

Evolutionary dynamics in non-Markovian models of microbial populationsFarshid Jafarpour^{1,*}, Ethan Levien^{2,*}, and Ariel Amir^{3,4}¹*Institute for Theoretical Physics, Utrecht University, 3584 CC Utrecht, The Netherlands*²*Mathematics Department, Dartmouth College, Hanover, New Hampshire 03755, USA*³*Department of Complex Systems, Faculty of Physics, The Weizmann Institute of Science, Rehovot 7610001, Israel*⁴*John A. Paulson, School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA* (Received 24 February 2023; revised 8 May 2023; accepted 7 June 2023; published 5 September 2023)

In the past decade, great strides have been made to quantify the dynamics of single-cell growth and division in microbes. In order to make sense of the evolutionary history of these organisms, we must understand how features of single-cell growth and division influence evolutionary dynamics. This requires us to connect processes on the single-cell scale to population dynamics. Here, we consider a model of microbial growth in finite populations which explicitly incorporates the single-cell dynamics. We study the behavior of a mutant population in such a model and ask: can the evolutionary dynamics be coarse-grained so that the forces of natural selection and genetic drift can be expressed in terms of the long-term fitness? We show that it is in fact not possible, as there is no way to define a single fitness parameter (or reproductive rate) that defines the fate of an organism even in a constant environment. This is due to fluctuations in the population averaged division rate. As a result, various details of the single-cell dynamics affect the fate of a new mutant independently from how they affect the long-term growth rate of the mutant population. In particular, we show that in the case of neutral mutations, variability in generation times increases the rate of genetic drift, and in the case of beneficial mutations, variability decreases its fixation probability. Furthermore, we explain the source of the persistent division rate fluctuations and provide analytic solutions for the fixation probability as a multispecies generalization of the Euler-Lotka equation.

DOI: [10.1103/PhysRevE.108.034402](https://doi.org/10.1103/PhysRevE.108.034402)**I. INTRODUCTION**

In recent decades, advances in single-cell technology and single-molecule biophysics have shed new light on the molecular mechanisms and dynamics underlying single-cell growth and division [1–3]. In some instances, aspects of cell growth are preserved across all domains of life, such as theadder mechanisms for cell-size regulation, which appears in bacteria, archaea, and eukaryotes [4]. In other cases, even closely related organisms have found distinct ways to achieve the same objective, such as the regulation of mating type genes in yeast [5]. These observations have led to questions about how evolution has shaped single-cell physiology and vice versa. While phylogenetic methods can, to some extent, help us understand the evolutionary history of physiological mechanisms, their applications are limited to cases where we have access to a clear mapping between the physiological mechanisms of interest and their associated genotypes. In order to understand cellular physiology from an evolution perspective, we must therefore develop a mechanistic understanding of how evolution acts on properties of single cells.

A standard model used in theoretical population genetics is the continuous-time Moran process. In this model, cells of genotype i divide at a constant rate ρ_i (the fitness). In order to keep the population size fixed, every time a cell divides, a cell is randomly selected from the entire population (including the two newborn cells) to be expelled [see Fig. 1(a)]. The Moran

process can be thought of as a mathematical idealization of a turbidostat culture; in such cultures the rate at which fresh media is continuously pumped into a vial is modulated in order to keep the optical density (and hence the number of cells) fixed. Since real organisms do not divide at a constant rate, ρ is typically thought of as a proxy for the viability of a genotype, rather than the literal division rate of a cell. Suppose that two genotypes are present in the population with fitnesses ρ_r (the resident) and ρ_m (the mutant). It can be shown that for sufficiently large populations, the fraction of mutant cells, denoted ϕ , will obey the stochastic differential equation (SDE)

$$\frac{d\phi}{dt} = (\rho_m - \rho_r)\phi(1 - \phi) + \sqrt{\frac{(\rho_m + \rho_r)\phi(1 - \phi)}{N}}\xi(t), \quad (1)$$

where the noise term should be interpreted in the Itô sense. This is the stochastic version of a replicator equation [6]. We will derive this equation in Sec. II. It is from this SDE that all the fundamental results concerning the fate of rare mutations in an evolving population can be derived.

In the neutral case ($\rho_r = \rho_m = \rho$), the frequencies have mean $\langle\phi\rangle = \phi(0)$, while the variance in ϕ will evolve as

$$\frac{d}{dt}\text{var}(\phi) = \frac{2\rho}{N}[\phi_0(1 - \phi_0) - \text{var}(\phi)]. \quad (2)$$

Therefore, the $2\rho/N$ factor sets the timescale for the frequency of neutral mutations to change by a significant amount. We will derive Eq. (2) more generally in Appendix C 1. In the case

*These authors contributed equally to this work.

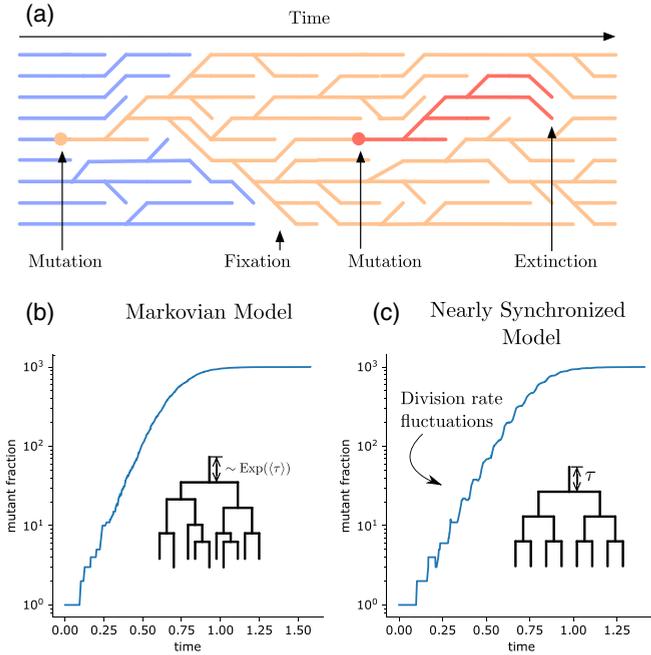


FIG. 1. Panel (a) shows a schematic of the Moran process. After each division, a cell is expelled from the population to keep the total population size constant. Evolution proceeds by mutations that eventually either fix or go extinct. Panels (b) and (c) show trajectories of an initial mutant population which goes on to establish in the culture within two different models. The null model (b) is Markovian division where the probability of division for each cell is constant regardless of its size or age, corresponding to exponentially distributed generation times τ . In this model, the rate of divisions of cells in the population quickly reaches a constant value. In a more realistic model of growth and division, panel (c), cell divisions remain nearly synchronized due to the narrow distribution of generation times τ .

of a beneficial mutation ($\rho_m > \rho_r$), when $\phi > 1/Ns$ [where $s = (\rho_m - \rho_r)/\rho_r$ is the *selection coefficient*], Eq. (1) is well approximated by a logistic growth [7]. However, since the mutant population initially consists of one cell, we are interested in the chance that the new mutant will eventually overtake the resident, or the *fixation probability*. A central result in population genetics states that in a large population the fixation probability is given by [8,9]

$$p_{\text{fix}} \approx \frac{1 - e^{-2s}}{1 - e^{-2Ns}}. \quad (3)$$

When s and sN are both sufficiently small, we obtain the simple relation $p_{\text{fix}} = s + O(1/N + s^2)$. Importantly, both the rate of genetic drift and the probability for a mutant to fix depend only on the long-term growth rates of the resident and mutant populations, and therefore, one might conclude that the single-cell dynamics are relevant only insofar as they affect the long-term growth rate.

In order to develop a predictive theory of microbial evolution, we need to link the relevant parameters ρ_r and ρ_m to properties of single cells. The problem of understanding how single-cell dynamics map to population fitness dates back to the work of Euler and Lotka [10]. By extending the work of these authors, recent studies have developed an understand-

ing of how variation in traits such as cell size, cell growth rates, and generation times combine to shape the long-term growth of a population [11–21]. These studies have shown that under conditions of constant exponential growth, the distribution of ages in a population converges to a steady state that determines the growth rate of a population. Moreover, the long-term fitness in the context of the Moran process (defined as the division rate, ρ) can be inferred from the distribution of generation times $\psi(\tau)$ measured among newborn cells using the Euler-Lotka equation:

$$\frac{1}{2} = \int_0^\infty \psi(\tau) e^{-\rho\tau} d\tau = \tilde{\psi}(\rho), \quad (4)$$

where $\tilde{\psi}$ is the Laplace transform of ψ . This relation has proven in both experimental and theoretical studies to be a powerful tool for probing the effects of single-cell dynamics on growth. For example, it has been used to elucidate the effects of cell-size regulation in bacteria on population growth [13].

Equation (4), when combined with Eq. (1), appears to provide a link between the single-cell dynamics and the evolutionary dynamics; specifically, Eq. (4) can be used to obtain the long-term growth rates of each species, which can then be used to find the rate of genetic drift and the fixation probability via Eq. (1). However, this reasoning ignores a crucial subtlety of this problem: Eq. (1) is derived under the assumption that there exists a constant division rate, ρ , which characterizes the fitness of a genotype in a finite culture. This assumption is valid in situations where $\phi \gg 1/(Ns)$, so that demographic fluctuations can be neglected. For rare mutants with small selection coefficients, or for neutral mutants, the division rate will fluctuate as a result of fluctuations in the distribution of cell ages. In this paper, we seek to understand the extent to which these fluctuations influence population genetics.

Our focus on the division rate fluctuations is in contrast to previous efforts to understand evolutionary dynamics in the presence of age-structure and nongenetic variability (see [22,23]), which consider models with discrete age classes and assume the age distribution to be in steady state in order to obtain an effective SDE. We demonstrate that such approximations are not valid for physiologically realistic models of microbial growth. Therefore, it cannot be assumed that dynamics can be characterized by the constant growth rate terms obtained from the *steady-state phenotype* distribution and constant division rate [see Figs. 1(b) and 1(c)].

The remainder of this paper is structured as follows. In Sec. II, we start with a generalization of the Moran process derived from physiological models of growth, division, and cell-size control, and derive a Fokker-Plank equation for the genotype frequencies. Section III discusses neutral dynamics in the Moran process. We will show that the rate of neutral genetic drift depends not only on the long-term fitness of a genotype, but also the variability in generation times, with larger variability leading to larger genetic drift. In Sec. IV, we derive the fixation probabilities of a mutant arising in a large population, showing that the selection coefficient appearing in Eq. (3) is insufficient to predict the fate of a mutant. In particular, we will find that the fixation probability can decrease with the variability in generation times in the mu-

tant population when long-term fitness is kept constant but is unaffected by that of the resident cells. In contrast, it is well known that variation in generation times tends to be beneficial from the perspective of long-term fitness (assuming constant mean generation time). Thus, there are situations where decreasing the selection coefficient can actually increase the fixation probability and vice versa. As we will show, this can influence qualitative features of long-term evolutionary dynamics, leading to nonlinear fitness trajectories in situations where classical theory predicts linear increase in fitness with time.

II. CONTINUOUS AGE-STRUCTURED MORAN PROCESS

Consider a population consisting of n cells of genotype m (the mutants) and $N - n$ cells of genotype r (the residents) growing according to the Moran process model, with the exception that instead of cells dividing at a constant probability per unit time, their division times are determined by a physiological model. To be precise, we assume that after living for a (possibly random) time τ , called the *generation time*, a cell divides to produce two new offspring that themselves go on to produce offspring in the same manner. All the information about the underlying cellular physiology is encoded in how we select the generation time. For any microbe where the sizes of cells are controlled, the time at which a cell divides must be coupled to the cell's volume; otherwise, small variations in doubling times result in unbounded fluctuations in size [3]. As a result, the model of generation time must account for cell sizes and growth rates, so the probability for a cell to divide in a given time interval after having survived to an age a depends on its size at birth s_b , its growth rate λ , and its age a . We denote this probability by $\gamma(a, \lambda, s_b) dt$. Letting $f(\tau, \lambda, s_b)$ be the joint distribution of generation times, growth rates, and initial cell sizes, we have $\gamma(a, \lambda, s_b) = f(a, \lambda, s_b) / \int_a^\infty f(a', \lambda, s_b) da'$. In Appendix A, we discuss a common framework for modeling cell-size regulation.

For each genotype i , we can define the per capita population division rate $\rho_i(t)$ at time t as the population average of single-cell division rates γ_i :

$$\rho_i(t) = \frac{1}{N_i} \sum_{j=1}^{N_i} \gamma_i(a_j, \lambda_j, s_{b,j}), \quad (5)$$

where the subscript j represents an individual cell in the population, and N_i is given by n for the mutant cells and $N - n$ for the resident cells. Notice that only in the special case of exponentially distributed generation times [i.e., $f(\tau, \lambda, s_b) = \gamma e^{-\tau\gamma}$] do we obtain a time-independent per capita division rate $\rho_i = \gamma_i$. However, it has previously been shown that in a deterministic treatment of the population dynamics that assumes selection is dominant over genetic drift [$\phi \gg 1/(Ns)$] there is a well-defined phenotype distribution, $p_i^\infty(a, \lambda, s_b)$. As was shown in Ref. [18], this steady-state distribution is the same as the steady-state distribution of single-cell variables in an exponentially growing population. This suggests it might be reasonable to neglect fluctuations in $\rho_i(t)$ and replace Eq. (5) with the average over the steady-state distribution:

$$\rho_i = \iiint da d\lambda ds_b p_i^\infty(a, \lambda, s_b) \gamma_i(a, \lambda, s_b), \quad (6)$$

where now, ρ_i is a time-independent property of genotype i . In this paper, we use the explicit time-dependent notation $\rho_i(t)$ to refer to the population division rate from Eq. (5), and whenever the time-independent notation ρ_i is used, it refers to the long-term fitness in Eq. (6).

We consider the dynamics of the fraction of mutant cells, $\phi(t) \equiv n/N$, under the assumption that ρ_i are good approximations for the population division rates $\rho_i(t)$. The number fraction ϕ increases when the division of a mutant cell is followed by a resident cell getting discarded and decreases when the division of a resident cell is followed by a mutant cell getting discarded. Therefore, the probability $P(\phi, t)$ of observing $\phi(t) = \phi$ at time t satisfies the master equation

$$\frac{dP(\phi, t)}{dt} = \sum_{\phi'} [T(\phi' \rightarrow \phi)P(\phi, t) - T(\phi \rightarrow \phi')P(\phi, t)], \quad (7)$$

where the transition rate $T(\phi \rightarrow \phi')$ from ϕ to ϕ' can be written in terms of ρ_i 's as

$$T(\phi \rightarrow \phi') = \begin{cases} N\phi\rho_m(1-\phi) & \phi' = \phi + 1/N \\ N(1-\phi)\rho_r\phi & \phi' = \phi - 1/N \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

For large N , the Kramers-Moyal expansion of Eq. (7) leads to a Fokker-Planck equation whose corresponding stochastic differential equation is the continuous Moran process described in Eq. (1) (see Appendix B for derivation).

The continuous Moran process described in Eq. (1) is known to hold for a Markovian model of cell division where cells divide with equal probability at any time independent of their age or size, but now, it is derived for an arbitrary model of growth, division, and cell-size regulation, assuming fluctuations in the division rates can be neglected.

In the next section we analyze the special case of neutral genetic drift, where $\rho_r = \rho_m$, and show that simulation results disagree with the prediction from constant division rate assumption. As we will show, the rate of neutral genetic drift, as observed in simulations, is slower than what is predicted by the standard theory, even in the limit $N \rightarrow \infty$. We will understand this in terms of coalescent theory as discussed in the next section.

III. NEUTRAL EVOLUTION

We start by analyzing the model introduced in the previous section for the case of neutral mutations. Consider a population of N total cells of two genotypes of resident and mutant cells, r and m , both dividing with the same division rate ρ . From Eq. (1), the dynamics of the number fraction of the mutant species m in the classical Moran (i.e., assuming generation times are exponentially distributed), is

$$\frac{d\phi_M}{dt} = \sqrt{\frac{2\rho}{N}} \phi_M(1-\phi_M)\xi(t), \quad (9)$$

where the subscript M is used to distinguish these dynamics from the age-structured case. The dynamics of variance of ϕ in this case is given by Eq. (2) (see Appendix C 1 for derivation); thus, the variance increases at a rate $2\rho/N$ (see Fig. 2). In the previous section, we showed that Eq. (1) [and therefore

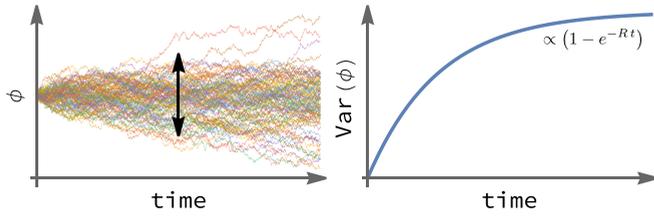


FIG. 2. Left: Stochastic trajectories of number fraction ϕ of a mutant starting from an initial number fraction ϕ_0 . In neutral genetic drift, the expected value of ϕ remains constant, but the variance increases over time. Right: The exponential rate R with which the variance of ϕ increases over time is defined as the rate of neutral genetic drift.

Eq. (9)] should hold for the age-structured case, if we neglect the fluctuations in the division rate of the population. To compare this to the age-structured setting, we simulate a finite population of cells each growing exponentially in size with their elongation rates independently and identically distributed.

We start with a simple model of division, where all cells are born with the same size s_b and divide with a fixed size $s_d = 2s_b$. Assuming cells grow exponentially, $s_d = s_b e^{\lambda\tau}$ and hence $\tau = \ln(2)/\lambda$. Therefore, all the variation in the generation times comes from the growth rate λ , which we assumed to be uncorrelated between cells. For mathematical simplicity, we select an inverse-gamma distribution for λ , which results in a gamma distribution for τ with probability density

$$f(\tau) = \frac{\beta^{-\alpha} e^{-\tau/\beta} \tau^{\alpha-1}}{\Gamma(\alpha)}. \quad (10)$$

The gamma distribution fits the empirical distribution of cell-size doubling times at least as well as a Gaussian distribution in various growth conditions, but also has many theoretical advantages: A gamma distribution has positive support, the Euler-Lotka equation [Eq. (4)] is analytically solvable for gamma distribution, and the gamma distribution naturally interpolates between exponential distribution (Markovian limit) and δ distribution (deterministic limit).

Simulations of the neutral model were performed using gamma distributed generation times for different coefficients of variation (CV is defined as the ratio of standard deviation to the mean). Figure 3 shows the ratio of the rate of neutral genetic drift, measured from these simulations, to the predicted value from Eq. (9). Here it can be seen that the variability in generation time increases the rate of genetic drift independently of how it affects the long-term fitness ρ , contradicting Eq. (9).

At the Markovian limit $CV = 1$, where the generation time distribution is exponential, the simulated result approaches the prediction from Eq. (9) for large system sizes. For $CV < 1$, the fact that the rate of genetic drift at large system sizes approaches a different limit indicates that the division rate cannot be replaced by its average ρ . Note that interesting dynamics may emerge with $CV > 1$, such as nondiffusion limiting processes, which have been explored in the context of the Cannings model [7]. However, we will focus on the cases $CV \leq 1$ in the present study.

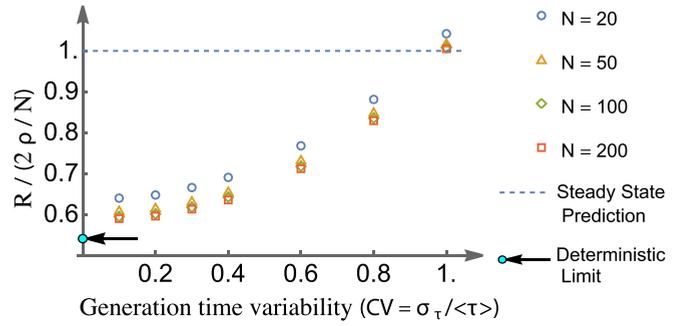


FIG. 3. The ratio of the rate R of neutral genetic drift (as defined in Fig. 2) from simulation to its predicted value $2\rho/N$ from the constant division rate assumption, Eq. (9), showing significant deviation from the predicted results. Different markers indicate different system sizes showing that the deviation is clearly not a finite-size effect. As the coefficient of variation of generation times approaches 1, the distribution of generation times approaches an exponential (Markovian limit) recovering the predicted results at large system sizes. For a smaller variability in generation times, fluctuations in the division rates have a significant effect on the rate of neutral genetic drift causing the deviation from the analytically predicted result.

A. Persistent division rate fluctuations emerge from coalescence

To understand why division rate fluctuations are important, we consider a population consisting of only one species. While the number of cells is kept constant, the division rate and distribution of sizes and ages of the cells in the population can vary over time. Figure 4 shows the division rate for such a population, which appears to oscillate in populations of various sizes even after 1000 generation times. To ensure that these oscillations are not side effects of an oversimplified model, we have performed these simulations using a realistic model and parameter values for growth, division, and cell-size control of *Escherichia coli* previously used in Refs. [12,13], although the same oscillations can be observed in the model with independent, gamma distributed generation times. They are reminiscent of the oscillations in growth rates of exponentially growing populations starting from a single cell [12,21] that are due to cells in the population sharing a common ancestor and dividing more or less at the same time. In growing populations, however, oscillations decay over time, because as more generations pass since the cells share a common ancestor, the noise in the doubling time accumulates giving rise to less synchronized populations. Why does that not happen in populations with constant population size? The key to answer this question lies in coalescent theory.

As shown in Fig. 5, starting from a population of cells at time $t = 0$ the descendants of all cells except one will go extinct after some time t_1 . At t_1 , the population will have a most recent common ancestor at some time $t_1 - T_1$. Since these cells share a common ancestor, their division times are more or less synchronized. If we wait some additional time until time t_2 , one might expect that the population would be less synchronized because it has been more generations since the cells shared a common ancestor, but that is not true. As highlighted in the bottom panel of Fig. 5, the population at time t_2 has a more recent common ancestor at the time $t_2 - T_2$. Crucially, the two coalescence times T_1 and T_2 are

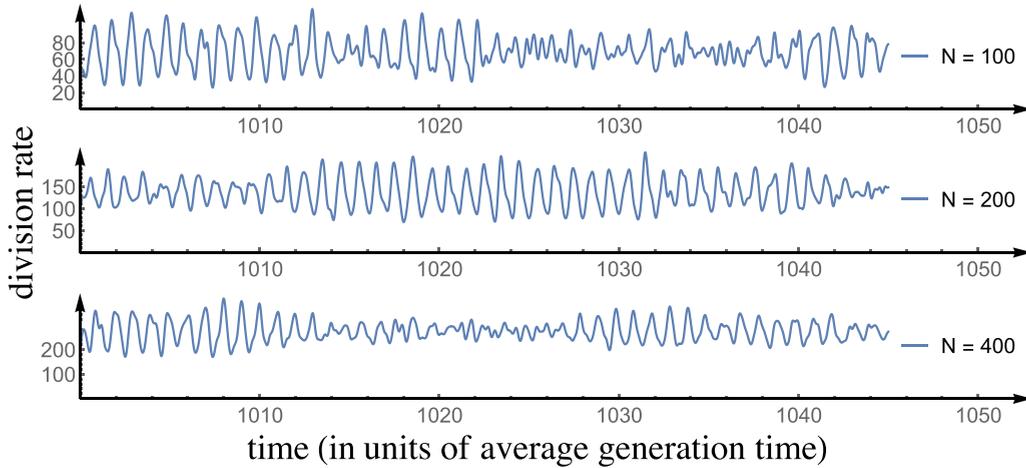


FIG. 4. Stochastic spontaneous oscillations of division rates in simulations of constant-size populations of cells with a realistic model of growth, division, and cell-size regulation for *E. coli* from Ref. [12] for three different system sizes $N \in \{100, 200, 400\}$, after 1000 generations from the start of the simulations. Simulation details: 10% division noise (size additive), 7% growth rate noise, adder model of cell-size regulation. The division rates are estimated using Gaussian kernel estimation on the division events of single simulations.

statistically identical random variables (with expected values on the order of $N\langle\tau\rangle$, where $\langle\tau\rangle$ is the average generation time). This means that over time, the population maintains the same degree of synchrony, giving rise to oscillations in the division rate shown in Fig. 4.

B. Neutral genetic drift at low division-time variability

In the context of genetic drift, memory in the division rate prevents us from representing the effects of all single-cell variables by a single fitness variable. In particular, the variability in single-cell doubling times associated with a genotype

affects the fate of an organism independently from how it affects the fitness variable ρ . In many bacteria such as *E. coli* this variability is much smaller than its mean, with the CV between 0.05 and 0.2 [24]. The rate of neutral genetic drift at these CV values is approximately the same as the zero-variability limit of Fig. 3. In this limit, the number fractions still experience a stochastic drift that can be approximated by (see Appendix C 2 for the derivation)

$$\frac{d\phi}{dt} = \sqrt{\frac{3\rho}{4N \ln(2)}} \phi(1-\phi) \xi(t), \quad (11)$$

where ρ is the average division rate. The rate of neutral genetic drift in this case is $R = 3\rho/4N \ln(2)$, which is a factor $3/8 \ln(2) \approx 0.54$ smaller than predicted by neglecting division rate fluctuations. This discrepancy can be traced back to the fact that cells more or less preserve the order of their divisions at different generations when the variability in their elongation rates is very small.

IV. ADAPTIVE EVOLUTION

In the previous section we found that the rate of genetic drift is determined not only by the long-term growth rates of species, but also by the details of the single-cell dynamics. We now shift our focus to adaptive evolution, starting with the question of how details of single-cell dynamics influence fixation probabilities. Consider a new mutant arising in a resident population described by the Moran process and let ρ_r and ρ_m be the long-term growth rates of the mutant and resident, respectively. It will be assumed throughout that $\rho_m > \rho_r$. If the mutant population survives to become sufficiently abundant, then its frequency ϕ will eventually be described by the deterministic logistic equation $d\phi/dt = (\rho_m - \rho_r)\phi(1-\phi)$. Genetic drift plays a negligible role in this phase of the dynamics, and fluctuations in the division rates can be ignored. However, when the mutant clone size is smaller than $1/s$, the stochastic effects dominate and we need to consider how these may be influenced by details of the single-cell dynamics.

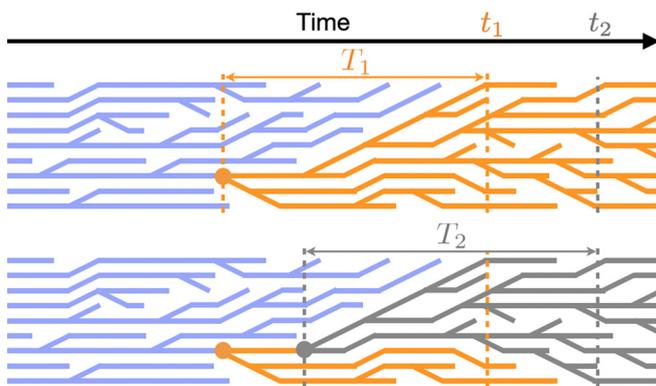


FIG. 5. Top panel shows a sketch of the lineage history of a constant-size population. The population eventually reaches a state where all cells share a common ancestor (at time t_1 in this illustration). The cells at this time share a most recent common ancestor from the time $t_1 - T_1$, and therefore have some level of synchrony in their division times. One would expect that if we wait until some later time t_2 , cells would be less synchronized, as it has been more generations since they shared a common ancestor. However, as shown in the bottom panel, the cells at time t_2 share a more recent common ancestor at time $t_2 - T_2$. The two stochastic variables T_1 and T_2 are coalescence times for the population at times t_1 and t_2 and have the same expected values.

In order to obtain analytical results, we once again assume that the mutant population is described by the independent generation time model with a gamma distribution of division times, and consider the limits $CV = 1$ and $CV \approx 0$. In the limit $CV = 1$, the mutant cells initially divide at a rate ρ_m so there is no distinction between the initial and long-term dynamics. In this case, the probability that the mutant fixes is obtained from the classical result of Haldane and Kimura given by Eq. (3). Now consider $CV \approx 0$. The long-term growth rate of the mutant population is $\rho_m = \ln(2)/\langle\tau\rangle$; however, initially the mutant cells divide approximately synchronously at a rate $1/\langle\tau\rangle$. Since the convergence to the steady-state growth rate is extremely slow, we expect that the quantity $\rho_m/\ln(2) = 1/\langle\tau\rangle > \rho_m$ might replace ρ in the formulas for the fixation probability, leading a higher chance of fixation than the predictions from Kimura’s formula. This is indeed the case, in fact, as we show in the next section the fixation probability of the synchronously dividing mutant populations is increased by a factor of $2 \ln(2)$ compared to the prediction of Kimura’s formula. We will then generalize this to any value of CV and discuss the implications for long-term evolutionary dynamics.

A. Derivation of fixation probabilities

We now derive a general formula for the fixation probability in terms of the generation time distribution. Assuming cells have independent generation times and the population is large enough that mutants are very unlikely to be expelled by the division of the other mutants, the probability that both of the cells offspring go extinct is simply p_{ext}^2 . Let q denote the probability that the mutant is expelled before it divides:

$$q = \mathbb{P}(\tau_e < \tau), \tag{12}$$

where τ is the doubling time of the mutant and τ_e the time until the mutant is expelled. In terms of q , the extinction probability can be written as

$$p_{\text{ext}} = q + (1 - q)p_{\text{ext}}^2. \tag{13}$$

The extinction probability p_{ext} is the smaller root of the quadratic equation (approaching the fixed point from below) given by

$$p_{\text{ext}} = \begin{cases} \frac{q}{1-q}, & q < \frac{1}{2} \\ 1, & q \geq \frac{1}{2}. \end{cases} \tag{14}$$

Since we are assuming the population is so large that nearly all removal events result from cells getting replaced by resident cells, the time τ_e until the mutant is expelled is simply set by the division rate of resident cells, ρ_r . In contrast to the mutant population, the division rate fluctuations of the resident population can safely be neglected (in contrast to the neutral case discussed above), since they make only a higher order contribution to the expulsion of mutant cells. Hence τ_e is exponentially distributed with $\langle\tau_e\rangle = 1/\rho_r$. The distribution of τ is given by the generation time distribution, $f_m(\tau)$. It follows that

$$q = \rho_r \int_0^\infty \int_0^\tau f_m(\tau) e^{-\rho_r \tau_e} d\tau_e d\tau = 1 - \tilde{f}_m(\rho_r), \tag{15}$$

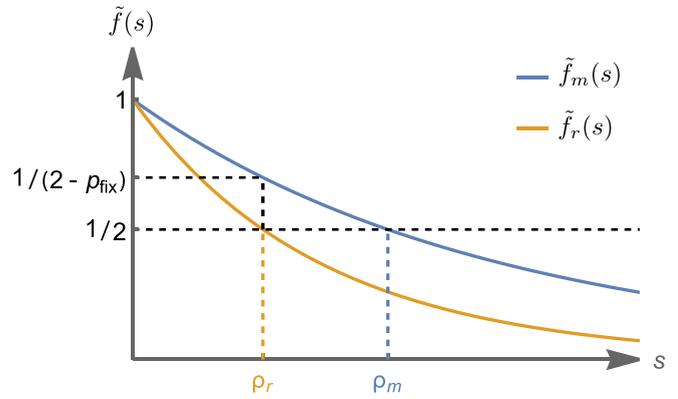


FIG. 6. The relation between fixation probability and the Euler-Lotka equation: The Euler-Lotka equation states that for both mutant and resident species m and r , the Laplace transforms of the generation-time density distributions $\tilde{f}_m(s)$ and $\tilde{f}_r(s)$ evaluated at their steady-state reproductive rates ρ_m and ρ_r are equal to $1/2$. Equation (16) states that the Laplace transform of the generation-time density distribution of the mutant cells $\tilde{f}_m(s)$ evaluated at the reproductive rate of the resident cells ρ_r is given by $1/(2 - p_{\text{fix}})$.

where $\tilde{f}_m(\rho_r) = \int_0^\infty f_m(\tau) e^{-\rho_r \tau} d\tau$ is the Laplace transform of f_m evaluated at ρ_r . In order to make the connection to the Euler-Lotka equation, it is convenient to write Eq. (14) in terms of \tilde{f}_m (see Fig. 6):

$$\tilde{f}_m(\rho_r) = \frac{1}{2 - p_{\text{fix}}}. \tag{16}$$

Since \tilde{f}_m is equal to the right-hand side of the Euler-Lotka equation [Eq. (4)] with ρ replaced by ρ_r , $\tilde{f}_m(\rho_m) = 1/2$ and $\tilde{f}_m(\rho_r) > 1/2$ if and only if $\rho_r < \rho_m$. Hence, as we would expect, $p_{\text{fix}} > 0$ (the mutant is beneficial) only when $\rho_m > \rho_r$. The connection between this result and the derivation of extinction probabilities by Bellman and Harris (see [25]) is discussed in Appendix D.

From this expression we can make three key observations: (1) As $N \rightarrow \infty$ the fixation probabilities are coupled to the resident population’s growth dynamics only through the population growth rate, ρ_r ; however, (2) the fixation probabilities may depend in more complex ways on the mutant population’s growth dynamics because of the appearance of $f_m(\tau)$. (3) Whether a mutant is deleterious or beneficial is determined by the long-term growth rates. That is, the fixation probability is positive if and only if $\rho_m > \rho_r$.

This last statement can also be deduced from the following intuitive argument: Suppose $\rho_m > \rho_r$. Let $\rho_m(t)$ be the transient population division rate. We know there is some n_c such that when the mutant population size $n > n_c$, $\rho_m(t) > \rho_r$ and at this point the population size n will increase approximately deterministically. There is some finite probability that n will exceed n_c even in an infinitely large population (since n_c can be reached through a finite sequence of cell divisions), thus there is a finite probability the mutants population will fix. Now suppose $\rho_m < \rho_r$. In this case there is an n_c such that $\rho_m(t) < \rho_r$, thus by a similar argument the mutant will never fix. Since the mutant never fixes, it eventually goes extinct with probability 1. Notice that we have not used the

independence of generation times in this argument, thus this result applies to more general models in which generation times are correlated.

B. Markovian case

One case that has been well studied is when the distribution of generation times is exponentially distributed:

$$f_m(\tau) = \rho_m e^{-\rho_m \tau}. \quad (17)$$

In this case the per unit time, per capita division rate is completely independent of the distribution of ages. Using Eq. (16), it is straightforward to derive the well-known result originally obtained by Haldane: $p_{\text{fix}} \approx s$, where s is the selection coefficient defined in Eq. (3).

C. Deterministic case

We now consider the limiting case where the distribution of generation times is nearly a delta function around τ . Thus, all cells in the mutant population divide at almost exactly the same age. In this case, Eq. (15) becomes

$$q \approx 1 - e^{-\rho_r \tau}. \quad (18)$$

Here, we are neglecting any contribution from the variation around τ . Using that the long-term growth rate of the mutant is $\rho_m \approx \ln(2)/\tau$, we obtain

$$p_{\text{fix}} \approx 2 \ln(2)s. \quad (19)$$

In this instance there is an additional factor of $2 \ln(2)$ compared to the previous example. As we have suggested earlier, this discrepancy results from the fact that in the present case, the initial age distribution of the mutant population, and hence the division rate, is distinct from that of the same genotype under stable exponential growth. It is this transient division rate that determines the ultimate fate of the mutant population.

D. General results for gamma distributed division times

The examples above can be seen as special cases of a more general result obtained assuming a gamma distribution of generation times. Then Eq. (15) becomes

$$q = 1 - (1 + \beta \rho_r)^{-\alpha}. \quad (20)$$

Using Eq. (4), we find that

$$\rho_m = \frac{2^{1/\alpha} - 1}{\beta}, \quad (21)$$

and q can be rewritten as

$$q = 1 - \left(1 + \frac{\rho_r}{\rho_m} [2^{1/\alpha} - 1]\right)^{-\alpha} \quad (22)$$

$$= 1 - (1 + (1 - s)[2^{1/\alpha} - 1])^{-\alpha}. \quad (23)$$

Using that $\text{CV}^2 = 1/\alpha$, we can express p_{fix} solely in terms of the population growth rates and the coefficient of variation of mutant generation times:

$$p_{\text{fix}} = \frac{2}{\text{CV}^2} \left(1 - 2^{-\text{CV}^2}\right) s + O(s^2). \quad (24)$$

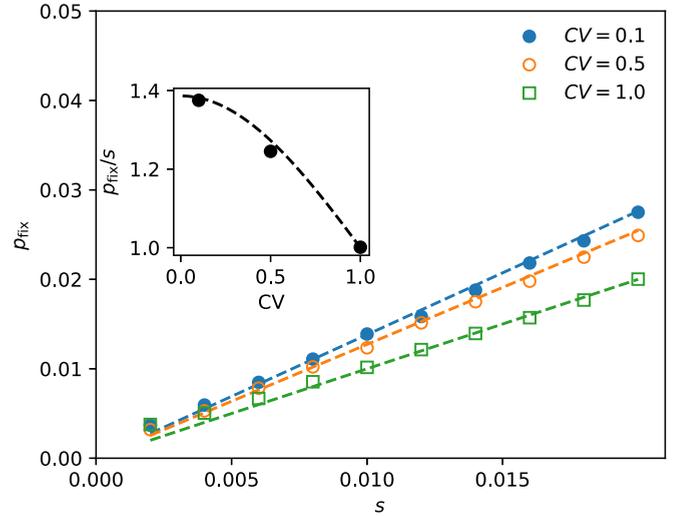


FIG. 7. Numerical simulations of the fixation probability. Populations starting with one cell were run for 10^8 runs and the fixation probability was computed from the fraction of populations that reaches a critical size where genetic drift ceased to dominate the dynamics. Dashed lines show the theoretical prediction of Eq. (24).

For $\text{CV}^2 = 1$ we obtain the Markovian case, $p_{\text{fix}} = s$, while in the limit $\text{CV}^2 \rightarrow 0$ we retrieve the deterministic case $p_{\text{fix}} = 2 \ln(2)s$. Agreement with numerical simulations is shown in Fig. 7. For most bacterial cells, $\text{CV}^2 \ll 1$, which makes the fixation probability larger than the selection coefficient s . Fixation probability would be less than s if $\text{CV}^2 > 1$.

E. Implications for long-term evolution

In the previous section we considered the dynamics following the emergence of a single mutation. We now consider evolutionary dynamics over long periods of time in which multiple mutations accumulate. To this end, we introduce the mutation probability μ , defined as the probability per cell division for a mutation to occur. We assume, for simplicity, that each mutation increases the growth rate by an amount $\delta\rho$. If a mutation does eventually fix, it will take a time τ_{sw} on the order of $1/\delta\rho \ln Ns$ to sweep through the population [26]. We assume $\tau_{\text{sw}} \ll 1/(\mu N \rho s)$, or $\mu N \ln Ns \ll 1$ [27]. This regime is often referred to as the strong selection–weak mutation regime and it ensures mutations are unlikely to compete with each other and can therefore be treated as independent events.

A basic question one can ask regarding the evolutionary trajectories is: How will the growth rate increase over time in such a population? Consider growth rate averaged over many realizations of an evolving population, which we denote $\langle \rho \rangle$. The per unit time increase in $\langle \rho \rangle$ is simply the rate at which mutations emerge multiplied by the probability of fixation; that is,

$$\frac{d}{dt} \langle \rho \rangle = \mu N \delta\rho \langle \rho p_{\text{fix}} \rangle. \quad (25)$$

Using the classical formula $p_{\text{fix}} = s$, we find $d\langle \rho \rangle/dt = \mu N (\delta\rho)^2$ and hence under these assumptions the long-term growth of $\langle \rho \rangle$ is linear.

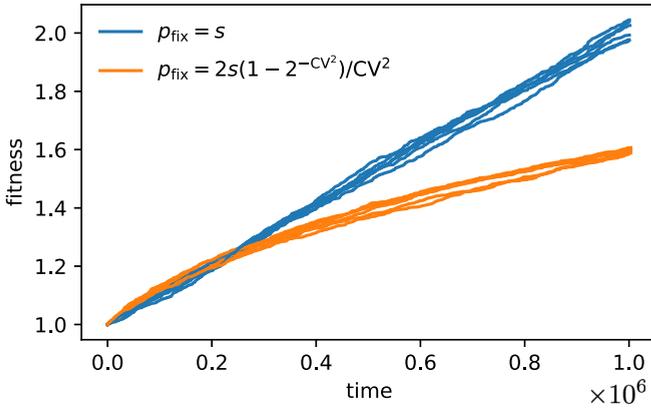


FIG. 8. Numerical simulations of the long-term evolutionary dynamics using Kimura's formula (blue line) and the corrected formula that accounts for division rate fluctuations (orange). Each line is a separate simulation. When accounting for the effects of fluctuations on fixation probabilities, evolution slows down considerably. This is because the coefficient of variation increases as the population adapts, decreasing the chance for new mutants to fix. In our simulations, the times between mutations are drawn from an exponential distribution with rate parameters $\mu N \delta \rho^2 p_{\text{fix}}$. Parameter values used are $s = 0.01$, $\mu = 10^{-6}$, and $N = 10^3$, ensuring we are in the strong selection–weak mutation regime.

We would now like to understand how, in principle, phenotypes (e.g., growth rate variation) can shape long-term evolution without directly altering the selection coefficients. As a contrived, but illustrative example, suppose each mutation increases CV^2 by a fixed increment δCV^2 and the generation time by an increment $\delta \tau$, which is selected to ensure $\delta \rho$ remains constant. As the population adapts, the CV increases. This decreases the fixation probabilities and as a result adaptation slows down. In Fig. 8 we show simulations of this model that illustrate this point. In reality the growth rates in an evolving population will not increase linearly for many other reasons, such as epistasis. However, our results suggest that the transient dynamics of mutant populations are also a factor that influences the long-term evolutionary dynamics. We leave it to future work to explore this effect within the context of more realistic models.

V. DISCUSSION

In the past decade, the use of microfluidic devices has allowed for detailed measurements of growth and division statistics of bacteria at the single-cell level [24,28]. Phenomenological models of bacterial growth, division, and cell-size regulation proposed based on these single-cell statistics [29] have sparked new interest in understanding how single-cell statistics affect population dynamics [11–14,18,30–33]. At the core of most of these studies is the question of how different sources of variability at single-cell level affect population growth rate. Meanwhile, high-throughput continuous-culture experiments allow for the precise measurement of evolutionary dynamics over thousands of generations [34–36]. By growing evolved strains in microfluidic devices, it is in-principle possible to investigate the adaptation of single-cell traits, such as the generation time distribution. It

is therefore natural to ask how different sources of variability at the single-cell level affect population genetics.

Several population dynamics studies have shown that the details of single-cell elongation dynamics (but not details of the division process) affect the growth rate of populations [12–14]. Recent work has drawn attention to the persistence of out-of-steady-state dynamics of a population of cells starting from a single cell [12,21,37,38]. These out-of-steady-state dynamics manifest in forms of oscillations in the population growth rate and traveling waves in distributions of sizes and ages of the cells, and they exist because descendants of a cell maintain a degree of synchrony in their growth and division. Their persistence time in exponentially growing populations also only depends on single-cell elongation dynamics and not the details of cell-division and cell-size regulation [12]. Here, we have shown that for finite populations, similar fluctuations in the division rate persist infinitely long (also observed in Ref. [39]), and they decrease the rate of genetic drift of neutral mutants. We leave it to future work to develop an analytical understanding of how the generation time distribution influences the rate of genetic drift in the Moran model.

In the context of beneficial mutants, we have shown that variability in doubling time reduces the probability of fixation, but it is always larger than the classical results from the Moran process. The analytic expression obtained for fixation probability has an interesting relationship to the Euler-Lotka equation as shown in Fig. 6. The Euler-Lotka equation states that the Laplace transform of the doubling-time distribution evaluated at the population growth rate ρ_m is $1/2$, while in our case, the Laplace transform of the doubling-time distribution evaluated at the population growth rate of the *resident* cells ρ_r is $1/(2 - p_{\text{fix}})$. Finally, we have shown that the small correction to the fixation probability due to out-of-steady-state effects can have potentially large effects on the long-term evolutionary trajectories. For example, if populations adapt by increasing the variability in their generation times, the rate of adaptation will decrease.

Section II highlights the role of fluctuations in population division rate in population genetics. This is done in the context of a very general model of growth, division, and cell-size regulation. However, for simplicity, the rest of the paper is organized around a simpler model in which the only source of noise is the elongation rate of single cells. We leave it to future work to investigate more detailed models of cell-size regulation (such as the ones discussed in Appendix A) and their effects on population genetics.

ACKNOWLEDGMENTS

We thank Oren Raz, David Lacoste, and Simone Pigolotti for helpful comments on this paper. This work was initiated at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-1607611. E.L. was supported by NSF Grant No. DMS-1902895.

APPENDIX A: CELL-SIZE CONTROL MODEL

We assume cells attempt to divide when they reach a target division size $g(s_b)$, which in general is a function of their initial size s_b . The actual division size s_d is the target division

size plus a noise term, $s_d = g(s_b) + \xi$. We assume cells grow exponentially with a possibly random elongation rate λ such that the generation time τ satisfies $s_d = s_b e^{\lambda \tau}$.

Commonly considered candidates for the function g are $g(s_b) = 2s_b$ (the timer model), $g(s_b) = s_b + \Delta$ (the adder model), and $g(s_b) = 2\bar{s}_b$ (the sizer model). The functional form of g is only relevant when the variance of the division noise ξ is nonzero; when this variance is zero, all three of these models are equivalent. With the exception of the simulation shown in Fig. 4 (where the adder model of cell-size regulation is used), we have neglected the division noise ξ , and considered the case where the only source of variability in the generation time τ comes from the variability in λ .

We refer to Ref. [3] for an in-depth discussion of the cell-size control model and its implications for population growth.

APPENDIX B: DERIVATION OF THE FOKKER-PLANCK EQUATION FOR THE NUMBER FRACTION OF MUTANT SPECIES

Here we derive a Fokker-Planck equation by Kramers-Moyal expansion of Eq. (7) whose corresponding SDE is given in Eq. (1). We start by plugging the transition rates from Eq. (8) into the master equation, Eq. (7). Defining $\epsilon \equiv 1/N$, we have

$$\begin{aligned} \frac{d}{dt}P(\phi, t) = & N\rho_m((\phi - \epsilon)(1 - \phi + \epsilon)P(\phi - \epsilon, t) \\ & - \phi(1 - \phi)P(\phi, t)) \\ & + N\rho_r((\phi + \epsilon)(1 - \phi - \epsilon)P(\phi + \epsilon, t) \\ & - \phi(1 - \phi)P(\phi, t)). \end{aligned} \quad (\text{B1})$$

We can recognize this master equation as a difference equation by defining $f(\phi, t) \equiv \phi(1 - \phi)P(\phi, t)$:

$$\begin{aligned} \frac{d}{dt}P(\phi, t) = & \rho_m \frac{f(\phi - \epsilon, t) - f(\phi, t)}{\epsilon} \\ & + \rho_r \frac{f(\phi + \epsilon, t) - f(\phi, t)}{\epsilon}. \end{aligned} \quad (\text{B2})$$

In the continuum limit, we can expand $f(\phi \pm \epsilon, t)$ up to second order in ϵ to derive the Fokker-Planck equation for the dynamics of the probability density $P(\phi, t)$ of the number fraction ϕ in time:

$$\begin{aligned} \frac{\partial}{\partial t}P(\phi, t) \approx & (\rho_r - \rho_m) \frac{\partial}{\partial \phi}[\phi(1 - \phi)P(\phi, t)] \\ & + \frac{1}{2N}(\rho_r + \rho_m) \frac{\partial^2}{\partial \phi^2}[\phi(1 - \phi)P(\phi, t)]. \end{aligned} \quad (\text{B3})$$

This equation describes the time evolution of the probability density of ϕ obeying the SDE given in Eq. (1).

APPENDIX C: RATE OF NEUTRAL GENETIC DRIFT

We discussed in Sec. III that the simulation results deviate from the steady-state approximation for the rate of neutral genetic drift (which we derive below) except when the CV approaches one, which corresponds to exponentially distributed doubling time or the Markovian limit. At the opposite limit,

$\text{CV} \rightarrow 0$, cells are expected to perfectly synchronize. Below, we derive the rate of neutral genetic drift at these two limits.

1. Steady-state approximation

At the steady state, where the division rates ρ_r and ρ_m are time independent, we can use the Fokker-Planck equation given in Eq. (B3) to derive a differential equation for the rate of change of the variance of ϕ starting from some initial value $0 < \phi_0 < 1$. In the neutral case $\rho_r = \rho_m = \rho$, we have

$$\begin{aligned} \frac{d}{dt}\text{var}(\phi) = & \frac{d}{dt} \int_0^1 (\phi - \phi_0)^2 P(\phi, t) d\phi \\ = & \int_0^1 (\phi - \phi_0)^2 \frac{\partial}{\partial t} P(\phi, t) d\phi \\ = & \frac{\rho}{N} \int_0^1 (\phi - \phi_0)^2 \frac{\partial^2}{\partial \phi^2} [\phi(1 - \phi)P(\phi, t)] d\phi. \end{aligned} \quad (\text{C1})$$

After integrating by parts twice, we have

$$\begin{aligned} \frac{d}{dt}\text{var}(\phi) = & \frac{2\rho}{N} \int_0^1 \phi(1 - \phi)P(\phi, t) d\phi \\ = & \frac{2\rho}{N} (\langle \phi \rangle(t) - \langle \phi^2 \rangle(t)) \\ = & \frac{2\rho}{N} [\phi_0(1 - \phi_0) - \text{var}(\phi)], \end{aligned} \quad (\text{C2})$$

where we have used $\langle \phi \rangle(t) = \phi_0$. Solving for $\text{var}(\phi)$, we have

$$\text{var}(\phi) = \phi_0(1 - \phi_0)(1 - e^{-2\rho t/N}), \quad (\text{C3})$$

with the rate of neutral genetic drift given by $R = 2\rho/N$.

2. Deterministic limit

At the limit where the CV of the cell-size doubling time goes to zero, we expect all the division events of each generation to happen at the same time. One might think that in this limit, the age-structured Moran process would approach the Wright-Fisher model where in discrete time intervals, all cells divide and half of the population is discarded. That is not the case here. In the age-structured Moran process, cells divide one at a time and after each division, one cell is removed, and as a result, not all cells get to divide. This distinction survives even at $\text{CV} \rightarrow 0$ limit where all the division events have to take place at the same time. To predict the rate of neutral genetic drift in this limit, we imagine the population of n mutant cells and $N - n$ resident cells at generation g (note that in this limit, division times are discrete, and we can use generation numbers instead of time). In the next generation, we randomly choose previous generation cells to divide and discard cells to keep the population constant. We continue until no cells of generation g are present. Our goal is to find the expected value and the variance of number n' of mutant cells in the next generation $g + 1$. This is a purely combinatorics problem. Below, we will show that the expected value of n' given n is given by $\langle n'|n \rangle = n$ using a symmetry argument. Also, for large N , we will show that the variance of n'

given n is

$$\text{var}(n'|n) = \frac{3}{4N}n(N-n). \quad (\text{C4})$$

Given the conditional mean and variance of n' , at the continuum limit, we can approximate this process as the following stochastic differential equation:

$$\begin{aligned} \frac{d\phi}{dt} &= \sqrt{\frac{3}{4N\tau}}\phi(1-\phi)\xi(t) \\ &= \sqrt{\frac{3\rho}{4N\ln(2)}}\phi(1-\phi)\xi(t). \end{aligned} \quad (\text{C5})$$

The rate of neutral genetic drift for this case is given by

$$R = \frac{3\rho}{4N\ln(2)}, \quad (\text{C6})$$

which is a factor $3/8\ln(2) \approx 0.54$ smaller than the rate of neutral genetic drift in the Markovian case $R = 2\rho/N$.

Note that the continuous-time approximation made here does not become exact at the large N limit, given that the divisions clearly happen in discrete-time intervals. Nevertheless, the estimated ratio of the rate of genetic drift matches the simulation fairly well as seen in the $\text{CV} \rightarrow 0$ limit of Fig. 3.

Derivations of the mean and the variance of n' given n

We use the symmetry that in the neutral case, all the cells are identical, and therefore, the numbers $X_i \in \{0, 1, 2\}$ of their surviving progeny at each generation are identically distributed (albeit dependent) random variables. Since the total population is constant, we have $\sum_{i=1}^N X_i = N$. Therefore, $\sum_{i=1}^N \langle X_i \rangle = N \langle X_i \rangle = N$ or $\langle X_i \rangle = 1$ for each cell i . The number n' of the mutant cells in the new generation is given by the sum of the numbers of progenies of the previous generation mutant cells, $n' = \sum_{i=1}^n X_i$, and therefore we have $\langle n' \rangle = n$, showing that the expected value of the number of mutant cells stays constant throughout the neutral dynamics.

Finding the variance of n' is a bit more tricky. We first show that the variance of the number of progenies of each cell $\text{Var}(X_i)$ is $3/4$. But since X_i 's are dependent, we also need to find the $\text{Cov}(X_i, X_j)$ (for $i \neq j$) to find the variance of $n' = \sum_{i=1}^n X_i$. We will show that it is given by $\text{Cov}(X_i, X_j) = -\text{Var}(X_i)/(N-1)$. Finally, we will express the variance of n' in terms of the variance and covariance of the variables X_i as

$$\begin{aligned} \text{Var}(n') &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= n\text{Var}(X_i) + n(n-1)\text{Cov}(X_i, X_j) \\ &= \frac{3}{4}\frac{n(N-n)}{N-1}, \end{aligned} \quad (\text{C7})$$

recovering Eq. (C4) in the large N limit.

The variance of X_i is given by

$$\text{Var}(X_i) = \sum_{X_i=0}^2 (X_i - \langle X_i \rangle)^2 = p_0 + p_2, \quad (\text{C8})$$

where $p_0 = \mathbb{P}(X_i = 0)$ and $p_2 = \mathbb{P}(X_i = 2)$. Since $\sum_{i=1}^N X_i = N$, $p_0 = p_2$; that is, for every cell with zero progeny, there is exactly one cell with two progenies. So let us find p_2 .

At each step of the process, a cell divides and another is discarded. We enumerate these steps with a discrete-time variable t . We define the variable f to be the expected value of the fraction of generation g cells at time t . At each step, f decreases by $1/N$ due to division. Additionally, f decreases by $1/N$ if the discarded cell is from the previous generation (with the probability f). This gives the dynamics of f as

$$\frac{df}{dt} = -\frac{1}{N} - \frac{1}{N}f = -\frac{1+f}{N}, \quad (\text{C9})$$

with solution $f = 2e^{-t/N} - 1$. This fraction goes to zero after $t = N\ln(2)$ steps. For a cell that divides when the fraction of the previous generation cells is f , the probability that it ends up with both of its progenies in the next generation is given by product of the probabilities $1 - 2/N$ of them not getting discarded in each step of division after f . The number of steps after fraction f is given by $N\ln(2) - N\ln[2/(f+1)] = N\ln(1+f)$. This gives the conditional probability of a cell ending with two surviving daughter cells given that it divided at f as

$$\left(1 - \frac{2}{N}\right)^{N\ln(1+f)} \approx e^{-2\ln(1+f)} = (1+f)^{-2}. \quad (\text{C10})$$

Since each cell of generation g eventually either divides or gets replaced as f decreases from 1 to 0, for each cell, f at which it divides or gets replaced is uniformly distributed. Conditioned on the value of f , the probability of division is $(1/fN)/(1/fN + 1/N) = 1/(1+f)$ (at f , the probability of division of each one of the remaining fN cells is $1/fN$ and the probability of them being discarded is approximately $1/N$). Putting it all together, the probability p_2 that a given cell ends up with two progenies is the integral over f of the probability that it divides at f multiplied by the conditional probability that neither of the daughter cells gets discarded given that it divided at f :

$$p_2 = \int_0^1 \frac{df}{(1+f)^3} = \frac{3}{8}. \quad (\text{C11})$$

This gives $\text{Var}(X_i) = p_0 + p_2 = 2p_2 = 3/4$.

Next, we find the $\text{Cov}(X_i, X_j)$ for $i \neq j$. Let us define $\delta X_i = X_i - \langle X_i \rangle$. The covariance can be written as

$$\text{Cov}(X_i, X_j) = \langle \delta X_i \delta X_j \rangle = \langle \langle \delta X_j | \delta X_i \rangle \delta X_i \rangle, \quad (\text{C12})$$

where $\langle \delta X_j | \delta X_i \rangle$ is the conditional expected value of δX_j given δX_i . Since $\sum_{i=1}^N \delta X_i = 0$, by symmetry, $\langle \delta X_j | \delta X_i \rangle = -1/(N-1)\delta X_i$. This gives

$$\text{Cov}(X_i, X_j) = -\frac{1}{N-1} \langle \delta X_i^2 \rangle = -\frac{1}{N-1} \text{Var}(\delta X_i), \quad (\text{C13})$$

which is the last ingredient needed in Eq. (C7).

APPENDIX D: RELATIONSHIP TO EXTINCTION IN BELLMAN-HARRIS PROCESS

Here, we discuss the relationship between our formula for the fixation probability, Eq. (14), in the age-structured Moran,

and the extinction probability in the Bellman-Harris process that was derived in [25]. In the Bellman-Harris process with binary fission and cell death, cells have a generation time drawn from a distribution $f(\tau)$. At age τ , they divide in two with probability $1 - \mu$, or are killed with probability μ . Since we study fixation probabilities in the large population size limit where cell dilutions can be treated as independent events, many of our formulas are related to those derived [25]. However, there are two key conceptual differences between the Moran model and the Bellman-Harris model that prevent us from obtaining the key results Eqs. (16) and (24) directly from [25]: First, the role of the resident species generation time distribution is lost in the Bellman-Harris process because only a single genotype is present. Second, in the Bellman-Harris process the chance that a cell dies (i.e., is expelled from the culture) is independent of its generation time, while in the Moran model a cell with a longer lifetime is more likely to be expelled.

To make these differences explicit, we provide a brief summary of the derivation in [25]. Let $N(t)$ be the number of cells at time t in the Bellman-Harris branching process. Set $w = \lim_{t \rightarrow \infty} N(t)/N(0)e^{\tilde{\Lambda}t}$ where $\tilde{\Lambda}$ is the long-term growth

rate of the process, accounting for cell deaths. Note that $\tilde{\Lambda}$ differs from Λ in our notation, which is the long-term growth rate if cells are never expelled from the culture. Bellman and Harris consider the generating function

$$\phi(s) = \langle e^{sw} \rangle, \quad (\text{D1})$$

which is shown to satisfy

$$\phi(s) = \mu + (1 - \mu) \int_0^\infty \phi(se^{-\tilde{\Lambda}\tau})^2 f(\tau) d\tau. \quad (\text{D2})$$

Since $\lim_{s \rightarrow -\infty} \phi(s)$ is the probability for $w = 0$, and hence the probability for the population to eventually go extinct, we immediately obtain an equation for the extinction probability:

$$p_{\text{ext}} = \mu + (1 - \mu)p_{\text{ext}}^2. \quad (\text{D3})$$

This is identical to our Eq. (13), with q replaced with μ . However, since the chance for cells to be expelled is independent of their generation time, the generation-time distribution of the mutant species plays no role in this equation and the extinction probability is determined solely by the independent parameter μ . In contrast, in our derivations it is the generation-time distributions of *both* the mutant and the resident that determine q .

-
- [1] Y.-J. Eun, P.-Y. Ho, M. Kim, S. LaRussa, L. Robert, L. D. Renner, A. Schmid, E. Garner, and A. Amir, Archaeal cells share common size control with bacteria despite noisier growth and division, *Nat. Microbiol.* **3**, 148 (2018).
- [2] C. Cadart, S. Monnier, J. Grilli, P. J. Sáez, N. Srivastava, R. Attia, E. Terriac, B. Baum, M. Cosentino-Lagomarsino, and M. Piel, Size control in mammalian cells involves modulation of both growth rate and cell cycle duration, *Nat. Commun.* **9**, 3275 (2018).
- [3] P.-Y. Ho, J. Lin, and A. Amir, Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects, *Annu. Rev. Biophys.* **47**, 251 (2018).
- [4] A. Amir and N. Q. Balaban, Learning from noise: How observing stochasticity may aid microbiology, *Trends Microbiol.* **26**, 376 (2018).
- [5] C. R. Baker, L. N. Booth, T. R. Sorrells, and A. D. Johnson, Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification, *Cell* **151**, 80 (2012).
- [6] P. Schuster and K. Sigmund, Replicator dynamics, *J. Theor. Biol.* **100**, 533 (1983).
- [7] O. Hallatschek, Selection-like biases emerge in population models with recurrent jackpot events. *Genetics* **210**, 1053 (2018).
- [8] J. B. S. Haldane, A mathematical theory of natural and artificial selection, in *Mathematical Proceedings of the Cambridge Philosophical Society* (Cambridge University Press, Cambridge, UK, 1927), Vol. 23, pp. 607–615.
- [9] M. Kimura, Some problems of stochastic processes in genetics, *Ann. Math. Stat.* **28**, 882 (1957).
- [10] N. Bacaër, *A Short History of Mathematical Population Dynamics* (Springer, London, 2011).
- [11] F. Jafarpour, C. S. Wright, H. Gudjonson, J. Riebling, E. Dawson, K. Lo, A. Fiebig, S. Crosson, A. R. Dinner, and S. Iyer-Biswas, Bridging the Timescales of Single-Cell and Population Dynamics, *Phys. Rev. X* **8**, 021007 (2018).
- [12] F. Jafarpour, Cell Size Regulation Induces Sustained Oscillations in the Population Growth Rate, *Phys. Rev. Lett.* **122**, 118101 (2019).
- [13] J. Lin and A. Amir, The effects of stochasticity at the single-cell level and cell size control on the population growth, *Cell Syst.* **5**, 358 (2017).
- [14] J. Lin and A. Amir, From single-cell variability to population growth, *Phys. Rev. E* **101**, 012401 (2020).
- [15] J. Lin, J. Min, and A. Amir, Optimal Segregation of Proteins: Phase Transitions and Symmetry Breaking, *Phys. Rev. Lett.* **122**, 068101 (2019).
- [16] R. García-García, A. Genthon, and D. Lacoste, Linking lineage and population observables in biological branching processes, *Phys. Rev. E* **99**, 042413 (2019).
- [17] P. Thomas, Single-cell histories in growing populations: Relating physiological variability to population growth, *bioRxiv* 100495 (2017), doi:10.1101/100495.
- [18] E. Levien, J. Kondev, and A. Amir, The interplay of phenotypic variability and fitness in finite microbial populations, *J. R. Soc., Interface* **17**, 20190827 (2020).
- [19] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, Noise-driven growth rate gain in clonal cellular populations, *Proc. Natl. Acad. Sci. USA* **113**, 3251 (2016).
- [20] G. Arthur and D. Lacoste, Fluctuation relations and fitness landscapes of growing cell populations, *Sci. Rep.* **10**, 11889 (2020).
- [21] Y. Hein and F. Jafarpour, Asymptotic decoupling of population growth rate and cell size distribution, *arXiv:2209.14683*.
- [22] M. Shpak, Selection against demographic stochasticity in age-structured populations, *Genetics* **177**, 2181 (2007).

- [23] L. Demetrius, The sensitivity of population growth rate to perturbations in the life cycle components, *Math. Biosci.* **4**, 129 (1969).
- [24] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, Cell-size control and homeostasis in bacteria, *Curr. Biol.* **25**, 385 (2015).
- [25] R. Bellman and T. E. Harris, On the theory of age-dependent stochastic branching processes, *Proc. Natl. Acad. Sci. U.S.A.* **34**, 601 (1948).
- [26] M. M. Desai, D. S. Fisher, and A. W. Murray, The speed of evolution and maintenance of variation in asexual populations, *Curr. Biol.* **17**, 385 (2007).
- [27] R. A. Neher and O. Hallatschek, Genealogies of rapidly adapting populations, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 437 (2013).
- [28] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Robust growth of *Escherichia coli*, *Curr. Biol.* **20**, 1099 (2010).
- [29] A. Amir, Cell Size Regulation in Bacteria, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [30] E. Levien, J. Min, J. Kondev, and A. Amir, Non-genetic variability in microbial populations: Survival strategy or nuisance? *Rep. Prog. Phys.* **84**, 116601 (2021).
- [31] E. Levien, T. GrandPre, and A. Amir, Large Deviation Principle Linking Lineage Statistics to Fitness in Microbial Populations, *Phys. Rev. Lett.* **125**, 048102 (2020).
- [32] K. Pichór and R. Rudnicki, Cell cycle length and long-time behaviour of an age-size model, *Math. Meth. Appl. Sci.* **45**, 5797 (2022).
- [33] L. Luo, Y. Bai, and X. Fu, Master equation approach to the stochastic accumulation dynamics of bacterial cell cycle, *New J. Phys.* **23**, 083029 (2021).
- [34] L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez, Mutation dynamics and fitness effects followed in single cells, *Science* **359**, 1283 (2018).
- [35] L. Robert, J. Ollion, and M. Elez, Real-time visualization of mutations and their fitness effects in single bacteria, *Nat. Protoc.* **14**, 3126 (2019).
- [36] B. G. Wong, C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A. S. Khalil, Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER, *Nat. Biotechnol.* **36**, 614 (2018).
- [37] G. Chiorino, J. A. J. Metz, D. Tomasoni, and P. Ubezio, Desynchronization rate in cell populations: Mathematical modeling and experimental data, *J. Theor. Biol.* **208**, 185 (2001).
- [38] E. Gavagnin, S. T. Vittadello, G. Gunasingh, N. K. Haass, M. J. Simpson, T. Rogers, and C. A. Yates, Synchronized oscillations in growing cell populations are explained by demographic noise, *Biophys. J.* **120**, 1314 (2021).
- [39] T. Nozoe and E. Kussell, Cell Cycle Heritability and Localization Phase Transition in Growing Populations, *Phys. Rev. Lett.* **125**, 268103 (2020).