Microbial populations hardly ever grow logistically and never sublinearly

José Camacho-Mateu,[1,2] Aniello Lampo,[1,2] Mario Castro,[1,3] and José A. Cuesta[1,2,4,*]

[1] *Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain*
[2] *Universidad Carlos III de Madrid, Departamento de Matemáticas, Leganés, Spain*
[3] *Universidad Pontificia Comillas, Instituto de Investigación Tecnológica, 28015 Madrid, Spain*
[4] *Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, Zaragoza, Spain*

(Dated: November 13, 2024)

We investigate the growth dynamics of microbial populations, challenging the conventional logistic model. By analyzing empirical data from various biomes, we demonstrate that microbial growth is better described by a generalized logistic model, the θ-logistic model. This accounts for different growth mechanisms and environmental fluctuations, leading to a generalized gamma distribution of abundance fluctuations. Our findings reveal that microbial growth is never sublinear, so they cannot endorse—at least in the microbial world—the recent proposal of this mechanism as a stability enhancer of highly diverse communities. These results have significant implications for understanding macroecological patterns and the stability of microbial ecosystems.

I. INTRODUCTION

Ever since Verhulst proposed it in 1838 [1], the logistic growth equation has become a standard in modeling populations in ecology—so much so that it has become a common default assumption when devising any population framework. Its rationale is simple—the per capita growth is assumed to be proportional to the fraction of resources left by the rest of the population. However, in Murray's words [2], the logistic equation "is more like a metaphor for a class of population models with density-dependent regulatory mechanisms ... and must not be taken too literally". In fact, it is not difficult to figure out mechanistic benchmarks leading to other types of density-dependent growth [3, 4]. Tumors, for instance, are better modeled with a Gompertz's law than with a logistic model [5, 6].

Whether population growth is logistic or not may have important consequences for ecological communities. For instance, in a recent work, Hatton et al. [7] show that a sublinear growth (a growth scaling with biomass x as x^μ, with $0 < \mu < 1$) might resolve the longstanding stability-diversity controversy [8, 9]. Even though this result has been contested [10], there is no doubt that the growth law may influence important aspects of ecological communities. Furthermore, empirical evidence accumulated over decades [11, 12] suggests that the growth laws governing populations of birds, mammals, fish, and insects are indeed heterogeneous.

The question becomes particularly relevant for microbial communities because simple mechanisms of cellular growth that consider both excluded volume and cell diffusion [3] generate effective growth laws other than the logistic model. Until the arrival of metagenomics, determining how microbial populations grow was out of the question. At present though, the unprecedented amount of available data [13–15] renders the problem amenable to analysis. Whether or not microbial populations grow logistically can be crucial to explain macroecological patterns observed in these communities [16–23] because some of them seem to depend critically on the logistic assumption [16].

In this work we will show (*i*) that the microbial growth in several real biomes [24] is better described by a generalization of the logistic model; (*ii*) that it does affect the emergent macroecological patterns; (*iii*) that, in spite of this, the proposed explanation for these patterns [16] can be 'saved' if properly stated; and (*iv*) that microbial growth is never sublinear—discarding any alleged benefit on the stability of microbial communities [7, 10].

II. GENERALIZED GROWTH MODEL FOR MICROBES

A generalization of the logistic model that covers a wealth of different growth mechanisms is the θ-logistic model [25], defined by the equation

$$\dot{x} = \frac{x}{\tau\theta}\left[1 - \left(\frac{x}{K}\right)^\theta\right], \tag{1}$$

where x represents the population of a species (also referred to as abundance, or density of individuals), τ is its intrinsic growth time, and K is the carrying capacity. The parameter θ governs the dependence of the per capita growth rate with the abundance. The exponent θ encapsulates the effect of the different growth processes [3, 4]. A value $\theta > 0$ describes a linear growth for small abundances ($\dot{x} \propto x$)—with $\theta = 1$ representing the logistic model—whereas $-1 < \theta < 0$ describes a sublinear growth ($\dot{x} \propto x^{1-|\theta|}$). The transition from $\theta > 0$ to $\theta < 0$ is smooth because in the limit $\theta \to 0$ the equation recovers Gompertz's model [26]

$$\dot{x} = -\frac{x}{\tau}\log\left(\frac{x}{K}\right), \tag{2}$$

commonly employed to describe tumor growth [5, 6].

* cuesta@math.uc3m.es

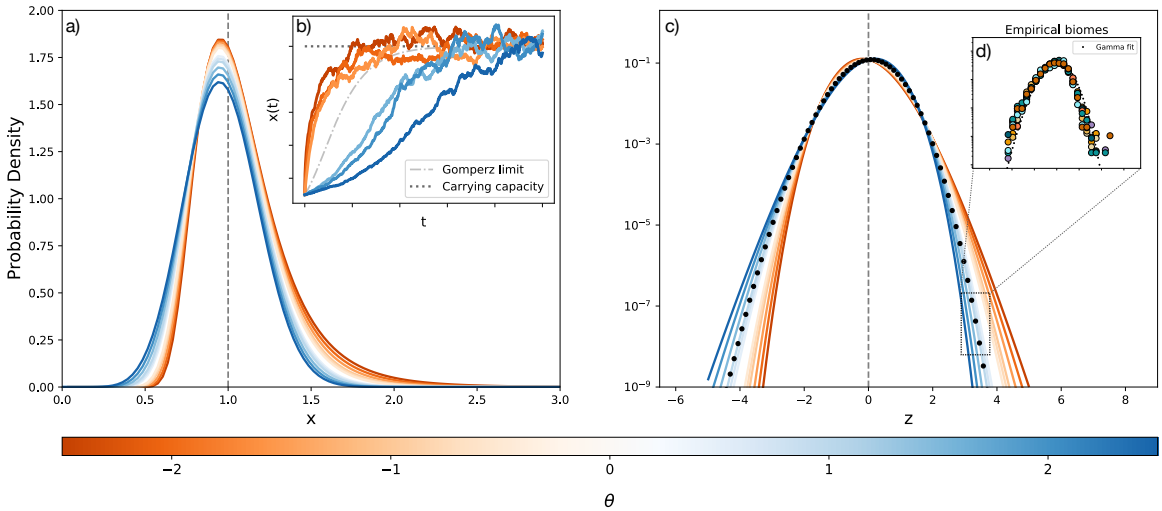

FIG. 1. a) Generalized gamma corresponding to the abundance fluctuation distribution $P(x)$, Eq. (7), for several values of θ (indicated by the colorbar). b) Realizations of the SθLM (3) for $\theta = -2.5, -2, -1.5, 1.5, 2, 2.5$, were obtained using an Euler-Maruyama integration scheme [27] with initial condition $x(0) = 0.1$ and integration time-step $\Delta t = 10^{-3}$. The black dotted line signals the value of the carrying capacity K, while the grey dot-dashed line provides the solution of Eq. (2) (corresponding to Gomperz's limit $\theta \to 0$). c) Graphical representation of $\log \bar{P}(z)$ [c.f. Eq. (16)], i.e. the logarithm of the abundance fluctuation distribution expressed as a function of the standardized abundance z [c.f. Eq. (10)], for several values of θ. The black dots portray the fit to the biomes of inset d), where we illustrate $\log \bar{P}(z)$ for nine real biomes: Seawater, Soil, Sludge, Oral1, Gut1, Gut2, Glacier, River, Lake [24]. All plots have been performed with $\tau = 0.1, K = 1, w = 0.5$.

However, real biomes are subject to environmental changes. The rapid fluctuations they cause in species abundance calls for a stochastic modeling [16, 18, 28, 29]. This is achieved by adding a noise term to the θ-logistic model (1). Analysis of the abundance fluctuations in empirical data suggests introducing a multiplicative noise proportional to x, as in Grilli's stochastic logistic model (SLM) [16, 18, 30, 31] (see [32], Fig. S1). The result is the stochastic θ-logistic model (SθLM) [4], defined by the Langevin equation

$$\dot{x} = \frac{x}{\tau\theta}\left[1 - \left(\frac{x}{K}\right)^{\theta}\right] + x\xi(t), \tag{3}$$

where $\xi(t)$ is a zero-mean, Gaussian, white noise such that $\langle \xi(t)\xi(t')\rangle = w\delta(t-t')$, w being the noise variance. Figure 1b) illustrates typical realizations of the SθLM for different values of θ ($\theta = 1$ reproduces the SLM).

Equation (3) describes the intrinsic stochastic dynamics of the abundance of different microbial taxa, each of which is expected to be described by specific parameters. For the sake of clarity though, no taxonomic identity is explicitly included in the equation.

III. ABUNDANCE FLUCTUATION DISTRIBUTION

The probability distribution of the SθLM (3) in the steady state (the so-called abundance fluctuation distribution, AFD) can be readily obtained through its associated Fokker-Planck equation $P_t + J_x = 0$, where

$$J \equiv \frac{x}{\tau\theta}\left[1 - \left(\frac{x}{K}\right)^{\theta}\right] P - \frac{w}{2}\frac{\partial}{\partial x}(x^2 P). \tag{4}$$

In the steady state, the current is constant. This constant is zero because of the boundary conditions $P(0) = P(x \to \infty) = 0$, so we can find the stationary probability $P(x)$ as the solution of

$$\frac{x}{\tau\theta}\left[1 - \left(\frac{x}{K}\right)^{\theta}\right] P = \frac{w}{2}(x^2 P)'. \tag{5}$$

Denoting $q(x) = x^2 P(x)$, this equation becomes

$$(\log q)' = \alpha\left(\frac{1}{x} - \frac{x^{\theta-1}}{K^{\theta}}\right), \qquad \alpha \equiv \frac{2}{w\tau\theta}, \tag{6}$$

hence

$$P(x) = \frac{x^{\alpha-2}}{Z}\exp\left[-\frac{\alpha}{\theta}\left(\frac{x}{K}\right)^{\theta}\right], \tag{7}$$

where

$$\begin{aligned} Z &= \int_0^{\infty} x^{\alpha-2}\exp\left[-\frac{\alpha}{\theta}\left(\frac{x}{K}\right)^{\theta}\right] dx \\ &= \left(\frac{\theta}{\alpha}\right)^{\beta}\frac{K^{\alpha-1}}{|\theta|}\Gamma(\beta), \qquad \beta \equiv \frac{\alpha-1}{\theta}. \end{aligned} \tag{8}$$

It is important to notice that $\beta > 0$ for any $\theta < 0$, as well as for any $0 < \theta < 2/w\tau$.

Probability density (7)—whose shape for different values of θ is illustrated in Fig. 1—describes how the abundance of a species fluctuates as a consequence of environmental noise. In the context of macroecology [33, 34] this is one of the patterns that characterize microbial communities. As a matter of fact, it has been recently proposed that this AFD has the shape of a gamma distribution [16], which corresponds to the case $\theta = 1$ (logistic growth). Instead, when $\theta \neq 1$, the AFD (7) has the shape of the generalized gamma distribution [4], whose moments are given by

$$\langle x^n \rangle = K^n \left(\frac{\theta}{\alpha}\right)^{\frac{n}{\theta}} \frac{\Gamma(\beta + \frac{n}{\theta})}{\Gamma(\beta)}. \tag{9}$$

A generalized gamma is a versatile distribution. Apart from recovering the gamma distribution for $\theta = 1$, it can also adopt other familiar shapes for different parameter choices. For instance, it can be a stretched exponential ($\alpha = 2$), an ordinary exponential ($\alpha = 2$, $\theta = 1$), a (positive) half-normal ($\alpha = \theta = 2$), or a Weibull ($\alpha = \theta$). In the limit $\theta \to 0$ it becomes a lognormal (see Appendix A). Furthermore, for $\theta < 0$ it abruptly drops to zero for small x whereas it decays as a power law for large x.

As the gamma shape of the AFD is an empirically well-established macroecological pattern in microbial communities [16, 21–23], the previous result seems to rule out any other growth mechanism but the logistic one. Before we jump to a wrong conclusion though, let us notice that all empirical analyses of the AFD of microbial species use the standardized variable

$$z = \frac{\log x - \langle \log x \rangle}{\sqrt{\mathrm{Var}(\log x)}}, \tag{10}$$

so it is $\bar{P}(z)$, the probability density of this standard variable z that we need to use for comparison.

In order to obtain $\bar{P}(z)$ let us introduce the intermediate variable

$$u = \log\left(\frac{\alpha}{\theta}\right) + \theta \log\left(\frac{x}{K}\right), \tag{11}$$

whose probability density relates to (7) through $\tilde{P}(u) = P(x)|dx/du| = P(x)x/|\theta|$. Thus,

$$\tilde{P}(u) = \frac{1}{\Gamma(\beta)} \exp\left[\beta u - e^u\right]. \tag{12}$$

We can now relate u to z by noticing that

$$\frac{u - \langle u \rangle}{\sqrt{\mathrm{Var}(u)}} = \varepsilon z,$$

where $\varepsilon \equiv \mathrm{sign}(\theta)$. In order to calculate $\langle u \rangle$ and $\mathrm{Var}(u)$ we use the identity

$$\Gamma(\beta) = \int_{-\infty}^{\infty} \exp\left[\beta u - e^u\right] du, \tag{13}$$

in terms of which

$$\langle u \rangle = \frac{\partial}{\partial \alpha} \log \Gamma(\beta) = \psi(\beta), \tag{14}$$

$$\mathrm{Var}(u) = \frac{\partial^2}{\partial \alpha^2} \log \Gamma(\beta) = \psi'(\beta), \tag{15}$$

where $\psi(x)$ denotes the digamma function [35]. Thus, as $\bar{P}(z) = \tilde{P}(u)\sqrt{\psi'(\beta)}$, we conclude that $\bar{P}(z)$ takes the form of an exp-gamma distribution, namely

$$\begin{aligned} \bar{P}(z) &= \frac{\sqrt{\psi'(\beta)}}{\Gamma(\beta)} \exp\left[\beta u - e^u\right], \\ u &\equiv \psi(\beta) + \varepsilon\sqrt{\psi'(\beta)} z. \end{aligned} \tag{16}$$

This distribution is depicted in Fig. 1c for different choices of θ. Figure 1d also shows good agreement with empirical AFDs.

Two remarkable features of distribution (16) are worth noticing: first of all, it is independent of K, and secondly, except for the sign of θ it depends on the rest of the parameters (τw and θ) only through the combination β. Hence, the value of β obtained by fitting $\bar{P}(z)$ to a set of data, can be interpreted in very different ways—either we assume $\theta = 1$ and use β to determine τw, or we fix τw and use β to determine θ. In other words, the fact that (16) is a good description of empirical AFDs does *not* rule out a θ-logistic growth. Previous macroecological patterns are perfectly compatible with alternative growth laws.

We still have to discuss the dependence on $\varepsilon = \pm 1$. This parameter discriminates the two classes of growth at low abundances: linear ($\varepsilon = +1$) and sublinear ($\varepsilon = -1$). The skewness coefficient of $\bar{P}(z)$ is (see Appendix B)

$$\gamma = \langle z^3 \rangle = \frac{\varepsilon \psi''(\beta)}{\psi'(\beta)^{3/2}} \sim -\frac{\varepsilon}{\sqrt{\beta}} \quad (\beta \to \infty), \tag{17}$$

so $-\varepsilon$ determines the sign of the skewness of this distribution. Notably, all empirical distributions are negatively skewed, hence supporting the value $\varepsilon = +1$, i.e. linear growth (see Fig. 1c). This rules out—for microbes—the sublinear growth proposed by Hatton et al. [7] as a stability enhancer of highly diverse communities.

Unfortunately, the same feature that makes of $\bar{P}(z)$ a universal macroecological pattern (the dependence through the single parameter β) renders it useless to discriminate the kind of growth law followed by the different microbial species. In order to achieve this goal, we need to fit directly Eq. (7) to the data of microbial abundances.

IV. BAYESIAN INFERENCE OF THE GROWTH PARAMETER

Estimates of the parameter θ have been obtained in the past through least squares fits of the dynamic equation (1) to time series data [7, 12, 25, 36]. This method has many drawbacks—such as the poor estimation of derivatives produced by a finite-difference scheme applied to

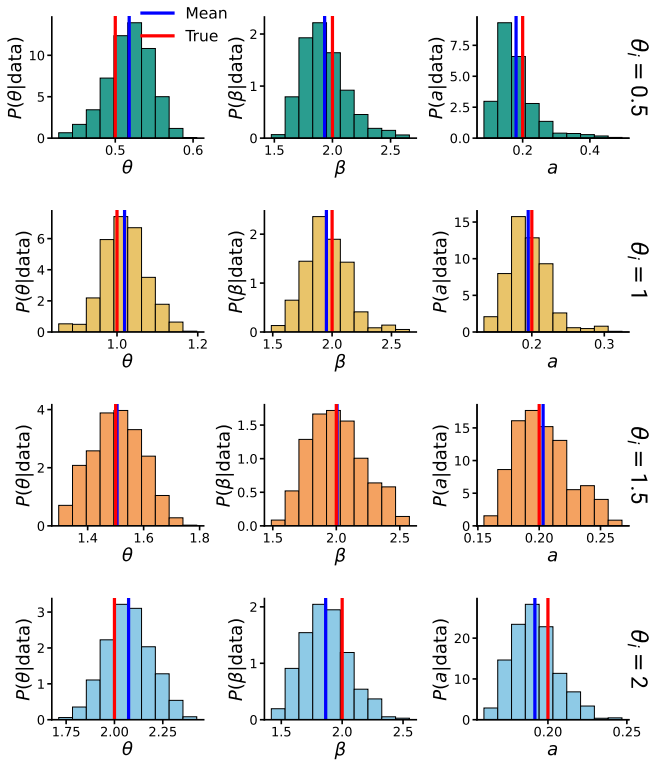

FIG. 2. Marginal posterior distributions for parameters θ, β, and a [see Eq. (19)] as obtained for a synthetic data set consisting of $S = 4$ taxa with $\theta = 0.5$, 1, 1.5, 2, and the same values of $\beta = 2$ and $a = 0.2$. The true parameter values are marked by red lines; the mean of the posterior predictive distributions [Eq. (26)] by blue lines. The narrow and peaked posteriors around the true values prove the identifiability of these parameters using Bayesian inference. Total number of samples $N_c = 15000$, sampling depth $N_s = 5000$.

noisy data, or the underlying assumption that the noise follows a normal distribution. The estimates of θ are thus unreliable. Instead, we took advantage of the knowledge of $P(x)$ [c.f. Eq. (7)] and performed a Bayesian inference of the parameter θ [37, 38].

Databases of microbiomes do not provide lists of species abundances. A typical sample is a vector of counts $\mathbf{n}_s = \left(n_{1,s}, n_{2,s}, \ldots, n_{S,s}\right)$, where $n_{i,s}$ is the number of reads attributed to taxon $i = 1, \ldots, S$ in sample s. The total number of reads across all taxa in each sample, $\sum_i n_{i,s} = N_s$, is referred to as 'sampling effort' or 'sampling depth'. Thus, we must add the variability introduced by the sampling process to the inherent variability of the abundances induced by the dynamics [captured by the distribution (7)]. We can describe this sampling by a Poisson distribution with mean $x_{i,s}N_s$. Therefore, the probability to observe n_i counts of the ith taxon in the given sample s is obtained by marginalizing over x_i, namely

$$P(n_i|N_s, \Omega_i) = \frac{N_s^{n_i}}{n_i!} \int_0^\infty dx_i \, P(x_i|\Omega_i) x_i^{n_i} e^{-x_i N_s}, \quad (18)$$

$P(x|\Omega)$ being the generalized gamma distribution (7)—which for the sake of simplicity we parametrize as in the R package `dgen.gamma`, i.e.

$$P(x|\Omega) = \frac{\theta a^{\beta\theta} x^{\beta\theta - 1}}{\Gamma(\beta)} e^{-(ax)^\theta}. \quad (19)$$

Here Ω denotes the set of parameters β, a, θ, where

$$a = \frac{1}{K} \left(\frac{\alpha}{\theta}\right)^{1/\theta}. \quad (20)$$

Now, the posterior distribution of parameters Ω_i can be obtained as

$$P(\Omega_i|N_s, n_i) = \frac{P(n_i|N_s, \Omega_i) P(\Omega_i|N_s)}{P(n_i|N_s)}. \quad (21)$$

Guided by the maximum entropy principle [37], we use exponential distributions as priors for the parameters:

$$P(\Omega|N_s) = \delta_\beta e^{-\delta_\beta \beta} \delta_a e^{-\delta_a a} \delta_\theta e^{-\delta_\theta \theta}, \quad (22)$$

with $\delta_\beta, \delta_a, \delta_\theta > 0$ small constants. These priors ensure the positiveness of these parameters but are otherwise quite agnostic with respect to what their values are.

The method just described produces an estimate of the value of θ for individual species. However, a comparison of the estimates for different species suggested that a gamma distribution

$$P(\theta|\nu_0, k_0) = \frac{\theta^{\nu_0 - 1} e^{-\theta/k_0}}{k_0^{\nu_0} \Gamma(\nu_0)}, \quad (23)$$

for an appropriate choice of the "hyper-parameters" ν_0, k_0, might be a reasonable model for the distribution of θ across species. Accordingly, to estimate this distribution more accurately we adopted a global strategy, inferring the parameters for all species at once, using a hierarchical Bayesian model. In other words, we calculate the global posterior

$$P(\mathbf{\Omega}|N_s, \mathbf{n}) = \frac{P(\mathbf{n}|N_s \mathbf{\Omega}) P(\mathbf{\Omega}|N_s)}{P(\mathbf{n}|N_s)}, \quad (24)$$

where $\mathbf{n}$ is the vector of counts of all species, and $\mathbf{\Omega}$ denotes the set of parameters of all species plus the hyper-parameters ν_0, k_0. Then we choose the prior

$$P(\mathbf{\Omega}|N_s) = \left[\prod_{i=1}^{S} \delta_\beta \delta_a e^{-\delta_\beta \beta_i - \delta_a a_i} \frac{\theta_i^{\nu_0 - 1} e^{-\theta_i/k_0}}{k_0^{\nu_0} \Gamma(\nu_0)}\right] \times \delta_{\nu_0} \delta_{k_0} e^{-\delta_{\nu_0} \nu_0 - \delta_{k_0} k_0}. \quad (25)$$

where we have also used exponential priors for the hyper-parameters.

This hierarchical Bayesian approach not only refines the estimates but also allows cross-learning among taxa, enhancing the overall robustness and accuracy of our parameter estimations while reducing overfitting at the

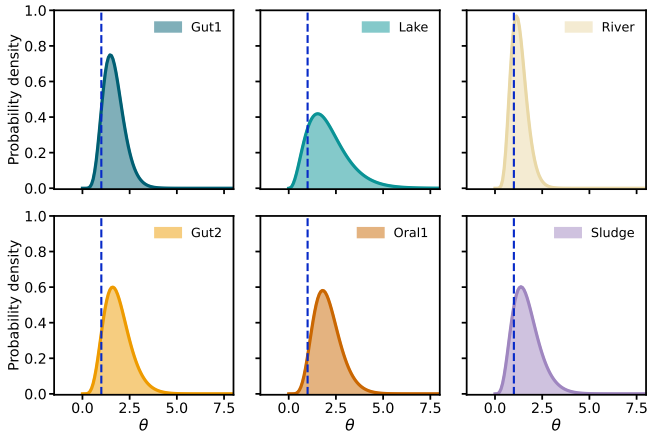

FIG. 3. Gamma distributions of θ across species for several biomes as obtained from the estimated maximum posterior hyperparameters. Dashed blue lines mark the value $\theta = 1$ corresponding to logistic growth. For computational limitations, in making this figure we only included species appearing in more than 95% of the samples.

same time. Implementation details can be found in Appendix C.

An important issue related to Bayesian inference is identifiability. Roughly speaking, we say that a parameter is identifiable if its posterior distribution is relatively narrow and peaked. We have conducted a synthetic test to prove that the parameters of the SθLM can actually be inferred by the described procedure. Thus, we generated a random sample of size N_c and fixed sampling depth N_s from a synthetic population consisting of $S = 4$ independent taxa. We then applied the Bayesian inference method just described, assuming that abundances follow the generalized gamma distribution (19) with equal parameters $\beta = 2$ and $a = 0.2$, but different θ parameters ($\theta = 0.5,\ 1,\ 1.5,\ 2$). Given that there are only four taxa, the θs are inferred individually—using a prior $\delta_\theta e^{-\delta_\theta \theta}$. The high reliability in recovering these parameters exhibited by the method (as illustrated in Figure 2) is reassuring.

Figure 3 shows the distributions for θ for several biomes estimated as we have just explained. The gamma functions plotted in this figure correspond to the inferred maximum posterior hyperparameters. For comparison, the value $\theta = 1$ (logistic growth) is marked with a dashed vertical line. The figure reveals that logistic growth is hardly the most probable growth law among microbes—the values of θ ranging from nearly zero to $\theta \approx 5$.

To assess the quality of our Bayesian inference, we checked that the outcome of the posterior predictive agrees well with the AFD of individual species (see [32], Figs. S6–S7). Posteriors predictive are obtained as

$$P(\tilde{\mathbf{n}}|N_s, \mathbf{n}) = \int P(\tilde{\mathbf{n}}|N_s, \mathbf{\Omega}) P(\mathbf{\Omega}|N_s, \mathbf{n})\, d\mathbf{\Omega}, \qquad (26)$$

where the integral covers the whole parameter space. These distributions can be sampled by Monte Carlo simulations if we interpret the integral above as the average of the likelihood $P(\tilde{\mathbf{n}}|N_s, \mathbf{\Omega})$ over the posterior distribution $P(\mathbf{\Omega}|N_s, \mathbf{n})$.

We can illustrate the accuracy of this posterior predictive by comparing averages (Fig. 4a) and variances (Fig. 4b), as obtained from (26) and from empirical data. As a reference, the insets in these figures provide the results obtained assuming that the AFD follows a gamma distribution (as derived from the logistic, $\theta = 1$, growth).

As a final test, we computed the distribution of the logarithmic fold-change $\lambda = \log(x_{s+1}/x_s)$ of two consecutive samples (x_s and x_{s+1}) of the abundance of a given species. This distribution has been proposed as another macroecological pattern and has been the focus of interest of several recent studies [17, 20, 39, 40]. It is straightforward to derive it from the AFD provided the two samples are sufficiently uncorrelated. The result is (see Appendix D)

$$\begin{aligned} P(\lambda) &= |\theta| \frac{\Gamma(2\beta)}{\Gamma(\beta)^2} \left[2\cosh\left(\frac{\theta\lambda}{2}\right) \right]^{-2\beta} \\ &\sim |\theta| \frac{\Gamma(2\beta)}{\Gamma(\beta)^2} e^{-\beta|\theta\lambda|} \qquad (|\lambda| \to \infty). \end{aligned} \qquad (27)$$

Interestingly, this is the distribution reported for λ when it is averaged across all samples and taxons [17]. In Fig. 5 we compare $P(\lambda)$ as obtained from the inferred parameters with the empirical one. For each taxon, we calculate the maximum a posteriori estimate (the mode of the posterior distributions) and the 95% confidence interval. A more thorough comparison can be found in [32], Figs. S8–S9.

V. DISCUSSION

We conclude from our previous analyses that there is sufficient evidence to rule out logistic growth among microbes in favor of more sophisticated mechanisms, in line with previous evidence obtained from birds, mammals, fish, and insects [11, 12]. Deviations from the logistic growth can originate in excluded volume effects [3], collective behavior (cooperation or competition) of bacteria [4], difficult access to external resources of microbes living at the interior of a colony, or any other complex dynamics. This fact brings about a crucial change in an important macroecological pattern: the AFD takes the form of a generalized gamma distribution instead of the gamma shape that had been claimed so far [16, 21–23]. This does not mean that previous analyses were incorrect, because all of them were performed using the variable z [c.f. Eq. (10)]. As we have shown in this paper, in terms of this variable all generalized gammas fall into two distinct patterns—that are otherwise indistinguishable. This is the reason why all previous research has misinterpreted the shape of the AFD as a gamma distribution. Interestingly, all microbes seem to fall into just

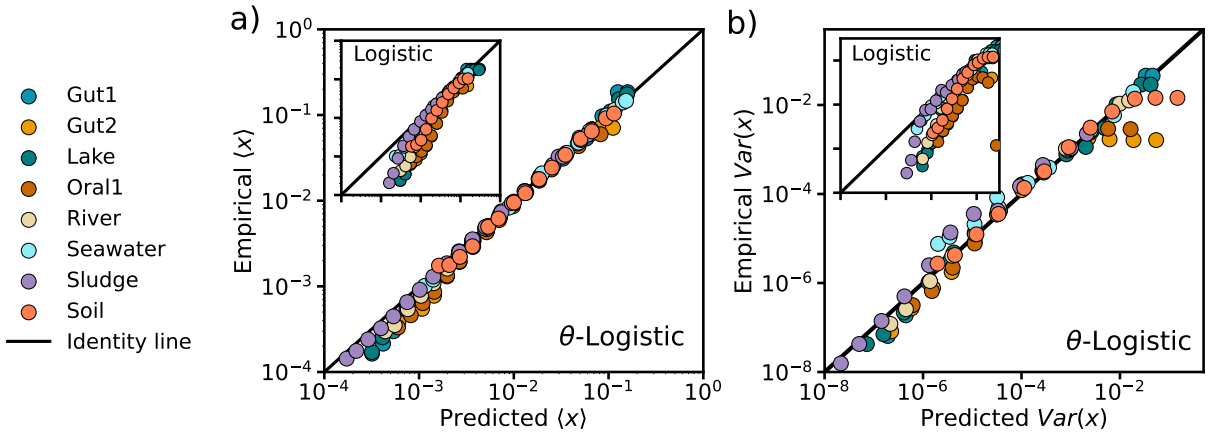

FIG. 4. Plot of a) the abundance average; and b) the variance for different biomes [24], as obtained from Eq. (9) with the inferred parameters, vs. their corresponding empirical values. Dots provide a coarse-grained visualization of these magnitudes, with error bars intentionally left aside for the sake of clarity (see [32], Figs. S2–S5). Insets show the same test applied to the SLM (with logistic growth). Its tendency to underestimate both magnitudes across most of the population enforces the conclusion that microbes rarely grow logistically.

one of the two patterns. This rules out the possibility of sublinear growth, and therefore of a possible mechanism to enhance the stability of highly diverse ecosystems [7]. Thus, whether this is a plausible mechanism or just an artifact as it has been recently argued [10], is a controversy that has no relevance for microbial communities.

Our findings impact another common macroecological pattern: Taylor's law. It has been argued [16] that the variance and the mean across microbial species are related by the power law $\sigma_x^2 \propto \langle x \rangle^2$. In a previous work [22] we showed that interactions introduce significant variability, albeit this law represents a general trend. Different growing mechanisms introduce an extra source of variability. As a matter of fact, given the expression (9) for the moments of the AFD, the ratio $\sigma_x^2/\langle x \rangle^2$ depends on θ, and this contributes an additional source of variability to Taylor's law. How much of the variability observed in empirical data is due to this effect or to the existence of interactions is a question that deserves to be explored.

We have performed our analysis on the available microbial datasets as if species were isolated from each other. We are aware that the presence of interactions might impact the growing behavior of each species and provide an alternative explanation for the non-logistic growth and it is not unreasonable to argue that the parameter θ could be an effective proxy for those interactions. Addressing this question theoretically is problematic because inferring interactions is already a complex problem by itself [23, 41, 42]. Accordingly, the best approach to decide on the growth behavior of microbes would be to carry out specific experiments to determine how populations of different microbial species grow in isolation from the rest of the species—as opposed to how they do within their communities.

We cannot resist adding a few thoughts on the problem of the stability of highly diverse microbial communities. Having ruled out the alleged mechanism based on a sublinear growth, and taking into account different alternatives proposed in the literature [43–48], as well as recent theoretical speculations on microbial ecosystems [49], we advocate for the spatial distribution as the most plausible mechanism favoring the increase of diversity observed in microbiomes. A spatial distribution of microbes can be regarded as a metacommunity, where many smaller communities coexist. Their small size reduces competition, and the distribution of species across them enhances biodiversity [48, 49]. Whether this can be proved to be a valid mechanism, theoretically and empirically, is a very interesting open problem.

We would like to finish with a word of caution. In line with Murray's remark on the logistic equation, and despite the analyses we have presented here, we do not think that Eq. (3) should be taken too literally either. The equation is nothing but an effective description of microbial dynamics that is more accurate than the simple logistic law. This means that we should not worry too much about attributing a biological meaning to the parameter θ. Microbial models are no more—but not less—than useful tools to describe real ecosystems. Nonetheless, they provide valuable insights about bacteria interactions and spatial organization, so we hope that this work helps trigger new experiments beyond the recollection of species abundances.

ACKNOWLEDGMENTS

We sincerely thank Dr. Pablo Catalán for very fruitful discussions, and for his help with the related bibliographic research on the biological significance of the θ parameter and the microscopic mechanisms behind population growth. This work has

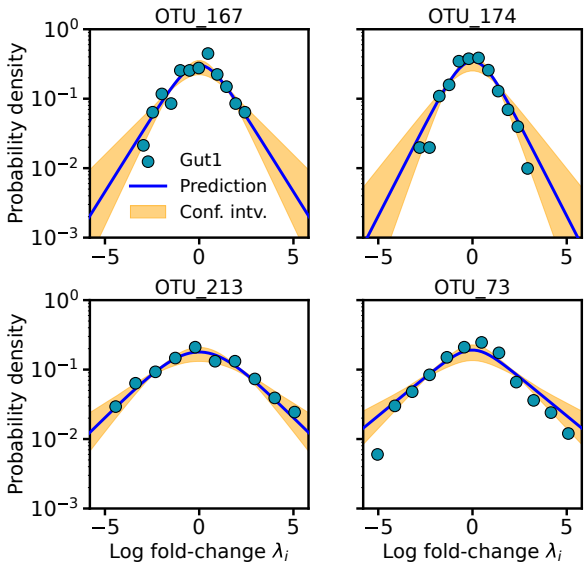

FIG. 5. Logarithmic fold-change (see main text) distribution (27) for four randomly chosen species of the Gut1 microbiome [24]. Lines are the inferred distributions; orange regions span 95% confidence intervals for each λ value; dots are otained from empirical data. The good agreement assesses the accuracy of the AFD in reproducing the data. (In cross-sectional data, where autocorrelations are negligible, this distribution follows from the AFD; see Appendix D.)

been supported by grants PID2022-141802NB-I00 (BASIC), PID2021-128966NB-I00, and PID2022-140217NB-I00, funded by MICIN/AEI/10.13039/501100011033 and by "ERDF/EU A way of making Europe".

Appendix A: The Gompertz limit

We shall derive this limit from Eq. (16). When $\theta \to 0$ we have $\beta \to \infty$. In this limit we can use the asymptotic expansions

$$\log \Gamma(\beta) = \left(\beta - \frac{1}{2}\right) - \beta + \log\left(\sqrt{2\pi}\right) + o(1), \tag{A1}$$

$$\psi(\beta) = \log \beta - \frac{1}{2\beta} + o\big(\beta^{-1}\big), \tag{A2}$$

$$\psi'(\beta) = \frac{1}{\beta} + \frac{1}{2\beta^2} + o\big(\beta^{-2}\big), \tag{A3}$$

from which u and e^u become

$$\begin{aligned} u &= \log \beta + \frac{\epsilon z}{\beta^{1/2}} - \frac{1}{2\beta} + o\big(\beta^{-1}\big), \\ e^u &= \beta \exp\left[-\frac{1}{2\beta} + \frac{\epsilon z}{\beta^{1/2}} + o\big(\beta^{-1}\big)\right] \\ &= \beta - \frac{1}{2} + \epsilon \beta^{1/2} z + \frac{z^2}{2} + o(1). \end{aligned} \tag{A4}$$

Therefore

$$\log \bar{P}(z) = -\log\left(\sqrt{2\pi}\right) + \frac{z^2}{2} + o(1), \tag{A5}$$

the equation of a lognormal distribution.

Appendix B: Skewness of the exp-gamma distribution

Given that $\langle z \rangle = 0$ and $\mathrm{Var}(z) = 1$, the skewness coefficient of $\bar{P}(z)$ can be obtained as

$$\gamma = \langle z^3 \rangle = \frac{\sqrt{\psi'(\beta)}}{\Gamma(\beta)} \int_{-\infty}^{\infty} z^3 \exp\left[\beta u - e^u\right] dz. \tag{B1}$$

In terms of the variable $u = \psi(\beta) + \varepsilon\sqrt{\psi'(\beta)}z$,

$$\gamma = \frac{\varepsilon}{\psi'(\beta)^{3/2}\Gamma(\beta)} \int_{-\infty}^{\infty} \left[u - \psi(\beta)\right]^3 \exp\left[\beta u - e^u\right] du. \tag{B2}$$

Finally, since [cf. Eq. (13)]

$$\int_{-\infty}^{\infty} u^k \exp\left[\beta u - e^u\right] du = \frac{\partial^k \Omega}{\partial \beta^k} = \Gamma^{(k)}(\beta), \tag{B3}$$

then

$$\begin{aligned} \gamma &= \frac{\varepsilon}{\psi'(\beta)^{3/2}} \left[\frac{\Gamma'''(\beta)}{\Gamma(\beta)} - 3\psi(\beta)\frac{\Gamma''(\beta)}{\Gamma(\beta)} + 2\psi(\beta)^3\right] \\ &= \frac{\varepsilon \psi''(\beta)}{\psi'(\beta)^{3/2}}. \end{aligned} \tag{B4}$$

Appendix C: Bayesian inference implementation

Bayesian inference was performed in *R* using *JAGS*, a software for analysis of Bayesian graphical models using Gibbs sampling [38]. *JAGS* is a very intuitive software that allows chaining distributions of different parameters that depend on each other, so it does not require us to know a closed expression for the likelihood (18). Figure 6 is a pseudo-code of the *JAGS* algorithm employed in the hierarchical Bayesian inference described in Sec. IV.

Appendix D: Log-fold change distribution

The log-fold change $\lambda = \log(x_{s+1}/x_s)$ is a frequently employed metric to measure these fluctuations in microbial abundances around a stationary state [17, 39]. In its definition, x_s and x_{s+1} denote the abundances of a taxon at consecutive samples. In our study, focused on cross-sectional data, the samples are collected from various communities: thus, s and $s+1$ merely label different samples. Thus, the so-called log-fold change distribution

```
for b = 1 to S do
    #Prior for parameters θ
    θ[b] ~ dgamma(ν0, k0)
    #Priors for hyper-parameters ν0, k0
    ν0 ~ dexp(0.1)
    k0 ~ dexp(0.1)
    #Priors for parameters a, β
    a[b] ~ dexp(0.1)
    β[b] ~ dexp(0.1)
    for t = 1 to Nc do
        for b = 1 to S do
            #Likelihood (Ns: sampling depth)
            n[t, b] ~ dpois(x[t, b] · Ns[t])
            #Generalized gamma AFD
            x[t, b] ~ dgen.gamma(β[b], a[b], θ[b])
        end for
    end for
end for
```

FIG. 6. JAGS pseudocode for the hierarchical Bayesian inference described in Sec. IV.

(LFD) is a result of the AFD. In order to show this, we used the variable u defined by Eq. (11), in terms of which

$$\lambda = \frac{1}{\theta}(u_{s+1} - u_s). \tag{D1}$$

Coming from two different samples, u_s and u_{s+1} are independent random variables with probability density (12). Therefore,

$$\begin{aligned} P(\lambda) &= \int_{-\infty}^{\infty} du\, \tilde{P}(u) \int_{-\infty}^{\infty} du'\, \tilde{P}(u')\delta\left(\lambda - \frac{u'-u}{\theta}\right) \\ &= |\theta| \int_{-\infty}^{\infty} \tilde{P}(u)\tilde{P}(u+\theta\lambda)\, du \\ &= \frac{|\theta| e^{\beta\theta\lambda}}{\Gamma(\beta)^2} \int_{-\infty}^{\infty} \exp\left[2\beta u - \left(1+e^{\theta\lambda}\right)e^{u}\right]\, du. \end{aligned}$$

Changing to the variable $t = \left(1+e^{\theta\lambda}\right)e^{u}$,

$$P(\lambda) = \frac{|\theta|}{\Gamma(\beta)^2}\left(\frac{e^{\theta\lambda/2}}{1+e^{\theta\lambda}}\right)^{2\beta} \int_0^{\infty} t^{2\beta-1}e^{-t}\, dt, \tag{D2}$$

which immediately yields (27).

[1] P.-F. Verhulst, Corresp. Math. Phys. **10**, 113 (1838).
[2] J. D. Murray, *Mathematical Biology: I. An introduction* (Springer, New York, 2002).
[3] G. J. Kimmel, J. West, M. Damaghi, A. R. Anderson, and P. M. Altrock, arXiv preprint arXiv:2108.10000 (2021).
[4] N. C. Petroni, S. De Martino, and S. De Siena, J. Phy. A: Math. Theor. **53**, 445005 (2020).
[5] A. K. Laird, Brit. J. Cancer **18**, 490 (1964).
[6] C. Vaghi, A. Rodallec, R. Fanciullino, J. Ciccolini, J. P. Mochel, M. Mastri, C. Poignard, J. M. Ebos, and S. Benzekry, PLoS Comp. Biol. **16**, e1007178 (2020).
[7] I. A. Hatton, O. Mazzarisi, A. Altieri, and M. Smerlak, Science **383**, eadg8488 (2024).
[8] R. M. May, Nature **238**, 413 (1972).
[9] K. S. McCann, Nature **405**, 228 (2000).
[10] G. Aguade-Gorgorio, I. Lajaaiti, J.-F. Arnoldi, and S. Kefi, bioRxiv 10.1101/2024.06.03.597151 (2024).
[11] B.-E. Sæther, S. Engen, and E. Matthysen, Science **295**, 2070 (2002).
[12] R. Sibly, D. Barker, M. Denham, J. Hone, and M. Pagel, Science **309(5734)**, 607 (2005).
[13] J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman, Chem. Biol. **5**, R245 (1998).
[14] H. L. Steele and W. R. Streit, FEMS Microbiol. Lett. **247**, 105 (2005).
[15] A. V. Mardanov, V. V. Kadnikov, and N. V. Ravin, in *Metagenomics: Perspectives, Methods, and Applications*, edited by M. Nagarajan (Academic Press, Cambridge, Massachusetts, 2018) pp. 1–13.
[16] J. Grilli, Nat. Commun. **11**, 4743 (2020).
[17] B. W. Ji, R. U. Sheth, P. D. Dixit, K. Tchourine, and D. Vitkup, Nature Microbiol. **5**, 768 (2020).
[18] L. Descheemaeker and S. De Buyl, eLife **9**, e55650 (2020).
[19] S. Zaoli and J. Grilli, Science Adv. **7**, eabj2882 (2021).
[20] J. Hu, D. R. Amor, M. Barbier, G. Bunin, and J. Gore, Science **378**, 85 (2022).
[21] W. R. Shoemaker and J. Grilli, eLife **12**, RP89650 (2024).
[22] J. Camacho-Mateu, A. Lampo, S. Ares, and J. A. Cuesta, Phys. Rev. E **110**, 044402 (2024).
[23] J. Camacho-Mateu, A. Lampo, M. Sireci, M. A. Muñoz, and J. A. Cuesta, Proc. Natl. Acad. Sci. (USA) **121**, e2309575121 (2024).
[24] A. L. Mitchell, M. Scheremetjew, H. Denise, S. Potter, A. Tarkowska, M. Qureshi, G. A. Salazar, S. Pesseat, M. A. Boland, F. M. I. Hunter, P. Ten Hoopen, B. Alako, C. Amid, D. J. Wilkinson, T. P. Curtis, G. Cochrane, and R. D. Finn, Nucleic Acids Res. **46**, D726–D735 (2018).
[25] F. J. Richards, J. Exp. Bot. **10**, 290 (1959).
[26] B. Gompertz, Phil. T. R. Soc. Lond. **115**, 513 (1825).
[27] R. Toral and P. Colet, *Stochastic Numerical Methods* (Wiley-VCH, Weingheim, Germany, 2014).
[28] S. M. Gibbons, S. M. Kearney, C. S. Smillie, and E. J. Alm, PLoS Comput. Biol. **13**, e1005364 (2017).
[29] R. Wolff, W. Shoemaker, and N. Garud, mBio **14**, e02502 (2023).
[30] S. Redner, Am. J. Phys. **58**, 267 (1990).
[31] M. A. Muñoz, F. Colaiori, and C. Castellano, Phys. Rev. E **72**, 056102 (2005).
[32] See Supplemental Material at [URL to be inserted by publisher].
[33] B. J. McGill, Global Ecol. Biogeogr. **28**, 6 (2019).
[34] A. Shade, R. R. Dunn, S. A. Blowes, P. Keil, B. J. Bohannan, M. Herrmann, K. Küsel, J. T. Lennon, N. J. Sanders, D. Storch, and J. Chase, Trends Ecol. Evol. **33**, 731 (2018).
[35] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1965).

[36] M. E. Gilpin and F. J. Ayala, Proc. Natl. Acad. Sci. (USA) **70**, 3590 (1973).

[37] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. (Chapman and Hall/CRC, Boca Raton, 2013).

[38] K. Kellner, M. Meredith, and M. K. Kellner, A Wrapper Around 'rjags' to Streamline 'JAGS' Analyses. R Package Version **1** (2019).

[39] A. B. George and J. O'Dwyer, Proc. Natl. Acad. Sci. (USA) **120**, e2215832120 (2023).

[40] X.-W. Wang and Y.-Y. Liu, Phys. Rev. Res. **5**, 013004 (2023).

[41] A. Carr, C. Diener, N. S. Baliga, and S. M. Gibbons, ISME J. **13**, 2647 (2019).

[42] S. Pinto, E. Benincà, E. H. van Nes, M. Scheffer, and J. A. Bogaards, PLoS Comput. Biol. **18**, e1010491 (2022).

[43] P. Chesson, Annu. Rev. Ecol. Syst. **31**, 343 (2000).

[44] A. Mougi and M. Kondoh, Science **337**, 349 (2012).

[45] S. Allesina, J. Grilli, G. Barabás, S. Tang, J. Aljadeff, and A. Maritan, Nat. Commun. **6**, 7842 (2015).

[46] J. Grilli, T. Rogers, and S. Allesina, Nat. Commun. **7**, 12031 (2016).

[47] E. Bairey, E. D. Kelsic, and R. Kishony, Nat. Commun. **7**, 12285 (2016).

[48] D. Gravel, F. Massol, and M. A. Leibold, Nat. Commun. **7**, 12457 (2016).

[49] M. Mattei and A. Arenas, Phys. Rev. E **110**, 014404 (2024).