Macroecological patterns 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 Institute of Functional Biology & Genomics, CSIC - Universidad de Salamanca, 37007, Salamanca, Spain.

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

Manuscript details

- Author emails: William R. Shoemaker, `williamrshoemaker@gmail.com`; Álvaro Sánchez, `alvarosanchezda@gmail.com` ; Jacopo Grilli, `grilli.jacopo@gmail.com`
- Running title (42 characters): Macroeco. patterns in experimental communities
- Keywords: Macroecology, experimental ecology, microbial ecology, migration, metacommunities, metacommunity ecology, community ecology, microcosms, theoretical ecology.
- Article type: Perspective
- Word count: Abstract = 195, Main text = 7,683.
- 93 references.
- 7 figures, one table in the main manuscript.
- Correspondence: William R. Shoemaker, `williamrshoemaker@gmail.com`, Str. Costiera, 11, 34151 Trieste TS, Italy. +39 351 702 5501.

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.

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`. Raw data was previously generated and can be accessed through the original manuscript [1]. Processed data for this project is available on Zenodo (DOI: `10.5281/zenodo.8393848`).

Abstract

Ecology has historically benefited from the characterization of statistical patterns of biodiversity within and across communities, an approach known as macroecology. Within microbial ecology, macroecological approaches have identified universal patterns of diversity and abundance that can be captured by effective models. Experimentation has simultaneously played a crucial role, as the advent of high-replication community time-series has allowed researchers to investigate underlying ecological forces. However, there remains a gap between experiments performed in the laboratory and macroecological patterns documented in natural systems, as we do not know whether these patterns can be recapitulated in the lab and whether experimental manipulations produce macroecological effects. This work aims at bridging the gap between experimental ecology and macroecology. Using high-replication time-series, we demonstrate that microbial macroecological patterns observed in nature exist in a laboratory setting, despite controlled conditions, and can be unified under the Stochastic Logistic Model of growth (SLM). We find that demographic manipulations (e.g., migration) impact observed macroecological patterns. By modifying the SLM to incorporate said manipulations alongside experimental details (e.g., sampling), we obtain predictions that are consistent with macroecological outcomes. By combining high-replication experiments with ecological models, microbial macroecology can be viewed as a predictive discipline.

Introduction

Microbial communities inhabit virtually every environment on Earth. Through their ubiquity, abundance, and diversity, microorganisms regulate the biogeochemical processes that sustain life on the biosphere. Furthermore, host-associated microbial communities play a crucial role in maintaining the health of many macroscopic forms of life, including humans, whose health is impacted by their microbiome [2, 3]. Given their environmental, medical, and economic importance, it is necessary to develop quantitative theories of ecology that allow 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 patterns of biodiversity, an approach known as macroecology [4–10]. The macroecological approach is, fundamentally, statistical in nature, allowing for quantitative predictions and extrapolations to be made about the typical features of ecological communities without having to specify microscopic ecological forces. The generalized nature of this approach has allowed researchers to successfully characterize a diverse array of microbial ecological patterns [11–20] and spurred the development of mathematical models of microbial ecology grounded in statistical physics [21–29].

Through the macroecological approach, disparate patterns were recently unified by the observation that the typical microbial community follows three macroecological patterns: 1) the abundance of a given community member across communities follows a gamma distribution, 2) the mean abundance of a given community member is not independent of its variance (i.e., Taylor's Law [30–32]), and 3) the mean abundance of a community member across communities follows a lognormal distribution [33]. These three general patterns can be captured by an intuitive mathematical model of density-dependent growth with environmental noise, the stochastic logistic model (SLM) of growth [33–35]. 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 [36], alternative stable-states [37], correlations in abundance between community members [38], community-level measures across taxonomic and phylogenetic scales [39], and dynamics within and across human hosts at the sub-species level (i.e., strains) [19, 40]. The results of these studies demonstrate that a minimal mathematical model of ecological dynamics can capture a broad assemblage of microbial macroecological patterns.

The generality of macroecology is a useful feature that permits the construction of a working theory of microbial ecology [41, 42]. However, there is a trade-off between this generality and the ability of macroecology to provide causal connections between ecological mechanisms and observed patterns (e.g., the exponent of Taylor's Law) [43]. This lack of causality is due to the inherent difficulty of manipulating macroecological patterns in natural communities. This study addresses this issue by expanding the scope of the macroecological approach from natural systems to experimental communities, providing the means to investigate how mechanistic manipulations alter macroecological patterns.

Experimentation has played a crucial role in the documentation and manipulation of ecological forces [44, 45]. The advent of 16S rRNA amplicon sequencing allows researchers to investigate the ecological dynamics of a large number of replicate microbial communities in a laboratory setting. This controlled approach has proven to be highly successful, allowing researchers to determine the extent that the assembly of

microbial communities is reproducible and predictable [46–53]. Surprisingly, seemingly simple environments can maintain highly dissimilar microbial communities, even in systems harboring a single exchangeable resource [1]. This observation has led researchers to investigate the susceptibility of microbial communities to experimentally imposed ecological forces. A prominent example is the migration of individuals between communities, an ecological force that is amenable to experimental manipulation [54, 55] and can alter the heterogeneity of communities across replicates [1, 56, 57].

In this study we sought to bridge the gap between microbial macroecological patterns observed in nature and experiments performed in the lab. We consider experimentally assembled microbial communities exposed to qualitatively different forms of migration [1]. We first demonstrate that macroecological patterns can be recapitulated in a laboratory setting. By connecting these patterns to the predictions of the SLM, we developed a model of microbial macroecology that incorporates experimental details. Specifically, we focus on two migration treatments. The first treatment (referred to as regional migration) corresponds to a classical mainland-island scenario [58], where migrants from the community from which replicate communities were originally assembled (known as the progenitor community) continued to migrate over time. The second case, referred here as global migration, corresponds to a fully-connected metacommunity model [59], where migration occurs between communities that were assembled from the same progenitor community. We examine the ways in which the ecological force of migration can be manipulated, their resulting macroecological outcomes, and the degree that these outcomes can be predicted by the SLM. Using these results, we identify when and how the SLM succeeded in predicting the effects of experimental manipulation. By leveraging high-throughput ecological experiments alongside robust statistical patterns, we can strengthen the predictive and quantitative elements of microbial ecological theory.

Materials and methods

Experimental data

We identified an appropriate dataset to investigate the macroecology of experimental microbial communities. A recent experimental study maintained ~ 100 replicate communities assembled from a single progenitor soil community and imposed a variety of demographic treatments [1]. Here, a given microcosm was inoculated from a single progenitor community. All microcosms contained a single supplied carbon source (i.e., glucose). Each community was allowed to grow for 48 hours. A small fraction of the volume (i.e., an aliquot) was then sampled and used to inoculate a new microcosm with the same profile of supplied resources. This procedure of inoculating a microcosm with an aliquot and allowing it to grow for a given period of time is known as a transfer cycle (Fig. 1a). Transfer cycles were repeated 18 times per replicate community.

The demography of replicate communities was systematically altered by manipulating the form of migration. For the first manipulation, aliquots of the progenitor community was mixed with aliquots from the previous transfer cycle, so that the community composition of a microcosm at the start of a new transfer cycle was comprised of both the composition of the progenitor and the previous transfer cycle. This manipulation is referred to as regional migration. In the second case, migration was manipulated by sampling aliquots of each community at the end of a given transfer cycle. These aliquots were mixed and redistributed at the start of the subsequent transfer cycle. This manipulation is referred to as global migration. Migration manipulations were only performed for the first 12 dilution cycles (out of the total 18 transfer cycles). This experimental design provides a series of time points where

migration was and was not performed within each replicate community (Fig. 3a).

A number of communities were also inoculated with a large aliquot of the progenitor community. Macroecological patterns of these communities were examined but were excluded from subsequent analyses since this manipulation only occurred at the first transfer cycle and did not induce lasting macroecological effects. Community data for this experiment was obtained via 16S rRNA sequencing. We reprocessed all raw FASTQ data from the original study to obtain Amplicon Sequence Variants (ASVs) using the package DADA2 [60]. Table S1 reports the number of replicate communities sampled at each transfer. S1 Text summarizes technical details about the experiment, which can can also be found in the original manuscript [1].

The macroecological predictions of the stochastic logistic model

Populations at low abundance often grow at an initially high rate that proceeds to decrease as they approach the maximum abundance that an environment can support. This is the central idea underlying logistic growth and it can be captured by two parameters: 1) the minimal length of time required to reproduce (inverse of the maximum growth rate), τ and 2) the carrying capacity of an environment, K. This simple picture assumes a fully deterministic environment. Environmental stochasticity can be added as fluctuations in the growth rate, where the strength of environmental noise is controlled by the parameter σ, representing the coefficient of variation of growth rate fluctuations. The Stochastic Logistic Model (SLM) captures this two fundamental ecological processes under the form of a Langevin equation [61,62]

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

where x_i is the relative abundance of ASV i. The term $\eta_i(t)$ introduces stochasticity into the equation. It is defines as white noise: the expected value of $\eta(t)$ is $\langle \eta(t) \rangle = 0$ and the time correlation is a delta function $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$ (implying that the noise at time t is uncorrelated with the noise at time t'). This model can be extended to time-correlated environmental fluctuations [33].

The SLM reproduces the three empirical macroecological patterns that are found in microbial communities across natural environments: 1) gamma distributed abundances, 2) a power law relationship between mean relative abundance and the variance of relative abundance (i.e., Taylor's Law), and 3) a lognormally distributed distribution of mean relative abundances [33]. To provide the necessary context for the study, we briefly summarize these three patterns and their connection to the SLM.

The Abundance Fluctuation Distribution

The Langevin equation defined in Eq. 1) defines a stochastic trajectory $x_i(t)$. The probability $P(x_i, t|x_i(0), 0)$ that the abundance of species i equals $x_i(t)$ at time t when starting from initial condition $x(0)$ at time $t = 0$ can be obtained from the Fokker-Planck equation corresponding to Eq. 1 [61]. For large times, this probability converges to the stationary probability distribution $P^*(x_i)$, which in the case of SLM takes the form of a Gamma distribution

$$P^*(x_i|\bar{x}_i, \beta_i) = \frac{1}{\Gamma(\beta_i)} \left(\frac{\beta_i}{\bar{x}_i}\right)^{\beta_i} \exp\left[-x_i \frac{\beta_i}{\bar{x}_i}\right] x_i^{\beta_i - 1} \tag{2}$$

where the two parameters $\bar{x}_i$ and β_i can be expressed in terms of the statistical moments of the distribution and also in terms of parameters appearing in Eq. 1

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

$$\frac{\mathrm{Var}(x)}{\langle x_i \rangle^2} = \frac{\sigma_i}{2 - \sigma_i} = \frac{1}{\beta_i} \tag{3b}$$

The expression above hold for absolute abundances, while we have access to finite number of reads obtained by sequencing, which are affected by both compositionality and sampling effects. We can model our sampling process as the Poisson limit of a multinomial distribution, obtaining a form of the AFD that explicitly accounts for the effect of sampling [33]. Using this distribution, we can calculate the probability of obtaining n reads out of a total sampling depth N for the ith ASV as

$$P(n_i|\bar{x}_i, \beta_i, N) = \frac{\Gamma(\beta_i + n_i)}{n_i!\Gamma(\beta_i)} \left(\frac{\bar{x}_i N}{\beta_i + \bar{x}_i N}\right)^{n_i} \left(\frac{\beta_i}{\beta_i + \bar{x}_i N}\right)^{\beta_i} . \tag{4}$$

Often in ecology we are interested in the relationship between the mean relative abundance of a community member and the fraction of communities where it is present (i.e., its occupancy) [14, 63, 64]. We can derive a prediction of occupancy using the sampling form of the AFD by setting $n = 0$ and noticing that probability that an ASV is present is the complement of its absence. Averaging over M communities, one obtains a prediction of occupancy

$$\langle o_i \rangle = 1 - \frac{1}{M} \sum_{m=1}^{M} P(0|\bar{x}_i, \beta_i, N_m) . \tag{5}$$

Prediction error was estimated as described in Text S2.

Taylor's Law

Taylor's Law describes the empirical relationship between the mean and variance of the relative abundance [30, 31] calculated across ASVs

$$\mathrm{Var}(x_i) \propto \langle x_i \rangle^{2b} . \tag{6}$$

By comparing this expression to eq. 3 one can see that the case where β_i^{-1} is independent of $\bar{x}_i$ corresponds to $b = 1$. In the context of the SLM, this choice implies that the strength of environmental noise is constant for all ASVs ($\sigma_i = \sigma$). Estimates of b were obtained by log transforming both axes and performing ordinary least squares regression with `SciPy v1.4.1`.

The Lognormally distributed Mean Abundance Distribution

In observed data, the mean relative abundances (calculated either across communities or over time) of ASVs, known as the Mean Abundance Distribution (MAD), frequently follows a lognormal probability distribution

$$P(\bar{x}_i)^{\mathrm{lognorm}} = \frac{1}{\sqrt{2\pi s^2}\bar{x}_i} \exp\left[-\frac{(\log \bar{x}_i - \mu)}{2s^2}\right] \tag{7}$$

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

$$P(\bar{x}_i)^{\text{lognorm}}_{\text{emp}} = \frac{\theta(\bar{x}_i - c)P(\bar{x}_i)^{\text{lognorm}}}{\int \theta(z-c)P(z)^{\text{lognorm}}dz} \tag{8}$$

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 [33]).

Incorporating migration into the SLM

The original form of the SLM is not directly applicable to microbial community assembly experiments. As discussed above, such experiments are performed as transfer cycles, where growth occurs over a period of time within a microcosm before an aliquot is transferred to an environment with replenished resources. Phenomenological models of logistic growth are appropriate for modeling dynamics within a single transfer cycle under certain conditions (S3 Text). In addition, the time-dependent form of Eq. 9 has been derived (i.e., $P(x_i(t), t|x_i(0), 0)$ and could be used to model the temporal dynamics within a given transfer cycle [65] (description of solution in S5 Text). However, no attempt has been made to incorporate experimental details into the SLM so that it can be used to obtain predictions for the macroecological effects of experimental manipulations.

First, we extended the SLM to incorporate the transfer cycle process. We consider a piece-wise form of the SLM to describe the dynamics within the kth transfer cycle

$$\frac{dx_i^{(k)}}{dt} = \frac{x_i^{(k)}}{\tau_i}\left(1 - \frac{x_i^{(k)}}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}}x_i^{(k)} \cdot \eta_i(t) \tag{9}$$

where the dynamics start at $t = 0$ and continue until time $t = T$. At time T an aliquot of the community is sampled. The volume of the aliquot relative to the total volume of the microcosm is known as the transfer dilution rate D_{transfer}. If the total abundance of the community is known at time T, then the total abundance at the start of the next transfer cycle can be calculated as $N^{(k)}(0) = D_{\text{transfer}} \cdot N^{(k-1)}(T)$. The process of sampling community members at the end of a transfer cycle can be modeled as multinomial sampling process (Text S6). With this sampling process and our piece-wise form of the SLM, we have a minimal dynamical model of a microbial community assembly experiment.

While they represent different forms of migration, both treatments are similar in that they are implemented by altering the initial abundance of a community member at the start of a transfer cycle. We can formulate the number of cells sampled due to migration at the start of a transfer cycle for a given ASV as follows

$$n_{i,\text{regional}}(0)^{(k)} \propto x_{i,\text{progenitor}} \tag{10a}$$

$$n_{i,\text{global}}(0)^{(k)} \propto \frac{\sum_{m=1}^{M} n_{i,m}^{(k-1)}(T)}{\sum_{i=1}^{S}\sum_{m=1}^{M} n_{i,m}^{(k-1)}(T)} \tag{10b}$$

The sampling process of migrants can again be modeled using a multinomial distribution (S6 Text). We can now examine how the mean initial *relative* abundance of a given ASV depends on migration. In this experiment the total abundance of a community at the end of a transfer cycle did not considerably vary from transfer to transfer, meaning that we can assume that the total abundance at time T within a transfer cycle remained the same for all transfer cycles k,

$(N^{(k-1)}(T) \approx N^{(k)}(T) \equiv N^*(T))$. Using this result, we divide the abundance of each ASV by the total abundance in a community to obtain relative abundances $(x_i^{(k-1)}(T) = \frac{n_i^{(k-1)}(T)}{N^*(T)})$, obtaining the following prediction for the mean abundance at the start of a transfer cycle

$$\left\langle x_i^{(k)}(0) \right\rangle = \begin{cases} \left\langle \frac{D_{\text{transfer}} x_i^{(k-1)}(T) N^*(T) + D_{\text{regional}} x_{i,\text{regional}} N_{\text{regional}}}{D_{\text{transfer}} N^*(T) + D_{\text{regional}} N_{\text{regional}}} \right\rangle, & \text{Regional} \\ \left\langle \frac{D_{\text{transfer}} x_i^{(k-1)}(T) N^*(T) + D_{\text{global}} x_{i,\text{global}}^{(k-1)} N_{\text{global}}}{D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}}} \right\rangle, & \text{Global} \\ \left\langle x_i^{(k-1)}(T) \right\rangle, & \text{No migration} \end{cases} \tag{11}$$

The impact that migration manipulations have on the initial relative abundance of the AFD and how they compare to the typical view of migration as a process that occurs at a constant rate is illustrated in Fig. S3. Parameter values are calculated using experimental details and can be found in Table 1 and simulation details of Eq. 9 can be found in S6 Text.

Treatment-specific migration statistics

We identified appropriate statistics to capture macroecological outcomes specific to the regional and global migration treatments. Significance was established in all instances using null distributions of statistics obtained via permutation.

Regional migration statistics

We first predicted that the mean relative abundance of an ASV in the assembled communities ($\langle x_i(t) \rangle$) should be affected by its relative abundance in the progenitor community ($x_{i,\text{progenitor}}$) due to a mismatch in abundances. This prediction is justified, as correlation was weak between these two quantities (Fig. S4). Based on this observation, we sought to identify 1) how the correlation in $\langle x_i(t) \rangle$ between no migration and migration treatments increased after the cessation of migration and 2) how the dependency of $\langle x_i(t) \rangle$ on $x_{i,\text{progenitor}}$ dissipated once migration manipulations ceased. The change in correlation coefficients (ρ) between transfers 12 and 18 was assessed using Fisher's Z statistic [66]

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

where $z_t = \log\left[\frac{1+\rho_t}{1-\rho_t}\right]$ and M_k is the number of replicate communities at transfer k. The denominator represents the standard error of the numerator. Dependency on the progenitor was evaluated by fitting a regression between $x_{i,\text{progenitor}}$ and the ratio of mean abundances in the regional and no migration treatments $\frac{\langle x_i \rangle_{\text{regional}}}{\langle x_i \rangle_{\text{no mig}}}$. Regression fits were obtained at transfers 12 and 18 and the statistical significance of the change in slope was evaluated using a permutation-based t-test.

Global migration statistics

For the global migration treatment we predicted that fluctuations in abundance would decrease under migration among assembled communities. This prediction can be justified based on experimental details and the properties of statistical moments of distributions. To perform the global migration treatment, aliquots were taken of each

replicate community on the $k-1$th transfer cycle after T hours, pooled together and intermixed, then redistributed among the same communities at the start of the kth transfer cycle (Fig. 3). Ignoring fluctuations driven by sampling, the relative abundance of an ASV in the intermixed pool before it is redistributed will be $x_{i,\text{global}}^{(k-1)} \equiv \frac{1}{M}\sum_m^M x_{i,m}^{(k-1)}(T)$, where the M^{-1} prefactor accounts for an equal volume of each aliquot having been sampled. From this definition, we obtain the expected relative abundance in the global migration pool

$$\left\langle x_{i,\text{global}}^{(k-1)} \right\rangle = \left\langle \frac{1}{M}\sum_m^M x_i^{(k-1)}(T) \right\rangle \tag{13a}$$

$$= \left\langle x_i^{(k-1)}(T) \right\rangle \tag{13b}$$

and the variance of relative abundance in the global migration pool

$$\text{Var}\left(x_{i,\text{global}}^{(k-1)}\right) = \text{Var}\left(\frac{1}{M}\sum_m^M x_i^{(k-1)}(T)\right) \tag{14a}$$

$$= \frac{1}{M^2}\sum_m^M \text{Var}\left(x_i^{(k-1)}(T)\right) \tag{14b}$$

$$= \frac{1}{M}\text{Var}\left(x_i^{(k-1)}(T)\right) \tag{14c}$$

When an aliquot of the intermixed global migration pool is sampled and added to a community at the start of transfer cycle k, we find that the mean relative abundance at the start of a transfer cycle remains unchanged

$$\left\langle x_i^{(k)}(0) \right\rangle = \left\langle \frac{D_{\text{transfer}} x_i^{(k-1)}(T) N^*(T) + D_{\text{global}} x_{i,\text{global}}^{(k-1)} N_{\text{global}}}{D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}}} \right\rangle \tag{15a}$$

$$= \left\langle x_i^{(k-1)}(T) \right\rangle \frac{D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}}}{D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}}} \tag{15b}$$

$$= \left\langle x_i^{(k-1)}(T) \right\rangle \tag{15c}$$

In contrast, when deriving the variance as fluctuations around the expected relative abundance, we obtain the following inequality

$$\text{Var}\left(x_i^{(k)}(0)\right) = \frac{(D_{\text{transfer}} N^*(T))^2 \text{Var}\left(x_i^{(k-1)}(T)\right) + (D_{\text{global}} N_{\text{global}})^2 \text{Var}\left(x_{i,\text{global}}^{(k-1)}\right)}{(D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}})^2} \tag{16a}$$

$$= \frac{(D_{\text{transfer}} N^*(T))^2 \text{Var}\left(x_i^{(k-1)}(T)\right) + (D_{\text{global}} N_{\text{global}})^2 \frac{1}{M} \text{Var}\left(x_i^{(k-1)}(T)\right)}{(D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}})^2} \tag{16b}$$

$$= \text{Var}\left(x_i^{(k-1)}(T)\right) \frac{(D_{\text{transfer}} N^*(T))^2 + (D_{\text{global}} N_{\text{global}})^2 \frac{1}{M}}{(D_{\text{transfer}} N^*(T) + D_{\text{global}} N_{\text{global}})^2} \tag{16c}$$

$$< \text{Var}\left(x_i^{(k-1)}(T)\right) \tag{16d}$$

Therefore, under the SLM we expect global migration to alter the CV of relative abundances while leaving the mean unchanged.

We performed two types of analyses to test our fluctuation predictions: examining whether 1) the correlation in CV_{x_i} between no migration and global migration treatments and 2) the CV of the log-ratio of relative abundances between consecutive transfers increased after the cessation of migration. The log-ratio of relative abundances between two timepoints is defined as $\Delta\ell_i^{(k)} = \log\left[x_i^{(k+1)}/x_i^k\right]$. We elected to use $\Delta\ell_i^{(k)}$ because 1) it can be interpreted as a discretized form of the per-capita growth rate and 2) as a ratio it cancels out any potential multiplicative time-independent sample biases [67,68]. We first calculated the CV of $\Delta\ell_i^{(k)}$ across communities at each transfer for each ASV, $\mathrm{CV}^{(k)}_{\Delta\ell_i}$. To contrast this measure of fluctuations *over replicates* with a measure of fluctuations over time *within a replicate*, we calculated the CV before and after the cessation of migration *for each replicate*, $\mathrm{CV}^{<}_{\Delta\ell_{i,m}}$ and $\mathrm{CV}^{>}_{\Delta\ell_{i,m}}$. The effect of the cessation of migration on a per-replicate basis was examined using an F statistic designed to compare two CVs [69]

$$F^{(i,m)}_{\mathrm{CV}} = \frac{(\mathrm{CV}^{>}_{\Delta\ell_{i,m}})^2}{1+(\mathrm{CV}^{>}_{\Delta\ell_{i,m}})^2\frac{M^{>}-1}{M^{>}}} \cdot \frac{1+(\mathrm{CV}^{<}_{\Delta\ell_{i,m}})^2\frac{M^{<}-1}{M^{<}}}{(\mathrm{CV}^{<}_{\Delta\ell_{i,m}})^2} \tag{17}$$

where $M^{<}$ and $M^{>}$ represent the number of transfers before and after the cessation of migration, respectively.

Evaluating SLM predictions

Using our experiment-informed simulation, we determined whether shifts in macroecological quantitities due to migration could be explained by the SLM (S6 Text). To briefly summarize, we first simulated the experiment for 10^4 uniformly drawn sets of $\{\tau, \sigma\}$. We used Approximate Bayesian Computation (ABC) to identify the parameters where the simulated macroecological pattern had the lowest Euclidean distance with the observed pattern. We then used this optimal set of parameters to generate distributions of the macroecological pattern or summary statistic using 10^3 SLM simulations. To evaluate whether migration was capable of altering a pattern under the SLM, we repeated our simulations across a grid of $\{\tau, \sigma\}$ combinations.

Results

Using a high-replication community assembly experiment, we first examine whether universal macroecological patterns observed in natural communities could be recapitulated in a laboratory setting. We then examine whether the patterns we observed were susceptible to experimental manipulation. We rationalize our observations by extending the SLM to incorporate experimental details, allowing us to balance the effectiveness of the SLM as a minimal model of ecology with experimental realism. Using simulations, we obtain quantitative predictions and identify whether the macroecological consequences of a given form of migration could be captured by the SLM, establishing its capacity to predict patterns of microbial biodiversity.

Macroecological patterns emerge in experimental systems

Predicting the outcomes of experimental manipulation is a vital goal of microbial ecology. With this goal in mind, we first determined the degree to which empirical macroecological patterns documented in observational data held in an experimental system (Figure 1a). The number of potential patterns one can examine is potentially

very large. From previous works on natural microbial communities, we know that three main patters are sufficient to recapitulate several others [33]: 1) the mean and variance of species abundances displaying non-independence (Taylor's Law), 2) the Abundance Fluctuation Distribution across independent sites (AFD) following a gamma distribution, and 3) the Mean Abundance Distribution across independent sites (MAD) following a lognormal distribution. Other commonly studied macroecological patterns (e.g., the Species Abundance Distribution, abundance-occupancy relationship), can be obtained and predicted as a consequence of these three patterns. Furthermore, these three patterns can be rationalized and predicted by a mean-field model, the Stochastic Logistic Model of growth [33].

We first quantified the variation in abundance that emerged in experimental communities. The average relative abundance varied over four orders of magnitude, meaning that a considerable degree of community-level variation can be maintained in a laboratory setting. We found that the shape of the relationship between the mean and variance of relative abundance across ASVs did not qualitatively vary across treatments, implying that Taylor's Law can be applied [30]. Figure 1b shows that the variance of relative abundance was proportional to the square of the mean, corresponding to Taylor's Law with an exponent equal to two. This value of the exponent implies that the coefficient of variation of relative abundances (CV) remained constant across ASVs. Fitting the exponent to each treatment for each transfer, we found a mean exponent of 2.1 ± 0.06, suggesting that despite the variation in typical abundance, the CV of relative abundances remained roughly constant across ASVs.

Given the existence of substantial fluctuations of abundance across replicate communities, we focused on the full distribution of abundances, known as the Abundance Fluctuation Distribution (AFD). To facilitate comparisons across ASVs and treatments, we rescaled the logarithm of the AFD for each ASV by its mean and variance (i.e., the standard score). We found that rescaled AFDs from different treatments tended to collapse on a single curve, implying that despite differences in experimental details, the general shape of the AFD remained invariant. Figure 2a shows that the bulk of the distributions generally followed the gamma distribution as observed in empirical data and predicted by the SLM (Eq. 2).

Abundance fluctuations offer only a partial view of community variation since patterns of presence/absence of community members could display non-trivial patterns. We therefore studied the fraction of communities where a given ASV was present, a measure known as occupancy. Given that the process of sampling ASVs within a community can be modeled as a Poisson process, the distribution of read counts can be derived from a gamma AFD which can be used to obtain occupancy predictions (Eq. 5). We find that the predictions of the SLM generally hold across treatments, with slight deviations at high values of observed occupancies (Fig. S1a). This trend is reflected in the distribution of relative errors in our occupancy predictions, where certain treatments appeared to have higher error values than others (Fig. S1b). However, upon examining ASVs that were present in global, regional, and no migration communities, we found that the error was not affected by migration (Fig. S1c). A permutation-based test was performed to establish the statistical significance of this observation. We found a significant, if slight, effect of migration, where it reduced the error of our predictions for both regional ($\overline{\Delta\varepsilon}_{\text{regional}} = -0.24$, $P = 0.001$) and global migration ($\overline{\Delta\varepsilon}_{\text{global}} = -0.2$, $P = 0.01$). Regardless, Eq. 5 predicts the occupancy of the typical ASV across treatments with a high level of accuracy.

Leveraging this result, we investigate the relationship between the mean relative abundance across replicates and the occupancy of an ASV, more commonly known as the abundance-occupancy relationship [14, 63, 64]. Fig. 2b shows that all treatments follow the predictions of the gamma distribution. The small deviations from the

prediction are likely driven by a combination of 1) averaging over a finite number of ASVs and 2) variation of β_i across ASVs [37]. The existence of a relationship between average abundance and occupancy in experimental communities is particularly striking, as it implies that the probability of observing a given community member is primarily determined by sampling effort (i.e., # reads) and mean relative abundance, despite differences in experimental details.

The success of the AFD in predicting occupancy and the observation that Taylor's Law holds implies that differences in abundances are primarily driven by differences in their mean relative abundance. The distribution of the mean relative abundance across taxa (known as Mean Abundance Distribution, MAD) in observational data is well described by a lognormal distribution ([33]; Eq. 8). This observation can be seen as an across-community extension of the observation that the distribution of abundances within a single microbial community can be captured by a lognormal distribution [12]. The low richness of experimental communities (~ 10 ASVs per community) makes it difficult to study the MAD of a given treatment, requiring ASVs to be pooled across treatments and transfers. Given the low sampling level, the resulting empirical MAD can be captured by a lognormal (Fig. 2c).

It is important to note 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.

In order to assess the effect of migration on the average abundances, we compare the MAD between transfer 12 (end of migration treatment) and transfer 18 (the last transfer in the experiment). Fig. 2d shows that the two MADs have a similar shape, suggesting that the MAD as a macroecological pattern consistently converged to a similar form, irrespective of the migration treatment.

Testing the macroecological effects of migration

Our analysis shows that demographic manipulations of experimental communities do not induce *qualitative* changes in macroecological patterns (e.g., altering the form of Taylor's Law). This result is consistent with the observation that natural microbial communities display similar patterns [22, 33]. However, the *quantitative* effects of these experimental manipulations is unclear. To determine whether these effects exist and the degree that they can be predicted, we examined deviations in macroecological quantities, incorporated experiment-specific forms of migration into the SLM, and tested their predictive capacity.

It is clear from the previous section that the SLM is a useful descriptive model of experimental microbial communities. Therefore, it is reasonable to expect that a form of the SLM that incorporates migration could serve as a predictive model. While a constant rate of migration appears intuitive and can be incorporated into the SLM (S4 Text), this model does not reflect the details of the experiment. Rather, in this experiment migrants only entered communities at the start of a given transfer cycle (Fig. 3a). This detail corresponds to a model where the effect of migration can be modeled as an experimentally-induced perturbation on the initial abundance of an ASV within the kth transfer cycle ($n_i^{(k)}(0)$; Materials and Methods, S6 Text). The full time-dependent solution of the SLM which depends on initial conditions has previously been solved [65], allowing one to model the temporal evolution of the AFD in response to experimental perturbations that alter initial abundances and compare their effects to AFDs with a constant rate of migration (S5 Text; Fig. S2).

While a time-dependent SLM appears appropriate, additional decisions must be

made about how parameters are determined and the number of variables (i.e., community richness). Using rarefaction curves and previously established richness estimation procedures [70], we found that the richness of the progenitor community was ~ 100 fold greater than the richness of a typical assembled community (Fig. S3). Migration has little effect on richness, as estimates are only slightly higher among communities that experienced migration while remaining orders of magnitude lower than the richness of the progenitor community. This result suggests that the carrying capacity of a given ASV played a primary role in determining its survival, raising the question of how to 1) specify the carrying capacity of an ASV in the assembled communities (K_i) and 2) determine how K_i relates to the abundance of an ASV in the progenitor community. The observation that Taylor's Law holds in the no migration treatments implies that the coefficient of variation of growth fluctuations was constant across ASVs (i.e., $\sigma_i \approx \sigma$), meaning that the mean relative abundance of an ASV across communities is proportional to its carrying capacity ($\bar{x}_i \propto K_i$). Given that $\bar{x}_i$ follows a lognormal distribution, K_i should also follow a lognormal.

The weak correlations between the mean relative abundance after the cessation of migration ($\bar{x}_i(t = 18)$) and the relative abundance observed in the progenitor ($x_{i,\text{progenitor}}$) for all treatments supports the view that the K_i of a given ASV in the assembled communities is independent of its abundance in the progenitor community (Fig. S4). However, this observation suffers from survivorship bias, since we can only compare values of $\bar{x}_i$ and $x_{i,\text{progenitor}}$ for ASVs that were actually present in both the progenitor and assembled communities (i.e., ASVs with $K_i > 0$). We found that distributions of $x_{i,\text{progenitor}}$ differed depending on whether a given ASV was present in the assembled communities, with ASVs that were present in assembled communities having a higher relative abundance in the progenitor (Fig. S5a). A permutational Kolmogorov–Smirnov test found that this shift was significant, a result that held for all experimental treatments (Fig. S6). We modeled the dependency between the occupancy of an ASV is non-zero and its abundance in the progenitor using logistic regression (Fig. S5b). This statistical dependence between the progenitor and assembled communities, along with details related to sampling (i.e., number of reads, Fig. S7), were incorporated into our numerical simulations of the SLM (Materials and Methods, S6 Text). Information about the statistical inference of the two free SLM parameters can be found in S7 Text.

Experiment-agnostic macroecological patterns

We first examined the quantitative effect of migration using macroecological patterns that have been unified by the SLM: the AFD and Taylor's Law. Fig. 4a-c shows that the AFDs of ASVs rescaled by their mean and variance differ between transfers 12 (migration present) and 18 (migration absent). This shift is particularly notable in the regional migration treatment, whereas the global migration treatment AFDs appear to be the most similar.

To examine the shift in the AFDs we calculated the KS distance between AFDs from transfer 18 and 12 for each ASV, focusing on ASVs present in every replicate community, and then calculated the mean KS over ASVs. Using Approximate Bayesian Computation (ABC), we identified the $\{\tau, \sigma\}$ pair that best explained the observed mean KS statistic (Materials and Methods; Text S6). We then used the optimal set of parameters to generate distributions of mean KS using 10^3 SLM simulations. This significant discrepancy can be explained by the existence of multiple attractors, a scenario where certain ASVs had high abundances in a subset of communities and low abundances in others, a qualitative deviation from the SLM [1]. Focusing on no migration communities belonging to the major attractor ($\sim 70\%$ of communities, Table S2), we found that the SLM could predict the observed mean KS (insert of Fig. 4d).

The effect of regional migration was also captured by the SLM (Fig. 4e), with the observed statistic for global migration communities lying slightly outside predicted range (Fig. 4f). However, manipulating the SLM parameters revealed that the AFD exhibited the largest systematic deviation via migration for the regional migration treatment (Fig. S8). This result suggests that the inability of the SLM to fully capture the change in AFDs for the global migration treatment was due to its uninformative nature for this particular form of migration.

We then analyzed how migration impacted the exponent of Taylor's Law (Fig. 5a-f). Under regional micgration, the exponent is lower than that of the no migration at transfer 12 (Fig. 5a,b), but approaches the result of the no migration treatment by transfer 18 (Fig. 5d,e). Similarly, the intercept of regional migration treatment is initially higher but then approaches the value of the no migration treatment at treatment 18, as seen by the range of values on the y-axis. In contrast, there is little change in the values of the exponent or intercept for the global migration treatment. In line with our predictions, there is no significant change in the exponent ($t_{\text{exponent}} = -1.1$, $P = 0.2$) or the intercept ($t_{\text{intercept}} = -0.6$, $P = 0.6$) between transfers 12 and 18 for the communities that did not undergo migration. We found that the global migration did not alter the exponent ($t_{\text{exponent}} = -0.4$, $P = 0.7$) or the intercept ($t_{\text{intercept}} = -0.4$, $P = 0.7$). Contrastingly, regional migration significantly altered both the exponent ($t_{\text{exponent}} = 3.3$, $P = 0.02$) and the intercept ($t_{\text{intercept}} = 2.75$, $P = 0.03$).

By repeating the same ABC procedure outlined for the AFD we found that the distribution of exponents generated for an optimal set of parameters could predict the observed values of t_{exponent} across all migration treatments (Fig. 5d-f). However, by examining t_{exponent} 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). Our results demonstrate that the AFD and Taylor's Law were primarily informative of the effects of regional (island-mainland) migration.

Experiment-specific macroecological patterns

The macroecological patterns we have examined up to this point can be explained by the SLM, though the AFD was not substantially altered by migration. It is useful to supplement our analyses by considering novel macroecological patterns that are likely to be altered by migration. Given the details of the experiment, expected regional migration to alter typical abundance, whereas global migration would alter fluctuations in abundance (Materials and Methods).

Regional migration

Given the difference in abundance between assembled communities and the progenitor, we predicted that regional migration would alter the mean abundance of an ASV. We examined paired MADs for the regional and no migration treatments *before* ($k = 12$) and *after* ($k = 18$) the cessation of migration. The correlation between MADs was initially low and non-significant at transfer 12 but had significantly increased by transfer 18 (Fig. 6a,b; $Z_\rho = 2.7$, $P = 0.01$) [66]. Our SLM simulations recapitulate this observation, as we are able to predict the observed value for the optimal parameter combination of σ and τ from ABC (Fig. 6c) and over a specific range of parameter combinations (Fig. S11a,b). These results are consistent with the interpretation that ASVs revert to their previous typical abundance (captured by the carrying capacity) once island-mainland migration ceased.

We then sought to determine whether the effect of regional migration depended on the abundance of an ASV in the progenitor community. To examine this dependence,

we studied the relationship between 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 its abundance in the progenitor community. We found that at transfer 12 there was a strong relationship between the two quantities (statistically significant via permutation). By transfer 18 the slope is statistically indistiguishable from zero (Fig. 6d,e). The observed change in slopes is significant ($t = -2.6$, $P = 0.03$) and can be reproduced using the SLM for certain σ, τ parameter combinations. These combinations overlapped with those that reproduced observed estimates of Z_ρ (Fig. 6f; Fig. S11c,d), providing further validation of the SLM.

Global migration

Accourding to the prediction of the model, global migration would strictly alter fluctuations in ASV abundance across replicate communities while leaving the MAD unchanged (Materials and Methods). Specifically, the correlation in the MAD between global and no migration treatments should remain unchanged before and after the cessation of migration manipulations. This prediction holds, as there is no significant change in the correlation between transfers 12 and 18 (Fig. S12; $Z_\rho = 0.2$, $P = 0.4$). However, there was also no significant increase in the strength of correlation for the distribution of CVs ($Z_\rho = 0.3$, $P = 0.6$), meaning that the cessation of migration did not considerably alter fluctuations in abundance.

It is possible that two time points are insufficient to detect the impact of global migration. Detecting changes in fluctuations often requires more observations than detecting changes in typical values. As a solution, we leverage the higher sampling resolution of the global migration treatment to examine the change in abundance between time points ($\Delta\ell$), as a subset of global migration communities were sequenced at each of the 18 transfers (Table S1). The mean change in abundance across replicate communities at transfer k ($\left\langle \Delta\ell^{(k)} \right\rangle$) tends to relax towards a value of zero around the sixth transfer and remained there throughout the rest of the experiment (Fig. S14). This trend indicates that the abundances of ASVs reached stationarity with respect to the start of the experiment by transfer 6, allowing us to examine equal intervals of transfer cycles before ($7 - 12$) and after ($13 - 18$) the cessation of migration.

Using permutational t-tests we determined whether $\left\langle \Delta\ell^{(k)} \right\rangle$ was altered after the cessation of migration, controlling for ASV identity. There was no evidence that $\left\langle \Delta\ell^{(k)} \right\rangle$ changes in either the no migration treatment ($\bar{t} = 0.1$, $P = 0.3$) nor in the global migration treatment ($\bar{t} = 0.09$, $P = 0.1$). This result is consistent with that prediction that global migration would not alter typical abundances.

Turning to fluctuations, we examined the CV of $\Delta\ell^{(k)}$. As predicted, the distribution of $\text{CV}_{\Delta\ell}^{(k)}$ does not increase for the no migration treatment ($\bar{t} = 0.09$, $P = 0.8$; Fig. 7a). In the global migration treatment, the coefficient of variation $\text{CV}_{\Delta\ell}^{(k)}$ significantly increases after the cessation of migration ($\bar{t} = 1.3$, $P < 10^{-3}$; Fig. 7b). This result is consistent with our prediction that global migration dampens fluctuations across communities. However, while our SLM simulations succeeds in predicting $\bar{t}$ for the no migration treatment, they fail to capture the effects of the global migration treatment (Fig. 7c,d).

Experimental details may explain why the effect of the global migration treatment, while significant, was not particularly large, as the size of the inoculum was nearly two orders of magnitude smaller than that of regional migration (Table 1). However, this experimentally-imposed parameter does not explain why the observed value of $\text{CV}_{\Delta\ell}^{(k)}$ lay outside the predictions of the SLM. Attractor status also does not explain this discrepancy, as the global migration communities were previously classified as belonging to the same attractor (Table S2, [1]). Instead of evaluating fluctuations *across* communities, we focused on fluctuations over time *within* individual communities to

evaluate the effect, if any, of global migration. We see that estimates of the CV before ($\mathrm{CV}^{<}_{\Delta\ell}$) and after the cessation of migration ($\mathrm{CV}^{>}_{\Delta\ell}$) are similar for both no and global migration treatments (Fig. 7e,f). Using a measure of the change in CV, we find that distributions of both treatments highly overlap, meaning that the change in the CV *within a given community* does not considerably increase in the global migration treatment (Fig. 7g). Splitting said distribution into high occupancy ASVs reinforces this conclusion, as there is no systematic increase in the change in CV for the global migration treatment (Fig. 7h).

Discussion

The main result of this study is the demonstration that experimental microbial communities grown on a single carbon source can sustain large ecological variation. 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 and using a macroecological approach. Meeting this criterion allowed us to document the existence of macroecological patterns that have only been documented in naturally occurring communities, suggesting a quantitative equivalence between experimental and observational studies despite the controlled nature of artificially maintained communities. 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 [71–73], the absence of which would preclude the possibility of macroecological investigations.

Characterizing robust empirical patterns in artificial communities is a key step toward identifying predictive ecological models. 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 extending the SLM, an empirically validated model of microbial community composition, using experimental details proved to be a useful framework for identifying treatments that were capable of generating macroecological effects. We examined the effects of two different forms of migration: regional (island-mainland) and global (fully interconnected metacommunity) [1]. 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 the SLM can link experimental communities and macroecological patterns.

Regarding global migration, we predicted that the treatment would primarily alter ASV fluctuations around their typical abundance [59]. We observed no change in the abundance fluctuations of community members after the cessation of migration *within individual communities*. However, we found that variation *across communities* tended to increase after the cessation of migration. This trend is consistent with our hypothesis regarding the effect of global migration, but one that cannot be reproduced by the SLM. This disagreement between fluctuations *across* and *within* communities suggests that experimental communities readily reach a state where their statistical properties do not change over transfer cycles, but can vary from community-to-community. In physics parlance, experimental communities are stationary, but may not necessarily be ergodic [74]. The existence of alternative stable-states are a likely explanation, as they are common in experimental microbial communities [56, 75, 76] and can be induced via fluctuations intrinsic to serial dilution setups [77]. In addition, their ergodicity-breaking effects have been used to explain invasion dynamics in experimental microbial communities [78]. The SLM can incorporate such states [37], though prior efforts required extensive temporal samples. Regardless, we can make two claims about the

impact of alternative stable-states. First, the quantity of migrants chosen for the global migration treatment was sufficient to alter attractor status, but insufficient to alter fluctuations in abundance within a given community. Second, macroecological patterns captured by the SLM persisted despite the pervasiveness of alternative stable-states, implying that heterogeneous outcomes of community assembly do not qualitatively alter macroecological patterns.

In this study we focused on patterns relating to the typical abundance and fluctuations in abundance across communities as well as over time. Noticeably, we did not investigate patterns that likely require the explicit invocation of interactions between community members (e.g., the correlations of abundance fluctuations), as the addition of interaction terms into the SLM would require several assumptions about the network of interactions and their magnitude. However, the absence of interaction terms in the SLM does not mean that interactions did not contribute to the macroecological patterns we observed. Models such as the SLM are known as "mean-field" models, where interactions between community members are permitted, but their effects on the dynamics of a given community member are determined by the mean dynamics of all community members [79]. Specifically, the SLM as presented in this study will continue to hold if the effect of community interactions is restricted to the parameters of the SLM (e.g., carrying capacity). Investigating correlations in abundance between community members, which can be understood as an outcome of specific ecological interactions rather than the mean effect of the community, likely requires models that go beyond phenomenology by explicitly considering mechanisms such as resource consumption [13, 80]. Indeed, consideration of resource consumption has proven critical for investigating the evolutionary dynamics of microorganisms in an ecological context [23].

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 [81–84]. Extending studies of microbial macroecology beyond patterns of abundance and community composition to the level of function would fully realize the physiological and energetic breadth that allowed macroecology vital to advance our understanding of macrobial life [5].

It is worth discussing how the experimental implementation of migration relates to natural environments. Microbial community assembly experiments often examine ecological dynamics as a "boom-and-bust" phenomenon, where abundances are initially low and proceed to reach a carrying capacity. This decision is often made due to the unwieldy nature of managing an array of continuous cultures (i.e., chemostats). However, this experimental design may reflect pulse-like forms of migration in nature, as boom-and-bust dynamics can be found across diverse ecosystems. Examples of environments where boom-and-bust dynamics occur include pitcher plants [85], particles of organic matter in the open ocean [86, 87], and even the human gut [88, 89]. Boom-and-bust dynamics can serve as a useful model for environments where resources are periodically supplied, as migrants must persist until the environment becomes favorable to growth (e.g., surviving in a metabolically inactive state [90–92]). Migration in such systems is difficult to quantify, though recent evidence suggests that it is sufficiently high such that microbes are rarely endemic to the environment in which they are found [93]. Appropriate models of migration may be necessary to investigate the macroecology of natural microbial communities.

Finally, considering 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 [5]. 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 were often impractical and 2) producing generalities from experiments has proven to be difficult [94]. Our results demonstrate that these two constraints are ameliorated by the features of microbial communities, providing counter evidence to recent claims that statistical distributions do not provide information about ecological mechanisms [95]. 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.

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 B.H. Good, O. Mazzarisi, M. Sireci, and N. I. Wisnoski for their comments on the manuscript. This work was supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2010885 (W.R.S.). Á.S. acknowledges support from Grant PID2021-125478NA-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". Conceptual diagrams made with BioRender.com.

References

1. 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`.

2. 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/`.

3. 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/`.

4. 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.

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

6. 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.

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

8. 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).

9. 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.

10. Lawton, J. H. Patterns in Ecology. *Oikos* **75**, 145–147 (1996). URL `https://www.jstor.org/stable/3546237`. Publisher: [Nordic Society Oikos, Wiley].

11. 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.

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

13. 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.

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

15. 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`.

16. 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].

17. George, A. B. & O'Dwyer, J. Universal abundance fluctuations across microbial communities, tropical forests, and urban populations. *Proceedings of the National Academy of Sciences* **120**, e2215832120 (2023). URL `https://pnas.org/doi/10.1073/pnas.2215832120`.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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].

25. 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.

26. 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.

27. 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.

28. 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.

29. Camacho-Mateu, J., Lampo, A., Sireci, M., Muñoz, M. A. & Cuesta, J. A. Sparse species interactions reproduce abundance correlation patterns in microbial communities. *Proceedings of the National Academy of Sciences* **121**, e2309575121 (2024). URL https://pnas.org/doi/10.1073/pnas.2309575121.

30. 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.

31. 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.

32. Shoemaker, W. R., Chen, D. & Garud, N. R. Comparative Population Genetics in the Human Gut Microbiome. *Genome Biology and Evolution* **14**, evab116 (2022). URL `https://doi.org/10.1093/gbe/evab116`.

33. 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.

34. 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.

35. 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.

36. 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.

37. Zaoli, S. & Grilli, J. A macroecological description of alternative stable states reproduces intra- and inter-host variability of gut microbiome. *Science Advances* **7**, eabj2882 (2021).

38. Sireci, M., Muñoz, M. A. & Grilli, J. Environmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance. *Proceedings of the National Academy of Sciences* **120**, e2217144120 (2023). URL `https://www.pnas.org/doi/abs/10.1073/pnas.2217144120`. Publisher: Proceedings of the National Academy of Sciences.

39. Shoemaker, W. R. & Grilli, J. Macroecological patterns in coarse-grained microbial communities. *eLife* **12** (2023). URL `https://elifesciences.org/reviewed-preprints/89650`. Publisher: eLife Sciences Publications Limited.

40. Wolff, R., Shoemaker, W. & Garud, N. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. *mBio* **14**, e02502–22 (2023). URL `https://journals.asm.org/doi/10.1128/mbio.02502-22`. Publisher: American Society for Microbiology.

41. 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].

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

43. 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.

44. 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.

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

46. 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.

47. 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).

48. 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.

49. 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.

50. 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.

51. 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.

52. Vila, J. C. C. *et al.* Metabolic similarity and the predictability of microbial community assembly (2023). URL https://www.biorxiv.org/content/10.1101/2023.10.25.564019v1. Pages: 2023.10.25.564019 Section: New Results.

53. Chang, C.-Y., Bajić, D., Vila, J. C. C., Estrela, S. & Sanchez, A. Emergent coexistence in multispecies microbial communities. *Science* **381**, 343–348 (2023). URL https://www.science.org/doi/10.1126/science.adg0727.

54. 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.

55. 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.

56. 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.

57. 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.

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

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

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

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

67. Keitt, T. H. & Stanley, H. E. Dynamics of North American breeding bird populations. *Nature* **393**, 257–260 (1998). URL https://www.nature.com/articles/30478. Publisher: Nature Publishing Group.

68. Keitt, T. H., Amaral, L. A. N., Buldyrev, S. V. & Stanley, H. E. Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**, 627–633 (2002). URL https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2001.1013. Publisher: Royal Society.

69. 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.

70. 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.

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

72. 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.

73. 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.

74. Balakrishnan, V. *Elements of Nonequilibrium Statistical Mechanics* (Springer International Publishing, Cham, 2021). URL http://link.springer.com/10.1007/978-3-030-62233-6.

75. 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.

76. 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.

77. 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.

78. Hu, J., Barbier, M., Bunin, G. & Gore, J. Collective dynamical regimes predict invasion success and impacts in microbial communities (2024). URL https://www.biorxiv.org/content/10.1101/2024.02.05.579032v1. Pages: 2024.02.05.579032 Section: New Results.

79. O'Dwyer, J. P. & Chisholm, R. A mean field model for competition: from neutral ecology to the Red Queen. *Ecology Letters* **17**, 961–969 (2014). URL https://onlinelibrary.wiley.com/doi/10.1111/ele.12299.

80. Ho, P.-Y., Nguyen, T. H., Sanchez, J. M., DeFelice, B. C. & Huang, K. C. Resource competition predicts assembly of gut bacterial communities in vitro. *Nature Microbiology* 1–13 (2024). URL https://www.nature.com/articles/s41564-024-01625-w. Publisher: Nature Publishing Group.

81. Diaz-Colunga, J. *et al.* Global epistasis on fitness landscapes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **378**, 20220053 (2023). URL https://royalsocietypublishing.org/doi/10.1098/rstb.2022.0053. Publisher: Royal Society.

82. 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.

83. Skwara, A. *et al.* Statistically learning the functional landscape of microbial communities. *Nature Ecology & Evolution* **7**, 1823–1833 (2023). URL https://www.nature.com/articles/s41559-023-02197-4.

84. Sanchez, A. *et al.* The community-function landscape of microbial consortia. *Cell Systems* **14**, 122–134 (2023). URL https://www.sciencedirect.com/science/article/pii/S2405471222004999.

85. Goyal, A., Bittleston, L. S., Leventhal, G. E., Lu, L. & Cordero, O. X. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. *eLife* **11**, e74987 (2022). URL https://doi.org/10.7554/eLife.74987. Publisher: eLife Sciences Publications, Ltd.

86. Ebrahimi, A., Goyal, A. & Cordero, O. X. Particle foraging strategies promote microbial diversity in marine environments. *eLife* **11**, e73948 (2022). URL https://doi.org/10.7554/eLife.73948. Publisher: eLife Sciences Publications, Ltd.

87. Enke, T. N. *et al.* Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. *Current biology: CB* **29**, 1528–1535.e6 (2019).

88. Vincent, C. *et al.* Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. *Microbiome* **4**, 12 (2016).

89. Johnson, K. E. *et al.* Human milk variation is shaped by maternal genetics and impacts the infant gut microbiome (2023). URL https://www.biorxiv.org/content/10.1101/2023.01.24.525211v1. Pages: 2023.01.24.525211 Section: New Results.

90. Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology* **9**, 119–130 (2011). URL https://www.nature.com/articles/nrmicro2504. Number: 2 Publisher: Nature Publishing Group.

91. Shoemaker, W. R. & Lennon, J. T. Evolution with a seed bank: The population genetic consequences of microbial dormancy. *Evolutionary Applications* **11**, 60–75 (2018).

92. Wisnoski, N. I. & Shoemaker, L. G. Seed banks alter metacommunity diversity: The interactive effects of competition, dispersal and dormancy. *Ecology Letters* **25**, 740–753 (2022). URL https://onlinelibrary.wiley.com/doi/10.1111/ele.13944.

93. Louca, S. The rates of global bacterial and archaeal dispersal. *The ISME Journal* **16**, 159–167 (2022). URL https://www.nature.com/articles/s41396-021-01069-8. Number: 1 Publisher: Nature Publishing Group.

94. 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.

95. 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.

Figures

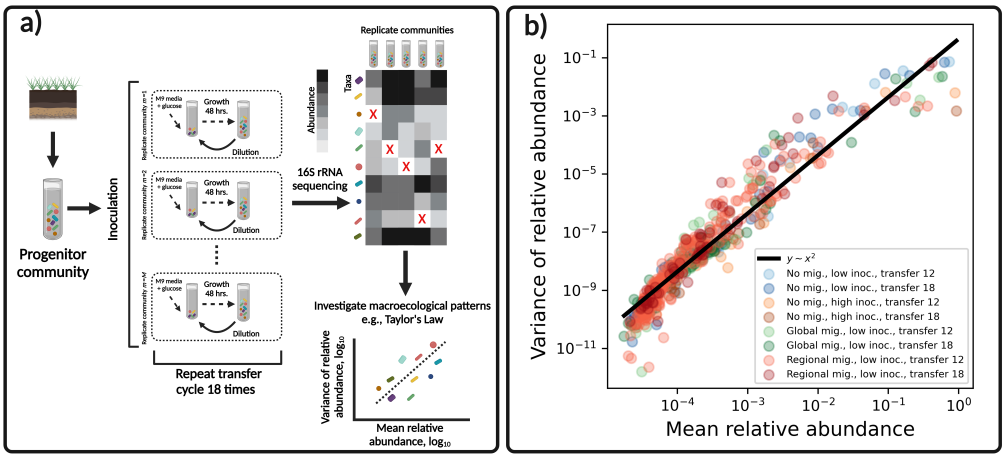

Figure 1. Emergent variation in experimental communities. a) Microbial community assembly experiments provide a means to maintain diversity in a laboratory setting. These experiments are commonly performed by growing a community sampled from a given environment (e.g., soil) and using it to inoculate a large number of replicate microcosms containing the same resource (e.g., M9 minimal media with glucose). These replicate communities are allowed to grow for a period of time (e.g., 48 h.) before aliquots are taken and diluted into microcosms containing replenished resource. This process known as a "transfer cycle" is repeated a given number of times (e.g., 18) and the resulting communities are sequenced via 16S rRNA sequencing. The abundances of community members have the potential to provide the variation necessary to investigate the existence of macroecological patterns in an experimental setting (e.g., Taylor's Law), a goal of this study. Red "X" symbols represent the absence of a community member in a sample. **b)** 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. Each data point represents statistical moments calculated across replicate communities for a single ASV from a single experimental treatment.

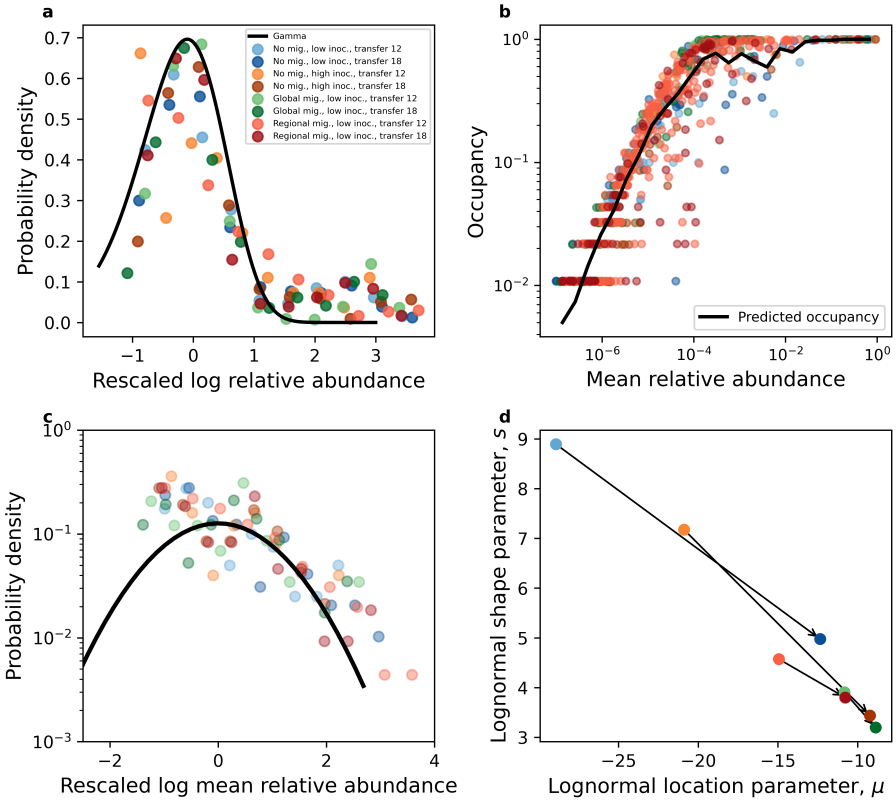

Figure 2. Macroecological patterns hold in experimental communities. Empirical macroecological patterns that were previously identified in natural microbial systems consistently arise in experimental communities [33]. **a)** The Abundance Fluctuation Distribution (AFD) tends to follow a gamma distribution 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). The prediction of the gamma (black line) was obtained by averaging over ASVs within a given mean relative abundance bin. **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.

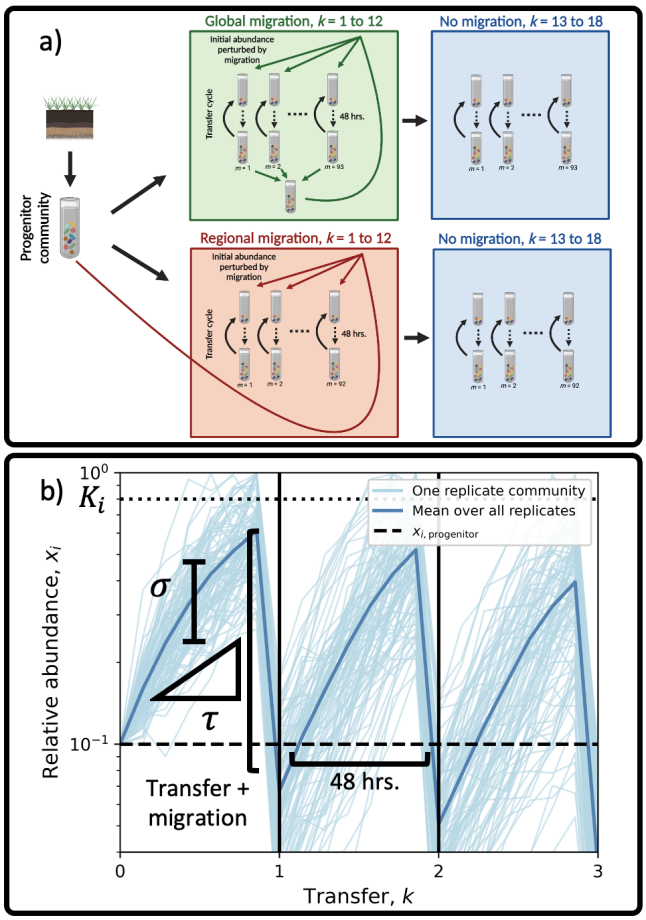

Figure 3. Incorporating experimental details into the Stochastic Logistic Model. a) Replicate communities were initiated from a single progenitor community that was isolated from soil. Communities were grown in microcosms containing a single carbon source for 48 hours, where they were then transferred to a microcosm with replenished resourced. This process constitutes a single transfer cycle. Migration was manipulated by altering the abundances of ASVs at the start of a transfer cycle. Two forms of migration were performed: regional and global. Regional migration represents a form of island-mainland migration, where aliquots of the progenitor community were added at the start of a transfer cycle. Global migration was manipulated by mixing aliquots of each replicate community at the end of a transfer cycle, which was then redistributed at the start of the subsequent transfer cycle. In both cases migration manipulations were performed for the first 12 transfers (1 to 12) and ceased for the remaining six transfers (13 to 18). Community assembly experiments were also performed with no migration manipulations for the entirety of the 18 transfers (not pictured). **b)** We modeled the ecological dynamics of each ASV within a given transfer cycle by incorporating the forms of migration performed in the experiment into the Stochastic Logistic Model of growth (Eq. S19) . Under the SLM, the parameter σ controls the strength of environmental noise, represented here as the variation in trajectories within a transfer cycle. The parameter τ represents the minimum timescale of growth (inverse of the maximum rate of growth), controlling the rate that the population approaches its carrying capacity. Migration alters the relative abundance of an ASV at the start of a transfer cycle as a perturbation of initial conditions (Eq. 11). 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 then proceeds to increase in abundance at a rate set by the timescale of growth τ. Conceptual diagram **a** was modified from Estrela et al. [1].

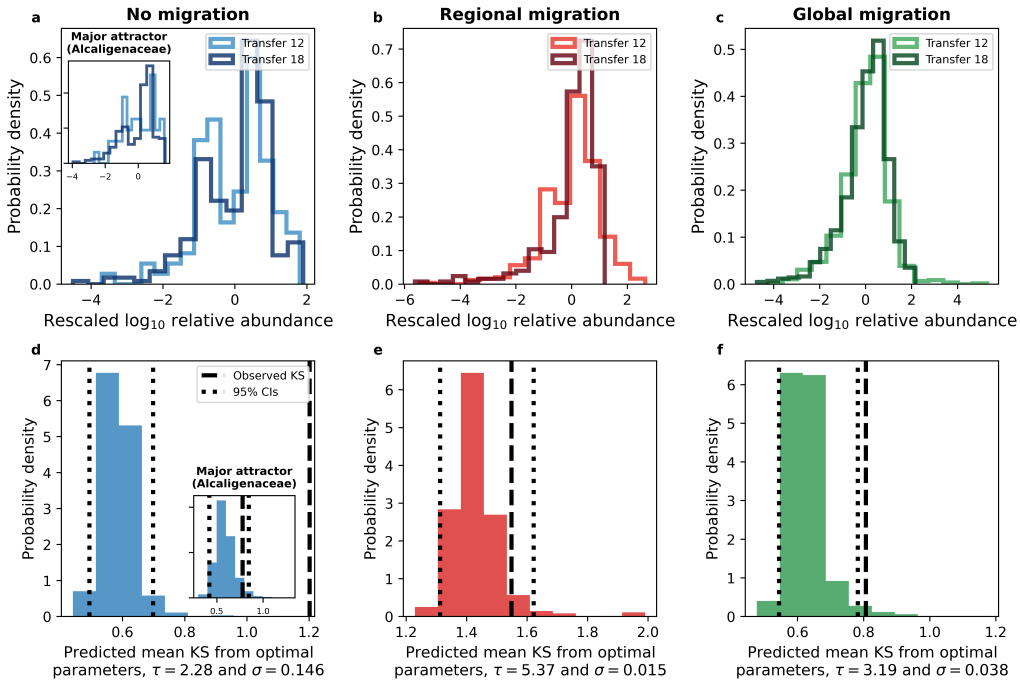

Figure 4. Migration impacts the shape of AFDs. a-c) By rescaling $\log_{10}$ transformed AFDs for each ASV, we can examine how the shape of the AFD changes before and after the cessation of migration. We quantified this shift by estimating the KS distance between ASVs before (transfer 12) and after (transfer 18) the cessation of the migration treatment for each ASV, then taking the mean over ASVs. **d-f)** Using optimal parameter combinations identified by ABC, we found that the SLM predicted reasonable KS statistics for regional migration as well as within the major attractor of the no migration treatment (inset of sub-plots **a,d**), with borderline successful predictions for the global migration treatment.

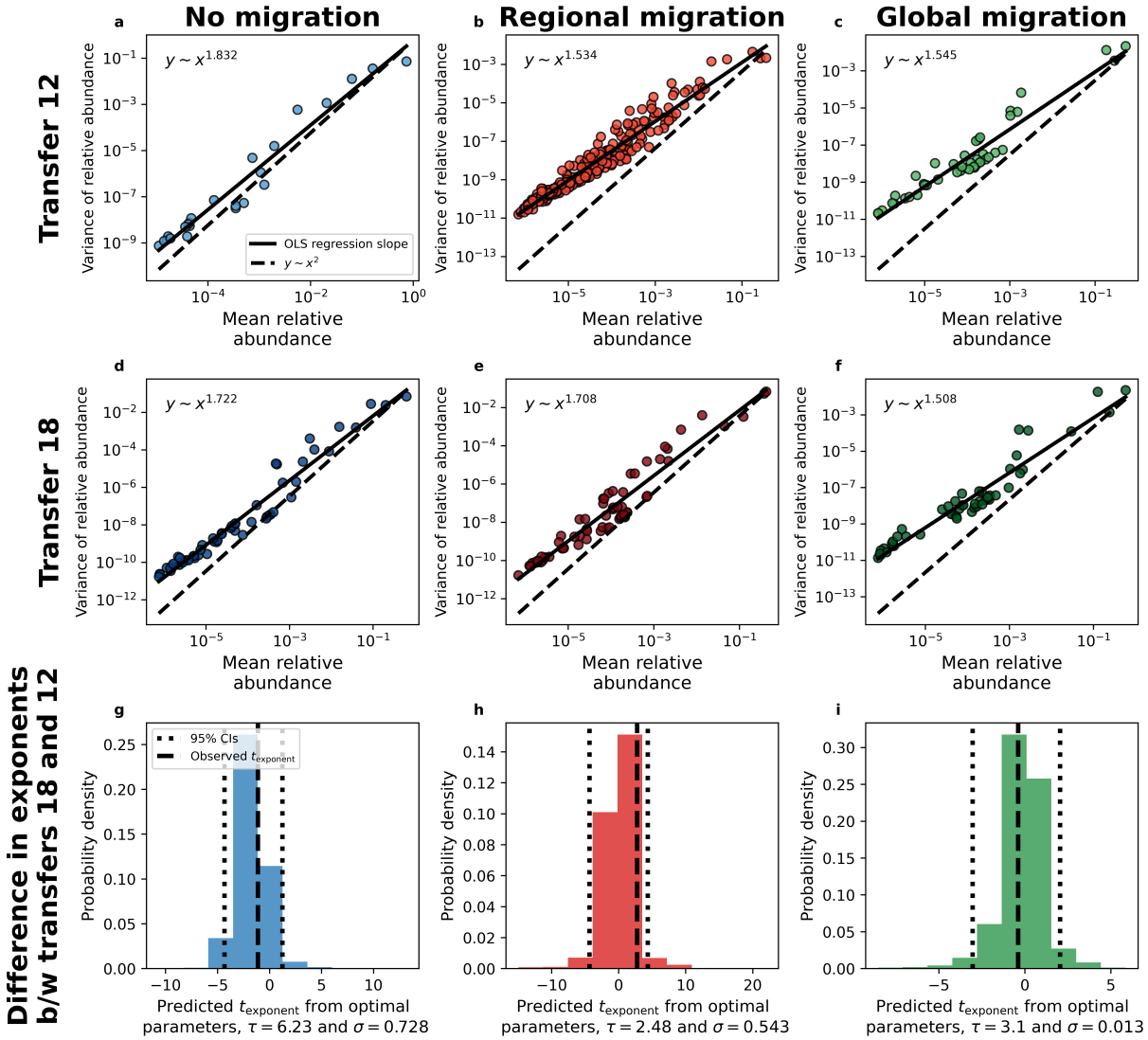

Figure 5. Regional migration impacts the exponent of Taylor's Law. a-f) By examining the exponent of Taylor's Law for each treatment we found that the exponent only considerably changed after the cessation of migration for the regional migration treatment. Each data point represents statistical moments calculated across replicate communities for a single ASV from a single experimental treatment. **g-i)** Using our ABC estimation procedure, the SLM succeeds in predicting this increase in the exponent for regional migration.

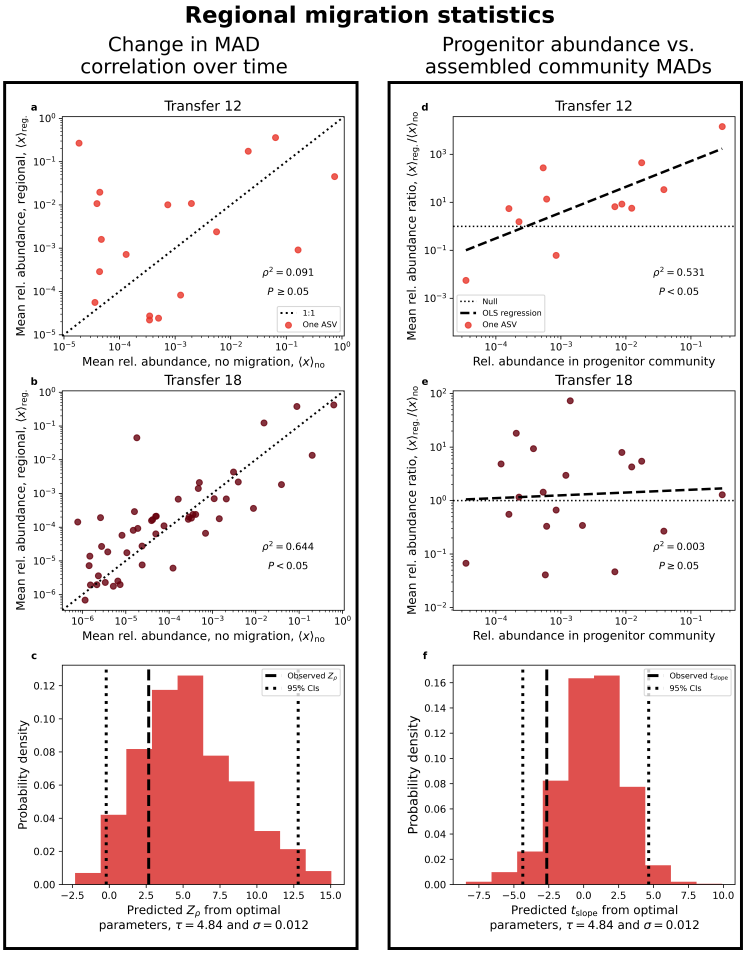

Figure 6. Regional migration alters typical abundance in a manner consistent with SLM simulations. The properties of the MAD were examined to evaluate the impact of migration over time. **a)** At transfer 12, the last transfer with migration, we found no apparent correlation between the MADs of the regional 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 distributions and identify reasonable parameter regimes **d)** as well as the relative error of our predictions. **e)** The ratio of the MAD between regional and no migration provides a single variable that can 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. Only sub-plots **c** and **f** contain simulated data.

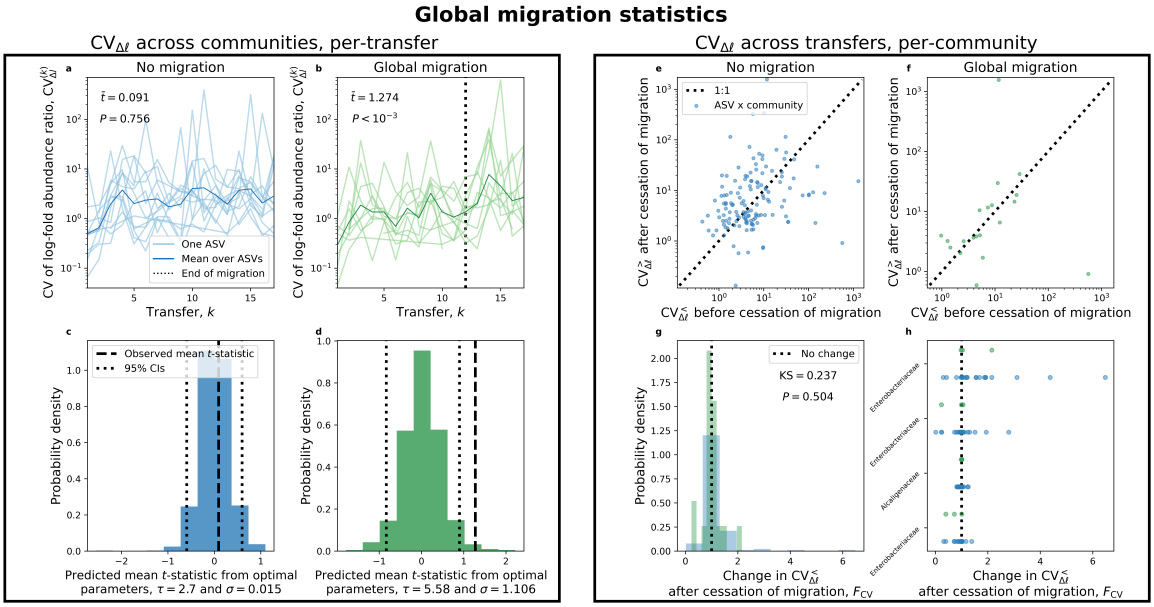

Figure 7. SLM simulations are unable to reproduce log-fold fluctuations *across* communities under global migration. By leveraging the entire timeseries we can examine how the coefficient of variation of $\Delta\ell$ was altered by migration for two scenarios: 1) fluctuations across replicate communities at a given transfer ($\mathrm{CV}^{(k)}_{\Delta\ell}$) and 2) fluctuations before ($\mathrm{CV}^{<}_{\Delta\ell}$) and after ($\mathrm{CV}^{>}_{\Delta\ell}$) the cessation of migration within a given community. **a)** As expected, there is no change in $\mathrm{CV}^{(k)}_{\Delta\ell}$ for the no migration treatment. **b)** However, $\mathrm{CV}^{(k)}_{\Delta\ell}$ tends to slightly increase after transfer 12 for global migration, a shift that was found to be significant using a t-test where transfer labels were permuted for each ASV. **c, d)** Our SLM predictions succeeded for the no migration treatment, but was unable to capture observed values of $\bar{t}$ under global migration. These results pertaining to the fluctuations across communities can be contrasted with fluctuations within a community. **e, f)** We did not observed a systematic shift in the CV before and after the cessation of migration for either treatment. **g)** By calculating the difference between two CVs (F_{CV}) for each ASV in each replicate community, we do not observe a significant difference between migration treatments using a KS test constrained on ASV identity. **h)** This conclusion is confirmed by investigating the change in the CV for high occupancy ASVs, as there is no systematic difference between migration treatments.

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}	$M \cdot (8 \cdot 10^{-5}) \cdot D_{\text{transfer}} \cdot N^* \approx 3.07 \cdot 10^4$
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 [1]. Estimates of total community size were previously obtained. 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.