

Large Deviation Principle Linking Lineage Statistics to Fitness in Microbial Populations

Ethan Levien^{1,*}, Trevor GrandPre^{2,*}, and Ariel Amir¹

¹*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, Harvard 02138, USA*

²*Department of Physics, University of California, Berkeley, California, Berkeley 94720, USA*



(Received 11 February 2020; accepted 18 June 2020; published 22 July 2020)

In exponentially proliferating populations of microbes, the population doubles at a rate less than the average doubling time of a single-cell due to variability at the single-cell level. It is known that the distribution of generation times obtained from a single lineage is, in general, insufficient to determine a population's growth rate. Is there an *explicit* relationship between observables obtained from a single lineage and the population growth rate? We show that a population's growth rate can be represented in terms of averages over isolated lineages. This lineage representation is related to a large deviation principle that is a generic feature of exponentially proliferating populations. Due to the large deviation structure of growing populations, the number of lineages needed to obtain an accurate estimate of the growth rate depends exponentially on the duration of the lineages, leading to a nonmonotonic convergence of the estimate, which we verify in both synthetic and experimental data sets.

DOI: 10.1103/PhysRevLett.125.048102

A key determinant of fitness in microbial populations is the population growth rate [1–3]. For organisms such as *Escherichia coli* which undergo binary fission, the exponential growth rate of the population is determined by single-cell properties such as generation time, defined as the time from cell birth to division. In reality, any clonal population of bacteria will exhibit a distribution of generation times due to a combination of intrinsic and environmental factors [4–12] resulting in a distribution of generation times, $\psi(\tau_d)$. The relationship between this distribution and the population growth rate, Λ , has been the subject of numerous studies. A key result is the Euler-Lotka equation [1–3,13–15],

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

which establishes a link between $\psi(\tau)$ and Λ . Equation (1) is a variant of a relation originally obtained by Euler [16,17] and later rediscovered by Lotka [18].

Despite the elegant simplicity of the Euler-Lotka equation, it obscures the underlying relationship between the stochastic dynamics along single lineages and the population growth rate. The reason is that $\psi(\tau_d)$, like Λ , is a property of the population rather than an intrinsic property of individual cells and it is therefore unclear how differences in single cell dynamics are reflected in $\psi(\tau_d)$. Only in the special case where the generation time of a newborn cell is completely uncorrelated with its immediate ancestor, or *mother*, does $\psi(\tau_d)$ correspond to the distribution of generation times along a single lineage [1,14]. When generation times are correlated between mother and daughter cells, the distribution of generation times, $f(\tau_d)$,

along a single lineage no longer contains enough information to deduce the growth exponent Λ using Eq. (1). Such correlations emerge naturally through feedback mechanisms and are required to maintain homeostasis of cell sizes [3,19,20].

The discrepancy between the statistics of a lineage and those from the entire population, which determine the population's fitness, raises the question of how to quantify fitness from data obtained from a single lineage, or a collection of independent lineages (see Fig. 1). Such data is typically obtained from mother machine experiments [21], where independent lineages are tracked for long periods of times in controlled conditions. Mother machine experiments enable detailed measurements of single-cell dynamics that would be impossible in bulk conditions. In contrast, bulk experiments can be used to probe population-level dynamics and measure fitness, but they are blind to the physiological details at the microscopic level [21]. Here, we present a lineage representation of the population growth rate that connects the population dynamics to the statistics along a single lineage, or a collection of independent lineages.

Lineage representation.—A lineage-based representation of the population growth rate that is independent of the model specifics can be derived using the division distribution, denoted $p_T(n)$, which can be obtained from an exponentially growing population as follows. Suppose a population of cells is grown for a time T and assume that we have access to the generation times of individual cells and the genealogical relationships between cells, as shown in Fig. 1. We can randomly sample a lineage from the tree by starting from the ancestral cell in the population and randomly selecting one of its daughter cells with equal

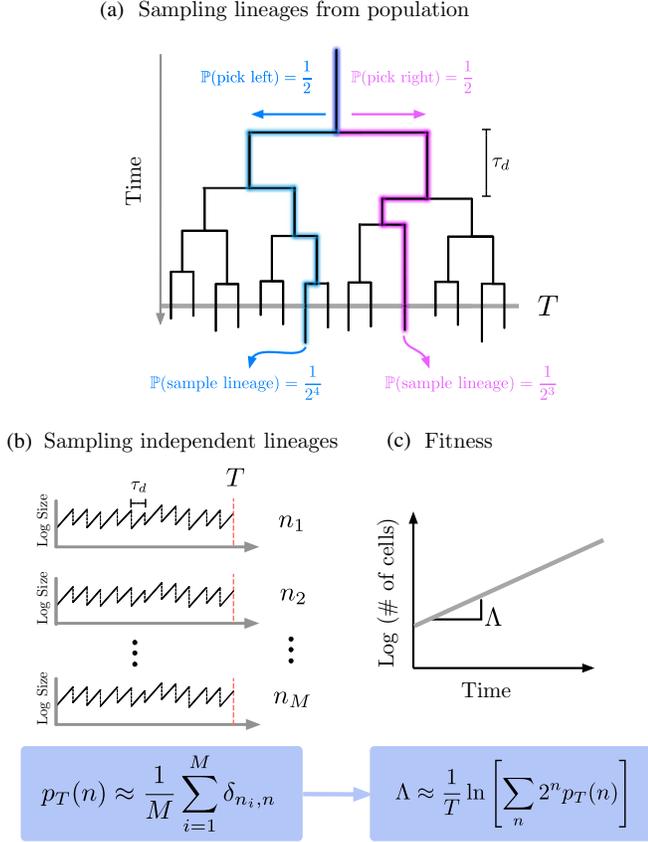


FIG. 1. (a) A population tree starting from a single ancestor. The distinction is made between single lineages (highlighted) and a population (black). Lineages can be sampled by traveling down the tree and randomly selecting a daughter cell at the end of each branch. The probability of selecting any specific lineage with n divisions is 2^{-n} . (b) M independent lineages of length T . For the i th lineage, n_i is the number of cell divisions along that lineage. For each lineage we have shown the cell size, which typically increases exponentially between divisions as a function of time. The lineage division distribution can be approximated from these independent lineages by recording the division events and using the highlighted formula. Here, $\delta_{n,m} = 1$ if $m = n$ and 0 otherwise. (c) A growing population of cells from which one can compute the fitness directly by counting the number of cells as a function of time [or utilizing Eq. (1)]. Using the lineage distribution of divisions we can obtain the fitness from independent lineages.

probability to obtain the next cell in the lineage. Repeating this procedure yields a single lineage, as shown by the highlighted paths in Fig. 1.

If $N(T)$ is the number of cells in the population at time T , then there are exactly $N(T)$ lineages, as each cell in the final population corresponds to a distinct lineage. However, by randomly selecting a lineage in the *forward* manner described above, lineages with more divisions are less likely to be selected, since each division decreases the chance that we will travel down that specific path through the tree. In particular, the probability of drawing any specific lineage from the tree is 2^{-n} . It follows that the

empirical distribution of divisions from lineages sampled in this way, denoted $\hat{p}_T(n)$, is given by [22,23] $\hat{p}_T(n) = 2^{-n}N(n, T)$. Here, $N(n, T)$ is a random variable representing the number of lineages with n divisions in a specific realization of a growing population. Note that $\hat{p}_T(n)$ is also a random variable, and will therefore differ between different realizations of the population tree. By averaging over many realizations of the tree, we obtain the division distribution: $p_T(n) \equiv \langle \hat{p}_T(n) \rangle_{\text{trees}}$. It is important to remember that $p_T(n)$ is distinct from what has been called the *retrospective distribution*, defined as the probability of observing n divisions in a lineage obtained by uniformly sampling a cell from the population at time T and following its ancestors back in time [23].

We define the long term population growth rate as

$$\Lambda \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \ln N(T). \quad (2)$$

This definition of Λ is justified in Supplemental Material [24], where we show that this limit is self-averaging. In order to express Λ in terms of lineage statistics, we note that a population tree in which every lineage has n divisions has 2^n cells. Intuitively, averaging over the contribution of each lineage to the total population therefore gives $N(T) = \langle 2^n \rangle_p$, so Eq. (2) can be expressed in terms of a *lineage representation* (see the Supplemental Material [24] for a proof),

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \langle 2^n \rangle_p. \quad (3)$$

Here, the angular brackets denote an average over $p_T(n)$. The lineage representation in Eq. (3) establishes a relationship between the lineage dynamics and the population fitness. A similar formulation was used in Ref. [23] to quantify how selection acts on an observable in a growing population; however, our formulation differs in that the average is taken over independent lineages rather than lineages from a single growing population.

In order to apply the lineage representation of the population growth rate to real data, we must develop an understanding of how quickly it converges in the number of lineages, M , and the duration of each lineage, T . However, before presenting our convergence analysis, we establish a relationship between the lineage representation and the large deviation principle underlying the growth process. This structure is best introduced with the example below.

Explicit calculation of $p_T(n)$ for discrete Langevin model.—We now perform an explicit calculation of $p_T(n)$ for a specific model in which generation times undergo a discrete Langevin process along a lineage, referred to as the *random generation time model* [1,2]. In this model, the generation time τ of a cell is related to its mother's generation time, τ' , according to $\tau = \tau_0(1 - c) + \tau'c + \xi$, where ξ is a Gaussian with mean

zero and variance σ_ξ^2 . It can be seen that the average generation time along a lineage is $\langle \tau \rangle = \tau_0$. The parameter c controls the strength of correlations between mother and daughter cells. By evaluating a path integral (see the Supplemental Material [24]), we obtain

$$p_T(n) = K e^{-n[(1-c)^2/2\sigma_\tau^2](\tau_0 - T/n)^2}, \quad (4)$$

where K is a normalization constant independent of n and $\sigma_\tau^2 = \sigma_\xi^2/(1-c^2)$ is the variance in τ taken over a single lineage.

The exponential form of Eq. (4) along with Eq. (3) suggests that the population growth rate is dominated by a particular value of n which maximizes the exponent of $2^n p_T(n)$. Treating n as a continuous variable and solving for n in $\partial/\partial n[n \ln 2 + \ln p_T(n)] = 0$ yields the dominant number of divisions

$$n_c = T / \sqrt{\langle \tau \rangle^2 - \frac{2 \ln(2) \sigma_\tau^2}{(1-c)^2}}. \quad (5)$$

Note that in the limit where $\sigma_\tau^2 \rightarrow 0$, we find $n_c = T/\langle \tau \rangle$, which is the number of divisions corresponding to the average generation time. Substituting the dominant value of n from Eq. (5) into Eq. (3) gives us the formula for the bulk population growth rate: $\Lambda = n_c \ln(2)/T + \ln p_T(n_c)/T$. After some simplification, we obtain

$$\Lambda = \frac{2 \ln(2)/\langle \tau \rangle}{1 + \sqrt{1 - 2 \ln(2) \frac{\sigma_\tau^2}{\langle \tau \rangle^2} \frac{1+c}{1-c}}}, \quad (6)$$

which is in agreement with previous computations using an alternative approach [2]. From Eq. (6), we can see how the three model parameters—namely the average generation time $\langle \tau \rangle$, the variance in generation times σ_τ^2 , and the mother-daughter correlations c —affect the population growth. In particular, growth is increased when σ_τ^2 and c are increased, while increasing $\langle \tau \rangle$ decreases growth. In Fig. 2, we compare Eq. (6) to the result of the lineage algorithm. We see that the two approaches to computing the population growth rate agree, illustrating that, given enough data, the lineage representation accurately predicts the population growth rate and captures the dependence on the model parameters, such as the mother-daughter correlations. This is true for any model of exponential growth, not only the Langevin model (see the Supplemental Material [24] where we have applied the algorithm to a different model).

Large deviation principle.—In order to connect the observation of the previous section to large deviation theory, we introduce the time averaged division rate $\gamma = n/T$, so that the distribution of division rates given by Eq. (4) can be expressed as

$$p_T(\gamma) \propto e^{-TI(\gamma)}, \quad (7)$$

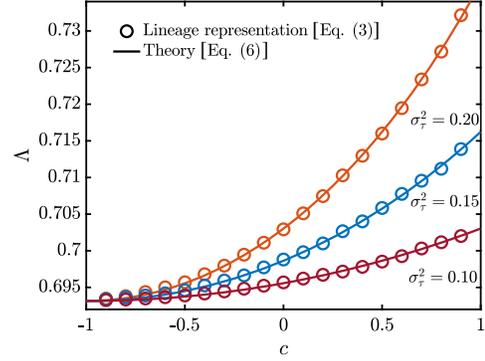


FIG. 2. Comparison of the analytical formula for the population growth rate [Eq. (6)] shown as solid lines with the lineage representation as a function of the mother-daughter correlations shown as open circles, c . Simulations were carried out by drawing the generation times of each daughter cell from a normal distribution with mean $\tau_0(1-c) + \tau c$ and variance σ_τ^2 . The parameters were $\tau_0 = 1$, and $\sigma_\tau^2 = 0.2, 0.15$, and 0.1 with $\sigma_\tau^2 = \sigma_\xi^2/(1-c^2)$. Error bars are smaller than the symbols. Desired accuracy was achieved with $M = 10^3$ – 10^6 lineages and $T = 10^3$ time duration.

with

$$I(\gamma) = \frac{\gamma(1-c)^2}{2\sigma_\tau^2} (\tau_0 - 1/\gamma)^2. \quad (8)$$

The exponential dependence of $p_T(\gamma)$ on T is known as a large deviation principle and suggests that for large T averages over $p_T(\gamma)$ are dominated by a single value of γ [30]. For the remainder of this Letter, we will assume this large deviation principle is satisfied. $I(\gamma)$ is known as the large deviation rate function and encodes all the information about the model details. For example, in the case of the random generation time model, Eq. (8) tells us that this function flattens as c approaches 1. It is convenient to work with the large deviation rate function because it is not model specific; while different forms of cell-to-cell variability (e.g., variability in generation times or growth rates) or correlations between mother and daughter cells can have model-dependent effects on the population growth rate, statements about how the large deviation rate function affect population growth are universal. In order to express the population growth rate in terms of $I(\gamma)$, we can make a saddle point approximation to the average in equation (3) (see the Supplemental Material [24]), leading to the variational formula

$$\Lambda = \max_\gamma [\gamma \ln 2 - I(\gamma)]. \quad (9)$$

This formula reflects a trade-off between larger values of γ leading to an exponentially larger progeny (as captured by $\gamma \ln 2$) while also being rare in the population (as captured by the large deviation rate function). For the random

generation time model, carrying out the maximization to obtain the optimal γ in Eq. (9) is equivalent to computing Eq. (5).

In Supplemental Material [24], we show that when $I''(\langle\gamma\rangle_p) \gg 1$ we can make a Gaussian approximation of $p_T(\gamma)$ to obtain

$$\Lambda \approx \frac{\ln(2)}{\langle\tau\rangle} + \frac{T \ln(2)^2 \sigma_\gamma^2}{2}, \quad (10)$$

with $\sigma_\gamma^2 = 1/TI''(\langle\gamma\rangle_p)$. The quantity $I''(\langle\gamma\rangle_p)$ plays an important role in determining our ability to estimate Λ from real data, but first, we illustrate how the approximation of Eq. (10) relates to the model parameters of the random generation time model. Combining Eqs. (10) and (8) yields

$$\Lambda \approx \frac{\ln(2)}{\langle\tau\rangle} + \frac{\ln(2)^2}{2} \frac{\sigma_\tau^2}{(1-c)^2 \langle\tau\rangle^3}. \quad (11)$$

This is equal to the exact growth rate given in Eq. (6) up to terms of order $\sigma^2 c$. If there is either a large amount of variability or strong mother-daughter correlations, the large deviation rate function becomes flatter and its behavior away from $\langle\gamma\rangle_p$ becomes relevant.

Convergence of lineage representation.—We now address the question: how accurately can we estimate Λ given M lineages with durations T ? To quantify the accuracy of an estimate of Λ , denoted $\hat{\Lambda}_{\text{lin}}$, we use the averaged squared deviation

$$\text{err}(\hat{\Lambda}_{\text{lin}})^2 = \langle (\hat{\Lambda}_{\text{lin}}/\Lambda - 1)^2 \rangle_\mathcal{E}. \quad (12)$$

Here, the average $\langle \cdot \rangle_\mathcal{E}$ represents the average over many realizations of the ensemble of M lineages of duration T , not to be confused with the averages elsewhere that are taken over the lineage representation. Two distinct factors contribute to the error: first, the estimate of Λ obtained from the lineage representation will be subject to a systematic error resulting from the fact that given an infinite number of lineages each with a finite duration T , the lineage representation produces the *arithmetic mean* fitness at time T : $\Lambda_{T,a} = 1/T \ln\langle N \rangle$ (this is distinct from Λ ; see the Supplemental Material [24] and Ref. [25]). We refer to this error as *finite duration error*, and as we have shown in the Supplemental Material [24], it will scale inversely with T .

The second factor contributing to $\text{err}(\hat{\Lambda})$ is sampling error in the approximation of the average $\langle 2^n \rangle_p$ from a finite number of lineages. As shown in the Supplemental Material [24], when

$$\frac{1}{\langle 2^n \rangle_p} \sqrt{\frac{\text{var}(2^n)}{M}} = \sqrt{\frac{2^{T \ln(2)/I''(\langle\gamma\rangle_p)} - 1}{M}} \ll 1, \quad (13)$$

the contribution of the sampling error to $\text{err}(\hat{\Lambda}_{\text{lin}})$ will grow exponentially in T for any fixed M . Eventually the sampling error will dominate the error resulting from finite lineage durations.

As T becomes large, the distribution of γ becomes much more narrow, so an ever-increasing number of lineages are needed to sample the variation in γ . In the long-time limit, all information about the variation is lost for finite M and the lineage representation simply retrieves the zeroth-order term in Eq. (10):

$$\lim_{T \rightarrow \infty} \hat{\Lambda}_{\text{lin}} = \ln(2) \langle\gamma\rangle_p. \quad (14)$$

This demonstrates that the T and M limits do not commute. As a result, there is a ‘‘Goldilocks effect’’: if T is too small the estimate will be inaccurate due to the finite duration error, while if T is too large we encounter the limit given by Eq. (14). The best estimate is in fact obtained by using an intermediate T where both effects are minimized. This prediction is validated numerically for the Langevin model in Fig. 3. We have also generated the same data for a more biophysically realistic model of cell growth (the cell-size regulation model [19]), and found the results are qualitatively similar (see the Supplemental Material [24]).

How much data do we need to be confident we are not encountering the limit given by Eq. (14)? The sampling error will have a negligible effect on the estimate when Eq. (13) is satisfied. This condition can be rewritten as $M \gg 2^{T \ln(2)/I''(\langle\gamma\rangle_p)}$. This implies that the number of lineages needed to avoid encountering the sampling error grows exponentially with the duration of the lineages and the generation time variance. In order to be confident that the finite duration error is small enough to resolve the second term in Eq. (10), we must select $T \gg I''(\gamma_c)$. This means that $T/I''(\langle\gamma\rangle_p)$ is necessarily a large quantity. For example, if we want to ensure that the finite duration error is an order of magnitude smaller than the generation time variance, a conservative choice of M will be much larger than $2^{10 \times \ln(2)} \approx 120$. However, in Fig. 3, we see that using only $M = 80$ lineages and $T = 120$ generations gives an error of $\text{err}(\hat{\Lambda})^2 \approx 2 \times 10^{-5}$. In this case $[\ln(2)/\langle\tau\rangle - \Lambda]/\Lambda - 1 \approx 2 \times 10^{-2}$, so the lineage algorithm already performs much better than the naive estimate. In the Supplemental Material [24], we have explored the applications of the lineage algorithm to mother machine data, where we have found that the dependence of $\hat{\Lambda}_{\text{lin}}$ on T is qualitatively consistent with the theory presented above.

Discussion.—Experimental advances over the last few decades have made it possible to observe the stochastic dynamics of growth and division in bacteria with increasing levels of precision [19–21,31–37]. These observations have revealed universal principles underlying microbial

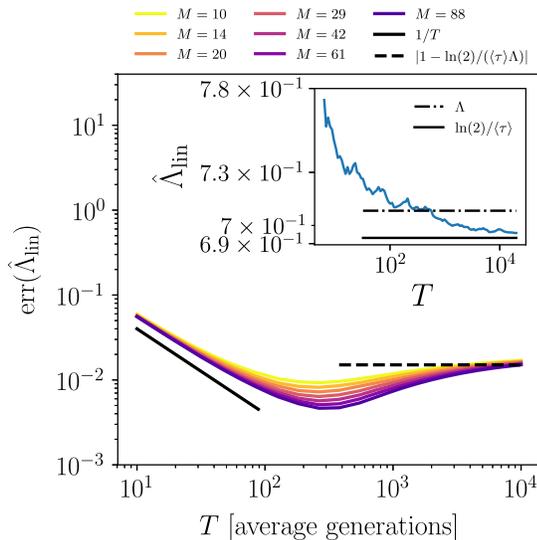


FIG. 3. Convergence of the error from the lineage representation as a function of the lineage durations, T , for different numbers of lineages, M . Data was generated from lineage simulations of the Langevin model with $\langle\tau\rangle = 1$, $c = 0.2$, and $\sigma_\tau = 0.2$. Here it can clearly be seen that the error initially scales as $1/T$, eventually increasing to approach the limit imposed on the sampling error by the large deviation rate function [see Eq. (14)]. The inset shows the lineage representation $\hat{\Lambda}_{\text{lin}}$ as a function of T using $M = 80$ lineages. This plot is noisy because only a single ensemble of lineages is used, in contrast the error in the main plot is computed by averaging over many ensembles of lineages.

growth. Bulk experiments in which bacteria are grown exponentially or in competition assays can be used to compare the fitness of different strains, and in principle elucidate how these physiological differences map to fitness. However, the equivalence between growth in bulk experiments and those used to observe single-cell traits remains unclear.

We have presented a lineage representation that links single-lineages to fitness by leveraging the large deviation structure of microbial population growth. This idea is reminiscent to the *optimal lineage* principle introduced by Wakamoto *et al.* which was used to calculate the population growth rate within the context of a model with uncorrelated generation times [22]. We have quantified exactly how much data is needed to resolve the effects of cell-to-cell variability on population growth from single lineages. We expect that this work will serve as a guide for future experimental studies seeking to link single-cell observations to fitness.

The code to generate Figs. 2 and 3 can be downloaded at [38].

We thank Jie Lin and Somenath Bakshi for helpful discussions related to this work. We thank Jane Kondev for providing feedback on the manuscript. We acknowledge

funding support from National Science Foundation under NSF CAREER Grant No. DMS-1902895 (E. L.), Californian Alliance Research Exchange with NSF Grants No. 1647273, No. 1742065, No. 1306595, No. 1306683, No. 1306747, and No. 1306760 (T. G.) and NSF CAREER 1752024 (A. A.).

*These authors contributed equally to this work.

- [1] J. Lin and A. Amir, *Cell Syst.* **5**, 358 (2017).
- [2] J. Lin and A. Amir, *Phys. Rev. E* **101**, 012401 (2020).
- [3] E. Levien, J. Kondev, and A. Amir, *J. R. Soc. Interface* **17**, 20190827 (2020).
- [4] Z. Wang and J. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* **108**, E67 (2011).
- [5] F. Duveau, A. Hodgins-Davis, B. P. Metzger, B. Yang, S. Tryban, E. A. Walker, T. Lybrook, and P. J. Wittkopp, *eLife* **7**, e37272 (2018).
- [6] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
- [7] Y. Sughiyama and T. J. Kobayashi, *Phys. Rev. E* **95**, 012131 (2017).
- [8] L. Chao, C. U. Rang, A. M. Proenca, and J. U. Chao, *PLoS Comput. Biol.* **12**, e1004700 (2016).
- [9] S. Vedel, H. Nunns, A. Košmrlj, S. Semsey, and A. Trusina, *Cell Syst.* **3**, 187 (2016).
- [10] A. Marantan and A. Amir, *Phys. Rev. E* **94**, 012405 (2016).
- [11] J. Min, J. Lin, and A. Amir, *Phys. Rev. Lett.* **122**, 116402 (2019).
- [12] B. Claudi, P. Spröte, A. Chirkova, N. Personnic, J. Zankl, N. Schürmann, A. Schmidt, and D. Bumann, *Cell* **158**, 722 (2014).
- [13] E. O. Powell, *Microbiology* **15**, 492 (1956).
- [14] J. L. Lebowitz and S. I. Rubinow, *J. Math. Biol.* **1**, 17 (1974).
- [15] R. Garcia-Garcia, A. Genthon, and D. Lacoste, *Phys. Rev. E* **99**, 042413 (2019).
- [16] N. Bacaër, *A Short History of Mathematical Population Dynamics* (Springer, London, 2011).
- [17] N. Keyfitz and D. Smith, *Mathematical Demography: Selected Papers* (Springer, New York, 1977).
- [18] A. J. Lotka, *Science* **26**, 21 (1907).
- [19] A. Amir, *Phys. Rev. Lett.* **112**, 208102 (2014).
- [20] P.-Y. Ho, J. Lin, and A. Amir, *Annu. Rev. Biophys.* **47**, 251 (2018).
- [21] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, *Curr. Biol.* **20**, 1099 (2010).
- [22] Y. Wakamoto, A. Y. Grosberg, and E. Kussell, *Evolution* **66**, 115 (2012).
- [23] T. Nozoe, E. Kussell, and Y. Wakamoto, *PLoS Genet.* **13**, e1006653 (2017).
- [24] See the Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.125.048102> for additional details of calculations, validation of the lineage representation on the cell-size regulation model, and application to experimental data, which includes Refs. [1,2,4,14,19,20,25–29].
- [25] R. C. Lewontin and D. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **62**, 1056 (1969).

- [26] R. Bellman and T. Harris, *Ann. Math.* **55**, 280 (1952).
- [27] F. Jafarpour, *Phys. Rev. Lett.* **122**, 118101 (2019).
- [28] F. Jafarpour, *Phys. Rev. X* **8**, 021007 (2018).
- [29] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, *Sci. Data* **4**, 170036 (2017).
- [30] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [31] S. Luro, L. Potvin-Trottier, B. Okumus, and J. Paulsson, *Nat. Methods* **17**, 93 (2020).
- [32] D. Camsund, M. J. Lawson, J. Larsson, D. Jones, S. Zikrin, D. Fange, and J. Elf, *Nat. Methods* **17**, 86 (2020).
- [33] C. Cadart, S. Monnier, J. Grilli, P. J. Sáez, N. Srivastava, R. Attia, E. Terriac, B. Baum, M. Cosentino-Lagomarsino, and M. Piel, *Nat. Commun.* **9**, 3275 (2018).
- [34] Y.-J. Eun, P.-Y. Ho, M. Kim, S. LaRussa, L. Robert, L. D. Renner, A. Schmid, E. Garner, and A. Amir, *Nat. Rev. Microbiol.* **3**, 148 (2017).
- [35] M. M. Logsdon, P.-Y. Ho, K. Papavinasasundaram, K. Richardson, M. Cokol, C. M. Sassetti, A. Amir, and B. B. Aldridge, *Curr. Biol.* **27**, 3367 (2017).
- [36] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, *Curr. Biol.* **25**, 385 (2015).
- [37] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, *Cell* **159**, 1433 (2014).
- [38] https://bitbucket.org/elevien/ldp_lineage_statistics/src/master/