} \sum_{m=1}^{M} P(0|\bar{x}_i, \beta_i, N_m) \tag{7}$$

The relative error of our predictions was estimated as

$$\varepsilon = \left|\frac{\text{Obs.} - \text{Pred.}}{\text{Obs.}}\right| \tag{8}$$

The difference in relative error between treatments was estimated as

$$\overline{\Delta\varepsilon} = \frac{1}{S_{\text{obs}}} \sum_{i=1}^{S_{\text{obs}}} \log\left[\frac{\varepsilon_{\text{mig.}}}{\varepsilon_{\text{no mig.}}}\right] \tag{9}$$

Taylor's Law

Taylor's Law describes the empirical relationship between the mean and variance of the relative abundance [27, 28]

$$\text{Var}(x) \propto \bar{x}^{2b} \tag{10}$$

When $b = 1$ the empirical coefficient of variation is constant across ASVs, which in the context of the SLM means that the parameter controlling the strength of environmental noise is constant ($\sigma_i = \sigma$).

The Lognormally distributed Mean Abundance Distribution

We consider that the Mean Abundance Distribution (MAD) follows a lognormal distribution

$$P(\bar{x}_i)^{\text{lognorm}} = \frac{1}{\sqrt{2\pi s^2} \bar{x}_i} \exp\left[-\frac{(\log \bar{x}_i - \mu)}{2s^2}\right] \tag{11}$$

To account for finite sampling, we use a modified form of the lognormal that considers mean abundances with a coverage greater than c.

$$P(\bar{x}_i)_{\text{emp}}^{\text{lognorm}} = \frac{\theta(\bar{x}_i - c)P(\bar{x}_i)^{\text{lognorm}}}{\int \theta(z-c)P(z)^{\text{lognorm}}dz} \tag{12}$$

where $\theta(\cdot)$ is the Heaviside step function. Parameters of this modified lognormal were fit to the empirical MAD as previously described (see Supplementary Note 7 in [29]).

Simulating the SLM with migration

In experimental microbial communities, after a period of time resources will become depleted and an aliquot of the community will be transferred to a new environment with replenished resources. This process of serial dilution introduces boom-bust cycles of microbial growth. To account for this experimental detail, we considered a piece-wise form of the SLM to describe the dynamics within the kth transfer

$$\frac{\partial x_i^{(k)}}{\partial t} = \frac{x_i^{(k)}}{\tau_i}\left(1 - \frac{x_i^{(k)}}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}}x_i^{(k)} \cdot \eta(t) \tag{13}$$

where the dynamics occur over a period $t \in [0, T]$. A transfer then occurs and the dynamics start anew.

The time-dependent form of Eq. 13 can be derived (i.e., $P(x_i(t)|x_i(t_0), t_0)$ and could feasibly be extended to incorporate the effect of migration on initial conditions [57] (description of solution in S2 Text). However, there are additional experimental details and observations we would like to incorporate that can difficult to investigate analytically.

We begin by modeling the relative abundance at the start of a given transfer cycle. The abundances of ASVs at the start of the kth transfer cycle can be modeled as a multinomial sampling process

$$\Pr\left[\vec{n}^{(k)}(0)\right] = \begin{cases} \frac{N_{\text{inoc}}!}{n_1!\cdots n_S!}x_{1,\text{progenitor}}^{n_1}\cdots x_{S,\text{progenitor}}^{n_S}, & \text{if } k = 0 \\ \frac{N_{\text{transfer}}!}{n_1!\cdots n_S!}(x_1^{(k-1)}(T))^{n_1}\cdots(x_S^{(k-1)}(T))^{n_S}, & \text{if } k > 0 \end{cases} \tag{14}$$

Where the vector of S abundances at the start of the kth transfer cycle ($\vec{n}^{(k)}(0)$) are drawn from the progenitor community for the first transfer cycle ($k = 0$) and are then drawn from the previous transfer at $T = 48$ hr. for all subsequent transfer cycles ($k > 0$).

We then turn to modeling migration manipulations performed by the experimentor. Despite their different forms, both global and regional migration treatments are similar in that migrants are not supplied at a constant rate within a given transfer cycle (stationary distribution of the SLM with constant migration derived in S1 Text). Rather, migration is manipulated by adding an aliquot at the start of each of the first twelve transfer cycles. This experimental detail means that both regional and global migration treatments alter the initial abundances of community members, which subsequently grow according to the SLM. Our task then is to model the initial conditions at the start of each transfer cycle as an outcome of migration.

The regional migration treatment can be viewed as a form of island-mainland migration, where the ancestral community is propagated into the descendant communities during the first twelve transfers [52]. Contrastingly, global migration can be viewed as a fully interconnected metacommunity, where aliquots of each replicate are intermixed and equally distributed among replicates [53]. In mathematical terms, each

form of migration can be formalized for the ith ASV at transfer k as

$$n_{i,\text{regional}}(0)^{(k)} \propto x_{i,\text{progenitor}} \tag{15a}$$

$$n_{i,\text{global}}(0)^{(k)} \propto \frac{\sum_{m=1}^{M} n_{i,m}^{(k-1)}(T)}{\sum_{i=1}^{S}\sum_{m=1}^{M} n_{i,m}^{(k-1)}(T)} \tag{15b}$$

where abundances of migrants can again be modeled as a multinomial sampling process

$$\Pr\left[\vec{n}_{\text{migration}}^{(k)}\right] = \begin{cases} \frac{N_{\text{regional}}!}{n_1!\cdots n_S!} x_{1,\text{progenitor}}^{n_1} \cdots x_{S,\text{progenitor}}^{n_S}, & \text{Regional, } k \leq 12 \\ \frac{N_{\text{global}}!}{n_1!\cdots n_S!} (x_{1,\text{global}}^{(k-1)})^{n_1} \cdots (x_{S,\text{global}}^{(k-1)})^{n_S}, & \text{Global, } k \leq 12 \end{cases} \tag{16}$$

where $\vec{n}_{\text{migration}}^{(k)}(0) = \vec{0}$ for $k > 12$.

We can then examine how the typical initial *relative* abundance of a given ASV depends on the dilution rate of transfers. We begin by noticing that the final total abundance of a community does not considerably vary from transfer to transfer ($N^{(k-1)}(T) \approx N^{(k)}(T) \equiv N^*(T)$). Using the above definitions and dividing ASV abundances by the total abundances to obtain relative abundances ($x_i^{(k-1)}(T) = \frac{n_i^{(k-1)}(T)}{N^*(T)}$), we obtain

$$\left\langle x_i^{(k)}(0) \right\rangle = \begin{cases} \left\langle \frac{D_{\text{transfer}} x_i^{(k-1)}(T) N^*(T) + D_{\text{regional}} x_{i,\text{regional}} N_{\text{regional}}}{D_{\text{transfer}} N^*(T) + D_{\text{regional}} N_{\text{regional}}} \right\rangle, & \text{Regional} \\ \left\langle \frac{D_{\text{transfer}} x_i^{(k-1)}(T) N^*(T) + D_{\text{global}} x_{i,\text{global}}^{(k)} N_{\text{global}}}{D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}}} \right\rangle, & \text{Global} \\ \left\langle x_i^{(k-1)}(T) \right\rangle, & \text{No migration} \end{cases} \tag{17}$$

where parameters were set by the experimenter and are detailed in Table 1.

Beyond migration, there are dependencies between the final abundance of an ASV and its abundance in the progenitor community that we would like to incorporate. The set of K_i was drawn from the parameters obtained by fitting Eq. 12 to the empirical MAD. Dependence between $\vec{x}_{\text{regional}}$ and the MAD, which is proportional to $\vec{K}$, was evaluated by performing logistic regression from `scikit-learn v0.22.1` [58], obtaining the formula

$$\Pr[K_i > 0 | x_{i,\text{progenitor}}] \propto \left[1 + \exp(-(a + b * \log_{10} x_{i,\text{progenitor}}))\right]^{-1} \tag{18}$$

where a is the intercept and b is the slope of the regression. From this relationship we can model $\Pr[K_i > 0 | x_{i,\text{progenitor}}]$, allowing us to define the full distribution of carrying capacities as follows

$$K_i \sim \begin{cases} \text{Lognorm} & \text{with probability } \Pr[K_i > 0 | x_{i,\text{progenitor}}] \\ 0 & \text{with probability } 1 - \Pr[K_i > 0 | x_{i,\text{progenitor}}] \end{cases} \tag{19}$$

The total number of ASVs with $K_i > 0$ was drawn from the distribution $S_{\text{descendant}} \sim \text{Binomial}(S_{\text{progenitor}}, 0.016)$, where $S_{\text{progenitor}}$ is the number of ASVs in the progenitor community and 0.016 is the mean fraction of ASVs in descendant communities relative to the progenitor.

The task of grafting these empirical considerations onto the time-dependent analytic solution of the SLM is not straightforward. Therefore, the incorporation of experimental

details required the use of a simulation procedure. We obtain such a model by using an approximation of the numerical solution of the SLM within each transfer cycle which incorporates the above detail while requiring only two global parameters: σ and τ, the values of the latter being constrained by the amount of growth that can possibly occur over 48 hr (S3 Text). We set σ as a constant due to the existence of Taylor's Law and τ as a constant due to the unknown nature of the distribution of growth rates. We used logarithmically spaced parameter values from the following ranges: $\sigma \in [0.01, 1.9]$ and $\tau \in [1.7, 6.9]$. The upper bound on the range of σ was set by the observation that $\langle x_i \rangle = 0$ for $\sigma \leq 2$. The range on τ was set by the observation that the serial dilution factor sets the total number of generations per-transfer as $\log_2(D_{\text{transfer}}^{-1}) \sim 7$, assuming exponential growth. This translates to a maximum generation time of $\tau_{\max} = 48\text{hr.}/7 \sim 6.9\text{hr.}$. We identified a reasonable bound on the minimum generation time by noticing that prior research efforts have established that these communities have a maximum growth rate of $\sim 0.6\,\text{hr.}^{-1}$, translating to a minimum generation time of $\tau_{\min} \equiv 0.6^{-1} \approx 1.7\text{hr.}$ [44].

Simulated relative abundances were calculated from simulated true abundances and the process of sampling was modeled as a multinomial process

$$\Pr\left[\vec{n}_{\text{reads}}^{(k)}\right] = \frac{N_{\text{reads}}!}{n_1! \cdots n_S!} (x_1^{(k)})^{n_1} \cdots (x_S^{(k)})^{n_S} \tag{20}$$

The total number of reads N_{reads} were sampled from the empirical distribution of total read counts with replacement (Fig. S7).

Statistical analyses

We identified appropriate statistics to capture macroecological changes specific to each migration treatment. Standard tests were used when appropriate (e.g., change in slopes of Taylor's Law), but two additional statistics were used to assess patterns relating to change in correlation coefficients for the regional migration treatment and the change in the CV of relative abundances for the global migration treatment due to the cessation of migration manipulations. The change in correlation coefficients (ρ) between transfers 12 and 18 was assessed using Fisher's Z statistic [59]

$$Z_\rho = \frac{z_{18} - z_{12}}{\sqrt{\frac{1}{M_{12}-3} + \frac{1}{M_{18}-3}}} \tag{21}$$

where $z_t = \log\left[\frac{1+\rho_t}{1-\rho_t}\right]$, M_t is the number of replicate communities at transfer t, and the denominator represents the standard error of the numerator.

The difference in CVs between transfers 12 and 18 was calculated using the following previously described F_{CV} statistic [60]

$$F_{\text{CV}} = \frac{\text{CV}_{18}^2}{1 + \text{CV}_{18}^2 \frac{M_{18}-1}{M_{18}}} \cdot \frac{1 + \text{CV}_{12}^2 \frac{M_{12}-1}{M_{12}}}{\text{CV}_{12}^2} \tag{22}$$

where M_t represents the number of replicate communities at transfer t. Parameter combinations of τ, σ that best explained the statistics estimated from the data were identified using Approximate Bayesian Computation. For analyses with a single statistic we selected the simulation iteration with the lowest Euclidean distance between the observed and simulated statistic For analyses where a set of test statistics $\mathbf{c}$ were considered, we used a weighted measure of Euclidean distance where the simulated values of each test statistic were scaled by their standard deviation to prevent the distance from being dominated by a statistic with high variation [61]

$$d(\vec{c}_{\text{obs}}, \vec{c}_{\text{sim}}) = \left[\sum_{k=1}^{K} \left(\frac{c_{\text{sim}}^{(k)} - c_{\text{obs}}^{(k)}}{\sqrt{\text{Var}(c_{\text{sim}}^{(k)})}} \right)^2 \right]^{\frac{1}{2}} \tag{23}$$

All regressions were performed using ordinary least squares regression with `SciPy v1.4.1`.

Results

Using the results of a high-replication community assembly experiment, we examined the extent that universal patterns of microbial macroecology from observed data could be recapitulated in a laboratory setting. Focusing on forms of migration corresponding to island-mainland and metacommunity dynamics, we examined the extent that qualitatively similar macroecological patterns differed in their quantitative details. We then extended the SLM to incorporate experimental details without introducing additional free parameters, allowing us to balance the effectiveness of the SLM as a minimal model with experimental realism. Through simulation results, we identified whether the macroecological effects of a given form of migration could be captured by the SLM and established its limitations for modeling microbial biodiversity.

Macroecological patterns emerge in experimental systems

Predicting the macroecological effects of experimental manipulation is a vital goal of microbial ecology. With this goal in mind, a reasonable first step was to determine the degree that empirical macroecological patterns from observational data hold in an experimental system. The number of potential patterns one can examine is myriad, though three stand out: 1) the mean and variance of species abundances are not independent (Taylor's Law), 2) the Abundance Fluctuation Distribution across independent sites (AFD) is gamma distributed, and 3) the Mean Abundance Distribution across independent sites (MAD) follows a lognormal distribution [29]. These three patterns can be unified through the lens of a mean-field model, the Stochastic Logistic Model of growth [29].

To determine whether these patterns occur in an experimental system, we elected to examine the microbial communities from a prior experiment where qualitatively different forms of migration were manipulated [44]. Here a community grown from an environmental sample was divided among a large number (~ 100) of identical environments containing a defined media with glucose as the sole source of carbon and transferred after 48 hr. of growth 18 times. We first assessed the variation that emerged in experimental communities. The average relative abundance varied over four orders of magnitude across treatments and different transfers, meaning that a considerable degree of community-level variation was maintained in the experimental communities. By examining the relationship between the mean and variance of relative abundance, we found that the two measures clearly follow a similar relationship on a log-log scale across treatments, suggesting that Taylor's Law applies [27](Fig. 1). Under Taylor's Law a slope of two implies that the coefficient of variation (CV) remains constant across ASVs. By plotting this slope value, we found that it reasonably captured the data. Fitting a slope to each treatment for each transfer, we found that the mean slope is 2.08 ± 0.057, suggesting that despite the variation in typical abundance the CV of relative abundances remained roughly constant across ASVs.

We then turned towards examining the full distribution of abundances across replicate communities, known as the Abundance Fluctuation Distribution (AFD). To

facilitate comparisons, we rescaled the logarithm of the AFD for each ASV by its mean and variance (i.e., the standard score). We found that the rescaled AFDs tended to overlap across treatments, implying that despite differences in experimental details, the general shape of the AFD remained invariant. We found that the bulk of the distributions generally followed the gamma distribution predicted by the SLM (Eq. 2, Fig. 2a), though there remained a long tail which served as motivation for comparing the AFDs of different treatments.

We then turned to the fraction of communities where a given ASV was found, a measure known as occupancy. Assuming that ASVs are sampled as a Poisson process, the distribution of observed read counts can be analytically derived with a gamma AFD, from which a prediction of the expected occupancy can be obtained (Eq. 7). We found that the predictions of the SLM generally held across treatments, with slight deviations at high values of observed occupancies for treatments without migration and with regional migration (Fig. S1a). This trend was reflected in the distribution of relative errors of our prediction, where certain treatments appeared to have higher errors than others (Fig. S1b). However, a paired inspection using ASVs that were present in both treatments that underwent migration as well as the no migration control found that the error did not considerably change due to migration (Fig. S1c). A permutation-based test for the average change in the relative error of occupancy between treatments revealed a significant, if slight, effect where migration reduced the error of our predictions for both regional ($\overline{\Delta\varepsilon}_{\text{regional}} = -0.236$, $P = 0.00140$) and global migration ($\overline{\Delta\varepsilon}_{\text{global}} = -0.196$, $P = 0.0104$). Regardless, we were able to predict the occupancy of the typical ASV across different demographic treatments with high accuracy. Leveraging this result, we investigated the relationship between the mean relative abundance across replicates and the occupancy of an ASV, more commonly known as the abundance-occupancy relationship [12, 62, 63]. We found that all treatments tended to follow the same curve, which could be captured by binning and plotting the predicted occupancy for a given empirical mean abundance (Fig. 2b). The existence of this relationship in experimental communities is particularly striking, as it implies that the probability of observing a given community member is primarily determined by sampling effort, despite differences in experimental details.

Finally, we turned to the distribution of mean abundances. Prior work has demonstrated that the Mean Abundance Distribution (MAD) across sites can be captured by a lognormal distribution that accounts for sampling ([29], Eq. 12), an across-site extension of the observation that distribution of abundances within a single microbial community can be captured by a lognormal [10]. While this approach is difficult since the typical richness of a given treatment is ≈ 10, by pooling ASVs across treatments and transfers we found that the empirical MAD could be captured by a lognormal (Fig. 2c). A comparison of the two free parameters of the lognormal between transfers 12 and 18 indicated that MADs approached a similar shape after the cessation of migration manipulations (Fig. 2d). This result suggests that the MAD as a macroecological pattern consistently converges to a similar form over time. However, it should be noted that the lognormality of the MAD itself does not validate or invalidate the SLM. Rather, under the SLM the observation that the mean relative abundance and variance are not independent implies that the mean is proportional to the carrying capacity $\bar{x}_i \propto K_i$, so evidence of a lognormal MAD informs us of the distribution of parameters used in the SLM, not the SLM itself.

Testing the macroecological effects of migration

We have interpreted the results of our macroecological analyses in the broadest possible sense to convey that demographic manipulations of experimental communities are unlikely to induce *qualitative* changes to macroecological patterns (e.g., removing

Taylor's Law). Most microbial communities are similar and many of their features can be captured by minimal mathematical models [19, 29]. Instead, we propose that the utility of macroecology is in its role as a quantitative mediator between empirical observation and predictions obtained from mathematical models. Given that our two migration treatments can only be paired with a control that lacks migration, we omitted the high inoculation no migration treatment from the remainder of our study. By examining the quantitative deviations noted above we can identify how experiment-specific forms of migration can be incorporated into the SLM as well as the appropriate variables necessary to identify the quantitative outcomes of experimental treatments.

To start, it is clear from the previous section that the SLM is a useful tool and that a form of the SLM that incorporates migration would be appropriate. While a constant rate of migration can be readily incorporated into the SLM and a stationary solution of the AFD can be derived (S1 Text), this model does not reflect the details of the experiment. Rather, migration occurs in this experiment only at the start of a given transfer cycle. This detail corresponds to a model where the effect of migration can be captured as an experimentally-induced perturbation on the initial conditions of a system, namely the initial abundance of an ASV at the kth transfer ($n_i^{(k)}(0)$). The full time-dependent solution of the SLM which depends on initial conditions has previously been solved [57], meaning that the temporal evolution of the AFD in response to experimental perturbations and its subsequent relaxation to the stationary state of the gamma distribution can be modeled (S2 Text). We found that this is the case, where migration as a perturbation of initial conditions drastically altered the AFD relative to the constant migration case, with its effects then rapidly dissipating (Fig. S2).

The fact that we observe quantitative differences in macroecological patterns between different treatments as well as over time once migration manipulations have ceased tells us that the communities did not relax to their steady-state by the end of a given transfer cycle. This observation suggests that while the stationary solution of the SLM is sufficient to characterize certain macroecological patterns, the time-dependent solution of the SLM would likely be more appropriate to explain differences between treatments. However, there are additional experimental details that one would like to incorporate before predicting the effect of migration in order to strike a balance between the realism and tractability of a model.

The first observation is the number of ASVs in a given community, i.e., its richness. Through rarefaction curves and previously established richness estimation procedures [64], we found that the richness of the progenitor community was ~ 100 fold greater than the richness of a typical assembled community (Fig. S3). Migration had little effect on richness, as estimates are only slightly higher among communities that experienced migration but pale in comparison to the richness of the progenitor. This result suggests that the carrying capacity of a given ASV played a primary role in its survival, raising the question of how to specify the K_i of a community and how it relates to that of the progenitor. The observation that Taylor's Law holds in the no migration treatments implied that the coefficient of variation of growth fluctuations was constant across ASVs (i.e., $\sigma_i = \sigma$), meaning that $\bar{x}_i \propto K_i$ and, consequently, $K_i \sim$ Lognormal. This result allowed us to simulate carrying capacities as draws from a lognormal distribution. Weak correlations between the mean relative abundance after the cessation of migration ($\bar{x}_i(t = 18)$) and $x_{i,\text{progenitor}}$ for all treatments supports the view that K_i is independent of the abundances in the regional community (Fig. S4). However this observation represents a form of survivorship bias, as we can only compare values of $\bar{x}_i$ and x_i^{regional} for ASVs that were present in both the regional and descendant communities. We found that distributions of $x_{i,\text{progenitor}}$ differ depending on whether a given ASV was present in descendant communities, with the distribution for

ASVs that were present having shifted to the right (Fig. S5a). A permutational Kolmogorov–Smirnov test found that this shift was significant, a result that holds for all treatments (Fig. S6). This conditional dependence between the probability that a carrying capacity is non-zero and the relative abundance of an ASV in the progenitor was captured by logistic regression (Fig. S5b). This statistical dependence, along with the details regarding depth of sampling (i.e., number of reads, Fig. S7), were incorporated into our numerical simulations of the SLM (Materials and Methods).

Experiment-agnostic macroecological patterns

We first examined the quantitative effect of migration on macroecological patterns that have been previously unified by the SLM. Namely, we investigated the effect of migration on the AFD and Taylor's Law. We found that $\log_{10}$-transformed AFDs rescaled by their mean and variance only seem to differ between transfers 12 and 18 when migration is present (Fig. 4a-c). A KS test supports this observation, where significance was assessed by permuting the transfer label for a given ASV in a given replicate community.

Determining whether the observed KS statistic could be explained by the SLM required our experiment-informed simulation. We generated a null distribution of 10^4 values of KS with τ and σ drawn from a uniform distribution for each iteration. Using Approximate Bayesian Computation (ABC) we identified the optimal combination of parameters (Materials and Methods). The Euclidean distance was 0.003, $9 \cdot 10^{-5}$, and 0.02 for no, regional, and global migration treatments respectively. We then used the optimal set of parameters to generate null distributions of KS from 10^3 simulations. While the distance between AFDs for the no migration treatment was too small to be significant, its value lied outside the bulk of the simulated distribution (Fig. 4d). Regarding migration, the effect of regional migration was reasonably captured by the simulation (Fig. 4e) though the same could not be said for the global treatment (Fig. 4f). Before a conclusion can be reached it is worth evaluating whether the AFD is an appropriate measure for assessing the impact of migration. By repeating our simulations across a grid of τ, σ combinations we found that KS only changed in a systematic manner across parameter regimes for the regional migration treatment (Fig. S8). This result suggests that the inability of the SLM to capture the difference in AFDs for the no and global migration treatments was likely driven by the uninformative nature of the AFD for assessing the impact of migration.

We then turned towards Taylor's Law to determine whether the slope of the relationship changed after the cessation of migration (Fig. 5a-f). We found that the slope under regional migration was lower than that of the no migration at transfer 12 (Fig. 5b), but approached the result of the no migration treatment by transfer 18 (Fig. 5e). Similarly, the intercept of regional migration treatment was initially lower but then approached the value of the no migration treatment at treatment 18, as seen by the range of values on the y-axis. As a contrast, we observed little change in the values of the slope or intercept for the global migration treatment. In line with our predictions, there was no significant change in the slope ($t_{\text{slope}} = -1.12$, $P = 0.227$) or the intercept ($t_{\text{intercept}} = -0.605$, $P = 0.583$) between transfers 12 and 18 for the communities that did not undergo migration. We found that global migration failed to alter the slope ($t_{\text{slope}} = -0.390$, $P = 0.692$) or the intercept ($t_{\text{intercept}} = -0.380$, $P = 0.734$). Contrastingly, regional migration significantly altered both the slope ($t_{\text{slope}} = 3.34$, $P = 0.0178$) and intercept ($t_{\text{intercept}} = 2.75$, $P = 0.0271$).

By repeating the same ABC procedure outlined for the AFD we found that the distribution of slopes generated for an optimal set of parameters could capture the observed values of t_{slope} across all migration treatments (Fig. 5d-f). However, by examining t_{slope} and $t_{\text{intercept}}$ across a grid of parameter combinations we found that

the statistic was again only informative for the regional migration treatment (Figs. S9,S10). These results demonstrate that the AFD and Taylor's Law are primarily informative of the effects of regional (island-mainland) migration.

Experiment-specific macroecological patterns

The macroecological patterns we have examined up to this point did not considerably change under migration. It is useful to instead consider macroecological patterns that may not have received considerable attention in the past, but are likely altered by the presence of specific forms of migration. In this section, we consider how the experimental manipulations of each treatment should affect specific macroecological patterns. Assuming that the effect of migration was sufficient to alter the initial relative abundance of an ASV at the start of a given transfer, we would expect regional migration to primarily alter the mean abundance, whereas global migration would primarily alter fluctuations in ASV abundance.

Regional migration

We first considered regional migration and the macroecological patterns it might alter. Under the framework of the SLM, the effect of migration on the initial conditions should be detected if the timescale of growth was large enough such that the final state of the community was not dependent on its initial conditions.

To test this prediction, we examined the paired MADs for the regional and no migration treatments before and after the cessation of migration. We found that the correlation between MADs was initially weak and non-significant at transfer 12 but then considerably increased and was significant by transfer 18 (Fig. 6a,b), a result that is consistent with the interpretation that ASVs revert to their carrying capacity once migration has ceased. Using a permutation-based form of Fisher's Z-test for the difference in correlation coefficients, we found that the increase in MAD correlation was significant ($Z_\rho = 2.67$, $P = 0.0103$) [59]. Our SLM simulations support this conclusion, as we were able to recapture the observed value for the optimal parameter combination of σ and τ from ABC (Fig. 6c) and over a range of parameter combinations (Fig. S11a,b).

Our correlation analyses demonstrated how regional migration altered the MAD, but it remained necessary to examine how the effect of migration depended on ASV abundances in the progenitor community. To examine this dependence, we calculated the ratio of the mean abundance of an ASV in the regional and no migration treatments $\frac{\langle x_i \rangle_{\text{regional}}}{\langle x_i \rangle_{\text{no mig}}}$ and plotted its dependence on the its abundance in the progenitor community. We found that at transfer 12 there was a strong relationship between the two quantities that was statistically significant via permutation. By transfer 18 the slope has dissipated and is no longer significant (Fig. 6d,e). We found that the change in slopes is significant using a permutation-based t-test ($t = -2.64$, $P = 0.0255$). Similar to Z_ρ, these t statistics can be reproduced using a model of the SLM in certain σ, τ parameter regimes, which overlap with those found to reproduce observed estimates of Z_ρ (Fig. 6f) Similar to what we observed with Z_ρ, we found that successful predictions of σ, τ were restricted to certain parameter regimes (Fig. S11c,d).

Global migration

We examined relationships that are analogous to those examined in the regional migration case to evaluate the impact of global migration. First, under the framework of the SLM we predicted that global migration as a form of fully-connected migration in a metacommunity would strictly alter the fluctuations in ASV abundance across

replicate communities while leaving the MAD unchanged. In terms of measurables, we predicted that the correlation in the MAD would remain unchanged between transfers 12 and 18, whereas the correlation in ASV CVs would increase after the cessation of the migration treatment. We first found that within both transfers 12 and 18 for both the MAD and distribution CVs that ASVs were significantly correlated (Fig. S12). By repeating the permutation-based Z-test, we found that there was no significant change in the correlation of the MAD between transfers 12 and 18, consistent with our predictions ($Z_\rho = 0.203$, $P = 0.420$). However, there was also no significant increase in the strength of correlation for the distribution of CVs ($Z_\rho = 0.289$, $P = 0.621$), meaning that the cessation of migration did not considerably alter fluctuations in abundance.

It is possible that measurements taken at two sole timepoints is insufficient to detect the effect of global migration. Unlike regional migration, here we are primarily interested the effect that a treatment has in the fluctuation around a typical value, rather than the typical value itself. Situations like these where one is interested in the macroecology of fluctuations rather than typical values such as the mean often require additional observations. As a solution, we leveraged the higher temporal sampling resolution of the global migration treatment. Unlike regional migration communities, several global migration communities were sequenced at each of the 18 transfers, providing an opportunity to examine the fluctuation in abundance between subsequent timepoints $\Delta\ell = \log\left[\frac{x(t+1)}{x(t)}\right]$. As an analogous measure to those considered in Fig. S12, we calculated the ensemble mean and CV of $\Delta\ell$ for each ASV at each timepoint and examined how the distribution of each measure changed before and after the cessation of migration manipulations. On first examination, we found that $\langle\Delta\ell\rangle$ tended to relax towards a value of zero around the sixth transfer and remained there throughout the rest of the experiment (Fig. S14). This pattern indicated that after six transfers the abundances of ASVs after 48 hr. growth within a given transfer cycle have reached stationarity with respect to the initial conditions of the experiment, allowing us to examine equal intervals $[7, 12]$ and $[13, 18]$ as the period of time before and after the cessation of migration, respectively. We note that this pattern does not mean that the effect of the initial conditions at the start of a given transfer cycle are no longer relevant.

We used permutational KS tests to determine whether distributions of $\langle\Delta\ell\rangle$ varied before and after the cessation of migration while controlling for ASV identity. There is no evidence that $\langle\Delta\ell\rangle$ changed in the communities without any migration ($\mathrm{KS}_{\langle\Delta\ell\rangle} = 0.101$, $P = 0.862$) or within the global migration treatment ($\mathrm{KS}_{\langle\Delta\ell\rangle} = 0.215$, $P = 0.177$) with simulations returning similar estimates (Fig. S14). This result was consistent with our prediction that global migration would not affect the mean. Turning to the question of fluctuations, we examined how $\mathrm{CV}_{\Delta\ell}$ changed with respect to time. As predicted, the distribution of $\mathrm{CV}_{\Delta\ell}$ did not change between time windows for the no migration treatment ($\mathrm{KS}_{\mathrm{CV}_{\Delta\ell}} = 0.1$, $P = 0.789$; Fig. 7a). Furthermore, we observed that $\mathrm{CV}_{\Delta\ell}$ tended to increase after the cessation of migration, generating a significant difference between distributions ($\mathrm{KS}_{\mathrm{CV}_{\Delta\ell}} = 0.285$, $P = 0.0120$; Fig. 7b). This result was consistent with our predictions regarding the effect of global migration on ensemble fluctuations. However, our ABC SLM simulations were generally unable to reproduce values of $\mathrm{KS}_{\mathrm{CV}_{\Delta\ell}}$ that are consistent with what was observed (Fig. 7c,d). This result was consistent with our search for regions of parameter space that generated successful predictions (Fig. S13).

At face value, the inability of the SLM to reproduce the changes we observe in the CV for the global migration treatment is concerning. The size of the inoculum presents a reasonable explanation, as the global migration inoculum was nearly two orders of magnitude smaller than that of regional migration (Table 1). However, this experimental detail does not explain why we observed a higher CV after the cessation of migration. An experimental detail can, again, provide us with insight: the presence of

multiple attractors in the no migration treatment (i.e., alternative stable-states). Replicate communities that did not undergo migration tended to assemble such that their composition was dominated by one of two bacterial families: Alcaligenaceae or Pseudomonadaceae. Contrastingly, communities that were subjected to the global migration treatment remained consistently Alcaligenaceae-dominated over the course of the experiment. While a modified form of the SLM has been previously demonstrated to be capable of explaining the existence of alternative stable-states within a single community over an extended period of time, it is difficult to apply the same approach to a system with 18 timepoints, which effectively reduces to six for global migration if one is primarily concerned with the period before and after the cessation of the migration treatment [33].

While the existence of attractors confounds the appropriateness of the SLM as a model capable of describing macroecological properties across experimental communities in the absence of additional information, it does not restrict its ability to capture temporal properties within a community. Given this argument, we turn from estimates of $\mathrm{CV}_{\Delta\ell}$ calculated over the ensemble at a given time t to those calculated over time for a given replicate community. We again split the time-series into observations taken before and after the cessation of migration, obtaining two estimates of the CV for each ASV in each replicate community, $\mathrm{CV}^{<}_{\Delta\ell}$ and $\mathrm{CV}^{>}_{\Delta\ell}$, from which we calculated an F statistic describing the change in the CV [60]. By plotting the observed distributions alongside nulls obtained by permuting transfer labels, we found that the distributions of F generally overlaped with the null (Fig. S15).

While the observed distributions appear similar to the null, we performed a statistical analysis. We ran a one-tailed t-test on the distributions of F where the null was obtained by permuting treatment labels. We performed this test for all ASVs that were present in at least three replicate communities in both the global and no migration treatments. We found that the F-statistics in the global migration treatment were not significantly greater than those in the control, meaning that the CV with respect to time did not considerably increase after the cessation of migration (Table S2). We repeated the test to control for attractor status in the no migration treatment, obtaining similar results (Table S2). The inability of the SLM with demographic details to reproduce the observed increase in the CV of abundance fluctuations suggests that the existence of alternative stable-states is the primary detail that determined the macroecological effect of the global migration treatment.

Discussion

A vital result of our study is the demonstration that communities propagated in artificial environments harboring a single carbon source maintain the ecological variation necessary for macroecological investigations. When viewed as an ensemble, this variation provides the means to assess distributions of typical abundances across several orders of magnitude, a prerequisite for examining broad probabilistic patterns of diversity. The fact that this criterion was met allowed us to document the existence of macroecological patterns that were previously observed in naturally occurring communities, suggesting a quantitative equivalence between experimental and observational studies despite the controlled nature of artificially maintained communities and the presence of distinct demographic treatments. However, the repeatable maintenance of variation observed in the face of demographic manipulations was likely contingent on the high level of variation present in the progenitor community. This figurative raw material is analogous to the need for genetic variation to exist before selection can occur [65–67], the absence of which would preclude the possibility of macroecological investigations.

Characterizing robust empirical patterns in artificial communities is a key step toward identifying appropriate statistical null models in microbial ecology. The existence of patterns predicted by the SLM in artificial communities provided an opportunity to evaluate the macroecological consequences of experimental manipulations. Our approach of modifying the SLM, an empirically validated null model of microbial community composition, using experimental details provided a useful framework for identifying treatments that are capable of generating macroecological effects. We examined the effects of two different forms of migration: regional and global [44]. As expected, regional migration altered macroecological patterns of typical relative abundance in a manner that was captured by the SLM using minimal free parameters. The results of our regional migration analyses demonstrate that modeling frameworks can be employed to link experimental microbial communities and macroecological patterns.

Contrastingly, we predicted that global migration would alter fluctuations around typical abundance [53]. We observed no change in the temporal variation in community member fluctuations after the cessation of migration within a given replicate community. However, we found that variation across the ensemble of communities tended to increase after the cessation of migration. This trend was consistent with our hypothesis regarding the effect of global migration, but one that could not be reproduced by the SLM using experimentally-set parameters. This inconsistency can be explained by the observation that the abundance of certain taxonomic families exhibited considerable heterogeneity across replicates that did not undergo migration, likely due to the presence of alternative stable-states [47, 68, 69]. Alternative stable-states cannot be captured by the SLM as-is without additional assumptions, namely on the form of a potential function that depends on a community member's abundance [70]. Additional environmental details may be necessary to incorporate as well, as the fluctuations imposed in a batch culture design can induce alternative stable-states [71]. However, we are able to draw two key conclusions. First, the quantity of migrants chosen for the global migration treatment was sufficient to alter the attractor status, but insufficient to alter fluctuations in abundance within a given replicate. Second, macroecological patterns explained by the SLM persisted despite the pervasiveness of alternative stable-states, suggesting that forms of fine-scale heterogeneity cannot considerably alter certain macroecological patterns.

The results of our global migration analysis illustrate how one's choice of model limits the set of empirical patterns that can be sufficiently captured. We have primarily focused on patterns relating to the typical abundance and fluctuations in abundance across communities as well as over time. Noticeably, we did not examine patterns that are likely to be overwhelmingly determined by interactions between community members (e.g., the correlations of abundance fluctuations), as the addition of interactions into the SLM would require several assumptions about the network of interactions and their magnitude. The consideration of microscopic dynamics such as resource consumption may be necessary in order to address experimental questions relating to interactions such as correlations in abundance [11]. Indeed, consideration of resource consumption has proven critical for investigating the evolutionary dynamics of microorganisms in an ecological context [20]. However, recent developments on the predictability of community function (e.g., total biomass, polysaccharide hydrolysis, resource excretion, etc.) point towards new avenues of exploration for microbial macroecology. There is increasing evidence that the functional profiles of experimental communities tend to follow quantitative rules that are amenable to mathematical modeling [72–75]. Extending microbial macroecology beyond patterns of abundance to the level of function embodies the original physiological and energetic breadth that allowed macroecology to advance our understanding of macroorganisms [4].

Finally, it is worth taking a step back to consider how the results presented here shape the microbial view of macroecology. The discipline of macroecology was originally conceived as an explicitly non-experimental form of investigation [4]. Analysis of the origin and development of macroecology provides two historical explanations for the initial rejection of experimental approaches: 1) large-scale community-level experiments are often impractical and 2) producing generalities from experiments has proven to be difficult [76]. Our results demonstrate that these two constraints are ameliorated in the study of microbial communities, running contrary to recent claims that statistical distributions are uninformative of mechanism [77]. We, and others [78], propose that the timescales, abundance, and comparative ease with which ensembles of communities can be maintained and manipulated make microorganisms an ideal system for testing quantitative macroecological predictions.

Data and code availability

All code written for this study is available on GitHub under a GNU General Public License: `https://github.com/wrshoemaker/experimental_macroecology`.

Acknowledgments

We thank S. Estrela for her assistance in reprocessing the data. We thank S. Bubnovich, M. Dal Bello, A. Goyal and members of the qEcoEvo group at ICTP for helpful discussions. We thank N. I. Wisnoski for their comments and insights. This work was supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2010885 (W.R.S.). ÁS acknowledges support from Grant PID2021-125478NA-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Author contributions

W.R.S., Á.S., and J.G. conceptualized the project, developed the mathematical models, and wrote the manuscript. W.R.S. performed all analyses.

References

1. Schluter, J. & Foster, K. R. The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. *PLoS Biology* **10**, e1001424 (2012). URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3502499/`.

2. Eloe-Fadrosh, E. A. & Rasko, D. A. The Human Microbiome: From Symbiosis to Pathogenesis. *Annual review of medicine* **64**, 145–163 (2013). URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3731629/`.

3. Brown, J. H. & Maurer, B. A. Macroecology: The Division of Food and Space Among Species on Continents. *Science* **243**, 1145–1150 (1989). URL `https://www.science.org/doi/10.1126/science.243.4895.1145`. Publisher: American Association for the Advancement of Science.

4. Brown, J. H. *Macroecology* (University of Chicago Press, Chicago, 1995).

5. Azaele, S. *et al.* Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics* **88**, 035003 (2016). URL https://link.aps.org/doi/10.1103/RevModPhys.88.035003. Publisher: American Physical Society.

6. Brown, J. H. Macroecology: Progress and Prospect. *Oikos* **87**, 3–14 (1999). URL https://www.jstor.org/stable/3546991. Publisher: [Nordic Society Oikos, Wiley].

7. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)* (Princeton University Press, 2011). URL https://www.degruyter.com/document/doi/10.1515/9781400837526/html. Publication Title: The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32).

8. Harte, J. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford series in ecology and evolution (Oxford University Press, Oxford ; New York, 2011). OCLC: ocn701807862.

9. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* **113**, 5970–5975 (2016). URL https://www.pnas.org/doi/10.1073/pnas.1521291113. Publisher: Proceedings of the National Academy of Sciences.

10. Shoemaker, W. R., Locey, K. J. & Lennon, J. T. A macroecological theory of microbial biodiversity. *Nature Ecology & Evolution* **1**, 107 (2017).

11. Ho, P.-Y., Good, B. H. & Huang, K. C. Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. *eLife* **11**, e75168 (2022). URL https://doi.org/10.7554/eLife.75168. Publisher: eLife Sciences Publications, Ltd.

12. Shade, A. *et al.* Macroecology to Unite All Life, Large and Small. *Trends in Ecology & Evolution* **33**, 731–744 (2018).

13. Lennon, J. T. & Locey, K. J. More support for Earth's massive microbiome. *Biology Direct* **15**, 5 (2020). URL https://doi.org/10.1186/s13062-020-00261-8.

14. Zamponi, N., Grigera, T. S., Gudowska-Nowak, E. & Chialvo, D. R. Scale free density and correlations fluctuations in the dynamics of large microbial ecosystems (2022). URL http://arxiv.org/abs/2206.12384. ArXiv:2206.12384 [q-bio].

15. George, A. B. & O'Dwyer, J. Universal abundance fluctuations across microbial communities, tropical forests, and urban populations (2022). URL https://www.biorxiv.org/content/10.1101/2022.09.14.508016v1. Pages: 2022.09.14.508016 Section: New Results.

16. Ji, B. W., Sheth, R. U., Dixit, P. D., Tchourine, K. & Vitkup, D. Macroecological dynamics of gut microbiota. *Nature Microbiology* **5**, 768–775 (2020). URL https://www.nature.com/articles/s41564-020-0685-1. Number: 5 Publisher: Nature Publishing Group.

17. Shoemaker, W. R. A macroecological perspective on genetic diversity in the human gut microbiome. *PLOS ONE* **18**, e0288926 (2023). URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0288926. Publisher: Public Library of Science.

18. Batchelor, M. T., Burne, R. V., Henry, B. I. & Slatyer, T. Statistical physics and stromatolite growth: new perspectives on an ancient dilemma. *Physica A: Statistical Mechanics and its Applications* **350**, 6–11 (2005). URL https://www.sciencedirect.com/science/article/pii/S0378437105000373.

19. Marsland, R., Cui, W. & Mehta, P. A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Scientific Reports* **10**, 3308 (2020). URL https://www.nature.com/articles/s41598-020-60130-2. Number: 1 Publisher: Nature Publishing Group.

20. Good, B. H., Martis, S. & Hallatschek, O. Adaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources. *Proceedings of the National Academy of Sciences* **115**, E10407–E10416 (2018). URL https://www.pnas.org/content/115/44/E10407. Publisher: National Academy of Sciences Section: PNAS Plus.

21. Xu, L., Xu, X., Kong, D., Gu, H. & Kenney, T. Stochastic Generalized Lotka-Volterra Model with An Application to Learning Microbial Community Structures (2020). URL http://arxiv.org/abs/2009.10922. ArXiv:2009.10922 [stat].

22. D'Andrea, R., Gibbs, T. & O'Dwyer, J. P. Emergent neutrality in consumer-resource dynamics. *PLOS Computational Biology* **16**, e1008102 (2020). URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008102. Publisher: Public Library of Science.

23. Cui, W., Marsland, R. & Mehta, P. Diverse communities behave like typical random ecosystems. *Physical Review E* **104**, 034416 (2021). URL https://link.aps.org/doi/10.1103/PhysRevE.104.034416. Publisher: American Physical Society.

24. Barbier, M. & Arnoldi, J.-F. The cavity method for community ecology (2017). URL https://www.biorxiv.org/content/10.1101/147728v3. Pages: 147728 Section: New Results.

25. Advani, M., Bunin, G. & Mehta, P. Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model. *Journal of Statistical Mechanics: Theory and Experiment* **2018**, 033406 (2018). URL https://dx.doi.org/10.1088/1742-5468/aab04e. Publisher: IOP Publishing and SISSA.

26. Camacho-Mateu, J., Lampo, A., Sireci, M., Muñoz, M. & Cuesta, J. A. Species interactions reproduce abundance correlations patterns in microbial communities (2023). URL http://arxiv.org/abs/2305.19154. ArXiv:2305.19154 [math, q-bio, stat].

27. Taylor, L. R. Aggregation, Variance and the Mean. *Nature* **189**, 732–735 (1961). URL https://www.nature.com/articles/189732a0. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4766 Primary_atype: Research Publisher: Nature Publishing Group.

28. Xiao, X., Locey, K. J. & White, E. P. A Process-Independent Explanation for the General Form of Taylor's Law. *The American Naturalist* **186**, E51–E60 (2015). URL https://www.journals.uchicago.edu/doi/full/10.1086/682050. Publisher: The University of Chicago Press.

29. Grilli, J. Macroecological laws describe variation and diversity in microbial communities. *Nature Communications* **11**, 4743 (2020). URL https://www.nature.com/articles/s41467-020-18529-y. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Biodiversity;Community ecology;Ecological modelling;Macroecology;Microbial ecology Subject_term_id: biodiversity;community-ecology;ecological-modelling;macroecology;microbial-ecology.

30. Descheemaeker, L., Grilli, J. & de Buyl, S. Heavy-tailed abundance distributions from stochastic Lotka-Volterra models. *Physical Review E* **104**, 034404 (2021). URL https://link.aps.org/doi/10.1103/PhysRevE.104.034404. Publisher: American Physical Society.

31. Descheemaeker, L. & de Buyl, S. Stochastic logistic models reproduce experimental time series of microbial communities. *eLife* **9**, e55650 (2020). URL https://doi.org/10.7554/eLife.55650. Publisher: eLife Sciences Publications, Ltd.

32. Zaoli, S. & Grilli, J. The stochastic logistic model with correlated carrying capacities reproduces beta-diversity metrics of microbial communities. *PLOS Computational Biology* **18**, e1010043 (2022). URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010043. Publisher: Public Library of Science.

33. Zaoli, S. & Grilli, J. A macroecological description of alternative stable states reproduces intra- and inter-host variability of gut microbiome. Tech. Rep. (2021). URL https://www.biorxiv.org/content/10.1101/2021.02.12.430897v1. Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article.

34. Shoemaker, W. R. & Grilli, J. Macroecological patterns in coarse-grained microbial communities (2023). URL https://www.biorxiv.org/content/10.1101/2023.03.02.530804v1. Pages: 2023.03.02.530804 Section: New Results.

35. Wolff, R., Shoemaker, W. R. & Garud, N. R. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. Tech. Rep. (2021). URL https://www.biorxiv.org/content/10.1101/2021.09.30.462616v1. Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article.

36. Gause, G. F. Experimental Studies on the Struggle for Existence. *Journal of Experimental Biology* **9**, 389–402 (1932). URL https://journals.biologists.com/jeb/article/9/4/389/22702/Experimental-Studies-on-the-Struggle-for.

37. Resetarits, W. J. & Bernardo, J. (eds.) *Experimental ecology: issues and perspectives* (Oxford Univ. Press, New York, NY, 2001).

38. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1**, 1–7 (2017). URL https://www.nature.com/articles/s41559-017-0109. Number: 5 Publisher: Nature Publishing Group.

39. Dal Bello, M., Lee, H., Goyal, A. & Gore, J. Resource-diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nature Ecology & Evolution* **5**, 1424–1434 (2021).

40. Meroz, N., Tovi, N., Sorokin, Y. & Friedman, J. Community composition of microbial microcosms follows simple assembly rules at evolutionary timescales. *Nature Communications* **12**, 2891 (2021). URL https://www.nature.com/articles/s41467-021-23247-0. Number: 1 Publisher: Nature Publishing Group.

41. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* (2018). URL https://www.science.org/doi/abs/10.1126/science.aat1168. Publisher: American Association for the Advancement of Science.

42. Estrela, S., Sanchez-Gorostiaga, A., Vila, J. C. & Sanchez, A. Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments. *eLife* **10**, e65948 (2021). URL https://doi.org/10.7554/eLife.65948. Publisher: eLife Sciences Publications, Ltd.

43. Bittleston, L. S., Gralka, M., Leventhal, G. E., Mizrahi, I. & Cordero, O. X. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nature Communications* **11**, 1440 (2020). URL https://www.nature.com/articles/s41467-020-15169-0. Number: 1 Publisher: Nature Publishing Group.

44. Estrela, S. *et al.* Functional attractors in microbial community assembly. *Cell Systems* **13**, 29–42.e7 (2022). URL https://www.sciencedirect.com/science/article/pii/S2405471221003793.

45. Obadia, B. *et al.* Probabilistic Invasion Underlies Natural Gut Microbiome Stability. *Current Biology* **27**, 1999–2006.e8 (2017). URL https://www.sciencedirect.com/science/article/pii/S0960982217305705.

46. Vega, N. M. & Gore, J. Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. *PLOS Biology* **15**, e2000633 (2017). URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2000633. Publisher: Public Library of Science.

47. Amor, D. R., Ratzke, C. & Gore, J. Transient invaders can induce shifts between alternative stable states of microbial communities. *Science Advances* **6**, eaay8676 (2020). URL https://www.science.org/doi/full/10.1126/sciadv.aay8676. Publisher: American Association for the Advancement of Science.

48. Gokhale, S., Conwill, A., Ranjan, T. & Gore, J. Migration alters oscillatory dynamics and promotes survival in connected bacterial populations. *Nature Communications* **9**, 5273 (2018). URL https://www.nature.com/articles/s41467-018-07703-y. Number: 1 Publisher: Nature Publishing Group.

49. McGill, B. J. Towards a unification of unified theories of biodiversity. *Ecology Letters* **13**, 627–642 (2010). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2010.01449.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2010.01449.x.

50. Lawton, J. H. Are There General Laws in Ecology? *Oikos* **84**, 177–192 (1999). URL https://www.jstor.org/stable/3546712. Publisher: [Nordic Society Oikos, Wiley].

51. Maurer, B. A. *Untangling ecological complexity: the macroscopic perspective* (University of Chicago Press, Chicago, Ill, 1999).

52. MacArthur, R. H. & Wilson, E. O. *The theory of island biogeography* (Princeton University Press, Princeton, 2001).

53. Leibold, M. A. *Metacommunity ecology*. No. 59 in Monographs in population biology (Princeton University Press, Princeton, NJ, 2018).

54. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010). URL https://www.nature.com/articles/nmeth.f.303. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5 Primary_atype: Correspondence Publisher: Nature Publishing Group Subject_term: Microbial ecology;Next-generation sequencing;Software Subject_term_id: microbial-ecology;next-generation-sequencing;software.

55. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581–583 (2016). URL https://www.nature.com/articles/nmeth.3869. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Metagenomics;Software;Statistical methods Subject_term_id: metagenomics;software;statistical-methods.

56. Gardiner, C. W. & Gardiner, C. W. *Stochastic methods: a handbook for the natural and social sciences*. Springer series in synergetics (Springer, Berlin, 2009), 4th ed edn.

57. Otunuga, O. M. Time-dependent probability density function for general stochastic logistic population model with harvesting effort. *Physica A: Statistical Mechanics and its Applications* **573**, 125931 (2021). URL https://linkinghub.elsevier.com/retrieve/pii/S037843712100203X.

58. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011). URL http://jmlr.org/papers/v12/pedregosa11a.html.

59. Snedecor, G. W. & Cochran, W. G. *Statistical methods* (Iowa State University Press, Ames, 1989), 8th ed edn.

60. Forkman, J. Estimator and Tests for Common Coefficients of Variation in Normal Distributions. *Communications in Statistics - Theory and Methods* **38**, 233–251 (2009). URL https://doi.org/10.1080/03610920802187448. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03610920802187448.

61. Prangle, D. Adapting the ABC Distance Function. *Bayesian Analysis* **12** (2017). URL https://projecteuclid.org/journals/bayesian-analysis/volume-12/issue-1/Adapting-the-ABC-Distance-Function/10.1214/16-BA1002.full.

62. Burns, A. R. *et al.* Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal* **10**, 655–664 (2016). URL https://www.nature.com/articles/ismej2015142. Number: 3 Publisher: Nature Publishing Group.

63. Sloan, W. T., Woodcock, S., Lunn, M., Head, I. M. & Curtis, T. P. Modeling Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data. *Microbial Ecology* **53**, 443–455 (2007). URL https://doi.org/10.1007/s00248-006-9141-x.

64. Willis, A. & Bunge, J. Estimating diversity via frequency ratios. *Biometrics* **71**, 1042–1049 (2015). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12332. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.12332.

65. Levins, R. & Lewontin, R. C. *The dialectical biologist* (Harvard university press, Cambridge (Mass.) London, 1985).

66. Sanchez, A. *et al.* Directed Evolution of Microbial Communities. *Annual Review of Biophysics* **50**, 323–341 (2021). URL https://doi.org/10.1146/annurev-biophys-101220-072829. _eprint: https://doi.org/10.1146/annurev-biophys-101220-072829.

67. Chang, C., Osborne, M. L., Bajic, D. & Sanchez, A. Artificially selecting bacterial communities using propagule strategies†. *Evolution* **74**, 2392–2403 (2020). URL https://doi.org/10.1111/evo.14092.

68. Schröder, A., Persson, L. & De Roos, A. M. Direct experimental evidence for alternative stable states: a review. *Oikos* **110**, 3–19 (2005). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0030-1299.2005.13962.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0030-1299.2005.13962.x.

69. Fukami, T. & Nakajima, M. Community assembly: alternative stable states or alternative transient states? *Ecology Letters* **14**, 973–984 (2011). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2011.01663.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2011.01663.x.

70. Risken, H. *The Fokker-Planck equation: methods of solution and applications*. No. v. 18 in Springer series in synergetics (Springer-Verlag, New York, 1996), 2nd ed edn.

71. Abreu, C. I., Woltz, V. L. A., Friedman, J. & Gore, J. Microbial communities display alternative stable states in a fluctuating environment. *PLOS Computational Biology* **16**, e1007934 (2020). URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007934. Publisher: Public Library of Science.

72. Diaz-Colunga, J. *et al.* Global epistasis on fitness landscapes (2022). URL http://arxiv.org/abs/2210.03677. ArXiv:2210.03677 [q-bio].

73. Diaz-Colunga, J., Skwara, A., Vila, J. C. C., Bajic, D. & Sanchez, A. Global epistasis and the emergence of ecological function (2023). URL https://www.biorxiv.org/content/10.1101/2022.06.21.496987v2. Pages: 2022.06.21.496987 Section: New Results.

74. Skwara, A. *et al.* Learning the functional landscape of microbial communities (2023). URL https://www.biorxiv.org/content/10.1101/2023.03.24.534159v1. Pages: 2023.03.24.534159 Section: New Results.

75. Sanchez, A. *et al.* The community-function landscape of microbial consortia (2022). URL https://ecoevorxiv.org/repository/view/3721/. Publisher: EcoEvoRxiv.

76. McGill, B. J. The what, how and why of doing macroecology. *Global Ecology and Biogeography* **28**, 6–17 (2019). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12855. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.12855.

77. Warren II, R. J., Costa, J. T. & Bradford, M. A. Seeing shapes in clouds: the fallacy of deriving ecological hypotheses from statistical distributions. *Oikos* **2022**, e09315 (2022). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/oik.09315. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.09315.

78. Lennon, J. T. & Locey, K. J. Macroecology for microbiology. *Environmental Microbiology Reports* **9**, 38–40 (2017). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12512. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-2229.12512.

Figures

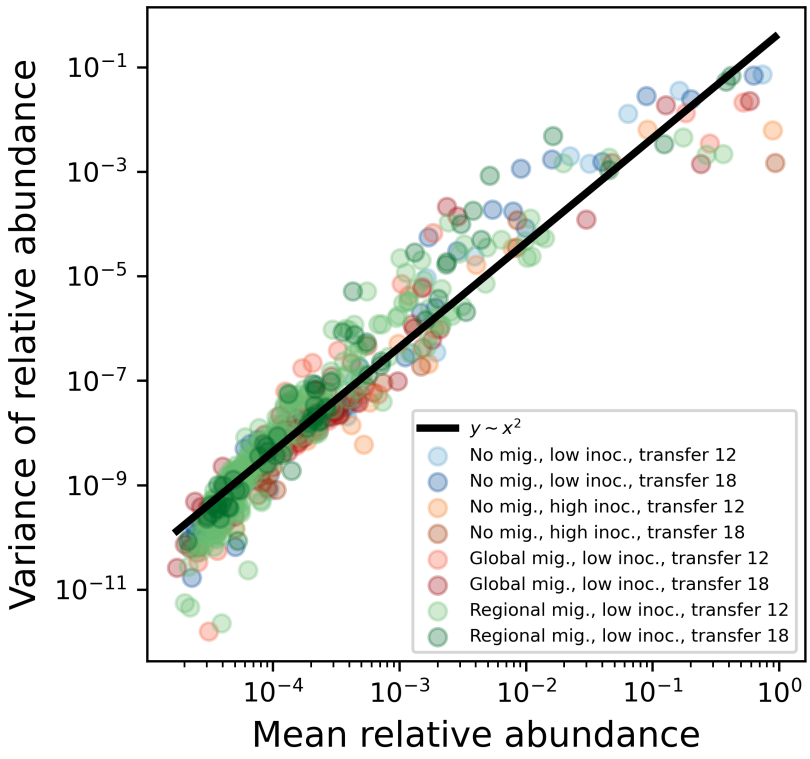

Figure 1. Emergent variation in experimental communities. Variation in abundance consistently arises across treatments and timescales in experimental communities. This variation can be captured by the relationship between the mean and variance of relative abundance, a pattern known as Taylor's Law.

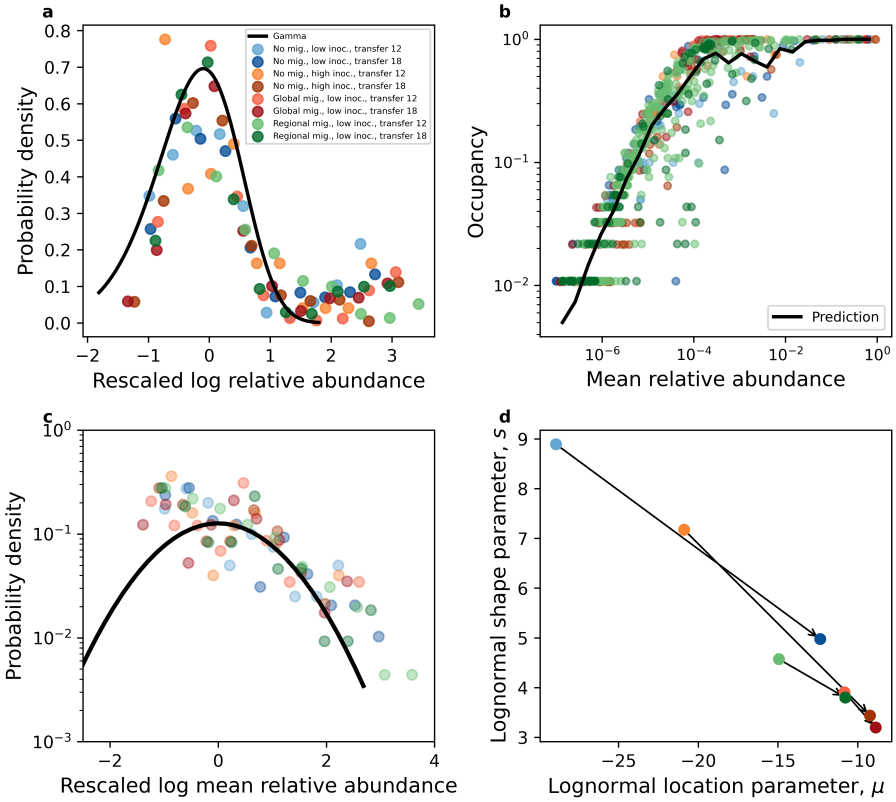

Figure 2. Macroecological laws hold in experimental communities. Empirical macroecological laws that were previously identified in natural microbial systems consistently arise in experimental communities [29]. **a)** The Abundance Fluctuation Distribution (AFD) tends to follow a gamma distribution across treatments. **b)** A gamma distribution that explicitly considers sampling successfully predicts the fraction of communities where an ASV is present (i.e., its occupancy). Measures of occupancy are primarily determined by the mean abundance of an ASV, implying that the presence or absence of an ASV in a given community with known $\bar{x}_i$ is primarily determined by sampling depth. **c)** The distributions of mean abundances are similar across treatments and largely follow a lognormal distribution and **d)** after the cessation of migration manipulations the two free parameters of the lognormal converged across treatments.

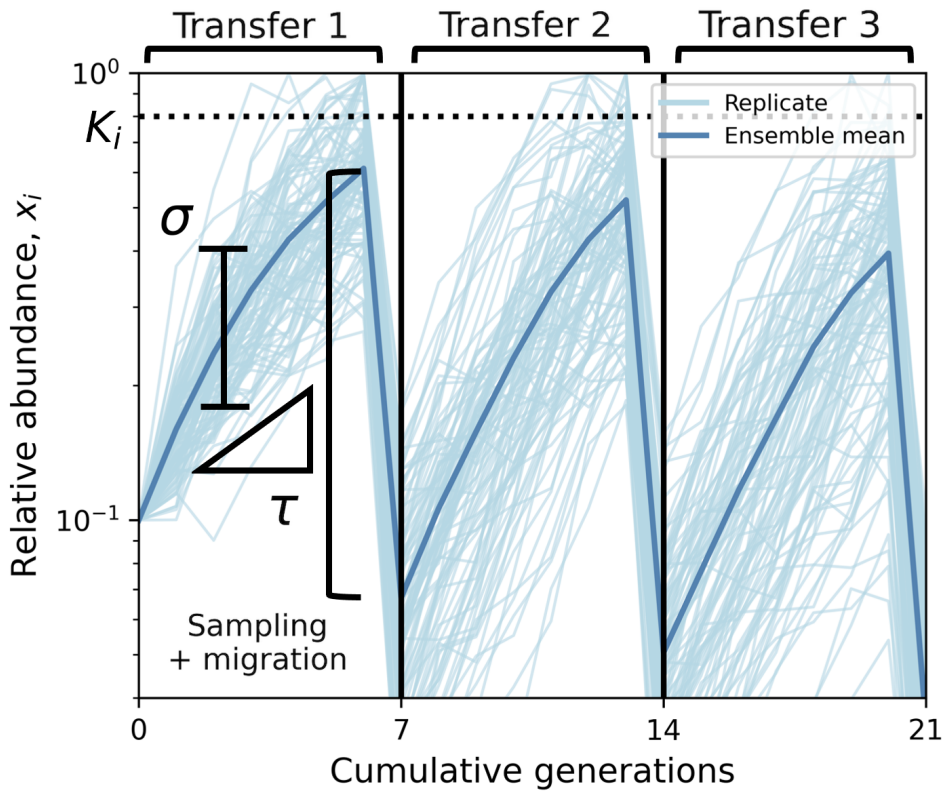

Figure 3. Incorporating experimental details into the Stochastic Logistic Model. We model the ecological dynamics within a given transfer cycle using the Stochastic Logistic Model of growth (Eq. 14). The random sampling of cells at the start of a transfer cycle does not alter the average relative abundance. However, migration alters the abundance of an ASV at the start of a transfer cycle as a perturbation (Eq. 17). This experimental detail reduces the relative abundance of a given ASV if the relative abundance in a migration inoculum was lower than the carrying capacity (K_i), where the ASV proceeds to increase abundance at a rate set by τ.

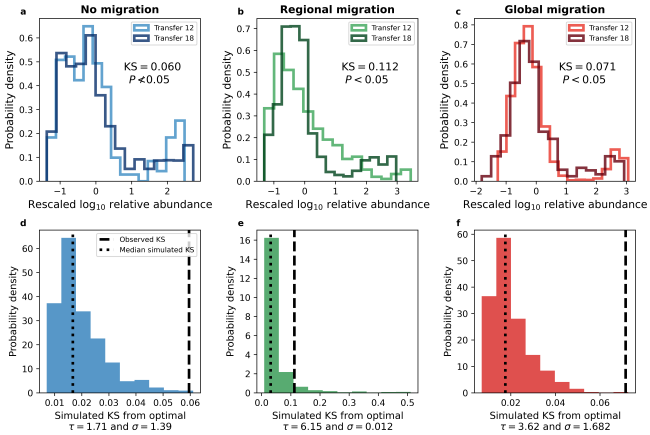

Figure 4. Migration impacts the shape of AFDs. a-c) By rescaling and merging $\log_{10}$ transformed AFDs of all ASVs for a given treatment we can examine how the shape of the AFD changes before and after the cessation of migration using a KS test. **d-f)** Using optimal parameter combinations identified by ABC, we found that the SLM predicted reasonable KS statistics (KS) for regional migration, but not for the global and no migration treatments.

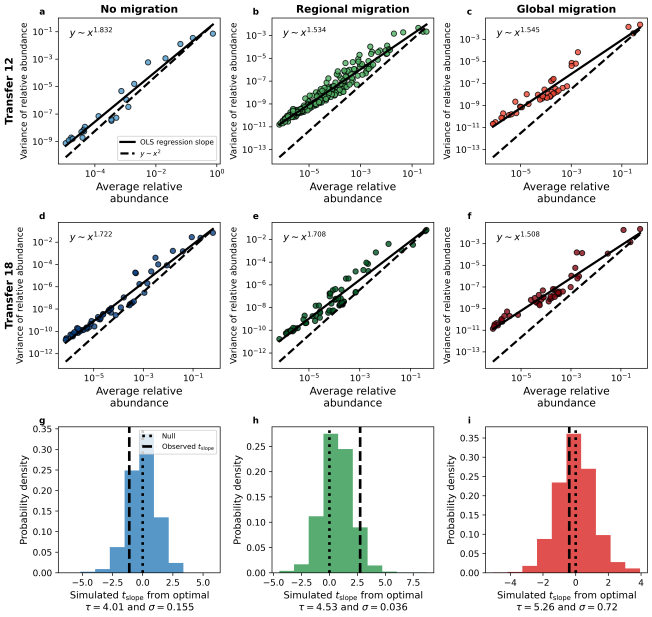

Figure 5. Regional migration impacts the slope of Taylor's Law. a-f) By examining the slope of Taylor's Law for each treatment we found that the slope only considerably changed after the cessation of migration for the regional migration treatment. **g-i)** Using our ABC estimation procedure, the SLM succeeds in predicting this increase in the slope for regional migration.

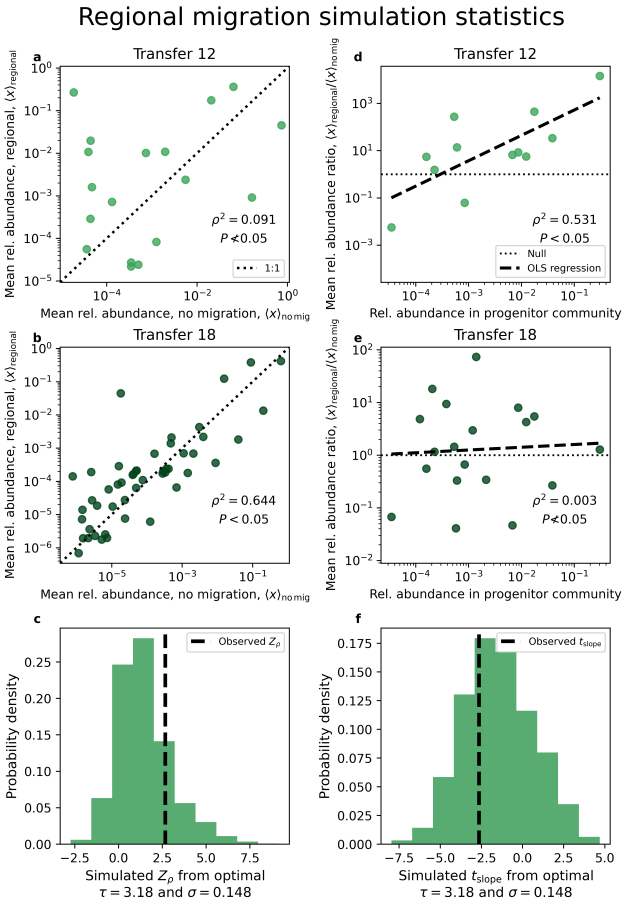

Figure 6. SLM simulations reproduce regional migration patterns. The properties of the MAD were examined to evaluate the impact of migration over time. **a)** At transfer 12, the last transfer with migration, we found no apparent correlation between the MADs of the regional and no migration treatments. **b)** Contrastingly, the strength of the correlation rapidly increased by transfer 18. The significance of this difference can be evaluated by calculating Fisher's Z-statistic, a statistic that can be applied to correlations calculated from simulated data. **c)** By performing simulations over a range of σ and τ, we can visualize how the observed value of Z_ρ compares to simulated results and identify reasonable parameter regimes **d)** as well as the relative error of our predictions. **e,f)** The ratio of the MAD between regional and no migration provides a single axis that can then be compared to the abundance of an ASV in the progenitor community. We found that there was a positive significant slope at transfer 12 that dissipated by transfer 18, reflecting the cessation of migration. **g,h)** Analogous heatmaps of the t-test for the difference in slopes can be constructed for simulated t statistics and their relative error.

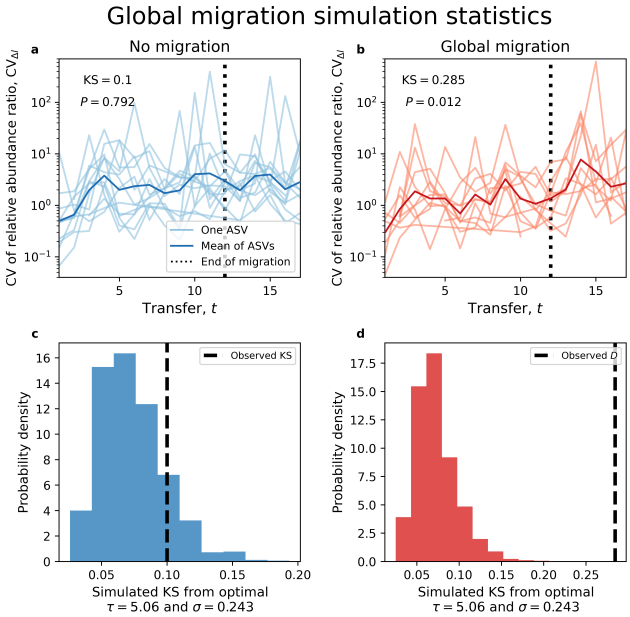

Figure 7. SLM simulations are unable to reproduce global migration patterns. By leveraging the entire timeseries we can examine how the coefficient of variation of $\Delta\ell$ changes after the cessation of migration at transfer 12 using a KS test (KS). **a)** As expected, there is no change in the distribution of $\mathrm{CV}_{\Delta\ell}$ for the no migration treatment. **b)** However, the typical $\mathrm{CV}_{\Delta\ell}$ tends to slightly increase after transfer 12 for global migration, a significant result. **c)** Our ABC estimation procedure for the SLM succeeds for the no migration treatment, but is unable to capture observed values of KS under global migration.

Tables

Quantity	Symbol	Value
Total abundance, transfer	$N^*(T)$	$\sim 10^8$
Total # of migrants, regional	N_{regional}	$7.92 \cdot 10^6$
Total # of migrants, global	N_{global}	$\#\,\text{reps.} \cdot (8 \cdot 10^{-5}) \cdot D_{\text{transfer}} \cdot N^* \approx 3.07 \cdot 10^4$
Dilution rate, transfer	D_{transfer}	$\frac{4\mu\text{L}}{500\mu\text{L}} = 0.008$
Dilution rate, global	D_{global}	$\frac{504\mu\text{L}}{60{,}000\mu\text{L}} \approx 0.0084$
Dilution rate, regional	D_{regional}	$\frac{504\mu\text{L}}{60{,}000\mu\text{L}} \approx 0.0084$

Table 1. Experimentally imposed parameters that were used in this study [44]. Estimates of total community size were previously obtained via CFU counts. The calculation of N_{global} is the result the procedure used to pool samples from replicate communities into a single global pool at the end of each transfer, which was then used to inoculate communities at the start of the subsequent transfer.