

RESEARCH ARTICLE

# Extremal events dictate population growth rate inference

Trevor GrandPre<sup>1,2,3</sup>, Ethan Levien<sup>4</sup>, Ariel Amir<sup>5\*</sup>

**1** Department of Physics, Washington University in St. Louis, St. Louis, Missouri, United States of America, **2** National Institute for Theory and Mathematics in Biology, Northwestern University and The University of Chicago, Chicago, Illinois, United States of America, **3** Department of Physics, Princeton University, Princeton, New Jersey, United States of America, **4** Department of Mathematics, Dartmouth College, Hanover, New Hampshire, United States of America, **5** Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel

☞ These authors contributed equally to this work.

\* [ariel.amir@weizmann.ac.il](mailto:ariel.amir@weizmann.ac.il)



**OPEN ACCESS**

**Citation:** GrandPre T, Levien E, Amir A (2026) Extremal events dictate population growth rate inference. *PLoS Comput Biol* 22(5): e1014088. <https://doi.org/10.1371/journal.pcbi.1014088>

**Editor:** Ville Mustonen, University of Helsinki, FINLAND

**Received:** March 30, 2025

**Accepted:** March 4, 2026

**Published:** May 13, 2026

**Copyright:** © 2026 GrandPre et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All data and code required to reproduce the results of this study are available at <https://github.com/elevien/LDPrediction>. This repository includes the full collection of Jupyter notebooks, Julia source files, and project dependencies used to generate the figures and analyses in the manuscript.

**Funding:** T.G. was supported in part by the National Science Foundation through the

## Abstract

Recent methods have been developed to map single-cell lineage statistics to population growth. Because population growth selects for exponentially rare phenotypes, these methods inherently depend on sampling large deviations from finite data, which introduces systematic errors. A comprehensive understanding of these errors in the context of finite data remains elusive. To address this gap, we study the error in growth rate estimates across different models. We show that under the usual bias-variance decomposition, the bias can be decomposed into a finite-time bias and nonlinear averaging bias. We demonstrate that finite-time bias, which dominates at short times, can be mitigated by fitting its monotonic behavior. In contrast, at longer times, nonlinear averaging bias becomes the predominant source of error, leading to a phase transition. This transition can be understood through the Random Energy Model, a mean-field model of disordered systems, where a few lineages dominate the estimator. Applying these methods to experimental data demonstrates that correcting for biases in lineage-based approaches yields consistent results for the long-term growth rate across multiple methods and enables the reverse-engineering of dynamic models. This new framework provides a quantitative understanding of growth rate estimators, clarifies the conditions under which they can be effectively applied to finite data, and introduces model-free approaches for studying the connections between physiology and cell growth.

## Author summary

Understanding how quickly a microbial population grows is a central question in biology, intimately linked to evolutionary fitness. While recent advances have made it possible to estimate growth rates from single-cell data, these estimates often vary widely in practice. In this work, we demonstrate that such

Center for the Physics of Biological Function (PHY-1734030) and by the Schmidt Science Fellowship. A.A. was supported by the European Research Council (101125981), the Israel Science Foundation (146873), and the Clore Center for Biological Physics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

inconsistencies arise from fundamental limitations imposed by the fact exponential growth selects for exponentially rare phenotypes, which dictate the growth rate. Here, we show that two widely used “model-free” approaches both suffer from tradeoffs between two sources of bias: at short timescales, limited observation windows lead to underestimation, while at longer timescales, a small number of exceptionally fast-growing cells disproportionately influence the growth rate. We present a unified framework that disentangles and corrects both sources of error, enabling robust growth rate estimates even from modest datasets. Our results clarify when lineage-based methods can be trusted and what kinds of data are required to accurately infer population growth from single-cell measurements.

## 1. Introduction

A central objective in biology is to understand the relationship between genotype, phenotype, and fitness [1–10]. In the context of microbes, a key component of fitness is the long-term growth rate [7], defined as

$$\Lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \left[ \frac{N(T)}{N(0)} \right], \quad (1)$$

where  $N(0)$  and  $N(t)$  are the population sizes at times  $t=0$  and  $t=T$ .

Experimentally,  $\Lambda$  can be measured by bulk fitness assays. When performed on libraries of different strains, such experiments can yield insight into the genotype-to-fitness map [11]. However, they do not reveal how individual single-cell traits contribute to fitness. The theoretical foundations of this question can be traced back to seminal work in demography, dating to Euler [12]. Most notably, the Euler-Lotka equation relates the population growth rate to the distribution of individual lifetimes within a population. For a population undergoing binary fission, this relationship is expressed as [12–15]

$$\mathbb{E}[e^{-\Lambda\tau}] = \frac{1}{2}. \quad (2)$$

In its modern form [16–18], which accommodates correlated generation times, the average  $\mathbb{E}[\cdot]$  is taken over all generation times  $\tau$  throughout the entire population tree. A key implication of Eq. 2 is that

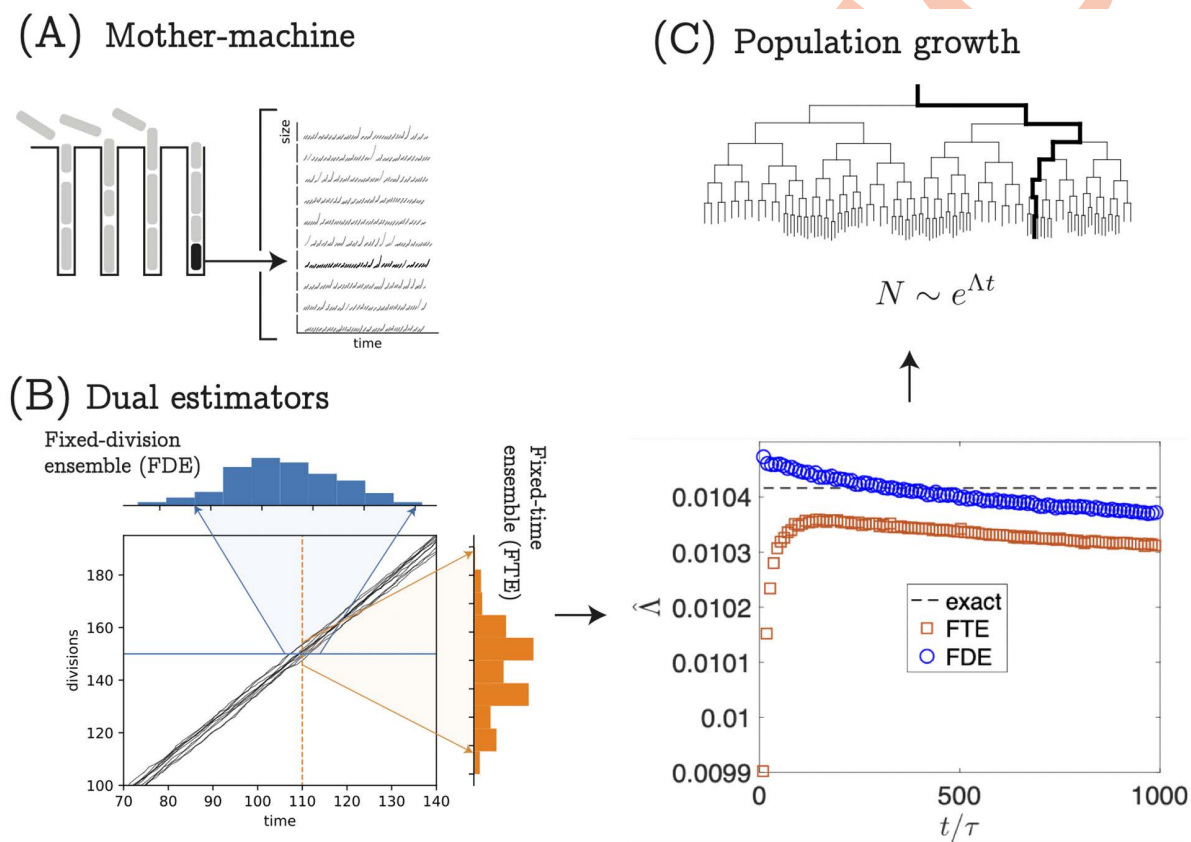
$$\Lambda \geq \frac{\ln(2)}{\mathbb{E}[\tau]} \quad (3)$$

which follows from Jensen’s inequality. In a population where every cell divides after a fixed time  $\tau_0$ , the growth rate is exactly  $\ln(2)/\tau_0$ . Equation (3) implies that introducing variability in generation times, while keeping the same mean generation time, cannot decrease the population growth rate below this bound. However, variability in single-cell growth rates, while fixing the mean growth rate, may lower the population

growth rate compared to a population without such variability [18,19]. In this case, the inequality in Eq. (3) still holds, but the impact of growth rate fluctuations is captured through changes in  $\mathbb{E}[\tau]$ , leading to a slower population growth rate.

Mother machine experiments offer a powerful investigative tool to explore the effects of mutations on single-cell lineage dynamics (see Fig 1A). In these experiments, cells are confined in microfluidic channels, with all but one lineage being expelled from the experiment in each channel [20,21]. Barcoding techniques have been developed to enable mother machine assays across multiple strains simultaneously [22]. A key question is how to connect the data from these experiments to data from bulk fitness assays. Although the Euler-Lotka equation provides a theoretical framework for connecting single-cell generation time statistics to bulk fitness, the average in Eq. 2 is taken over the entire population tree, whereas mother machines capture data from only single lineages. These two approaches coincide only when generation times are uncorrelated.

Recently, studies have explored the challenges in predicting the population growth rate from single-cell lineage statistics [5,10,19,24–26]. One estimator of the population growth rate from lineages is described in Ref. [24]:



**Fig 1. (A)** Schematic of the mother-machine microfluidic device used to collect single-cell lineage data, alongside an example of lineage data. **(B)** Dual ensembles can be used to extract large-deviation statistics relevant for growth-rate inference. This is illustrated using data generated from simulations of the cell-size control model defined in S1 Text, Cell-size control model [23], which produce an ensemble of single-cell trajectories relating division number to elapsed time. The blue histogram is obtained by taking a horizontal cross-section of this data at 150 divisions, while the orange histogram is obtained by taking a vertical cross-section at fixed elapsed time. Using these histograms to estimate the scaled cumulant generating function yields the fixed-divisions ensemble (FDE), Eq. (29), in which the number of divisions is fixed and the elapsed time fluctuates, and the fixed-time ensemble (FTE), Eq. (24), in which the total time is fixed and the number of divisions fluctuates, as defined in the SI [23]. **(C)** These lead to estimators of the exponential growth rate of a growing population whose lineages, when sampled by traveling forward in time from the root of the tree, have the same division statistics as the lineages in (A). The plot below the population tree shows the estimators applied to the same model of cell-size control as in (B).

<https://doi.org/10.1371/journal.pcbi.1014088.g001>

$$\hat{\Lambda}_1 = \frac{1}{T} \ln \left[ \frac{1}{M} \sum_{i=1}^M 2^{N_i^{(t)}} \right] \quad (4)$$

where  $N_i^{(t)}$  is the number of divisions along the  $i$ 'th lineage within a fixed time  $T$  and  $M$  is the number of lineages (see [Methods 4.1](#) for more details). In practice, this method is sensitive to extremal statistics of the sampled generation times or division counts, leading to non-monotonic convergence in the duration of time intervals. At short times, this estimator has a monotonic convergence with time. At longer times (and fixed  $M$ ), the sample averages become dominated by extremal statistics. As a result, the naive estimate on the right-hand side of [Eq. \(3\)](#) is obtained rather than the true growth rate [\[24\]](#).

Another estimator of the population growth rate from lineages is described in Ref [\[25\]](#):

$$\frac{1}{n} \ln \left[ \frac{1}{M} \sum_{i=1}^M e^{-T_n^{(i)} \hat{\Lambda}_2} \right] = -\ln(2). \quad (5)$$

Here,  $T_n^{(i)}$  represents the time of each lineage at a fixed number of divisions  $n$ , and we can solve for  $\hat{\Lambda}_2$  in the exponent. Unlike the standard Euler-Lotka equation in [Eq. \(2\)](#), this formulation is a generalized Euler-Lotka equation tailored to observables along lineages, making it particularly suitable for mother-machine data.

Both of these estimators can be related to a statistical physics theory called large deviation theory (see [Methods 4.1](#)), a generalization to the equilibrium free energy. Within this framework, they are considered to be two ensembles. [Equation 4](#) is to be considered a fixed divisions ensemble (FDE) and [Eq. 5](#) is a fixed time ensemble (FTE). In the limit that the time  $T$  goes to infinity in FTE and the  $n$  goes to infinity in FDE (e.g., an infinite amount of data), both estimators match  $\hat{\Lambda}_1 = \hat{\Lambda}_2$ . However, experimental data is finite, which leads to inconsistencies between the two estimates in practice. Our understanding of how these two methods perform in practice with finite data is incomplete and qualitative.

In this paper, we quantitatively study the systematic biases in both estimators, and show how to remove these biases to accurately compare the estimators. In [Sec. 2.1](#), we quantify the biases using the conventional bias-variance trade-off framework [\[27,28\]](#); the bias can be decomposed into two distinct components: a finite-time bias that persists even with an infinite number of lineages, and a nonlinear averaging bias arising when the lineage ensemble is not self-averaging (see [Methods 4.1](#)). In [Section 2.2](#), we show that finite-size scaling—a well-established concept in statistical physics [\[29\]](#)—can be employed to completely eliminate the finite-time bias.

In [Section 2.3](#), we establish that finite-lineage effects can be understood through their connection to the Random Energy Model (REM), a mean-field model of disordered systems. This connection clarifies how non-monotonic convergence arises from a phase transition into a “frozen state”. In the REM, this transition corresponds to the system entering a low-temperature phase dominated by a few energy states. Similarly, in the context of growth-rate estimators, this transition occurs when a few extremal lineages begin to dominate the ensemble, leading to analogous behavior. Our analysis draws parallels with previous studies on Jarzynski estimators of free energy differences [\[30–32\]](#), and also connects to research on estimating large deviations, particularly in the context of predicting the bandwidth of tele-traffic streams [\[33–35\]](#). We explore these connections further in SI [\[23\]](#). Our findings provide a framework for determining when growth rate estimation is feasible from single-cell data.

## 2. Results

### 2.1. Error decomposition

The performance of the estimators can be quantified by root mean square error,

$$\varepsilon(\hat{\Lambda}) = \sqrt{\mathbb{E}[(\hat{\Lambda} - \Lambda)^2]}. \quad (6)$$

Assuming perfect measurements of division times and counts, applying the standard bias-variance decomposition yields [27,28]:

$$\varepsilon^2(\hat{\Lambda}) = \underbrace{\mathbb{E}[(\hat{\Lambda} - \mathbb{E}[\hat{\Lambda}])^2]}_{\text{var}(\hat{\Lambda})} + \underbrace{(\mathbb{E}[\hat{\Lambda}] - \Lambda)^2}_{\text{Bias}(\hat{\Lambda})^2} \quad (7)$$

where  $\mathbb{E}[\cdot]$  is an ensemble average of  $\hat{\Lambda}$  from many realizations of the growth process. This formula follows from the definition of var and  $\varepsilon(\hat{\Lambda})$ .

To fully grasp the non-monotonic convergence, it is crucial to closely examine the bias term, which can be further decomposed as:

$$\text{Bias}(t, M) = \underbrace{(\mathbb{E}[\hat{\Lambda}] - \Lambda(t))}_{\text{Bias}_{\text{nl}}(t, M)} + \underbrace{(\Lambda(t) - \Lambda)}_{\text{Bias}_{\text{ft}}(t)} \quad (8)$$

where  $\Lambda(t) = t^{-1} \ln \mathbb{E}[2^{N_t}]$  and  $\Lambda = \lim_{t \rightarrow \infty} \Lambda(t)$ . Note that by taking the expectation, we assume the large  $M$  limit has been reached, with fixed lineage durations or division counts for the FTE and FDE, respectively. However, in order to compute  $\varepsilon(\hat{\Lambda})$  from Eq. (6) and its different contributions from Eq. (7) using simulations, we must replace all averages by their sample averages as detailed in Eqs. (26) and (30) over many realizations of the growth process.

The nonlinear averaging bias,  $\text{Bias}_{\text{nl}}(t, M)$ , is intimately connected to the concept of *quenched free energy* in disordered systems, a relationship that emerges when  $\langle \cdot \rangle_M$  is interpreted as a partition function. This connection will be clarified in Section 2.3, where we introduce a scaling of lineage durations with  $M$ , analogous to the approach used in [30,31] to estimate free energy differences. It is important to note that the dependence on  $M$  arises solely from  $\text{var}(\hat{\Lambda})$  and  $\text{Bias}_{\text{nl}}$ , and both go to zero in the limit that  $M$  goes to infinity. Hence, the only error at large  $M$  is the finite-time error:

$$\lim_{M \rightarrow \infty} \varepsilon^2(\hat{\Lambda}) = \text{Bias}_{\text{ft}}(t) \quad (9)$$

In the next section, we will present numerical and theoretical results and explore the differences in the finite-time bias between the FTE and FDE.

**2.1.1. Numerical results.** We now present numerical results that highlight the key features of the different terms in the error expression. We conduct simulations using a simple model where generation times follow a first-order autoregressive process (AR1) [18,19,36]:

$$\tau_{n+1} = \tau_0(1 - c) + \tau_n c + \xi_n, \quad (10)$$

where  $\tau_0$  is the average generation time,  $\tau_n$  is the parent generation time,  $c$  is the Pearson correlation coefficient between parent and offspring, and  $\xi_n$  represents noise with zero mean and a variance given by  $\sigma_\xi^2 = (1 - c^2)\sigma_\tau^2$ . This model incorporates correlations between generation times, which are essential for maintaining homeostasis, and captures key aspects of the convergence for the FTE and FDE estimators.

The long-time population growth rate for the model in Eq. (10) was first calculated in Ref. [19] to be

$$\Lambda = \frac{2 \ln(2)/\tau_0}{1 + \sqrt{1 - 2 \ln(2)(\sigma_\tau/\tau_0)^2}} \quad (11)$$

where

$$\sigma_T^2 = \sigma_\tau^2 \frac{(1+c)}{(1-c)}. \tag{12}$$

An alternative derivation of the population growth rate of this model using a *Second Cumulant Expansion* of FDE is discussed in [Methods 4.2](#).

In [Fig 2A](#), we show the total error for FTE and FDE from [Eq. \(6\)](#) from simulations over  $t/\tau_0 = 10^3$  generations. The parameters used were  $\tau_0 = 1$  and  $\sigma_\tau = 0.2$ , with 10, 20, and 40 lineages. Note that the error decreases over time for both methods, but the FDE has an order of magnitude smaller error for the first 100 generations. The methods have comparable error after the first 100 generations.

Next, we look at the different contributions to error. In [Fig 2B](#) and [Fig 2C](#), we show the four contributions to the total error based on [Eq. \(7\)](#) for FTE and FDE, respectively. At short times, the FTE estimator is dominated by the finite-time bias, while the FDE estimator is dominated by the variance of the estimator, though it retains a small, nonzero finite-time bias. For both methods, the variance of the estimators tends to zero as the number of generations tends to infinity, making the total error primarily driven by the nonlinear bias,  $\text{Bias}_{nl}(t, M)$ .

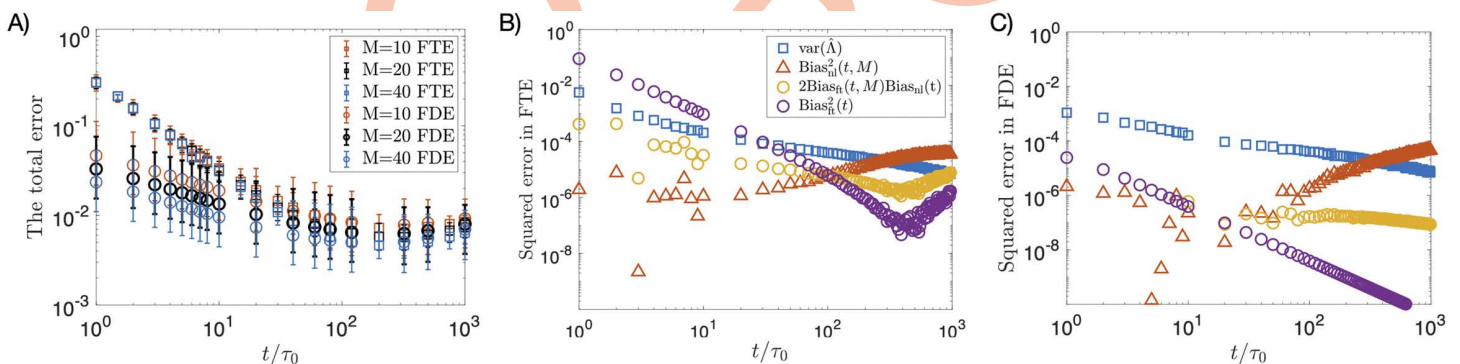
In the next section, we will discuss the monotonic convergence of  $\text{Bias}_{nl}(t)$  for both methods and why the FDE has a much smaller error than FTE. Then, in [Sec. 2.3](#) we will discuss the nonlinear averaging bias,  $\text{Bias}_{nl}(t, M)$ .

## 2.2. Finite duration bias ( $\text{Bias}_{ft}(t)$ )

In this section, we examine the behavior of the finite time and finite division number error in both the FTE and FDE. Specifically, we demonstrate that both ensembles exhibit inverse scaling with lineage duration given that the lineage number is large enough, while providing a justification for the significantly smaller prefactor observed in the FDE.

**2.2.1. Inverse time scaling.** In previous work [\[24\]](#), we demonstrated that for small noise in generation times,  $\hat{\Lambda}_1$  converges inversely with lineage length, as described by

$$\text{Bias}_{ft}(t, M) = \frac{A}{t} + o\left(\frac{1}{t}\right). \tag{13}$$



**Fig 2. Systematic error in growth rate estimates.** Simulations were done using the model in [Eq. \(10\)](#) with parameters given in [Sec. 2.1.1](#). The analytical solution of the long-time population growth rate of this model is shown in [Eq. 11](#). **(A)** The average total error for both estimators ([Eq. \(6\)](#)) for  $M=10$ ,  $M=20$ , and  $M=40$  lineages over  $10^3$  generations, calculated from  $10^3$  realizations, with error bars representing the standard deviation. **(B)** The absolute squared error ([Eqs. \(7\)](#) and [\(8\)](#)) and all contributions for FTE. **(C)** The absolute squared error ([Eqs. \(7\)](#) and [\(8\)](#)) and all contributions for FDE.

<https://doi.org/10.1371/journal.pcbi.1014088.g002>

Generally, deriving an exact expression for the finite-time bias is challenging, even when the large deviations are well understood. However, as we discuss below, in certain cases, this term can be approximately removed from the estimator using a finite-time scaling approach.

In this work, we demonstrate that a similar result can be obtained for the FDE estimator. We can compute the convergence of the FDE estimator exactly for all times for the model in [Eq. \(10\)](#) (see [Methods 4.1](#) and [S1 Text](#), Derivation of finite time growth rate for FDE [\[23\]](#) for derivations). The leading order correction to the FDE will be

$$\hat{\Lambda}_2 = \Lambda + \frac{B}{n} + o\left(\frac{1}{n}\right), \quad (14)$$

where  $B$  is the finite-time coefficient. In practice, the coefficients for both FDE and FTE are determined directly from the data by fitting the coefficients  $A$  and  $B$  at short times.

We demonstrate this in [Fig 3](#) for the model in [Eq. \(10\)](#) and show the fitting procedure and the finite-time coefficients as a function of the correlation strength,  $c$ . The finite-time coefficients are obtained by plotting the total error vs.  $\tau_0/t$  and extracting the slope at small times. In this format, fitting the linearly decreasing points from right to left gives the finite-time coefficients. As the number of generations gets too large, the nonlinear averaging bias dominates and the error becomes non-convex and begins to increase again. At the point that the error begins to increase, we stop the fitting. If there were an infinite amount of lineages, the estimators would linearly decrease until zero.

In the inset of [Fig 3](#), we show the value of the finite time coefficients over a range of  $c$  values. We find that the coefficient for the FDE is consistently about 10 times smaller than that of the FTE. In [SI Text](#), Transport equation derivation of FDE [\[23\]](#), we present a derivation using the van Foerster approach to explain this disparity. The key point is that the FTE exhibits a finite-time correction regardless of the initial conditions or the presence of correlations between generation times. In contrast, the FDE coefficient can vanish in the limit where such correlations are absent for certain initial conditions.

As shown in [Fig 3](#), there is a finite-time bias for both FTE and FDE, which decreases monotonically with time initially. At longer times, however, the total error begins to increase nonlinearly. This behavior arises because the nonlinear averaging bias starts to dominate the total error. The phenomenon of non-monotonic convergence for FTE was first identified in [Ref. \[24\]](#), where it was observed that, for large times and a fixed number of lineages, the estimator approaches the naive estimate on the right-hand side of [Eq. \(3\)](#). However, previous work did not quantitatively characterize the transition from monotonic to non-monotonic convergence in the total error for FTE, and this transition has not been observed before for FDE.

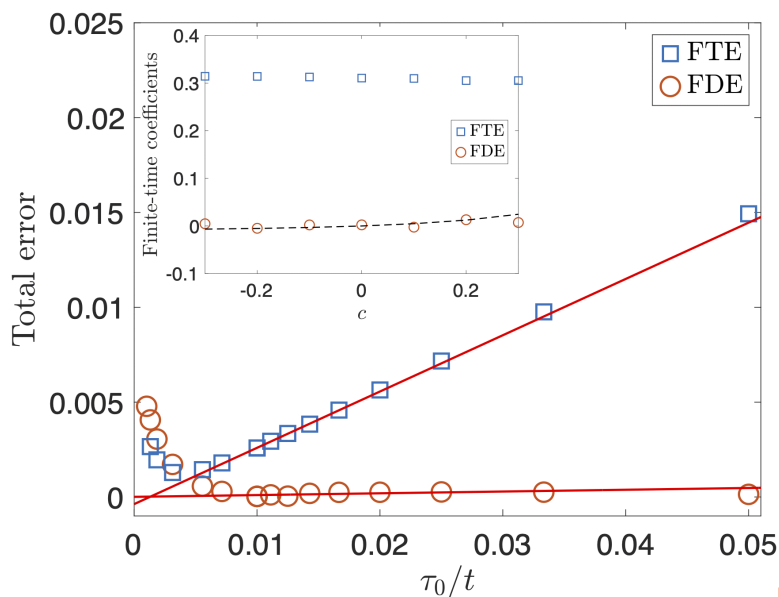
Next, we demonstrate that the crossover from monotonic to non-monotonic convergence in the total error represents a second-order phase transition. Furthermore, we quantitatively show how this systematic error can be avoided.

### 2.3. Nonlinear averaging bias ( $\text{Bias}_{n_t}$ ) and connection to the Random Energy Model

Here we discuss the bias resulting from the finite number  $M$  of lineages. Inspired by the approach in [Refs. \[30,31\]](#), we obtain an approximate expression for the bias in the FDE using a known formula for the free energy density of the REM.

As discussed in [Methods 4.1](#) and demonstrated by our numerical results in [Sec. 2.1.1](#), the empirical averages used to estimate the SCGF are influenced by a linearization effect, a phenomenon described in [Ref. \[32\]](#). This effect is analogous to the error seen in Jarzynski's Equality estimators of free energy differences [\[30\]](#), which can be understood through a connection to the Random Energy Model (REM). The REM is a mean-field model for disordered systems, which assumes that the energies of each state are sampled from an independent and identically distributed (iid) Gaussian variable [\[37\]](#). In this model, the partition function  $Z_N$  is simply an iid sum:

$$Z_N = \sum_{i=1}^{2^N} e^{-\beta \sqrt{N/2} X_i}, \quad (15)$$



**Fig 3. The finite-time coefficients for the model in Eq. 10.** The linear trend from right to left (red lines) is due to the finite-time error (see Eqs. (13) and (14)). The blue squares and red circles represent data from the FTE and FDE, respectively, obtained from 1,000 simulations using the same parameters as Fig 2. Error bars are omitted because they are significantly smaller than the symbols. For FDE, the x-axis represents the number of divisions,  $n = \tau_0/t$ . The inset shows the finite-time coefficients for a range of correlation coefficients. The black dotted line is the prediction of the FDE (see SI Text, Derivation of finite-time growth rate for FDE [23]). At larger times, the total error is no longer linearly decreasing in time so the fit is stopped when the error becomes nonconvex.

<https://doi.org/10.1371/journal.pcbi.1014088.g003>

where  $X_i$  are iid standard normal variables which would represent energy states in REM, but will represent individual lineages in our context. Additionally, within the original REM,  $N$  and  $\beta$  would be the system size and the inverse temperature. As we show below, in our case the  $N$  is related to the logarithm of the number of lineages and  $\beta$  is related to the ratio of the logarithm of the number of lineages to the number of divisions. Consequently, many properties of the REM can be derived using classical extreme value theory, without resorting to the more complex mathematical tools often required in the study of disordered systems (In many statistical mechanics papers,  $X_i$  are taken to have a variance of 1/2. Therefore, we have introduced a factor of 1/2 in the scaling of  $N$  to align our definition of the inverse temperature with the standard literature.). The free energy density of the model is given by

$$f_N(\beta) = -\frac{1}{\beta N} \ln Z_N, \tag{16}$$

and has an exact solution in the thermodynamic limit (see [Methods 4.3](#))

We begin by examining the connection of REM to the FDE, which offers two key advantages. First, (i) in this context, we can approximate  $T_n^{(j)}$  using a Gaussian distribution, whereas for the FTE, we must contend with the counting variables  $N_t^{(j)}$ . Second, (ii) the finite-time bias in the FDE is minimal and can be effectively eliminated in our simulations by setting the mother-daughter correlations to zero, thereby allowing us to isolate  $\text{Bias}_{n_i}$ .

The goal in this section is to understand when the system will not converge to the correct population growth rate due to the nonlinear averaging bias. We use the AR1 process defined in [Eq. \(10\)](#) for the simulations in this section. We express the exponent of [Eq. \(30\)](#) as

$$zT_n^{(i)} = z\tau_0 n + z\sigma\sqrt{n}X_i, \quad (17)$$

where  $X_i$  follows a standard normal distribution (It is important to note that this expression is valid only when the coefficient of variation,  $CV_T$ , is sufficiently small, ensuring that the likelihood of negative times remains negligible.). Next, we set  $M=2^N$  and fix

$$\alpha_{\text{FDE}} = \sqrt{\frac{2n}{N}}. \quad (18)$$

The free energy for the REM has an exact solution in the thermodynamic limit, but we cannot solve for  $\hat{\Lambda}$  analytically, since there is no closed-form expression for the finite-size free energy density. However, by approximating the finite-size free energy with its thermodynamic limit expression, the resulting equation can be solved to yield an estimate of the FDE estimator:

$$\hat{\Lambda}_{\text{REM}}(\alpha) = \begin{cases} \Lambda & \alpha < \alpha_c \\ \frac{2\ln(2)(1-\alpha^2/2)}{\tau_0(2\alpha\sigma/\tau_0\sqrt{\ln(2)-\alpha^2})} & \alpha \geq \alpha_c \end{cases} \quad (19)$$

with the critical value of  $\alpha$  given by

$$\alpha_c = \frac{2\sqrt{\ln(2)}}{\Lambda\sigma}. \quad (20)$$

Here  $\Lambda$  is given by [Eq. \(11\)](#). Note that a useful approximation to  $\alpha_c$  is

$$\alpha_c \approx 2/(\sqrt{\ln(2)}CV_T). \quad (21)$$

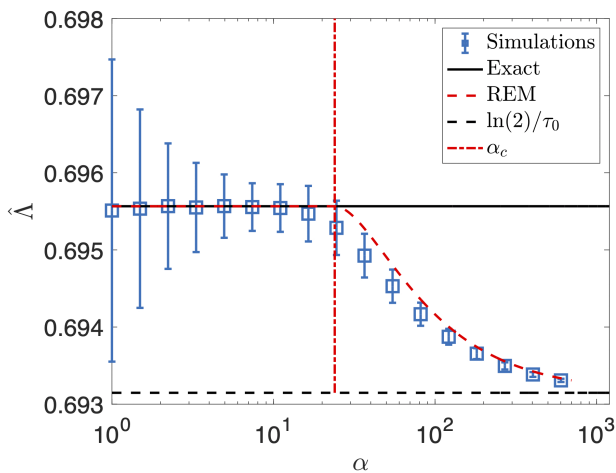
It can be checked that  $\hat{\Lambda}_{\text{REM}}(\alpha)$  is indeed continuous at  $\alpha_c$  and as  $\alpha \rightarrow \infty$  tends to  $\ln(2)/\tau_0$  since the transition to the frozen state is second order.

In [Fig 4](#), we show the phase transition as a function of  $\alpha_{\text{FDE}}$  for the FDE from simulations with  $c=0$ ,  $\tau = 1$ , and  $\sigma = 0.1$ . Small  $\alpha_{\text{FDE}}$  corresponds to the high temperature regime and large  $\alpha$  would be analogous to a small temperature regime which is predicted by [Eq. \(46\)](#). As the finite-time bias vanishes with  $n \rightarrow \infty$ , [Eq. \(46\)](#) provides insight into when  $\text{Bias}_{n_i}$  begins to increase, and thereby how small we need  $\alpha$  to be in order to obtain reliable growth rate estimates. Indeed, our numerical experiments demonstrate that the REM-based theory effectively captures the transition in  $\text{Bias}_{n_i}$ . In addition, the error bars decrease as  $\alpha$  increases because the variance between realizations—described by the first term on the right-hand side of [Eq. 7](#)—monotonically decreases over time. In [Methods 4.3](#), we show that, similar to the FDE, the FTE also quantitatively agrees with the REM framework. However, the decay after the phase transition differs slightly between the two ensembles.

As seen in [Fig 4](#), the analytical solution is slightly different on the right of the transition. Most likely, the deviations from the exact value seen in [Fig 4](#) are due to finite-size effects. For a given  $\alpha$ , both  $n$  and  $\ln(M)$  should go to infinity while leaving the ratio  $n/\ln M$  fixed. Deviations from REM were also seen in sampling Jarzynski's Equality [\[30\]](#) (see SI Text, Connection to other work).

#### 2.4. Application of estimators on real data

Now that we have an understanding of the convergence for both methods, we can apply this to experimental data. The ideal case is to have as long lineages as possible and avoid the linearization effect. Then, the estimators only have finite



**Fig 4. Estimates of  $\Lambda$  using the FDE compared to the theoretical prediction of the Random Energy Model (Eq. (45)).** The blue squares represent the mean of 1,900 realizations, with error bars indicating the standard deviation. These error bars monotonically decrease as  $\alpha$  increases, due to the monotonically decreasing variance term in the first term on the right-hand side of Eq. 7. Simulations are performed with independent Gaussian generation times with  $\sigma = 0.1$  and  $c=0$ . Different values of  $\alpha$  were realized by fixing  $M$  and modulating  $n$ . The black solid line is the exact population growth rate. The red dotted line is the prediction from Eq. 45, the dotted black line is  $\ln(2)/\tau_0$ , and the vertical red dashed-dotted line represents  $\alpha_c$  (Eq. (46)).

<https://doi.org/10.1371/journal.pcbi.1014088.g004>

time bias, which can be fitted and corrected. We can find how long lineages need to be to avoid the linearization effect for a given lineage size by using the equation for the  $\alpha_c$  in Eq. (46). For a given value for  $M$ , we can find the value of time durations before the nonlinear averaging bias takes over from the inequality:

$$n_c < \frac{\ln(M)\alpha_c^2}{2 \ln(2)}. \quad (22)$$

We would parse the data so that the number of divisions is below  $n_c$ , for a given lineage number.

We consider *E. coli* grown at 25 °C with data from Ref. [21]. The data contains 70 lineages with about 70 generations each. Hence, using Eq. 22 with  $M=70$  and  $\alpha_c \approx 7$ , we find that the linearization effect can be avoided if  $n < 150$ . To accurately extract the long-time population growth rate, we apply both methods to the lineage data and use the finite-time coefficients, as described in Sec. 2.2, to correct for Bias<sub>tt</sub> (see Methods 4).

In Fig 5, we show the performance of both FTE and the FDE on the data. Both methods asymptotically converge from below. The best fit lines for both methods are shown in Fig 7. When the two methods are fit and their finite-time error subtracted, the long-time growth rate estimates agree as shown in the red dotted (FTE) and blue (FDE) lines in Fig 5. The long-time population growth rate estimate for the FDE is approximately  $1.04 \times 10^{-2} \pm 1.97 \times 10^{-5} \text{ min}^{-1}$  for FDE and  $1.04 \times 10^{-2} \pm 3.11 \times 10^{-5} \text{ min}^{-1}$  for FTE, which are both distinct from the naive estimate,  $\ln(2)/\tau_0 \approx 1.025 \times 10^{-2} \text{ min}^{-1}$  (see Fig 7 for the fits). Since the two estimates agree, we have some confidence that the true population growth rate is near this value. Additionally, the finite time coefficient for FTE ( $A = -0.0060$ ) is about four times larger than FDE ( $B = -0.0016$ ), in agreement with our theory in Sec. 2.2.

Our analysis shows that the method enables inference of fitness differences as small as  $10^{-4}$ . This level of precision is biologically significant, as even small differences in growth rate can lead to substantial divergence in population sizes over time. To date, there has been no reliable way to infer such small growth rate differences from lineage data. Applying these methods naively leads to a finite-time bias that is on the order of  $10^{-2}$ . In contrast, the nonlinear averaging bias behaves more like an all-or-nothing effect: it can be entirely avoided if a sufficient number of lineages are included for a

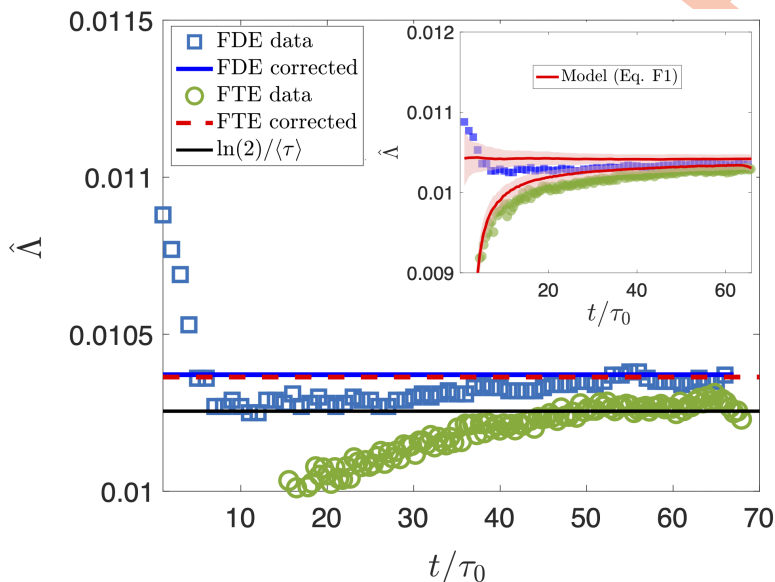
given observation time. One can reduce the finite-time error while avoiding the nonlinear averaging bias by exponentially increasing the number of lineages in the ensemble in tandem with the length of each lineage.

Most of our analysis has focused on understanding the convergence of model-free estimators of the growth rate. However, these estimators of the growth rate can also be used to reverse-engineer a model for the dynamics, albeit with a reconstruction that might not be unique. In the inset of Fig 5 we show that a two-component model for the generation time and log size dynamics that is fit to the correlations of the data quantitatively reproduces the convergence of both methods (see Methods 4.4 for model details).

### 3. Discussion

We have investigated the scaling behavior of errors in two estimators of population growth rate derived from lineage statistics. This problem is very similar to the estimation of free energy differences using Jarzynski's Equality [30,31] and to the methods for controlling buffer size in ATM networks [33] (see SI Text, Connection to other work [23]). These problems require sampling rare events, leading to a breakdown of traditional approaches to quantifying the sample distribution. Instead, it is essential to carefully account for extreme value statistics.

An effect of extremal statistics is the introduction of systematic error. Unlike the usual errors in statistical estimators, which stem from the lack of flexibility in the underlying model, this error arises because of poor sampling of the tails. In the context of growth rate estimations, there is an added complication which is the finite-time bias. We have demonstrated that there is, in fact, a trade-off between two types of bias: short-lineage lengths introduce a finite-time bias, while long-lineage lengths result in what we term a nonlinear averaging bias. This latter bias becomes significant when a few lineages dominate the sample averages  $\langle \cdot \rangle_M$ .



**Fig 5. Estimates of  $\Lambda$  obtained using FDE and FTE, including finite-size corrections, applied to data from Ref. [21].** The symbols represent the uncorrected FDE (blue squares) and FTE (green circles). The red dotted line and blue solid line correspond to the finite-time corrected population growth rates for FTE and FDE, respectively, which are both approximately  $0.0104 \text{ min}^{-1}$ . In the inset, we show that the cell-size control model described in Eqs. (54) and (55), which was fit to the data correlations, quantitatively reproduces the transient behavior observed in both methods. The symbols are the same data as FDE (blue squares) and FTE data (green circles). The solid red lines represent the mean of the estimators over 100 realizations, while the shaded regions denote their standard deviation.

<https://doi.org/10.1371/journal.pcbi.1014088.g005>

From the bias-variance decomposition of the total error, we find that at short times the finite length bias and the variance between realizations dominate the total error. At long times, there is a phase transition in which the linearization bias dominates. While the finite length bias can be mitigated through finite time-scaling, addressing the nonlinear averaging bias requires careful selection of lineage durations. By drawing a connection to the REM, we estimate the critical value of  $\alpha^2 = n \ln(2) / \ln M$  for FDE and  $\alpha^2 = t \ln(2) / \ln M$  for FTE at which the nonlinear averaging bias becomes dominant. This insight could be valuable in designing experiments that map single-cell data to population growth rates and has broader applications in contexts where one is interested in estimating large deviations of a counting process and its first passage time.

Another intriguing question is how accurately one can infer fitness using lineage-based methods. In principle, these methods can exactly recover the population growth rate from mother machine data. However, their practical accuracy is limited by noise in the data, which constrains the ability to reliably fit the finite-time coefficients. Moreover, perfectly fitting a line—even in principle—requires an infinite amount of data. Our analysis also depends on the location of the critical value  $\alpha_c$ , which is derived from a discrete-time Gaussian model. It is possible that non-Gaussian noise could lead to a different critical  $\alpha$ , or even to the absence of a well-defined threshold. Nevertheless, cell-size control models—such as those studied in this paper—have been shown to accurately capture replication dynamics, making this latter concern unlikely to pose a significant issue.

These results suggest a simple, practical protocol for applying lineage-based growth-rate estimators to finite datasets. To reliably infer population growth rates from lineage data, proceed as follows. *First*, determine the maximum usable lineage length before extremal statistics dominate by computing the critical duration  $n_c$  (or  $t_c$ ) from Eqs. (21) and (22), which depends on the number of independent lineages  $M$  and the variability of generation times. *Second*, partition the data so that lineage segments have lengths just below this critical value, ensuring that no small number of exceptionally fast-growing lineages dominates the estimator. *Third*, below the critical point, the finite time bias will dominate. Within this regime, compute the growth-rate estimator as a function of lineage length and verify that it increases approximately linearly when plotted versus the inverse duration (e.g.,  $1/n$  or  $1/t$ ). *Fourth*, fit this monotonic region to the expected inverse-length scaling: the slope gives the finite-duration bias coefficient (Eq. (13) or Eq. (14)), while the intercept yields the long-time population growth rate. When applied in this order, both fixed-time and fixed-division estimators converge to the same long-term growth rate, yielding robust and reproducible fitness estimates from finite single-cell datasets.

Although our focus is on inferring population growth rates from single-cell lineage data, it is important to note that bulk (population-level) growth-rate inference is itself nontrivial and subject to systematic uncertainties. In batch culture experiments, microbial populations typically exhibit a lag phase followed by exponential growth and eventual deviations from exponential behavior as resources become limiting and carrying capacity is approached [59,60]. Bulk growth rates are commonly inferred by fitting optical density, biomass, or colony-forming unit measurements to phenomenological growth models, such as exponential, logistic, or Gompertz forms, often using only a subset of the growth curve presumed to represent balanced growth [61]. However, the inferred growth rate depends sensitively on how the lag phase is treated, on deviations from pure exponential growth, and on the choice of fitting window and model [62]. Moreover, environmental shifts, diauxic transitions, and physiological heterogeneity can bias bulk estimates, complicating their interpretation as intrinsic fitness measures [63]. Acknowledging these limitations provides a more balanced comparison. Lineage-based approaches introduce systematic biases associated with rare-event sampling, but these biases can be explicitly identified and corrected within the inference framework developed here. In contrast, several dominant sources of error in bulk growth-curve analysis—such as lag-phase treatment and deviations from balanced exponential growth—do not admit a comparably systematic or model-independent correction.

In addition, methods exist to infer how different genotypes affect population growth. Single-variant fitness effects are often estimated in bulk by competing many genotypes in pooled batch culture and tracking their relative abundances over time, frequently using DNA barcodes and deep sequencing so that each genotype's log-frequency trajectory is

approximately linear in time, with slope set by its Malthusian growth-rate difference relative to a reference [64–66]. While powerful and scalable, these assays inherit the same growth-regime ambiguities as standard bulk curves (lag, non-exponential phases, density dependence) and add assay-specific artifacts from serial-dilution bottlenecks, sampling noise, and batch effects across replicates. In barcoded designs, a further dominant failure mode is systematic barcode processing bias: barcode sequence and sample-specific PCR/sequencing conditions can distort read-count trajectories and therefore misestimate slopes, even when underlying population dynamics are otherwise well behaved [66]. Recent correction methods (e.g., REBAR) can infer and remove substantial barcode-induced bias post hoc, but they do not eliminate the more fundamental ambiguities tied to growth-phase selection and history dependence in bulk cultures. In this sense, lineage-based inference offers a complementary route to variant fitness that connects directly to single-cell statistics, while making a different—and in our setting, explicitly correctable—set of finite-data assumptions.

Regarding the connection to of lineage-based methods to REM, it is intriguing to explore more rigorously whether the limiting behavior of  $\hat{\Lambda}$  can be understood using similar techniques. It is well-known (see, e.g., [38,39]) that the REM exhibits two phase transitions at  $\beta_2 = \sqrt{\ln 2}$  and  $\beta_1 = 2\sqrt{\ln 2}$ . For  $\beta > \beta_2$ , the partition function  $Z_N$  fails to obey the Law of Large Numbers, meaning that  $Z_N/E[Z_N]$  no longer converges to one in probability. For  $\beta > \beta_1$ , the Central Limit Theorem no longer holds, implying that  $(Z_N - E[Z_N])/var(Z_N)$  does not exhibit Gaussian fluctuations. A deeper understanding of these phase transition behaviors in  $\hat{\Lambda}$  could enable more accurate inference in the future.

Lastly, understanding the biases and their trade-offs is important for applications involving Jarzynski's Equality and ATM networks (see SI Text, SI Text, Connection to other work [23]). More broadly, such trade-offs arise in thermodynamic inference problems, including the estimation of entropy and entropy production rate (EPR) [40–43]. One widely used approach is the *plug-in* method, where empirical probabilities estimated from time-series data are directly substituted into information-theoretic expressions—in this case, the Kullback–Leibler divergence between forward and time-reversed trajectories. In Ref. [42], this method was shown to suffer from the same systematic errors highlighted here: finite-time biases arising from splitting trajectories into blocks of limited duration, and finite-sample biases due to limited statistics. However, unlike in the present work, no phase transition was observed. Also, there are some EPR methods that involve waiting time distributions [48,49], which have similarities to generation times. A precise understanding of biases inherent to finite data will allow for accurate inference of fundamental properties.

In addition, similar biases arise in importance sampling methods used to compute large deviation functions, such as cloning [44–47]. In this method, many independent copies (clones) of the system are simulated in parallel. By the law of large numbers, some clones naturally exhibit the desired rare behavior. The algorithm then amplifies these trajectories by duplicating the clones that realize the rare dynamics, while deleting those that do not. The cloning algorithm has a finite time bias as well as a finite clone bias, which is analogous to the finite lineage bias in our work.

## 4. Methods

### 4.1. Relation of growth rate estimators to large deviation theory

We consider the general setting in which single-cell generation times evolve stochastically according to some process  $\{\tau_k\}$ . This need not be a Markov process, but to be concrete we imagine there is some underlying phenotype (e.g., gene expression)  $x_k \in \mathbb{R}^d$  which evolves according to a Markov process with transition operator  $h(x_{k+1}|x_k)$ , and that generation times are deterministic functions of the phenotype  $\tau(x)$ . This modeling framework can capture all existing models of single-cell dynamics [50–52]. We let  $T_n = \sum_{k=1}^n \tau_k$  denote the time at which the  $n$ th cell in a lineage divides.

The long-term growth rate is related to the lineage-to-lineage fluctuations in the counting process,

$$N_t = \max \left\{ n : \sum_{i=1}^n \tau_i < t \right\}, \quad (23)$$

and  $\Lambda$  is given by [24]

$$\Lambda = \psi(\ln(2)) \quad (24)$$

where  $\psi(z)$  is the scaled-cumulant generating function,

$$\psi(z) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \mathbb{E}[e^{N_t z}]. \quad (25)$$

Here,  $z$  is the conjugate variable to  $N_t$ , and  $\mathbb{E}[\cdot]$  represents the expected value with respect to the lineage distribution within the *fixed-time ensemble (FTE)*, as illustrated in Fig 1B and C. The intuition behind Eq. (24) is that lineages with  $n$  divisions on average contribute  $2^{N_t}$  cells to the final population, hence the total population size is on average  $\mathbb{E}[2^{N_t}] \sim e^{\Lambda t}$ .

Given lineage samples  $\{N_t^{(j)}\}$  which come from repeated experiments or splitting a long lineage into blocks,  $\psi(z)$  and hence  $\Lambda$  can in principle be estimated by replacing the expectation with an empirical average:

$$\hat{\psi}_t(z) = \frac{1}{t} \ln \left\langle e^{N_t^{(j)} z} \right\rangle_M \quad (26)$$

where  $\langle \cdot \rangle_M$  is the empirical average over  $M$  samples. An estimator of  $\Lambda$  is then  $\hat{\Lambda}_1(t, M) = \hat{\psi}_t(\ln(2))$  which is equivalent to Eq. (4), and is always a biased estimator.

These formulas naturally connect to the large deviations of  $N_t$  through the Gärtner-Ellis Theorem, which states that

$$-\lim_{t \rightarrow \infty} \frac{1}{t} \ln P(N_t/t = x) = I(x), \quad (27)$$

where  $I(x)$  is the *large deviation rate function*, defined as the Legendre-Fenchel transform  $\psi^*(x)$  of  $\psi(x)$ . Informally, Eq. 27 tells us that the fluctuations in  $N_t/t$  decay exponentially with time:  $P(N_t/t = x) \approx A_t e^{-tI(x)}$ , where  $A_t$  is the normalization constant.

Equation 27 states a large deviation principle, which we assume in order to apply large deviation theory to infer the population growth rate. Specifically, we assume that for the FDE the sequence of division times  $\tau_k$  (or, equivalently for the FTE, the cumulative number of divisions up to time  $t$ ) obeys a large deviation principle. This condition is typically met when the  $\tau_k$  (or division counts) are only finitely correlated, ensuring the law of large numbers applies, and when the observation time  $T_n = \sum_{k=1}^n \tau_k$  (or equivalently the number of divisions  $n$ ) is much longer than the correlation time, i.e.,  $n \gg 1$ .

In Ref. [24], it was shown that due to the large deviation structure, the estimator  $\hat{\Lambda}_1$  exhibits a somewhat surprising non-monotonic convergence. This phenomenon is related to the so-called linearization effect [32], which can be understood as follows. The integral  $\mathbb{E}[e^{xz}] \approx \int p_t(x) e^{xz} dx$  is dominated by a value  $x^* = N_t^*/t$  for which  $\psi(z) = x^* z - I(x^*)$  is extremal. Therefore, to obtain an accurate estimate of  $\mathbb{E}[e^{N_t z}]$  from finite samples, we must have a high probability of sampling  $N_t = N_t^*$ . However, when  $t$  is large, this is an exponentially rare event, requiring an exponentially large number of samples.

As shown in Ref. [25], the growth rate can alternatively be expressed in terms of the scaled cumulant generating function (SCGF) of  $T_n$ ,  $\phi(z) = \lim_{n \rightarrow \infty} n^{-1} \ln \mathbb{E}[e^{T_n z}]$ , as

$$-\phi(-\Lambda) = \ln(2). \quad (28)$$

This result arises from the fact that  $\phi(z) = \psi^{-1}(z)$  for the SCGFs of a counting process and its dual first-passage time process [53], which we refer to as the *fixed-divisions ensemble (FDE)* (Note that in Ref. [25], this estimator was called the

Generalized Euler-Lotka (GEL) Equation.). A comparison of these ensembles is shown in Fig 1B and C. Additionally, the large deviation rate function for  $T_n/n$ , denoted by  $J$ , can be related to  $I$  through the expression  $J(y) = xI(1/x)$  [53].

Following Ref. [25] and Eq. (5), an estimator  $\hat{\Lambda}_2$  can be obtained as the positive solution to the nonlinear equation

$$-\hat{\phi}_n(-\hat{\Lambda}_2(n, M)) = \ln(2), \tag{29}$$

where  $\hat{\phi}_n(z)$  is the empirical SCGF for the dual process:

$$\hat{\phi}_n(z) = \frac{1}{n} \ln \left\langle e^{T_n^\theta z} \right\rangle_M. \tag{30}$$

Since  $\hat{\phi}_n(z)$  is convex,  $\hat{\Lambda}_2(n, M)$  is uniquely defined by Eq. (29).

Our primary goal is to analyze the convergence behavior of the FTE and FDE ensembles, denoted by  $\hat{\Lambda}_1$  and  $\hat{\Lambda}_2$ , respectively, as a function of the parameters  $M$  and  $t$  (for FTE) and  $M$  and  $n$  (for FDE). Although the dual estimator  $\hat{\Lambda}_2$  also appears to have systematic errors from the linearization effect, it is unclear whether one estimator consistently outperforms the other or how their convergence patterns are influenced by specific model details.

In this work, we demonstrate that a similar result can be obtained for the FDE estimator as follows: We can express the average on the left-hand side of Eq. (29) in terms of the large deviation function (see SI 4.1),

$$\left\langle e^{-T_n^\theta \hat{\Lambda}_2} \right\rangle = \frac{K_n}{n} \int p(t) e^{-nt\hat{\Lambda}_2} dt, \tag{31}$$

where  $t = T/n$  and  $K_n$  is a normalization constant.

#### 4.2 Derivation of the population growth rate of the simulation model SCE

If the large deviation rate function is quadratic (e.g.,  $T_n$  is Gaussian), then we can do a cumulant expansion of Eq. (30) and truncate at second order. Starting with the series expansion

$$\ln \mathbb{E}[e^{T_n z}] = \sum_{k=1}^{\infty} \kappa_k \frac{z^k}{k!}, \tag{32}$$

and dropping all but the first two terms, and substituting into Eq. (28) yields

$$\tau_0 \Lambda - \frac{\sigma_T^2}{2} \Lambda^2 = \ln(2) \tag{33}$$

where

$$\tau_0 = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[T_n]}{n} \tag{34}$$

$$\sigma_T^2 = \lim_{n \rightarrow \infty} \frac{\text{var}(T_n)}{n}. \tag{35}$$

Solving the quadratic equation for  $\Lambda$  yields Eq. (11). Since this formula is exact when the large deviation rate function is quadratic, and thus the statistics of  $T_n$  are Gaussian, we call it the *Second Cumulant Expansion (SCE)*. However, for real data, the accuracy of this approximation is not known a priori.

To illustrate the limitations of the SCE, we present a counter example where the method does not work. We examine a model in which generations are independent and  $\tau = t_0$  with probability 1/2, and  $\tau = 1$  otherwise. Because this is an independent generation time model, the tree and lineage distributions are identical, resulting in the same equation from both the Euler-Lotka and FDE:

$$e^{t_0\Lambda} = e^{(t_0-1)\Lambda} + 1 \tag{36}$$

which can be numerically solved for  $\Lambda$ .

Now, we can determine the growth rate for the SCE by substituting

$$\tau_0 = \frac{t_0 + 1}{2} \tag{37}$$

and

$$\sigma^2 = \frac{(t_0 - 1)^2}{4} \tag{38}$$

into [Eq. \(11\)](#), yielding

$$\Lambda = \frac{\frac{4 \ln(2)}{t_0 + 1}}{1 + \sqrt{1 - 2 \ln(2) \left(\frac{t_0 - 1}{t_0 + 1}\right)^2}}. \tag{39}$$

Note that the SCE simplifies to the naïve estimate  $\ln(2)/\tau_0$  when  $t_0 = 1$ , reflecting the zero-noise limit.

The predictions of [Eqs. \(36\)](#) and [\(39\)](#) are compared in [Fig 6](#), showing that the second cumulant method is inaccurate. Indeed, even for the case of independent generation times, the (exact) Euler-Lotka equation tells us that knowledge of the entire generation time distribution is needed; hence, the mean and variance are insufficient - therefore, the SCE cannot be guaranteed to yield accurate results for non-Gaussian generation time distributions.

### 4.3. The mapping of FDE and FTE to REM

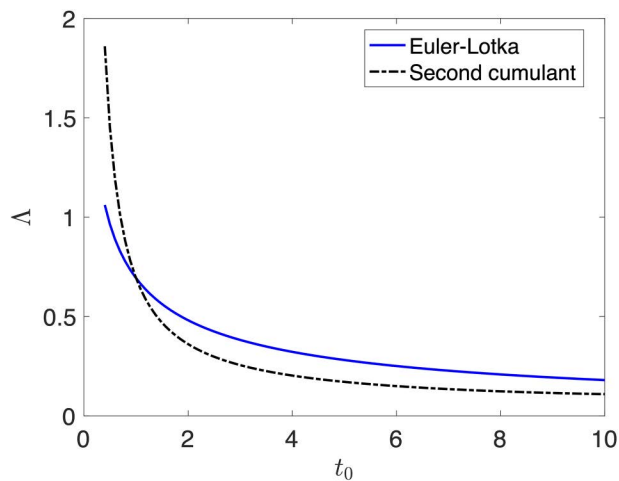
We will first show the connection of the FDE with REM. We focus on the asymptotic behavior of the free energy for REM density, defined as

$$f_N(\beta) = -\frac{1}{\beta N} \ln Z_N, \tag{40}$$

with the limit of the quenched average given by  $\bar{f}(\beta) = \lim_{N \rightarrow \infty} \mathbb{E}[f_N(\beta)]$ . In the thermodynamic limit, we find the free energy to be [\[37\]](#):

$$\bar{f}(\beta) = \begin{cases} -\frac{\beta}{4} - \frac{\ln 2}{\beta}, & \text{if } \beta \leq \beta_c, \\ -\sqrt{\ln 2}, & \text{if } \beta > \beta_c. \end{cases} \tag{41}$$

where  $\beta_c = 2\sqrt{\ln 2}$  is the critical inverse temperature. The free energy in the high temperature regime  $\beta < \beta_c$  is entropically dominated and can be obtained by replacing  $Z_N$  with  $\mathbb{E}[Z_N]$  in [Eq. \(40\)](#), avoiding the need to evaluate the quenched average. This is due to a concentration of the Gibbs measure, which is well established for the REM. In contrast, for the low



**Fig 6. The figure shows an illustrative example where the SCE approximation fails.** In this example, generation times are uncorrelated, and chosen to be  $\tau = t_0$  with probability 1/2 and  $\tau = 1$  otherwise. The (exact) result of the Euler-Lotka equation is given by the solid blue line (Eq. (36)), and is compared with the SCE method (black, dashed line), which relies only on the first and second cumulants of the distribution (Eq. (39)).

<https://doi.org/10.1371/journal.pcbi.1014088.g006>

temperature regime  $\beta \geq \beta_c$  the free energy density is determined by the extremal energy levels. It can be shown that the transition corresponds to the breakdown of the Law of Large Numbers for  $Z_N$  [39]. Within this phase, the partition function is dominated by a few energy levels. As we show below, within our context the growth rate estimators are dominated by a few extremal lineages past the phase transition.

We begin by examining the connection of REM to the FDE, which offers two key advantages. First, (i) in this context, we can approximate  $T_n^{(l)}$  using a Gaussian distribution, whereas for the FTE, we must contend with the counting variables  $N_t^{(l)}$ . Second, (ii) the finite-time bias in the FDE is minimal and can be effectively eliminated in our simulations by setting the mother-daughter correlations to zero, thereby allowing us to isolate  $\text{Bias}_{n_l}$ .

The goal in this section is to understand when the system will not converge to the correct population growth rate due to the nonlinear averaging bias. We use the AR1 process defined in Eq. (10) for the simulations in this section. We express the exponent of Eq. (30) as

$$zT_n^{(l)} = z\tau_0 n + z\sigma\sqrt{n}X_i, \quad (42)$$

where  $X_i$  follows a standard normal distribution (It is important to note that this expression is valid only when the coefficient of variation,  $CV_\tau$ , is sufficiently small, ensuring that the likelihood of negative times remains negligible.) Next, we set  $M = 2^N$  and fix

$$\alpha_{\text{FDE}} = \sqrt{\frac{2n}{N}}. \quad (43)$$

Equation (30) can then be rewritten in terms of the (finite-size) free energy density of the REM by introducing the temperature parameter  $\beta_{\text{FDE}}(z) = z\sigma\alpha_{\text{FDE}}$ :

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\phi}_n(-z)] = \frac{-\beta_{\text{FDE}}(z)}{\alpha_{\text{FDE}}^2} f_N(\beta_{\text{FDE}}(z)) - \frac{\ln(2)}{\alpha_{\text{FDE}}^2} - z\tau_0. \quad (44)$$

Equation (44) is an exact equation but we can't solve for  $\hat{\Lambda}$  analytically, since we don't have a closed formula for the finite-size free energy density  $f_N(\beta(\mathbf{z}))$ . However, we can study the large  $N$  behavior by replacing  $f_N(\beta(\mathbf{z}))$  with the formula for  $\bar{f}(\beta(\mathbf{z}))$  given in Eq. (41). The resulting equation can be solved to yield an estimate of the FDE estimator  $E[\hat{\Lambda}] \approx \hat{\Lambda}_{\text{REM}}(\alpha_{\text{FDE}})$  which has the explicit formula

$$\hat{\Lambda}_{\text{REM}}(\alpha) = \begin{cases} \Lambda & \alpha < \alpha_c \\ \frac{2 \ln(2)(1-\alpha^2/2)}{\tau_0(2\alpha\sigma/\tau_0\sqrt{\ln(2)-\alpha^2})} & \alpha \geq \alpha_c \end{cases} \quad (45)$$

with the critical value of  $\alpha$  given by

$$\alpha_c = \frac{2\sqrt{\ln(2)}}{\Lambda\sigma}. \quad (46)$$

Here  $\Lambda$  is given by Eq. (11). Note that a useful approximation to  $\alpha_c$  is

$$\alpha_c \approx 2/(\sqrt{\ln(2)}CV_T). \quad (47)$$

It can be checked that  $\hat{\Lambda}_{\text{REM}}(\alpha)$  is indeed continuous at  $\alpha_c$  and as  $\alpha \rightarrow \infty$  tends to  $\ln(2)/\tau_0$  since the transition to the frozen state is second order.

However, it is important to note that our definition of  $\text{Bias}_{\text{nl}}$  is based on  $E[\hat{\Lambda}]$ , where the expectation is taken after solving Eq. (30). In contrast, to derive Eq. (45), we employed Eq. (41), where the expectation is taken before solving Eq. (44). Namely, the ordering of limits is different which could lead to systematic differences. Most likely, the deviations from the exact value seen in Fig 4 are due to finite-size effects. For a given  $\alpha$ , both  $n$  and  $\ln(M)$  should go to infinity while leaving the ratio  $n/\ln M$  fixed. Deviations from REM were also seen in sampling Jarzynski's Equality [30].

In the FTE, we can similarly make the connection to the REM by viewing  $\ln(2)N_t^{(j)}$  as the scaled energy levels. Since  $N_t$  is a counting variable, Eq. (41) is no longer valid. The REM with discrete energy levels has been studied in Refs. [54–56], where both binomial and Poisson energy distributions have been studied. We found that a Gaussian approximation nevertheless seems to capture the convergence very well in the regimes we are interested in.

In our earlier work [24], we derived the rate function for the autoregressive generation time model:

$$I(y) = \frac{(1-c)^2}{2\sigma_T^2} (\tau_0 - 1/y)^2. \quad (48)$$

A Taylor expansion around  $y = 1/\tau_0$  yields a Gaussian approximation:

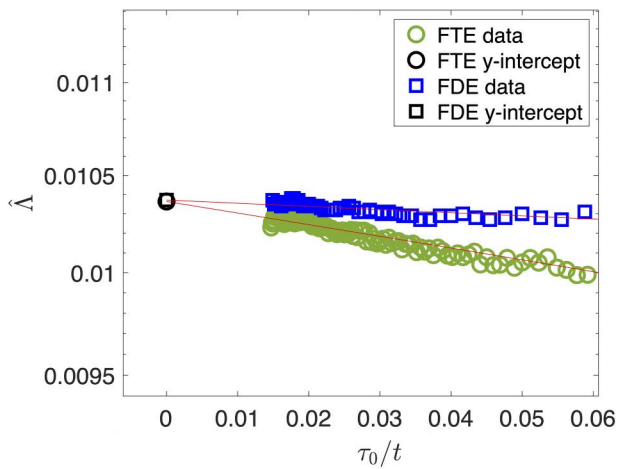
$$\ln(2)N_t \approx t \ln(2)/\tau_0 + \ln(2)\sqrt{t}\sigma_y X_t. \quad (49)$$

This motivates the following definitions:

$$\alpha_{\text{FTE}} = \sqrt{\frac{2t}{N}}, \quad (50)$$

and

$$\beta_{\text{FTE}} = \ln(2)\sigma_y\alpha_{\text{FTE}}. \quad (51)$$



**Fig 7. Estimates of  $\Lambda$  using the FDE and FTE with finite size corrections.** By plotting as a function of  $\tau_0/t$ , we can fit the convergence to a line. The symbols represent the FDE (blue squares) and FTE (green circles) applied with no corrections. The solid red solid lines represent the fit and the black symbols are y-intercepts which give the long-time population growth rate.

<https://doi.org/10.1371/journal.pcbi.1014088.g007>

where  $\sigma_y = \sigma_\tau^2/(1-c)^2$ . Similar to the approach taken for the FDE, we can rewrite Eq. (26) in terms of  $f_N(\beta)$ :

$$\tilde{\Lambda}_{\text{REM}}(\alpha) = \frac{\ln(2)}{\tau_0} + \frac{1}{\alpha^2} (-\beta f_N(\beta) - \ln(2)). \quad (52)$$

Once we replace  $f_N(\beta)$  with  $\bar{f}_N(\beta)$  we obtain

$$\tilde{\Lambda}_{\text{REM}}(\alpha) = \begin{cases} \tilde{\Lambda} & \alpha < \alpha_c \\ \frac{\ln(2)}{\tau_0} + \frac{1}{\alpha^2} (\sqrt{\ln(2)} - \ln(2)) & \alpha \geq \alpha_c \end{cases} \quad (53)$$

where  $\tilde{\Lambda} = \ln(2)/\tau_0 + \ln(2)^2\sigma_y^2/4$ , which is the Taylor expansion of  $\Lambda$  given by Eq. 11 in  $\sigma_y^2$ . Note that the transition occurs at the same critical  $\alpha$  for both estimators but the decay to the naive solution of  $\ln(2)/\tau_0$  is different.

In Fig 2 and in Ref. [24], we used  $\sigma = 0.2$ . For this value of the noise,  $\alpha_c \approx 12$ , and  $n < 300$  is needed to avoid the linearization effect. This analysis qualitatively agrees with a more heuristic approach in Ref. [24] where we obtained the criteria  $n < \frac{\ln(M)}{\sigma^2 \ln(2)} \approx 100$  divisions (Fig 7).

#### 4.4. Fits of the experimental data

Simulated data was generated by fitting autoregressive models to the experimental *E. coli* data. First, we fit the autoregressive model given by Eq. (10). This was achieved by performing a simple linear regression. Next, in order to account for the fact that cell size is regulated, we fit a multivariate autoregressive model, where log cell length was included as a predictor. This takes the form

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{b} + \xi_i \quad (54)$$

Here,  $\mathbf{A} \in \mathbb{R}^{2 \times 2}$  and  $\mathbf{b} \in \mathbb{R}^2$  are coefficients to be fitted and  $\xi_i \in \mathbb{R}^2$  is a noise vector. The regression variable is

$$\mathbf{x}_i = \begin{bmatrix} \ln s_i \\ \tau_i \end{bmatrix} \quad (55)$$

and  $s_i$  is the size of the  $i$ th cell at birth.

We used standard least squares for regression with multiple response variables [27] to fit the coefficients  $\mathbf{A}$  and  $\mathbf{b}$ , as well as the noise magnitudes.

Note that by simply including the additional variable of log cell-size, cell-size is automatically regulated. We could have alternatively included growth rates, instead of generation time and/or log fold change in size as predictors. This was the approach taken in [57]. However, we found that for the purpose of predicting growth rate and the convergence pattern of the FTE and FDE, simply adding the additional predictor of size was sufficient.

## Supporting information

### S1 Text. Supplementary methods and derivations for “Extremal events dictate population growth rate inference.

” Provides detailed model definitions and analytical results: (i) Cell-size control model used in simulations; (ii) Transport (von Foerster) equation derivation of the FDE estimator and discussion of finite-time bias; (iii) Connections to related work, including Jarzynski’s Equality–based estimators and admission control in ATM networks; (iv) Derivation of the finite-time growth rate for FDE under Gaussian/AR(1) lineage statistics. Includes full equations, assumptions, and references. (PDF)

## Author contributions

**Conceptualization:** Ethan Levien, Ariel Amir.

**Data curation:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Formal analysis:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Funding acquisition:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Investigation:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Methodology:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Project administration:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Resources:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Software:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Supervision:** Ariel Amir.

**Validation:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Visualization:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Writing – original draft:** Trevor GrandPre, Ethan Levien, Ariel Amir.

**Writing – review & editing:** Trevor GrandPre, Ethan Levien, Ariel Amir.

## References

1. Orr HA. Fitness and its role in evolutionary genetics. *Nat Rev Genet.* 2009;10(8):531–9. <https://doi.org/10.1038/nrg2603> PMID: [19546856](https://pubmed.ncbi.nlm.nih.gov/19546856/)
2. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet.* 2014;15(7):480–90. <https://doi.org/10.1038/nrg3744> PMID: [24913663](https://pubmed.ncbi.nlm.nih.gov/24913663/)
3. Cadart C, Venkova L, Recho P, Lagomarsino MC, Piel M. The physics of cell-size regulation across timescales. *Nat Phys.* 2019;15(10):993–1004.

4. Yamauchi S, Nozoe T, Okura R, Kussell E, Wakamoto Y. A unified framework for measuring selection on cellular lineages and traits. *Elife*. 2022;11:e72299. <https://doi.org/10.7554/eLife.72299> PMID: 36472074
5. Nozoe T, Kussell E, Wakamoto Y. Inferring fitness landscapes and selection on phenotypic states from single-cell genealogical data. *PLoS Genet*. 2017;13(3):e1006653. <https://doi.org/10.1371/journal.pgen.1006653> PMID: 28267748
6. Jafarpour F, Levien E, Amir A. Evolutionary dynamics in non-Markovian models of microbial populations. *Phys Rev E*. 2023;108(3–1):034402. <https://doi.org/10.1103/PhysRevE.108.034402> PMID: 37849168
7. Lin J, Manhart M, Amir A. Evolution of Microbial Growth Traits Under Serial Dilution. *Genetics*. 2020;215(3):767–77. <https://doi.org/10.1534/genetics.120.303149> PMID: 32366512
8. Lambert G, Kussell E. Quantifying selective pressures driving bacterial evolution using lineage analysis. *Phys Rev X*. 2015;5(1):011016. <https://doi.org/10.1103/PhysRevX.5.011016> PMID: 26213639
9. Fink JW, Held NA, Manhart M. Microbial population dynamics decouple growth response from environmental nutrient concentration. *Proc Natl Acad Sci U S A*. 2023;120(2):e2207295120. <https://doi.org/10.1073/pnas.2207295120> PMID: 36598949
10. Genthon A, Lacoste D. Fluctuation relations and fitness landscapes of growing cell populations. *Sci Rep*. 2020;10(1):11889. <https://doi.org/10.1038/s41598-020-68444-x> PMID: 32681104
11. Kinsler G, Geiler-Samerotte K, Petrov DA. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *Elife*. 2020;9:e61271. <https://doi.org/10.7554/eLife.61271> PMID: 33263280
12. Bacaër N. A short history of mathematical population dynamics. vol. 618. Springer; 2011.
13. Powell EO. Growth rate and generation time of bacteria, with special reference to continuous culture. *J Gen Microbiol*. 1956;15(3):492–511. <https://doi.org/10.1099/00221287-15-3-492> PMID: 13385433
14. Lotka AJ. Relation Between Birth Rates And Death Rates. *Science*. 1907;26(653):21–2. <https://doi.org/10.1126/science.26.653.21-a> PMID: 17754777
15. Smith DP, Keyfitz N. Estimates of fertility and mortality based on reported age distributions and reported child survival. *Math Demography*. 1977;:301–5.
16. Levien E, Kondev J, Amir A. The interplay of phenotypic variability and fitness in finite microbial populations. *J R Soc Interface*. 2020;17(166):20190827. <https://doi.org/10.1098/rsif.2019.0827> PMID: 32396808
17. Lebowitz JL, Rubinow S. A theory for the age and generation time distribution of a microbial population. *J Math Biol*. 1974;1(1):17–36.
18. Lin J, Amir A. The Effects of Stochasticity at the Single-Cell Level and Cell Size Control on the Population Growth. *Cell Syst*. 2017;5(4):358–367.e4. <https://doi.org/10.1016/j.cels.2017.08.015> PMID: 28988800
19. Lin J, Amir A. From single-cell variability to population growth. *Phys Rev E*. 2020;101(1–1):012401. <https://doi.org/10.1103/PhysRevE.101.012401> PMID: 32069565
20. Thiermann R, Sandler M, Ahir G, Sauls JT, Schroeder J, Brown S, et al. Tools and methods for high-throughput single-cell imaging with the mother machine. *Elife*. 2024;12:RP88463. <https://doi.org/10.7554/eLife.88463> PMID: 38634855
21. Tanouchi Y, Pai A, Park H, Huang S, Buchler NE, You L. Long-term growth data of *Escherichia coli* at a single-cell level. *Sci Data*. 2017;4:170036. <https://doi.org/10.1038/sdata.2017.36> PMID: 28350394
22. Camsund D, Lawson MJ, Larsson J, Jones D, Zikrin S, Fange D, et al. Time-resolved imaging-based CRISPRi screening. *Nat Methods*. 2020;17(1):86–92. <https://doi.org/10.1038/s41592-019-0629-y> PMID: 31740817
23. GrandPre T, Levien E, Amir A. Supporting Information for Extremal Events Dictate Population Growth Rate Inference. *PLOS Computational Biology*. 2025.
24. Levien E, GrandPre T, Amir A. Large Deviation Principle Linking Lineage Statistics to Fitness in Microbial Populations. *Phys Rev Lett*. 2020;125(4):048102. <https://doi.org/10.1103/PhysRevLett.125.048102> PMID: 32794821
25. Pigolotti S. Generalized Euler-Lotka equation for correlated cell divisions. *Phys Rev E*. 2021;103(6):L060402.
26. Rochman ND, Popescu DM, Sun SX. Ergodicity, hidden bias and the growth rate gain. *Phys Biol*. 2018;15(3):036006. <https://doi.org/10.1088/1478-3975/aab0e6> PMID: 29461250
27. Demidenko E. *Advanced statistics with applications in R*. John Wiley & Sons. 2019.
28. Breiman L. *Bias, variance, and arcing classifiers*. Statistics Department, University of California, Berkeley. 1996.
29. Goldenfeld N. *Lectures on phase transitions and the renormalization group*. CRC Press. 2018.
30. Suárez A, Silbey R, Oppenheim I. Phase transition in the Jarzynski estimator of free energy differences. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2012;85(5 Pt 1):051108. <https://doi.org/10.1103/PhysRevE.85.051108> PMID: 23004704
31. Palassini M, Ritort F. Improving free-energy estimates from unidirectional work measurements: theory and experiment. *Phys Rev Lett*. 2011;107(6):060601. <https://doi.org/10.1103/PhysRevLett.107.060601> PMID: 21902307
32. Rohwer CM, Angeletti F, Touchette H. Convergence of large-deviation estimators. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2015;92(5):052104. <https://doi.org/10.1103/PhysRevE.92.052104> PMID: 26651644

33. Lewis JT, Russell R, Toomey F, McGurk B, Crosby S, Leslie I. Practical connection admission control for ATM networks based on on-line measurements. *Comput Commun.* 1998;21(17):1585–96.
34. Duffy KR, Williamson BD. Estimating large deviation rate functions. In: 2015. <https://arxiv.org/abs/1511.02295>
35. Patch B, Zwart B. Ranking transmission lines by overload probability using the empirical rate function. In: 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2020. 1–6. <https://doi.org/10.1109/pmaps47429.2020.9183567>
36. Cerulus B, New AM, Pougach K, Verstrepen KJ. Noise and Epigenetic Inheritance of Single-Cell Division Times Influence Population Fitness. *Curr Biol.* 2016;26(9):1138–47. <https://doi.org/10.1016/j.cub.2016.03.010> PMID: 27068419
37. Derrida B. Random-energy model: Limit of a family of disordered models. *Phys Rev Lett.* 1980;45(2):79.
38. Bovier A. *Statistical mechanics of disordered systems: a mathematical perspective.* Cambridge University Press. 2006.
39. Ben Arous G, Bogachev LV, Molchanov SA. Limit theorems for sums of random exponentials. *Probab Theory Relat Fields.* 2005;132(4):579–612. <https://doi.org/10.1007/s00440-004-0406-3>
40. Strong SP, Koberle R, Van Steveninck RRDR, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett.* 1998;80(1):197.
41. Roldán E, Parrondo JMR. Estimating dissipation from single stationary trajectories. *Phys Rev Lett.* 2010;105(15):150607. <https://doi.org/10.1103/PhysRevLett.105.150607> PMID: 21230886
42. GrandPre T, Teza G, Bialek W. Direct estimates of irreversibility from time series. 2024. <https://arxiv.org/abs/241219772>
43. Teza G, Stella AL, GrandPre T. Coarse-graining via lumping: exact calculations and fundamental limitations. In: 2025. <https://arxiv.org/abs/2512.11974>
44. Guevara Hidalgo E, Nemoto T, Lecomte V. Finite-time and finite-size scalings in the evaluation of large-deviation functions: Numerical approach in continuous time. *Phys Rev E.* 2017;95(6–1):062134. <https://doi.org/10.1103/PhysRevE.95.062134> PMID: 28709321
45. Ray U, Chan GK-L, Limmer DT. Importance sampling large deviations in nonequilibrium steady states. I. *J Chem Phys.* 2018;148(12):124120. <https://doi.org/10.1063/1.5003151> PMID: 29604886
46. GrandPre T, Klymko K, Mandadapu KK, Limmer DT. Entropy production fluctuations encode collective behavior in active matter. *Phys Rev E.* 2021;103(1–1):012613. <https://doi.org/10.1103/PhysRevE.103.012613> PMID: 33601608
47. Angeli L, Grosskinsky S, Johansen AM, Pizzoferrato A. Rare event simulation for stochastic dynamics in continuous time. *J Stat Phys.* 2019;176(5):1185–210.
48. van der Meer J, Degünther J, Seifert U. Time-Resolved Statistics of Snippets as General Framework for Model-Free Entropy Estimators. *Phys Rev Lett.* 2023;130(25):257101. <https://doi.org/10.1103/PhysRevLett.130.257101> PMID: 37418719
49. Pietzonka P, Coghi F. Thermodynamic cost for precision of general counting observables. *Phys Rev E.* 2024;109(6–1):064128. <https://doi.org/10.1103/PhysRevE.109.064128> PMID: 39020906
50. Levien E, Min J, Kondev J, Amir A. Non-genetic variability in microbial populations: survival strategy or nuisance?. *Rep Prog Phys.* 2021;84(11):116601. <https://doi.org/10.1088/1361-6633/ac2c92>
51. Amir A. Cell size regulation in bacteria. *Phys Rev Lett.* 2014;112(20):208102.
52. Ho P-Y, Lin J, Amir A. Modeling Cell Size Regulation: From Single-Cell-Level Statistics to Molecular Mechanisms and Population-Level Effects. *Annu Rev Biophys.* 2018;47:251–71. <https://doi.org/10.1146/annurev-biophys-070317-032955> PMID: 29517919
53. Gingrich TR, Horowitz JM. Fundamental Bounds on First Passage Time Fluctuations for Currents. *Phys Rev Lett.* 2017;119(17):170601. <https://doi.org/10.1103/PhysRevLett.119.170601> PMID: 29219443
54. Moukarzel C, Parga N. The REM zeros in the complex temperature and magnetic field planes. *Phys A: Stat Mech Appl.* 1992;185(1–4):305–15.
55. Ogure K, Kabashima Y. On analyticity with respect to the replica number in random energy models: I. An exact expression for the moment of the partition function. *J Stat Mech.* 2009;2009(03):P03010. <https://doi.org/10.1088/1742-5468/2009/03/p03010>
56. Derrida B, Mottishaw P. Finite size corrections in the random energy model and the replica approach. *J Stat Mech.* 2015;2015(1):P01021. <https://doi.org/10.1088/1742-5468/2015/01/p01021>
57. Kohram M, Vashistha H, Leibler S, Xue B, Salman H. Bacterial Growth Control Mechanisms Inferred from Multivariate Statistical Analysis of Single-Cell Measurements. *Curr Biol.* 2021;31(5):955–964.e4. <https://doi.org/10.1016/j.cub.2020.11.063> PMID: 33357764
58. Rasmussen CE. *Gaussian Processes in Machine Learning.* Lecture Notes in Computer Science. Springer Berlin Heidelberg. 2004. p. 63–71. [https://doi.org/10.1007/978-3-540-28650-9\\_4](https://doi.org/10.1007/978-3-540-28650-9_4)
59. Monod J. The growth of bacterial cultures. *Annu Rev Microbiol.* 1949;3:371–94.
60. Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. Modeling of the bacterial growth curve. *Appl Environ Microbiol.* 1990;56(6):1875–81. <https://doi.org/10.1128/aem.56.6.1875-1881.1990> PMID: 16348228
61. Baranyi J, Roberts TA. A dynamic approach to predicting bacterial growth in food. *Int J Food Microbiol.* 1994;23(3–4):277–94. [https://doi.org/10.1016/0168-1605\(94\)90157-0](https://doi.org/10.1016/0168-1605(94)90157-0) PMID: 7873331
62. Peleg M, Corradini MG. Microbial growth curves: what the models tell us and what they cannot. *Crit Rev Food Sci Nutr.* 2011;51(10):917–45. <https://doi.org/10.1080/10408398.2011.570463> PMID: 21955092

63. Ram Y, Dellus-Gur E, Bibi M, Karkare K, Obolski U, Feldman MW, et al. Predicting microbial growth in a mixed culture from growth curve data. *Proc Natl Acad Sci U S A*. 2019;116(29):14698–707. <https://doi.org/10.1073/pnas.1902217116> PMID: [31253703](https://pubmed.ncbi.nlm.nih.gov/31253703/)
64. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res*. 2009;19(10):1836–42. <https://doi.org/10.1101/gr.093955.109> PMID: [19622793](https://pubmed.ncbi.nlm.nih.gov/19622793/)
65. Wisner MJ, Lenski RE. A Comparison of Methods to Measure Fitness in *Escherichia coli*. *PLoS One*. 2015;10(5):e0126210. <https://doi.org/10.1371/journal.pone.0126210> PMID: [25961572](https://pubmed.ncbi.nlm.nih.gov/25961572/)
66. McGee RS, Kinsler G, Petrov D, Tikhonov M. Improving the Accuracy of Bulk Fitness Assays by Correcting Barcode Processing Biases. *Mol Biol Evol*. 2024;41(8):msae152. <https://doi.org/10.1093/molbev/msae152> PMID: [39041198](https://pubmed.ncbi.nlm.nih.gov/39041198/)