Quantifying phenotypic variability and fitness in finite microbial populations

Ethan Levien¹, Jane Kondev², and Ariel Amir³

^{1,3} John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA,

^{1,2} Department of Physics, Brandeis University, Waltham, MA 02453 USA

Summary

In isogenic microbial populations, phenotypic variability is generated by a combination of intrinsic factors, specified by cell physiology, and environmental factors. Here we address the question: how does phenotypic variability of a microbial population affect its fitness? While this question has previously been studied for exponentially growing populations, the situation when the population size is kept fixed has received much less attention. We show that in competition experiments with multiple microbial species, the fitness of the population can be determined from the distribution of phenotypes, provided all variability is due to intrinsic factors. We then explore how robust the relationship between fitness and phenotypic variability is to environmental fluctuations. We find that this relationship breaks down in the presence of environmental fluctuations, and derive a simple formula relating the average fitness of a population to the phenotype distribution and fluctuations in the instantaneous population growth rate. Using published experimental data we demonstrate how our formulas can be used to discriminate between intrinsic and environmental contributions to phenotypic diversity.

Introduction

A central problem in microbiology is understanding how measurable phenotypic traits such as a cell’s size, growth rate, or expression of a gene affect population dynamics [1, 2, 3, 4]. This problem is made difficult by the fact that there is rarely a well-defined mapping from genotype to phenotype, instead, a single genotype can give rise to a distribution of phenotypic traits throughout a population’s history. This distribution can be generated at the cellular level due to *intrinsic* factors, such as the stochasticity of biochemical reactions [5, 6, 7, 8, 9, 10], or due to *environmental* fluctuations that affect the entire population [11, 12, 13, 14, 15]. It is now well established that a population’s long term growth rate can not be determined by averaging single-cell traits, such as the growth rates or generation times of cells, but is instead determined by the variability and heritability of phenotypes measured throughout a population. The discrepancy is an example of *non-ergodicity* and results from epigenetic heritability of phenotypic traits that is present in all populations. It is therefore necessary to quantify the distribution of a phenotypic trait, and not just the average, in order to deduce how phenotypes affect fitness.

Measurements of phenotypic variability are commonly obtained in microfluidic devices, such as the *dynamic cytometer* pictured in Figure 1 (A), where a fixed number of cells are grown for long periods of time in controlled conditions [16, 17, 18, 19]. Such experimental setups allow one to track single-cell dynamics over hundreds of generations, a task that would be impossible if the population continued to grow exponentially. By tracking cells in a finite population over many generations, detailed information about the variability and epigenetic heritability of phenotypic traits can be obtained. How does one relate this information to some measure of the population’s fitness? Many previous studies have explored how population growth is related to the single-cell dynamics of an exponentially growing population in a constant environment [1, 20, 21, 22, 23, 19, 24, 25]. In the setting of exponential growth, a proxy for fitness is the population growth rate Λ , defined by the relation $N \sim e^{\Lambda t}$, where N is the number of cells at time t . Most notably, it was shown that the population growth rate, Λ , can be computed from the distribution of single-cell generation times taken over the entire history of the population using the *Euler-Lotka equation*, shown in Figure 1 (B). The Euler-Lotka equation establishes a link between the distribution of phenotypes and the fitness of the population, but it generally requires knowledge of the entire *lineage tree* depicted in Figure 1 (B). In the finite population, one cannot obtain the entire lineage tree because most of the cells are expelled from the population. Instead, one must analyze data from a “pruned” lineage tree such as the one depicted

in Figure 1 (B). The objective of our study is to understand how the fitness of a finite population can be quantified from this, more limited, lineage data.

By studying the relationship between the statistical properties of lineage trees in exponentially growing and finite cultures, we derive an analog of the Euler-Lotka equation that requires only knowledge of the pruned lineage data. Just as one can compute the population growth rate in an exponentially growing culture by measuring the slope of $\ln N(t)$, in the finite culture the population growth rate can be estimated directly by observing the rate at which cells are expelled from the culture, or the *dilution rate*. Importantly, the accuracy of these estimates depends on whether the phenotypic variability comes from intrinsic or environmental factors. The former arises due to stochasticity at the cellular level that causes different cells throughout the population to have distinct phenotypic traits, while the latter is a result of environmental fluctuations affecting unrelated cells simultaneously; see Figure 1 (C). We find that in the presence of environmental fluctuations, the population growth rate cannot be determined solely from the phenotype distribution, and we derive a formula for the deviations of the population growth rate from the prediction of the Euler-Lotka equation in this setting.

Results

General model of intrinsic variability

We start by introducing a general modeling framework allowing us to study intrinsic variability in an arbitrary (and possibly multi-variate) phenotypic trait x . We assume each cell is born with a value of x drawn from a distribution $f(x|x')$ depending on the phenotype x' of the mother cell. For example, x might represent the initial concentration of a protein which is expressed stochastically over the course of a cell cycle, or some macroscopic phenotype, such as the growth rate. After a time $\tau(x)$, the cell divides and produces two new cells with generation times drawn from the same distribution. This framework can capture any possible form of *intrinsic* (as opposed to environmental) generation time variability through the appropriate choices of the phenotype x and distribution f . Therefore, existing models can be derived as specific cases of this framework. These include models where division is asymmetric [26, 27, 28, 25] and where the relationship between successive generation times is non-monotonic and nonlinear, such as the kicked cell-cycle model used to describe circadian rhythms [29].

As a concrete example, we will often refer to a specific model of cell-size regulation, which is described in detail in Methods and [1]. The central assumptions of this model are that single cells increase in volume exponentially with time and divide (up to some noise) at a specified volume. The strategy for selecting the volume at division is known as the *cell-size regulation strategy*. In the notation of our modeling framework, the phenotype $x = (\lambda, v_d, v_b)$ where λ , v_d and v_b are the cell's growth rate, volume at division and volume at birth respectively. The transition operator $f(x|x')$ is then dependent on the specific cell-size regulation strategy (see Methods).

Growth rate analysis in the finite culture

In the setting where a population described by our model grows exponentially, the problem of determining the population growth rate Λ from $f(x|x')$ has been explored in a number of recent studies for specific choices of the observable x [30, 20, 21]. In our general modeling framework, Λ , along with the distribution of phenotypes among all cells in the exponentially growing population tree, $\psi_{\text{tree}}(x)$, obeys the recursive equation (see Methods)

$$\psi_{\text{tree}}(x) = 2 \int_0^\infty f(x|x') \psi_{\text{tree}}(x') e^{-\Lambda \tau(x')} dx'. \quad (1)$$

Integrating with respect to x retrieves the Euler-Lotka formula [31],

$$\frac{1}{2} = \langle e^{-\Lambda \tau} \rangle_{\text{tree}} \quad (2)$$

Equations 1 and 2 determine the unknowns $\psi_{\text{tree}}(x)$ and Λ in terms of the model of phenotypic variability, which is encoded in $f(x|x')$. Given that $f(x|x')$ is sufficiently smooth in x and x' , uniqueness of these

Figure 1: Population dynamics, lineage data and sources of phenotypic variability

(A) (left) In a population growing exponentially the population growth rate, Λ , a proxy for fitness, can be computed from temporal dynamics of the number of cells. (right) In a finite culture containing two species, the fraction of one species, ϕ_1 , typically evolves according to the logistic growth equation with rate parameter $\Delta\Lambda$ representing the difference in growth rates of the individual species.

(B) Lineage data obtained from the different experiments in (A). Each vertical line represents a cell and the length of the line its generation time. In the exponentially growing populations, the lineage data can be related to the population growth rate, Λ , via the Euler-Lotka equation (orange box). The average $\langle \cdot \rangle_{\text{tree}}$ refers to the average over every branch in the tree. In finite culture experiments it is unclear how to obtain the distribution of all cells on the tree because most of the cells are expelled from the population.

(C) (left) A diagram of a device used to collect single cell lineage data. (middle) A model of cell proliferation with only intrinsic variability. In such a model, the generation times along a lineage obey a Markov process. (right) Temporal fluctuations in the environment (gray and white areas) may also induce phenotypic variability.

solutions can be established using the theory of integral equations [31], and we assume throughout that these conditions are satisfied. As discussed in the introduction, only in the simple case where phenotypes are uncorrelated across generations ($f(x|x') = f(x)$) is this distribution equal to the distribution of phenotypes along a lineage, which we refer to as the *lineage distribution* and denote by $\psi_{\text{lin}}(x)$; see Figure 3 (A). Hence while the lineage distribution may be the most natural distribution to consider because it does not require

us to grow the population exponentially, it is generally less informative if we are interested in the population growth.

We now consider a model where N , the number of cells in the population, is fixed. In order to achieve this, every division event must correspond to a cell being expelled from the culture, which is approximately the case in a microfluidic device such as the one pictured in Figure 1 (B). In a simulation, this can be implemented by picking a random cell from the population at each division event and removing it from the culture. The resulting process is known as a *Moran process* in the mathematics literature [32]. Since the population is not growing, we can no longer obtain Λ from the exponential growth curves (given by the equation in Figure 1 (A)). Instead, one can measure growth in terms of the rate per cell at which cells are expelled from the population, or the *dilution rate* Λ_D . Intuitively, if N is sufficiently large, the continuous expulsion of cells from the culture should have no influence on the relative frequencies of the phenotypes, and therefore the instantaneous rate at which cell volume is accumulated should be equal to Λ . On the other hand, since the size of the culture remains fixed, this must be balanced by the rate at which cells are expelled, implying $\Lambda = \Lambda_D$. It follows that Λ_D should satisfy the same relation as Λ , namely,

$$\psi_{\text{tree}}^F(x) = 2 \int_0^\infty f(x|x') \psi_{\text{tree}}^F(x') e^{-\Lambda_D \tau(x')} dx', \quad (3)$$

where we are now using ψ_{tree}^F to represent the distribution of phenotypes over all cells throughout the history of the culture; see Figure 3 (B). As discussed in Methods, ψ_{tree}^F can also be interpreted as the distribution of phenotypes over all cells in a snapshot that have just been born. In general we will use the superscript F to indicate a distribution in the finite culture, but we omit this when there is no ambiguity.¹ For a rigorous derivation of Equation 3 we refer to Methods. When phenotypes are not correlated across generations, $f(x|x') = f(x)$, hence it follows from Equation 3 that $\psi_{\text{tree}}^F(x) = f(x)$. We may integrate with respect to x and replace Λ_D with Λ in order to obtain Equation 2. This relation was previously obtained Powell in a model without correlations across generations. For this reason, his derivation [31] does not clarify which distribution the average should be taken with respect to.

Phenotype distribution predicts selective advantage in the finite culture

Having established that the Euler–Lotka equation can be generalized to the setting of the a finite culture, we now consider two species competing in a finite culture and show that the population growth rate is also a measure of fitness in this setting. We will once again assume that the number of cells (N) is sufficiently large so that any effects of order $1/N$ may be neglected. Before considering these competition dynamics directly, suppose that both species are grown separately in environments with unlimited space and nutrients, see Figure 1 (A). Under these conditions, the number N_i of cells in species i will then grow exponentially at some rate Λ_i , hence $N_i \sim e^{\Lambda_i t}$. A simple calculation reveals that the fraction of species 1, $\phi_1 = N_1/(N_1 + N_2)$, evolves according to the logistic growth equation

$$\frac{d}{dt} \phi_1 = \phi_1(1 - \phi_1) \Delta \Lambda, \quad \Delta \Lambda = \Lambda_1 - \Lambda_2. \quad (4)$$

The quantity $S = \Delta \Lambda / \Lambda_2$ is known as the *selective advantage* of species 1 over species 2 and the sign of the quantity determines which species eventually takes over the culture. We now imagine placing both species in a finite culture that can hold exactly N cells; see the left panel in Figure 1 (B). As before, this is implemented by removing a randomly selected cell from the culture every time a cell division occurs. We will assume that nutrients are not limiting growth so that the rate at which cells of a given species divide does not depend on the relative abundances of each species. Heuristically, we can obtain the evolution of the frequency $\phi_1(t)$ from the exponentially growing population by uniformly sampling N cells from the exponentially growing population at each time, therefore it is plausible to assume that Equation 4 describes the evolution of the frequency ϕ_1 in the finite culture as well. A systematic derivation of Equation 4 for two phenotypically heterogeneous species along with the generalization to the case of M species is given in Methods and the agreement between the logistic growth Equation 4 and numerical simulations of two species competing in a finite culture is shown in Figure 2.

¹We emphasize that when computing $\psi_{\text{tree}}(\tau)$ one needs to know the generation times of the cells that are expelled from the culture, not the age at which they are expelled.

In the context of the cell-size regulation model, previous work has shown that if the mean growth rate is fixed along a lineage, variability in single-cell growth rates decreases the population growth rate for biologically realistic parameter values [20]. Therefore, it follows from our analysis that in a finite culture variability in growth rates will reduce a species' competitive advantage. This is shown numerically in Figure 2, where we have simulated the competition between two species with different levels of growth-rate variability, but identical mean growth rates.

Figure 2: Population growth predicts fitness in the finite culture

(A) Sample path of $\phi_1(t)$ starting with a culture equal parts species 1 and 2. Each population is modeled using the cell-size regulation model. We have non-dimensionalized by taking $\langle \lambda \rangle_{\text{lin}} = 1$ and define a generation as the doubling time of a cell with the average growth rate, hence 1 generation = $\ln(2)$ units of time.

(B) $\phi_1(t)$ at $t = 20$ generations for two populations with the same average growth rate and different values of σ_i^2 , the variance in growth rates. We have fixed $\sigma_2^2 = 0.1$ and increased σ_1 . For both species growth rates are uncorrelated across generations. Each circle represents the average of 1000 simulations for the same parameter values.

The relationship between methods of sampling phenotype data

The tree distribution, which is required to compute the population growth rate using Equation 2, is only one of a number of ways to quantify phenotypic variability in a lineage tree, yet it is not always the most convenient to work with. For example, in the device pictured in Figure 1 (A) it is difficult to track the generation times of cells that are expelled from the culture, yet these are needed in order to compute $\psi_{\text{tree}}^F(\tau)$. Therefore, it is useful to understand the relationship between the different distributions that can be obtained from the lineage tree. A related question is how distributions of phenotype over subsets of a lineage tree differ in exponential and finite cultures. In some cases the exponential and finite cultures are equivalent. For example, consider the current cells distribution, $\psi_{\text{snap}}(x)$. This is the distribution of phenotypes in a snapshot of the population at a given time; see Figure 3. One can show that in both finite and exponentially growing populations taking a snapshot of phenotypes yields the same distribution (see Methods). However, as we will show below this equivalence of distributions does not hold for phenotypes sampled over the ancestors of the current cells in the population.

Another important distribution is given by the expression $2\psi_{\text{tree}}^F(x)e^{-\Lambda_D\tau(x)}$; this term is twice the integrand of the Euler-Lotka equation, which implies it must be a probability distribution, as it is normalized. As was previously noted by Powell [31] in the context of a model with uncorrelated generation times, this can be interpreted as the distribution of x over all cells that divide before being expelled from the culture. Following Powell's terminology, we define the *carrier distribution* as

$$\psi_{\text{car}}^F(x) = 2\psi_{\text{tree}}^F(x)e^{-\Lambda_D\tau(x)}. \quad (5)$$

(A) Lineage data from an exponentially growing culture

(B) Lineage data from a finite culture

Figure 3: **Methods of sampling lineage trees**

(A) (left) The various ways of sampling phenotypes from a lineage tree obtained in an exponential culture. The gray horizontal line on the right panel indicates the current time of observation. The blue highlighted lines correspond to cells that are sampled to obtain the distribution indicated above the lineage tree. For example, to obtain ψ_{ans} we sample all the cell in the tree excluding those that are currently alive. (B) The same image for a finite culture. The red dots represent cells that are expelled from the culture before they are able to divide.

This relationship is verified numerically for the size-control model in Figure S.1. The subset of the lineage tree which gives this distribution is depicted in Figure 3 (B). To derive Powell's interpretation of the carrier distribution in the more general setting, first note that $e^{-\Lambda_D \tau(x)}$ is the probability that a cell stays in the culture until it divides, given that it has phenotype x . Also, the probability that any cell throughout the history of the population divides in the culture must be $1/2$ in order for the number of cells to remain fixed. The interpretation of $\psi_{\text{car}}^F(x)$ then follows from an application of Bayes' Theorem where $\psi_{\text{tree}}^F(x)$ and $\psi_{\text{car}}^F(x)$ play the role of the *prior* and *posterior*, respectively. In the exponentially growing culture the distribution $2\psi_{\text{tree}}(x)e^{-\Lambda \tau(x)}$ has a different interpretation; it is equal to the distribution over all the current cell's ancestors, $\psi_{\text{ans}}(x)$ (this has been referred to as the *branch distribution* in [20]). This distribution is obtained by sampling all cells in the tree excluding those that are currently alive, as shown in Figure 3 (B). The fact that $\psi_{\text{car}}(x) = \psi_{\text{ans}}(x)$ suggests there may be an alternative interpretation of these distributions that is common to both types of cultures. In fact, in both the continuous culture and exponentially growing culture, the carrier (or branch) distribution can be interpreted as the distribution of phenotypes of cells that have just divided, or the *division distribution* $\psi_{\text{div}}(x)$ (see Methods for a detailed discussion of this distribution).

In practice the distribution $\psi_{\text{car}}^F(x)$ may be easier to obtain than $\psi_{\text{tree}}^F(x)$ because sampling $\psi_{\text{tree}}^F(\tau)$

requires knowledge of generation times of cells that are expelled from the culture. In a microfluidic device these cells may be difficult to track, and additionally, their growth may be not be statistically identically to those cells that remain in the culture due to the different conditions in and out of the microfluidic chamber where cells are tracked. This motivates us to develop a method for estimating the population growth rate, Λ , from the available lineage data. To this end, we utilize the normalization of $\psi_{\text{tree}}^F(x)$ to obtain

$$2 = \left\langle e^{\Lambda_{\text{car}} \tau(x)} \right\rangle_{\text{car}}, \quad (6)$$

where we will denote the estimate of Λ obtained from this equation as Λ_{car} in order to distinguish it from the dilution rate estimate Λ_D .

Figure 4: **The relationship between various methods for sampling lineage data**

(A) The relationship between distributions in the finite and exponentially growing culture.

(B) The relationship between different ways of sampling the phenotype distributions in the finite culture. An expanded table that includes the birth and death distributions is shown in Figure S.3.

(C) Plots of the various distributions of single-cell growth rates obtained from simulations of the cell-size regulation model with $N = 10^4$, $CV_\lambda^2 = 0.25$, $\langle \lambda \rangle = 1$, $v_{\text{div}} = 2$, $\alpha = 1$ and $\rho_\lambda = 0.2$. The lineage distribution, which can be computed analytically for this model, is also shown. For the ancestral distribution, we have plotted both the histogram of $\ln \lambda$ over all ancestral cells (red dashed line) as well as evaluated the right hand side of Equation 7 using the carrier and tree distributions (purple line).

The ancestral distribution

In the exponential culture the ancestral distribution can be interpreted as the distribution over all the current cell's ancestors. How might the distribution over the ancestral cells in the finite culture be related to $\psi_{\text{ans}}(x)$ in the exponential culture (or equivalently, $\psi_{\text{car}}(x)$)? Because all cells have a common ancestor we can compute $\psi_{\text{ans}}(x)$ by taking only the ancestor cells of this common ancestor; as we trace back in time far enough the more recent cells make up a negligible fraction of the observed phenotypes. $\psi_{\text{ans}}(x)$ is therefore

the distribution of phenotypes obtained by looking *backwards* along a lineage beginning from an arbitrary cell in the culture. If phenotypes are not correlated across generations, then the phenotypes sampled along the lineage are independent and this is therefore equivalent to sampling cells in the population that have divided in the culture; as we saw before, this is $\psi_{\text{car}}(x)$. However, if there are correlations across generations, we need to consider how the ancestor's phenotypes along a lineage are biased by the decedent's phenotypes. This is captured by the recursive formula for $\psi_{\text{ans}}(x)$ (derived in Methods):

$$\psi_{\text{ans}}^F(x) = \psi_{\text{car}}^F(x) \int_0^\infty f(x'|x) \frac{\psi_{\text{ans}}^F(x')}{\psi_{\text{tree}}^F(x')} dx' \quad (7)$$

From this equation, it is easy to see that $\psi_{\text{ans}}^F(x) = \psi_{\text{car}}^F(x)$ precisely when $f(x|x') = f(x)$, as expected from the argument above.

Our results concerning the relationship between phenotype distributions in exponential and finite cultures are summarized in Figure 4 (A) and (B). To explicitly demonstrate these relationships we performed simulations of the cell-size regulation model (discussed in Methods). Results of these simulations are shown in Figure 4 (C).

Environmental fluctuations limit the predictive power of the phenotype distribution

In the previous section we derived a relationship between the population growth rate and the lineage data under the assumption that there is only intrinsic variability of phenotypes in the populations. However, it is very rarely the case that this assumption is satisfied in natural biological settings, and even in highly controlled experiments it is difficult keep the environmental conditions constant. How do both environmental fluctuations and intrinsic variability interact to shape a population's growth rate? While the Euler-Lotka equation is no longer satisfied if the phenotype distribution is affected by environmental fluctuations, we will show that the generation time statistics can still be related to the population growth rate provided one also has information about the fluctuations in the instantaneous population growth rate.

To be concrete, we first consider an extension of the cell-size regulation model in which cells grow at a rate $\lambda[S_e]$ that depends on a time-dependent external nutrient concentration, denoted $S_e(t)$. A natural choice for the growth-rate function is Monod's law [33],

$$\lambda[S_e] = \frac{\lambda_{\text{max}} S_e}{K_m + S_e}. \quad (8)$$

Here λ_{max} is the maximum growth rate of a cell and K_m is the concentration at which a cell grows at half its maximum growth rate. As discussed in Methods, intrinsic variability is introduced by taking λ_{max} to be drawn from a normal distribution when a cell is born (with mean $\langle \lambda \rangle$ and variance σ_λ^2) while environmental fluctuations come from perturbations of $S_e(t)$ around an average value S_0 . To this end, we define the environmental perturbation $\xi(t)$ by $S_e(t) = S_0(1 + \xi(t))$. We will assume that $\langle \xi(t) \rangle = 0$ so that $\langle S_e(t) \rangle_T = S_0$ and use σ_ξ^2 to denote the strength of environmental fluctuations; σ_ξ^2 is given by the time average of $\xi(t)^2$ over a long interval. In our simulations, we take $\xi(t)$ to be an Ornstein-Uhlenbeck process; that is, $\xi(t)$ undergoes a continuous time random walk in a quadratic potential with minimum at the origin. This process is characterized by a diffusion coefficient D_ξ and a relaxation time-scale t_ξ . Importantly, $\xi(t)$ is a *global* source of variability affecting all cells equally, while any variation in λ_{max} is due to intrinsic variability in the cell; see Figure 1 (C). Recall that the assumption of the cell-size regulation model is that a cell divides upon reaching a (possibly random) size $\tilde{v}_d$; therefore, the generation time τ of a cell born at time t satisfies $\int_0^\tau \lambda[S_e(t+s)] ds = \ln(\tilde{v}_d/v_b)$. Adding environmental fluctuations to the cell-size regulation model in this way does not significantly affect the relationship between Λ_D and Λ . This can be seen numerically in Figure 5 (B) and an analytical justification is given in Methods. In contrast, the estimate Λ_{car} is a decreasing, linear function of σ_ξ^2 . The specific form of this linear relationship depends on the parameters of Monod's law according to

$$\Lambda_{\text{car}} \approx \Lambda \left(1 - \frac{\ln(2)}{2} \left[\frac{K_m}{K_m + S_0} \right]^2 \sigma_\xi^2 \right). \quad (9)$$

In Figure 5 (B) we have shown estimates of $\Lambda_{\text{car}}/\Lambda_D$ and Λ_D/Λ from numerical simulations, which are consistent with our theoretical prediction.

Equation 9 implies that the population growth rate cannot be predicted solely from the distribution of generation times, instead, one needs both generation time statistics and information about the environmental fluctuations to predict the population growth rate. It is therefore natural to ask whether there exists a general relationship that connects the population growth rate to the generation time statistics via the environmental fluctuations. In the limit where the environment evolves on a time-scale longer than the typical generation time of a cell, such a relationship is given by (see Methods)

$$\Lambda \approx \Lambda_{\text{in}} + \frac{\sigma_{\Lambda}^2 \langle \tau^2 \rangle_{\text{car}}}{2 \langle \tau \rangle_{\text{car}}}.$$
 (10)

Here, Λ_{in} is the rate at which the population would grow if it had the same distribution of generation times in the carrier distribution, but no environmental variability, while σ_{Λ}^2 is the variance in Λ . As we have already shown, Λ_{in} is simply Λ_{car} , and can therefore be computed from the carrier distribution. We have confirmed this relation numerically in Figure 5 (C). Equation 10 tells us that the phenotype distribution becomes less informative as environmental fluctuations increase. In Methods, we show that this relationship is also satisfied in a competition experiment between multiple species in a fluctuating environment, even if the environment has distinct effects on each species. In the next section we will discuss how σ_{Λ}^2 is related to the fluctuations in the dilution rate, which differ from the growth rate fluctuations in a small population.

Population size effects

Because many experimental setups used to probe bacterial physiology are performed in devices that can only support between $N = 15$ and 40 cells, it is import to quantify the order $1/N$ terms in predictions of the population growth rate. The difference between the population growth rate, Λ , and the estimate obtained from the carrier distribution, Λ_{car} , will be on the order of σ_{Λ}^2/N . To see this, observe that in the limit where there is no phenotypic variability, $\Lambda_{\text{car}} = \Lambda$ for all N . On the other hand, as $1/N$ approaches 0, Λ_{car} converges to Λ for all σ_{Λ}^2 . Therefore, we can express Λ_{car} as a Taylor expansion in σ_{Λ}^2/N where the zeroth order term is Λ . This implies Λ_{car} gives a relatively accurate approximation to Λ . On the other hand, Λ_D deviates from Λ by a factor of order $1/N$. Numerically, we have shown that

$$\Lambda_{\text{car}} \approx \Lambda, \quad \Lambda_D \approx \Lambda \left(1 + \frac{C_p}{N} \right)$$
 (11)

where C_p is a constant that is independent of the population size. For the cell-size control model, we have found $C_p \approx \ln(2)$; see Figure 5 (D) and Figure S.4. The discrepancy between Λ_D and Λ arises because, in small populations with little phenotypic variability, all the cells tend to have similar ages and generation times. Therefore, cells divide at approximately the same time. As a result, the dilution rate tends to measure the inverse doubling time of the populations, $T_d^{-1} = \Lambda/\ln(2)$. This can easily be seen in a population with no phenotypic variability, since in this case we will observe N cells expelled over a period T_d , and therefore $\Lambda_D = T_d^{-1} \ln(2)/N$.

Just as the average dilution rate depends on the populations size, the variance in growth rate, σ_{Λ}^2 , appearing in Equation 10 is distinct from the variance in the dilution rate, $\sigma_{\Lambda_D}^2 = \text{Var}(\Lambda_D(t))$. Formally, we can express the variance as a sum,

$$\sigma_{\Lambda_D}^2 = \sigma_N^2 + \sigma_{\text{env}}^2,$$
 (12)

where σ_{env}^2 represents any fluctuations that do not vanish in an infinite population. In other words, for a given population, the term σ_{env}^2 tells us how much the growth rate would fluctuate in a population with the same single-cell dynamics and environmental variability, but with an infinite number of cells. It follows immediately from this definition that $\sigma_{\text{env}}^2 = \sigma_{\Lambda}^2$, where σ_{Λ}^2 is the variance appearing in Equation 10. In many situations, it is desirable to determine σ_{env}^2 from data that is obtained in a small population where σ_N^2 is non-negligible. In these situations, Equation 10 can be used to estimate σ_{env}^2 indirectly. To do so, one must have both the lineage data and a reliable estimate of the population growth rate. As we have shown, Equation 11 can be utilized to obtain such as estimate from Λ_D . In the next section, we apply this procedure to experimental data.

Application to experimental data

We now apply our results to analyze the experimental data original reported by Hashimoto *et al.* in [16]. In different experimental conditions, *Escherichia coli* were grown in a dynamics cytometer that can support approximately 10 to 50 cells. In [16], the population growth rate was estimated by solving Euler-Lotka equation for the tree distribution (Equation 2), by using the distribution of all cells that divide in the culture (the carrier distribution) weighted by their survival probability. This procedure gives an excellent approximation to the dilution rate, but is problematic in a small population where, as the authors observe,

Figure 5: **The relationship between growth rate predictions**

(A) A summary of the different types of data and growth rate predictions that can theoretically be obtained from a finite culture. (top) The dilution rate is obtained by measuring the rate at which cells are expelled from the culture. (bottom) Single-cell tracking data can be used to obtain the carrier distribution (as well as other distribution of phenotypes).

(B) Ratio of different estimates for the population growth rate, obtained from a finite population (Λ_{car} , Λ_D), and the population growth rate of an exponentially growing population (Λ), as a function of the environmental fluctuations. In all simulations, the instantaneous dilution rate is measured by sampling the number of cell expelled from the culture over intervals of length $\Delta t = 0.2\langle\tau\rangle$ and dividing by $N\Delta t$. Here the environmental perturbation $\xi(t)$ evolves according to an Ornstein-Uhlenbeck (OU) process. In this case the best estimate is clearly Λ_D , which predicts the true population growth rate independently of the environmental fluctuations. Parameters are $N = 500$, $CV_\lambda = 0.1$, $\rho_\lambda = 0$ and $\alpha = 1$. (Continued on next page)

Figure 5: (C) The true population growth rate compared to the population growth rate obtain from Equation 10. The hue of each point represents the number of simulations with those measured growth rates (lighter points indicate more simulations). In order to compute the true value of Λ we have looked at the instantaneous change in volume over a small time interval $dt = 0.1$. Simulations were performed for the cell-size regulation model with $N = 25$, $CV_\lambda = 0.2$, $\rho_\lambda = 0.1$, $v_d = 1$, $\langle \lambda \rangle = 1$ and $\alpha = 1$. The environment was simulated using a time-discretized OU process with relaxation timescale $t_\xi = 10$ and D_ξ was varied between 0 and 0.01. In Figure S.2, we use Equation 10 on the same synthetic data to compute σ_Λ^2 instead of Λ . (D) The same as (B), but showing the predictions for the different estimates of the population growth rate as a function of the population size, N . The parameter values used are the same as (B).

the dilution rate deviates from the population growth rate. The authors also computed the mother-daughter generation time correlations. In *E. coli*, generation times are typically slightly negatively correlated with its mother’s generation time (see [1] and Methods for an explanation in terms of the cell-size regulation model). However, it was found that in certain experiments there were relatively large positive generation time correlations; see Figure 6. One explanation for such correlations is that environmental fluctuations are influencing the growth statistics. We therefore expect that the environmental variability should be correlated with the mother-daughter generation time correlations.

In order to test the predictive power of our theory, we used Equation 6 to directly solve for Λ_{car} and computed the dilution rate, Λ_D , by averaging the instantaneous dilution rate over time. Using Equation 11, we are able to transform Λ_D into Λ , giving us all the ingredients to solve for σ_Λ^2 in Equation 10, which is equivalent to σ_{env}^2 in Equation 12. We then plotted the mother-daughter correlations as a function of $\sigma_{\text{env}}^2/\Lambda_{\text{car}}^2$; see Figure 6. This suggests that the positive generation time correlations are consistent with a model including both environmental and intrinsic noise, because all experiments with positive generation time correlations correspond to positive values of σ_{env}^2 . For one data point, σ_{env}^2 is negative, which is inconsistent with our theory, although this experiment is an extreme outlier in the data set. We have also compared the experimental data to synthetic data and observed that, after excluding outliers, both data sets obey the same trend. Note that the generation time correlations caused by environmental fluctuations may be canceled by cell-size regulation or other intrinsic effects in some cases, hence the fact that the correlations do not strictly increase with σ_{env}^2 is consistent with our theory.

Discussion

In this paper, we have studied the relationship between growth of phenotypically heterogeneous populations in finite and exponential cultures. While a number of studies have explored the dynamics of phenotypically heterogeneous populations under exponential growth conditions and shown how to relate lineage statistics to the population growth rate (e.g., [1, 20, 21, 22, 23, 19, 24]), precise measurements of phenotypic variability often need to be obtained in cultures containing a fixed number of cells. Because the growth process is not ergodic (the population averages are not equivalent to single-lineage averages) it is essential to distinguish between different ways of sampling phenotypes when analyzing data from these experiments. For example, an erroneous result is obtained if the lineage statistics are used to compute the population growth rate instead of the carrier distribution (the distribution taken over all cells that are not expelled from the culture). Moreover, in some situations one is limited in what distributions can be sampled. This is the case in certain microfluidic devices where it is difficult to obtain the distribution of generation times of *all* cells because we do not have access to the generation times of cells that are expelled from the culture.

We have clarified the relationship between the different distributions in both the exponentially growing and finite culture, as well as derived a simple mapping between the various distributions in the two types of cultures. Importantly, distributions obtained in equivalent ways do not always have the same meaning in different types of cultures. In particular, the ancestral distribution in a finite culture is not equivalent to the ancestral distribution of the exponentially growing culture due to a bias towards surviving cells, yet both are obtained by looking at the history of all cells in the culture. A corollary of our mapping between the lineage statistics in different cultures is a formula for the population growth rate (given by Equation 6) that requires knowledge only of the lineage data obtained in a sufficiently large finite culture. In addition to

Figure 6: Predicting mother-daughter correlations in experimental data

The mother-daughter correlations as a function of σ_{env}^2 (computed by solving for σ_{Λ}^2 in Equation 10) for experimental data (colored points) and synthetic data (black points). The right panel shows a zoomed-in view of the region where most experimental data points lie. The synthetic data was generated by simulating the cell-size regulation model with $\text{CV}_{\lambda}^2 = 0.1$, $\langle \lambda \rangle = 1$, $\rho_{\lambda} = 0$, $t_{\xi} = 0.5$ and various values of D_{ξ} (the diffusion coefficient of the OU process describing the environmental fluctuations) ranging from 0 to 0.1. Measurement error has been added by adding Gaussian noise to the growth rate and generation time measurements. Note that this leads to some data points with negative σ_{env}^2 , which is also observed in the experimental data. The fact that $\sigma_{\text{env}}^2 > 0$ for many data points indicates that the phenotypic variability should not be interpreted as purely intrinsic. For additional information, see Figure S.3.

relating the lineage statistics to the population growth rate, we have shown that the population growth rate predicts the fitness of a species in a competition experiment.

We have also considered how environmental fluctuations influence population growth predictions, and established a model-independent relationship between the fluctuations in the population growth rate and the growth rate prediction based on the phenotype distribution. Since single cell growth rate variability has been observed to change growth rate estimates by around several percent in *E. coli* [20], even small environmental fluctuations can conceal the intrinsic effects of phenotypic variability. We have established that the phenotype distribution becomes increasingly less informative as the environmental fluctuations in the population growth rate increase in amplitude (Equation 10). Hence, in the presence of environmental variability the fitness of a population is not determined solely by the distribution of phenotypes. We have tested the predictive power of the relationship between the population growth rate, the phenotype distribution and the environmental fluctuations on synthetic and experimental data. For most experimental data points, we see that there is a relationship between the environmental fluctuations predicted from Equation 10 and the mother-daughter generation time correlations. This is consistent with the observation that environmental fluctuations tend to create positive correlations between mother and daughter cell's generation times.

Over the last decade there has been a growing interest in understanding the implications of phenotypic heterogeneity in the clinical context, yet understanding how to separate distinct sources of variability has proven to be a difficult problem [34, 11, 35]. For example, *Salmonella* grown in host macrophages exhibit increased phenotypic variability compared to *Salmonella* grown in *in vitro* cultures, yet the underlying causes of this variability, and in particular, whether it is induced by environmental fluctuations or stochastic gene expression is not known [11]. We have proposed one method for dissecting phenotypic variability in a growing population, and although we have motivated this study by data obtained in microfluidic devices, one could in principle apply these same techniques to populations grown in more complex environments. Specifically, by tracking single cell dynamics of bacteria growth in macrophages, it may be possible to deduced whether the increased phenotypic variability was caused by environmental or intrinsic factors by examining the

consistency of different growth rate predictions, as we have done here. This would require generalizing our framework to account for both spatial and temporal environmental fluctuations, which we have not included in our analysis, as much of the environmental variability is spatial in host environments [11]. Nonetheless, we believe this theory provides a promising starting point for future investigations of phenotypic variability.

In addition to data-driven applications of our theory, there are numerous theoretical questions concerning the fitness effects of phenotypic variability that we have left unanswered. For example, our analysis of the competition experiment has assumed that nutrients are not limiting growth and the dilution rate is adjusted to keep the population size fixed (such conditions can be created in a turbidostat [33]). We therefore do not account for chemostat-like conditions where species compete for limiting resources. In the future, it would be interesting to extend our analysis to account for such effects. It will particularly useful to explore how variability in different parameters effecting growth, such as maximal growth rate and yield, affect fitness. In addition, we have assumed that the competition occurs in a well-mixed environment, yet the geometry of the environment where the competition takes place can vastly influence the evolutionary dynamics [36]. It would be interesting to investigate the fitness effects of phenotypic variability in more complex environments where these factors play a role.

Acknowledgements

We acknowledge Jie Lin for helpful discussion related to this work. We thank Farshid Jafarpour for offering helpful feedback on this manuscript and Yuichi Wakamoto for sharing the experimental data. We acknowledge funding support from National Science Foundation grant DMR-1610737 (JK,EL), MRSEC at Brandeis University DMR-1420382 (JK), the Simons Foundation (JK), MRSEC at Harvard University under grant DMR-1420570 (EL,AA) and NSF CAREER 1752024 (EL,AA).

References

- [1] Po-Yi Ho, Jie Lin, and Ariel Amir. Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects. *Annual Review of Biophysics*, 47:251–271, 2018.
- [2] Avigdor Eldar and Michael B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167, 2010.
- [3] Zhi Wang and Jianzhi Zhang. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *PNAS; Proceedings of the National Academy of Sciences*, 108(16):E67–E76, 2011.
- [4] Fabien Duveau, Andrea Hodgins-Davis, Brian PH Metzger, Bing Yang, Stephen Tryban, Elizabeth A Walker, Tricia Lybrook, and Patricia J Wittkopp. Fitness effects of altering gene expression noise in *saccharomyces cerevisiae*. *eLife*, 7, 2018.
- [5] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584), 2002.
- [6] Lev S Tsimring. Noise in biology. *Reports on Progress in Physics*, 77(2):026601, 2014.
- [7] Mukund Thattai and Alexander Van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, 2004.
- [8] Jeremie. Roux, Marc Hafner, Samuel Bandara, J. Joshua Sims, Hannah Hudson, Diana Chai, and Peter K. Sorger. Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold. *Molecular Systems Biology*, 11(5):803–803, 2015.
- [9] Andreas Hilfinger and Johan Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences*, 108(29):12167–12172, 2011.
- [10] Johan. Paulsson, Otto. G. Berg, and Mans. Ehrenberg. Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. *Proceedings of the National Academy of Sciences*, 97(13):7148–7153, 2000.

- [11] Beatrice Claudi, Petra Spröte, Anna Chirkova, Nicolas Personnic, Janine Zankl, Nura Schürmann, Alexander Schmidt, and Dirk Bumann. Phenotypic variation of salmonella in host tissues delays eradication by antimicrobial chemotherapy. *Cell*, 158(4):722–733, 2014.
- [12] Giulia Manina, Neeraj Dhar, and John D McKinney. Stress and host immunity amplify mycobacterium tuberculosis phenotypic heterogeneity and induce nongrowing metabolically active forms. *Cell host & microbe*, 17(1):32–46, 2015.
- [13] Yuki Sughiyama and Tetsuya J. Kobayashi. Steady-state thermodynamics for population growth in fluctuating environments. *Physical Review E*, 95(1), 2017.
- [14] Stanislas Leibler and Edo Kussell. Individual histories and selection in heterogeneous populations. *Proceedings of the National Academy of Sciences*, 107(29):13183–13188, 2010.
- [15] Jose M. G. Vilar and J. Miguel Rubi. Determinants of population responses to environmental fluctuations. *Scientific Reports*, 8(1), 2018.
- [16] Mikihiro Hashimoto, Takashi Nozoe, Hidenori Nakaoka, Reiko Okura, Sayo Akiyoshi, Kunihiro Kaneko, Edo Kussell, and Yuichi Wakamoto. Noise-driven growth rate gain in clonal cellular populations. *PNAS; Proceedings of the National Academy of Sciences*, 113(12):3251–3256, 2016.
- [17] Mats Wallden, David Fange, Ebba Gregorsson Lundius, Özden Baltekin, and Johan Elf. The synchronization of replication and division cycles in individual e. coli cells. *Cell*, 166(3):729–739, 2016.
- [18] Manuel Campos, Ivan V. Surovtsev, Setsu Kato, Ahmad Paintdakhi, Bruno Beltran, Sarah E. Ebmeier, and Christine Jacobs-Wagner. A constant size extension drives bacterial cell size homeostasis. *Cell*, 159(6):1433–1446, Dec 2014.
- [19] Guillaume Lambert and Edo Kussell. Quantifying selective pressures driving bacterial evolution using lineage analysis. *Physical Review X*, 5(1), 2015.
- [20] Jie Lin and Ariel Amir. The effects of stochasticity at the single-cell level and cell size control on the population growth. *Cell Systems*, 5(4):358 – 367.e4, 2017.
- [21] Phillip Thomas. Single-cell histories in growing populations: relating physiological variability to population growth. *bioRxiv*, 2017.
- [22] Philipp Thomas. Making sense of snapshot data: ergodic principle for clonal cell populations. *Journal of The Royal Society Interface*, 14(136):20170467, 2017.
- [23] Farshid Jafarpour. Cell size regulation induces sustained oscillations in the population growth rate. *Physical Review Letters*, 122(11), 2019.
- [24] Yuichi Wakamoto, Alexander Y Grosberg, and Edo Kussell. Optimal lineage principle for age-structured populations. *Evolution*, 66(1):115–134, 2012.
- [25] Farshid Jafarpour. Bridging the timescales of single-cell and population dynamics. *Physical Review X*, 8(2), 2018.
- [26] Bree B. Aldridge, Marta Fernandez-Suarez, Danielle Heller, Vijay Ambravaneswaran, Daniel Irimia, Mehmet Toner, and Sarah M. Fortune. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science*, 335(6064):100–104, 2012.
- [27] Lin Chao, Camilla Ulla Rang, Audrey Menegaz Proenca, and Jasper Ubrajara Chao. Asymmetrical damage partitioning in bacteria: A model for the evolution of stochasticity, determinism, and genetic assimilation. *PLOS Computational Biology*, 12(1):e1004700, 2016.
- [28] Ariel Amir Jie Lin, Jiseon Min. Optimal segregation of proteins: Phase transitions and symmetry breaking. *Physical Review Letters*, 122(6), 2019.

- [29] Noga Mosheiff, Bruno M.C. Martins, Sivan Pearl-Mizrahi, Alexander Grünberger, Stefan Helfrich, Irina Mihalcescu, Dietrich Kohlheyer, James C.W. Locke, Leon Glass, and Nathalie Q. Balaban. Inheritance of cell-cycle duration in the presence of periodic forcing. *Physical Review X*, 8(2), 2018.
- [30] Jie Lin and Ariel Amir. Population growth with correlated generation times at the single-cell level. *arXiv preprint arXiv:1806.02818*, 2018.
- [31] E. O. Powell. Growth rate and generation time of bacteria, with special reference to continuous culture. *Microbiology*, 15(3):492–511, 1956.
- [32] Patrick Alfred Pierce Moran. Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 60–71. Cambridge University Press, 1958.
- [33] David Gresham and Maitreya J Dunham. The enduring utility of continuous culturing in experimental evolution. *Genomics*, 104(6):399–405, 2014.
- [34] Steven J. Altschuler and Lani F. Wu. Cellular heterogeneity: Do differences make a difference? *Cell*, 141(4):559 – 563, 2010.
- [35] Ofer Fridman, Amir Goldberg, Irine Ronin, Noam Shores, and Nathalie Q Balaban. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature*, 513(7518):418, 2014.
- [36] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.
- [37] Eric J Stewart, Richard Madden, Gregory Paul, and François Taddei. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology*, 3(2):e45, 2005.

Methods

Cell-size regulation model

In our simulations we use an established model of cell-size regulation that has proven to accurately describe microbial growth across a range of organisms [20]. The fundamental assumption of this model is that cells grow exponentially at the single cell level and divide upon reaching a size $v_{\text{div}}(v_{\text{birth}})$ depending on their size at birth, v_{birth} . We will assume that cells divide symmetrically so that v_{birth} is obtained by dividing the cell's mother's size at division by 2. Phenotypic variability is introduced by adding noise to both the single-cell growth rates as well as the volume at division. To this end, we take the growth rate λ of a cell to obey

$$\ln \lambda = (1 - \rho_\lambda)a_\lambda + \rho_\lambda \ln \lambda' + \eta_\lambda \quad (13)$$

where a_λ is the average log growth-rate along a lineage, ρ_λ can be shown to approximate the Pearson correlation coefficient between mother and daughter cells, λ' is the mother cell's growth rate, and η_λ is a normally distributed random variable. By taking $\ln \lambda$ to progress in time in this manner, we are assuming the daughter cell's growth rate depends only on its most recent ancestor (the mother). This will lead to a log-normal distribution of growth rates [37, 20]. Within this model, the generation time of a cell is

$$\tau(\lambda, v_{\text{birth}}, \tilde{v}_{\text{div}}) = \frac{1}{\lambda} \ln \left(\frac{\tilde{v}_{\text{div}}}{v_{\text{birth}}} \right) \quad (14)$$

where $\tilde{v}_{\text{div}} = v_{\text{div}} + \eta_v$ and η_v is another normally distributed random variable. The specific form of v_{div} is usually taken to be a linear combination of a size increment v_0 and the size at birth,

$$v_{\text{div}} = 2(1 - \alpha)v_b + 2\alpha v_0. \quad (15)$$

Here, the cell-size regulation strategy is controlled by the parameter α , for example $\alpha = 1$ corresponds to dividing at a critical size (known as a “sizer”), while $\alpha = 1/2$ corresponds to adding a constant size v_0 (known as an “adder”). We refer to [1] for an in-depth discussion of the cell-size control model and its implications for population growth. Note that in the notation of our general modeling framework the phenotype that controls generation times is $x = (\lambda, \tilde{v}_{\text{div}}, v_{\text{birth}})$ and it is straightforward to derive the distribution $f(x|x')$ from our modeling assumptions. In the special case where $\alpha = 1$ and $\eta_v = 0$, the phenotype is given by $x = \ln(\lambda)$, the generation time is simply $\tau = \ln(2)/\lambda$.

It has previously been shown that the cell-size regulations strategy typically induces negative mother-daughter correlations in generation times. In the specific case where there is no growth rate variability ($\eta_\lambda = 0$), the Pearson correlation coefficient between mother and daughter cell's generation time is $-\alpha/2$ [20]. Growth rate and volume fluctuations tend to suppress these correlations, but preserve the sign. Therefore, since bacteria regulate their size we expect negative mother-daughter generation time correlations in the absence of any environmental noise.

Analysis of the phenotype distributions

Here we derive the Euler-Lotka equation and present the detailed derivations of the various distributions in the finite culture.

Derivation of Euler-Lotka equation

Let $\psi(t, x, u)$ denote the joint of species with phenotypes x and ages u in an instantaneous observation of a population with $N \gg 1$ bacteria. We are mostly interested in the steady-state distribution $\psi(x, u)$. In this section we will omit the superscript F because we are only concerned with the finite culture. Over an interval dt , the change in the number of cells with phenotypes between x and $x + dx$ and ages between u and $u + du$ is

$$\begin{aligned} N\psi(t + dt, x, u)dxdu - N\psi(t, x, u)dxdu &\approx N\psi(t, x, u - dt)dxdu \\ &- N\psi(t, x, u)dxdu - N\Lambda_D dt dx du \end{aligned} \quad (16)$$

Taking the limit $dt \rightarrow 0$ and dividing by dt gives

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial u} \right) \psi(t, x, u) = -\Lambda_D \psi(t, x, u). \quad (17)$$

This is exactly the Von Foerster equation for an age structured population, although the role of the “death rate” is played by $\Lambda_D(t)$ and is no longer given as a model input. Instead, this must be derived from a boundary condition to ensure the population remains finite. At steady-state,

$$\frac{\partial}{\partial u} \psi(x, u) = -\Lambda_D \psi(x, u). \quad (18)$$

and the boundary condition is

$$\psi(x, u) = 2 \int_0^\infty f(x|x') \psi(x', \tau(x')) dx'. \quad (19)$$

Consider the distribution of phenotypes among cells that are just born, this is given by $\psi_{\text{birth}}(x) = \psi(x, 0) \Psi_0^{-1}$ where Ψ_0 is a normalization constant obtained by integrating $\psi(x, 0)$; $\Psi_0 = \int_0^\infty \psi(x, 0) dx$. Experimentally, $\psi_{\text{birth}}(x)$ is obtained by taking an instantaneous observation of the population and sampling only the cells that have just been born. By solving Equation 18 and inserting the solution into Equation 19, it is straightforward to obtain

$$\psi_{\text{birth}}(x) = 2 \int_0^\infty f(x|x') \psi_{\text{birth}}(x') e^{-\Lambda_D \tau(x')} dx. \quad (20)$$

Because this distribution is time invariant, sampling those cells that have just been born at any sequence of times $t_1, t_2, \dots, t_n$ yields the same distribution. Carrying out this sampling procedure at an infinite sequence of times samples every cell only once, and therefore $\psi_{\text{birth}}(x) = \psi_{\text{tree}}(x)$. This implies Equation 1 holds in the finite culture.

The current cells distribution

The distribution of phenotypes of the current cells is found by integrating $\psi(x, u)$ over all possible ages

$$\int_0^{\tau(x)} \psi(x, u) du = \psi(x, 0) \int_0^{\tau(x)} e^{-u \Lambda_D} du = \psi(x, 0) \frac{1 - e^{-\Lambda_D \tau(x)}}{\Lambda_D} \quad (21)$$

We can express the density of phenotypes among cells with age zeros as

$$\psi(x, 0) = \psi_{\text{tree}}(x) \Psi_0 \quad (22)$$

where Ψ_0 is the prefactor in the age distribution. This gives

$$\psi_{\text{snap}}(x) = 2 \psi_{\text{tree}}(x) \left(1 - e^{-\Lambda_D \tau(x)} \right), \quad (23)$$

which is identical to the current cells distribution in the exponentially growing culture [30].

The ancestral distribution

In the finite culture $\psi_{\text{ans}}(x)$ is the distribution over all cells that have descendents in the current culture. Let $\tilde{f}(x'|x) dx$ be the probability that a cell’s mother has phenotype in $[x', x' + dx]$ given that its phenotype is x . Bayes’ theorem tells us that

$$\tilde{f}(x'|x) = f(x|x') \frac{\psi_{\text{car}}(x')}{\psi_{\text{tree}}(x)} \quad (24)$$

It follows that

$$\psi_{\text{ans}}(x') = \int_0^\infty \tilde{f}(x'|x) \psi_{\text{ans}}(x) dx = \psi_{\text{car}}(x') \int_0^\infty f(x|x') \frac{\psi_{\text{ans}}(x)}{\psi_{\text{tree}}(x)} dx. \quad (25)$$

Division distribution

In both the continuous culture and exponentially growing culture, the carrier (or branch) distribution can be interpreted as the distribution of cells that have just divided, or the *division distribution* $\psi_{\text{div}}(x)$. In terms of the joint density of phenotypes and ages throughout the population, $\psi(x, u)$, the number of cells with phenotypes between x and dx that have just divided is

$$N\psi(x, \tau(x))dx = N\psi(x, 0)e^{-\Lambda_D \tau(x)}dx = N2\psi_{\text{tree}}(x)e^{-\Lambda_D \tau(x)}dx \quad (26)$$

Dividing by Ndx gives the distribution of phenotypes among recently divided cells, which is $\psi_{\text{div}}(x) = 2\psi_{\text{tree}}(x)e^{-\Lambda_D \tau(x)}$. As this argument works for both a finite and exponentially growing culture, the carrier distribution, branch distribution and division distributions are all the same.

Summary of distributions in the finite culture

To summarize, we have discussed the following ways to sample cells in the finite culture:

- *Tree distribution* ($\psi_{\text{tree}}^F(x)$): Sample all the cells throughout the history of the culture, including or excluding the current cells, but including those that are expelled before they divide.
- *Carrier distribution* ($\psi_{\text{car}}^F(x)$): Sample all the cells throughout the culture, excluding those that are expelled before they divide.
- *Ancestral distribution* ($\psi_{\text{ans}}^F(x)$): Sample all the cells in the history of the culture that have ancestors that are currently in the culture. Equivalently, select a random cell from the culture at any time and trace along a lineage backwards from this cell.
- *Current cells distribution* ($\psi_{\text{snap}}^F(x)$): Sample all the cells that are currently in the culture.
- *Birth distribution* ($\psi_{\text{birth}}^F(x)$): From all the cells that are in the culture at time t , sample only those that have just been born (meaning they have age ≈ 0), then repeat this for all t and average the results.
- *Division distribution* ($\psi_{\text{div}}^F(x)$): From all the cells that are in the culture at time t , sample only those whose age is equal to their generation time born, then repeat this for all t and average the results.

The relationship between these distributions is summarized in Figure S.1.

Derivation of Equation 4

Here we provide a more systematic derivation of Equation 4. Let n_i denote the number of species i in a culture containing M species. It will always be assumed that the culture is obtained by mixing homogenous cultures that have reached steady state. Let $f_i(x|x')$ denote distribution of phenotypes of a daughter cell of species i conditioned on the mother cell's, $\phi_i = n_i/N$ be the fraction of species i and $\psi_i(t, x, u)dudx$ be the fraction of cells of species i with ages between u and du and phenotype between x and dx . When a cell of species i reaches age $\tau_i(t, x)$ it divides, hence over a periods of time dt , the number of cells that divide is

$$\sum_{i=1}^M \phi_i N \int_0^\infty \psi_i(t, x, \tau_i(t, x)) dx dt. \quad (27)$$

This means the per capita dilution rate is

$$\Lambda(t) = \sum_{i=1}^M \phi_i \Lambda_i(t) \quad (28)$$

where $\Lambda_i(t) = \int_0^\infty \psi_i(t, x, \tau_i(t, x)) dx$ is the contributions of each species to the total dilution rate. At a given time, the number of bacteria of species i is $\phi_i(t) N \psi_i(t, x, u) dx du$, so over an interval dt , the change in the number of bacteria of species i with age $u < \tau_i(x)$ is

$$\begin{aligned} \phi_i(t+dt) N \psi_i(t+dt, x, u) dx - \phi_i(t) N \psi_i(t, x, u) dx &\approx \phi_i(t) N \psi_i(t, x, u-dt) dx \\ &- \phi_i(t) N \psi_i(t, x, u) dx - \phi_i(t) N \Lambda(t) dt dx \end{aligned} \quad (29)$$

Taking the limit $dt \rightarrow 0$ gives

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial u}\right) \psi_i(t, x, u) = -\left(\Lambda + \frac{1}{\phi_i} \frac{d}{dt} \phi_i(t)\right) \psi_i(t, x, u) \quad (30)$$

These equations depend on ϕ_i , for which the instantaneous growth is the difference between the rate at which cells are born ($\Lambda_i(t)$) and flushed out ($\Lambda(t)$). Hence,

$$\frac{d}{dt} \phi_i = (\Lambda_i(t) - \Lambda(t)) \phi_i. \quad (31)$$

It follows that Equation 30 becomes

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial u}\right) \psi_i(t, x, u) = -\Lambda_i(t) \psi_i(t, x, u) \quad (32)$$

This equation holds for $u < \tau_i(x)$ and is supplemented by the boundary conditions prescribing how the phenotypes are selected at birth,

$$\psi_i(t, x, 0) = 2 \int_0^\infty f(x|x') \psi_i(t, x', \tau_i(t, x')) dx. \quad (33)$$

Along with an initial distribution $\psi_i(0, x, u)$, Equations 32 and 33 describe how the distribution of phenotypes and ages evolves in the mixed culture. Importantly, these equations are decoupled from $\phi_i(t)$ and each other. Moreover, they turn out to be identical to the transport equations for a homogenous culture. This implies that the dynamics of the phenotype distribution are unchanged by the competition, and the fitness Λ_i are therefore determined by the population growth rates of the individual species in exponentially growing cultures. Since we have allowed for Λ_i and τ_i to depend on time, this result holds for both intrinsic and environmental variability.

Interpretation of the Euler-Lotka equation in a variable environment

In order to understand how the solution of the Euler-Lotka equation relates to the true population growth rate in a temporally varying environment, we derive a generalization of the Euler-Lotka equation which can be used to bound the true population growth rate. As before, the average expected number of offspring from a randomly selected cell must be 1 in a finite culture. The expected number of offspring is twice the probability that this cell divides in the culture, thus we need to calculate this probability in order to constrain the population growth rate. To this end, let

$$P_s(\tau) = \mathbb{P}[\text{A cell divides in the culture} \mid \text{generation time} = \tau]. \quad (34)$$

It follows that $P_s(\tau) \psi_{\text{tree}}(\tau)$ is the joint probability that a cell has generation time τ and divides in the culture. Integrating this quantity yields the probability that a randomly selected cell divides in the culture, and hence the expected number of offspring from a cell is

$$2 \int_0^\infty \psi_{\text{tree}}(\tau) P_s(\tau) d\tau = 1. \quad (35)$$

This statement assumes nothing about the underlying model of phenotypic variability, and in particular is independent of the environmental dynamics. We now introduce an explicit dependence on the state of the environment $\xi(t)$ by defining

$$P_s(\tau, \xi) = \mathbb{P}[\text{A cell divides in the culture} \mid \text{generation time} = \tau \text{ and cell is born when the environment is } \xi]. \quad (36)$$

Assuming the time-scale associated with the environmental fluctuations is large relative to the the average generation time of a cell, we can make the approximation

$$P_s(\tau, \xi) \approx e^{-\Lambda_D(\xi)\tau} \quad (37)$$

where $\Lambda_D(\xi)$ is the dilution rate in environment $\xi(t) = \xi$. Assuming the environment is a stationary process with stationary density $u(\xi)$,

$$P_s(\tau) = \int_{-\infty}^{\infty} u(\xi) P_s(\tau, \xi) d\xi = \left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u. \quad (38)$$

By combining Equations 35 with 38 we obtain a generalization of the Euler-Lotka equation,

$$\frac{1}{2} = \int_0^{\infty} \psi_{\text{tree}}(\tau) \left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u d\tau, \quad (39)$$

which relates the average survival probability to the tree distribution. In this context, the carrier distribution is therefore

$$\psi_{\text{car}}(\tau) = 2\psi_{\text{tree}}(\tau) \left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u. \quad (40)$$

Since $\psi_{\text{tree}}(\tau)$ is normalized to one, we have

$$\int_0^{\infty} \psi_{\text{car}}(\tau) \left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u^{-1} d\tau = 2. \quad (41)$$

An application of Jensen's inequality yields the bound

$$\left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u \geq e^{-\langle \Lambda_D(\xi) \rangle_u \tau}, \quad (42)$$

which holds for any τ ; therefore,

$$2 = \int_0^{\infty} \psi_{\text{car}}(\tau) \left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u^{-1} d\tau < \int_0^{\infty} \psi_{\text{car}}(\tau) e^{\langle \Lambda_D(\xi) \rangle_u \tau} d\tau. \quad (43)$$

Note that $\Lambda = \langle \Lambda_D(\xi) \rangle_u$ is the true population growth rate (assuming there are a large number of cells in the culture). On the other hand, since Λ_{car} satisfies Equation 6, we have

$$2 = \int_0^{\infty} \psi_{\text{car}}(\tau) e^{\Lambda_{\text{car}}\tau} d\tau < \int_0^{\infty} \psi_{\text{car}}(\tau) e^{\langle \Lambda_D(\xi) \rangle_u \tau} d\tau. \quad (44)$$

Since both integrals are decreasing functions of the exponent, this implies $\Lambda_{\text{car}} < \Lambda$.

We can obtain a more quantitative estimate by assuming the deviations in Λ_D from $\Lambda_D(\xi)$ and Λ_{car} are small. In this limit, Taylor expanding gives

$$\left\langle e^{-\Lambda_D(\xi)\tau} \right\rangle_u \approx e^{-\Lambda\tau} \left(1 + \frac{\tau^2}{2} \sigma_{\text{env}}^2 \right) \quad (45)$$

$$e^{-\Lambda_{\text{car}}\tau} \approx e^{-\Lambda\tau} (1 - (\Lambda_{\text{car}} - \Lambda)\tau). \quad (46)$$

where σ_{env}^2 is the environmental noise term defined by $\lim_{N \rightarrow \infty} \sigma_{\Lambda_D}^2$. Note that as N tends to infinity, σ_{env}^2 tends to σ_{Λ}^2 , where σ_{Λ}^2 is the total variance of Λ including the contribution from intrinsic noise. Averaging Equations 45 and 46 over $\psi_{\text{tree}}(\tau)$ gives

$$\Lambda - \Lambda_{\text{car}} = \frac{\sigma_{\text{env}}^2}{2} \frac{\langle e^{-\Lambda\tau} \tau^2 \rangle_{\text{tree}}}{\langle e^{-\Lambda\tau} \tau \rangle_{\text{tree}}}, \quad (47)$$

or in terms of $\psi_{\text{car}}(\tau)$,

$$\Lambda - \Lambda_{\text{car}} = \frac{\sigma_{\text{env}}^2}{2} \frac{\langle \tau^2 \rangle_{\text{car}}}{\langle \tau \rangle_{\text{car}}} \quad (48)$$

When the environmental noise is weak, Λ_D is approximately linear in the perturbation ξ , and therefore

$$\sigma_{\text{env}}^2 = \text{var}(\lambda[S(t)]) \approx \left[\frac{\lambda_{\text{max}} S_0 K_m}{(K_m + S_0)^2} \right]^2 \sigma_{\xi}^2 \quad (49)$$

Using that

$$\frac{\langle e^{-\Lambda\tau} \tau^2 \rangle_{\text{tree}}}{\langle e^{-\Lambda\tau} \tau \rangle_{\text{tree}}} = \frac{\ln(2)}{\Lambda} + O(\sigma_\lambda^2) \quad (50)$$

we have

$$\Lambda_{\text{car}} \approx \Lambda \left(1 - \frac{\ln(2)}{2} \left[\frac{K_m}{K_m + S_0} \right]^2 \sigma_\xi^2 \right) \quad (51)$$

where we are using $\approx$ to indicate that terms of order $\sigma_\xi^2 \sigma_\lambda^2$ are being neglected. Note that in general ψ_{tree} will depend on the environmental fluctuations; however, these affects vanish in the leading order analysis. Finally, when ξ is modeled by an OU process, σ_ξ^2 is related to the diffusion coefficient D_ξ and a relaxation time-scale t_ξ according to $\sigma_\xi^2 = D_\xi t_\xi$.

Supplementary figures

Figure S.1: The ratio of some of the distributions plotted in Figure 4 (C). Equation 5 predicts that $\psi_{\text{tree}}(x)/\psi_{\text{car}}(x) = e^{\Lambda\tau(x)}/2$. In this case $x = \ln(\lambda)$, hence $e^{\tau(x)} = e^{\ln(2)/e^x} = 2^{\Lambda/\lambda}/2$. On the other hand, there is no such relationship between ψ_{ans} and ψ_{tree} .

Figure S.2: (A) Using the data plotted in Figure 5 (C), we have computed the environmental variance σ_{Λ}^2 computed using Equation 10 to the true value obtained from the fluctuations in biomass accumulation rate. In order to solve for σ_{Λ}^2 in this equation, we have computed $\langle \tau \rangle$, $\langle \tau^2 \rangle$ and Λ_{in} using the carrier distribution, while Λ is obtained by taking the average dilution rate $\langle \Lambda_D(t) \rangle$ and using Equation 11 to correct for population size effects. (B) The variance in the instantaneous dilution rate compared to the environmental fluctuations. This gives a much poorer estimation of the environmental variance than Equation 10.

Figure S.3: (A) The ratio Λ_{car}/Λ obtained from different experiments as a function of the average number of cells in the culture. Each experiment corresponds to 2 - 5 biological replicates in the same experimental conditions. Λ has been obtained by computing the dilution rate and corrected for the population size. (B) Information about the different experiments, including the strain, growth medium and temperature.

Figure S.4: The ratios (A) $\Lambda_{\text{car}}/\Lambda_D$ and (B) $\Lambda_{\text{car}}/\Lambda_D$ for various parameters in the cell-size regulation model. The black dots indicate the experimental data shown in detail in Figure 6 (C). Other parameters are the same as Figure 5.