Quantifying Selective Pressures Driving Bacterial Evolution Using Lineage Analysis

Guillaume Lambert

The Institute of Genomics and Systems Biology, The University of Chicago, Chicago, Illinois 60637, USA

Edo Kussell

Department of Biology and Center for Genomics and Systems Biology,
and Department of Physics, New York University, New York, New York 10003, USA

Organisms use a variety of strategies to adapt to their environments and maximize long-term growth potential, but quantitative characterization of the benefits conferred by the use of such strategies, as well as their impact on the whole population's rate of growth, remains challenging. Here, we use a path-integral framework that describes how selection acts on *lineages*—i.e., the life histories of individuals and their ancestors—to demonstrate that lineage-based measurements can be used to quantify the selective pressures acting on a population. We apply this analysis to *Escherichia coli* bacteria exposed to cyclical treatments of carbenicillin, an antibiotic that interferes with cell-wall synthesis and affects cells in an age-dependent manner. While the extensive characterization of the life history of thousands of cells is necessary to accurately extract the age-dependent selective pressures caused by carbenicillin, the same measurement can be recapitulated using lineage-based statistics of a *single* surviving cell. Population-wide evolutionary pressures can be extracted from the properties of the surviving lineages within a population, providing an alternative and efficient procedure to quantify the evolutionary forces acting on a population. Importantly, this approach is not limited to age-dependent selection, and the framework can be generalized to detect signatures of other trait-specific selection using lineage-based measurements. Our results establish a powerful way to study the evolutionary dynamics of life under selection and may be broadly useful in elucidating selective pressures driving the emergence of antibiotic resistance and the evolution of survival strategies in biological systems.

I. INTRODUCTION

The theory of natural selection is based on the concept of fitness, which quantifies an organism's ability to reproduce and transmit its genetic information to its offspring [1]. The fitness of an organism, however, is notably difficult to measure directly [2–4], especially when the selective pressure is weak or depends indirectly on an organism's phenotype and external environment. In microorganisms, conventional measures of fitness typically rely on the instantaneous growth rate in a constant environment, but this approach fails to provide a complete measure of fitness that accounts for natural environmental fluctuations [5–7]. In order to optimally adapt to changing environments, cells have to optimize the fitness trade-offs involved in the use of strategies such as phase variation [8], phenotypic switching [9], or epigenetic inheritability [10], which stochastically generate heterogeneity within the population at typically

low rate in response to unstable environments [11–13]. Since traditional approaches to determine the fitness of a species require the analysis of population-wide properties (e.g., coalescence statistics [14], the variance of a trait in a population [15], or the rate of allele growth [16]), precise characterization of selection forces acting on biological systems remains challenging. This challenge is compounded by the importance of life-history traits, i.e., the dependence of growth rates and survival on organisms' ages and reproductive history [17–19]. Unless every reproduction and death event within a population is monitored for extended periods of time to account for low-probability events, an explicit and sensitive measure of selection is rarely obtained in biological systems.

Here, we propose a method to determine an organism's fitness that relies on the distribution of properties of *lineages*—i.e., the full history of a cell and of its ancestors—to extract the evolutionary pressures affecting a population. In the lineage formulation, a population is viewed as a collection of histories that reproduce and evolve over time [20]. As we discuss herein, due to exponential growth, the set of lineages present at any given time in a population (i.e., the surviving lineages) exhibits a statistical

distribution that maximizes fitness over the entire collection of possible lineages according to a variational principle. These optimal lineages balance a trade-off between (a) increasing their numbers by using individuals with more offspring than expected and (b) decreasing their numbers by trading the more numerous, typical individuals for the rare, high-performing individuals. This optimization occurs automatically, given sufficient time, in any continuously proliferating population and yields surviving lineages with a statistical distribution that differs from that of typical individuals within the population.

Remarkably, we find that a few of those surviving lineages provide sufficient information to determine the selective forces acting on the whole population. This result may seem particularly surprising, especially because the fitness of individuals that contribute to surviving lineages deviates substantially from that of typical individuals observed independently within the population. For example, individuals that die do not contribute to surviving lineages, yet mortality is a key determinant of fitness that seemingly should require explicit measurement. Likewise, individuals that by chance reproduce more often than others might be overrepresented in surviving lineages compared to their frequency in isolation.

In this paper, we show that while surviving lineages provide information about only a small fraction of all individuals in the population, they possess a statistical distribution that can be used to precisely infer selective forces and their magnitude. We demonstrate that the underrepresentation of certain individuals (e.g., those that die) along a lineage is as informative as the overrepresentation of others. For example, the absence of individuals above a certain age along lineages can be used to deduce a sharp decline in reproductive capacity with age (due to the combined effects of mortality and fecundity). In contrast, deducing the same using the conventional approach without lineages would require significantly more data: one would need to observe many individuals that aged and did not reproduce, whereas along a lineage their absence is by itself highly informative.

We apply the lineage approach to infer the life-history traits of *Escherichia coli* bacteria subjected to cyclical antibiotic treatments. Before describing our main results, we first review in the next sections the basic theory connecting lineage analysis and life-history traits (see Ref. [21] for further details).

A. Life-history traits and strength of selection

Our analysis examines two basic life-history traits, both of which are functions of the age x of an organism: the survival function $\ell(x)$, which is the probability that an individual reaches age x, and the age-dependent reproduction rate $b(x)$, also known as fecundity, which is the rate of offspring production at age x [22]. The product of these two functions is called the reproductive function

$k(x) = \ell(x)b(x)$, and it specifies the total reproductive output of individuals at age x [23]. We define $\kappa(x) \equiv k'(x)/k(x)$, which measures how quickly the reproductive rates change as a function of age; e.g., if fecundity is constant at all ages, then $\kappa(x)$ represents the age-specific mortality.

Quantifying these life-history traits provides detailed information on the reproductive potential of individuals (i.e., their fitness). However, information about the fitness of an organism alone cannot predict which features of the life history are under the strongest evolutionary pressures within a population. Measuring the strength of selection acting on life-history traits requires further analysis.

Hamilton [24] formulated an elegant approach that uses the sensitivity of the population's long-term growth rate Λ to changes in $\ell(x)$ and $b(x)$ to measure selection strength. The quantity $\delta\Lambda/\delta h(x)$ measures the impact on the population's growth rate of a small change in a life-history trait h acting at age x and provides a direct link between the behavior of individual cells and the population-wide properties of a biological system. As an example of this formalism, consider two small fitness boosts $\delta h(x_1)$ and $\delta h(x_2)$ that have the same magnitude but affect individuals at age x_1 and x_2, respectively [Fig. 1(a)]. This change in fitness may result from mutation, an epigenetic modification, a different survival strategy, etc. Each δh affects two different classes of individuals and, as a consequence, may have a very different impact on the population's overall growth rate $\delta\Lambda$ [Figs. 1(b) and 1(c)]. The impact of these small boosts in fitness depends on the relative contribution
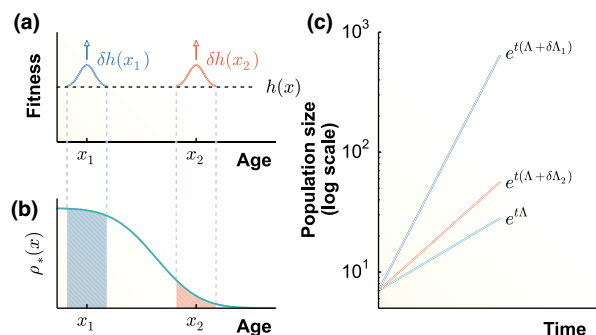

FIG. 1. Strength of selection on life-history traits and the optimal lineage relation. (a) A behavioral change (e.g., a mutation, an epigenetic modification, or a different survival strategy) causes a small fitness increase δh in a life-history trait $h(x)$ for cells at ages x_1 and x_2. (b) After a sufficiently long time, a population reaches steady-state growth and the population dynamics are dominated by cells whose ancestral lineage has an optimal age distribution $\rho_*(x)$. The distribution ρ_* determines what portion of lineages benefit from the small boosts in fitness at ages x_1 and x_2: the change in growth rate $\delta\Lambda$ resulting from the boost in fitness at age x_i is precisely equal to $\delta h \cdot \rho_*$ evaluated at x_i. (c) In this example, the change in growth rate is larger at age x_1 because it affects a larger fraction of the population's lineage history.

of the targeted subpopulation to the population's overall growth. For example, decreasing infant mortality in human populations has a much greater impact on a country's population growth rate than prolonging the life expectancy of individuals past reproductive age. In practice, Hamilton's sensitivity analysis yields a formula for computing $\delta\Lambda/\delta h(x)$, given precise measurements of life-history traits on a sufficiently large sample of individuals [23].

Next, we show how lineage analysis provides a much more direct measurement of Hamilton's strength of selection and uses the population's dynamics as a whole to infer the sensitivities.

B. Lineage dynamics and the optimal lineage relation

We consider a growing asexual population consisting of individuals that reproduce and survive according to given life-history parameters $b(x)$ and $\ell(x)$. At any given time $t > 0$, each individual in the population can be traced back to the initial cell present at time 0, and its lineage of ancestral cells can be recorded. Such lineage tracing analysis yields a function $\sigma(t')$, which specifies the age of the ancestral cells at each point in time $t' < t$ along the lineage. Each time a new ancestral cell is born along the lineage, the value of $\sigma(t')$ jumps back to zero. Since the set of all possible lineages in a population is astronomically large, for analysis purposes, it is useful to coarse grain lineages into large groups. For life-history traits such as $l(x)$ and $b(x)$ that depend on age alone and do not depend on the precise order of reproduction events in a lineage, one can group lineages by their lineage age distribution $\rho(x)$, which specifies the frequency of age x along a lineage. For each value of ρ, there exists a large number of distinct lineages having ρ as their lineage age distribution; their number grows exponentially with time with a rate $s(\rho)$. This exponential growth is purely entropic (i.e., due to the increasingly large number of possible lineages as the total time increases), and we call $s(\rho)$ the lineage entropy. Moreover, due to reproduction and differential survival, the frequency of each distinct lineage grows (or declines) in the population exponentially at a rate that we call the lineage fitness $f(\rho)$. Together, this implies that the number of cells in the population that have lineage age distribution ρ is given by $e^{ts(\rho)}e^{tf(\rho)}$.

Next, we consider the total size of a population $N(t)$ by summing over all possible values of the distribution ρ. We obtain

$$N(t) = \int e^{t[s(\rho)+f(\rho)]}\mathcal{D}\rho, \qquad (1)$$

where the integral is carried out by summing over all the possible age distributions ρ. This path-integral expression provides a key intuition about the population's lineage structure. As the total time t increases, lineages with age

distribution ρ that maximize $s(\rho) + f(\rho)$ will exponentially outcompete all other lineages. These optimal lineages therefore dominate the population at long times; we call their associated age distribution ρ_*. By the same reasoning, Eq. (1) also shows that the long-term growth rate of the population, $\Lambda \equiv (1/t)\log N(t)$, is equal to $s(\rho_*) + f(\rho_*)$. In fact, it is possible to calculate explicitly the lineage entropy and fitness of each age distribution ρ and maximize their sum over all possible ρ to find ρ_* as a function of life-history parameters [21]. Inverting this relationship, one can infer the life-history traits by measuring the lineage age distribution ρ_* (see below). Using our experimental data, we show that this procedure provides very accurate measurements using a relatively small amount of lineage data.

More than just a convenient tool, the lineage formulation makes deep connections between life-history analysis and evolutionary theory. In particular, it shows that Hamilton's sensitivity measure is fully determined by the optimal lineages [21] in the following way:

$$\frac{\delta\Lambda}{\delta\kappa(x)} = \rho_*(x) \quad \text{(optimal lineage relation).} \qquad (2)$$

This relation states that the change in long-term growth rate that would result from the specified change in life-history traits is, in fact, already imprinted in the optimal lineage distribution—it is simply the observed frequency of the given age along an optimal lineage. Moreover, this sensitivity, which reports on trait-specific selection acting on the *entire* population, is determined by a measurement made along a *single* optimal lineage. Using this relationship, any other sensitivity, such as, e.g., $\delta\Lambda/\delta\ell(x)$, can also be determined [21].

Measuring the optimal lineage age distribution may be challenging in biological systems: very few organisms possess a lineage distribution equal to $\rho_*(x)$ over short time scales, and an exponentially growing population cannot be accurately tracked over long time scales. However, even in a bounded population of continuously reproducing cells, the time-averaged rate of cell divisions converges to Λ, and the distribution of *surviving* lineages, denoted by $\rho_s(x)$, approaches the optimal lineage distribution at $t \gg 1$. These results have been demonstrated theoretically in Ref. [20]. Here, we confirm that (1) the surviving lineage ρ_s is a good approximation of ρ_* and (2) the optimal lineage relation described in Eq. (2) provides an accurate measurement of Hamilton's selection strength for life-history traits by applying the lineage formalism to cyclical antibiotic treatments known to affect *E. coli* bacteria in an age-dependent manner. Our approach demonstrates that measuring the properties of a single surviving lineage (i.e., the "path" that maximizes the survival probability) can be used to extract the selection forces acting on a population. These results have powerful implications: instead of measuring the

survival and growth rate of all members of a population, we show that monitoring the properties of the surviving lineage in a population is sufficient to determine individuals' life-history parameters, while simultaneously characterizing the selection pressures acting on the population as a whole.

II. RESULTS

A. Antibiotic treatments and age-dependent selection

We choose carbenicillin, a member of the carboxypenicillin antibiotic family that is known to interfere with cellular growth by preventing cross-linking of peptidoglycan units during cell-wall synthesis [25,26], as an age-dependent selective agent. Bacteria often survive antibiotic treatments by stopping cellular divisions, but not elongation, during exposures to cytotoxic agents [27]. Furthermore, analogs of carbenicillin have been shown to specifically target cells undergoing cellular division [28,29], suggesting that division rather than cell-wall synthesis is the main factor that increases the sensitivity to antibiotics.

To experimentally test whether carbenicillin affects bacterial cells in an age-dependent manner, we first monitor 18 different populations of *E. coli* bacteria growing inside a microfluidic device [Fig. 2(a)] and subject them to a 20-min pulse of a lethal dose of carbenicillin (100 μg/mL) every 90 min. A fluorescence micrograph of each population is recorded every minute, and each image undergoes processing to extract the physiological properties (e.g., position, length, age, elongation rate) of each cell. Cells grow and divide in a monolayer within the chambers, and are removed either by antibiotic-mediated killing or by "flow" death, i.e., departure from the growth chamber due to cellular growth [Fig. 2(a)].

Before the first antibiotic treatment, the average number of cells in each growth chamber, shown in Fig. 2(b), is stable, indicating that cell divisions and "flow" deaths are balanced. Then, the number of surviving cells systematically decreases following each antibiotic treatment, and the population recovers and stabilizes at a lower population size that is sustained over time (consisting of ~15–30 cells). While a small fraction of cells still undergo a flow death during and after the antibiotic treatments, most of the population decrease is attributed to cell lysis. The occurrence of every lysis event is recorded and the average lysis rate following each treatment is shown in Fig. 2(b), demonstrating that cell lysis is highly correlated with the antibiotic treatments [Fig. 2(b), lower panel]. No resistance to the carbenicillin treatments emerges, and cells remain susceptible to the antibiotic after 25 treatments.

Figure 2(c) shows the lysis rate averaged over each 90-min treatment. The lysis rate increases linearly with carbenicillin exposure and peaks approximately 10 min after external carbenicillin is removed. The survival probabilities under other fluctuation regimes are shown in Fig. S1 in the Supplemental Material [30]. Under these
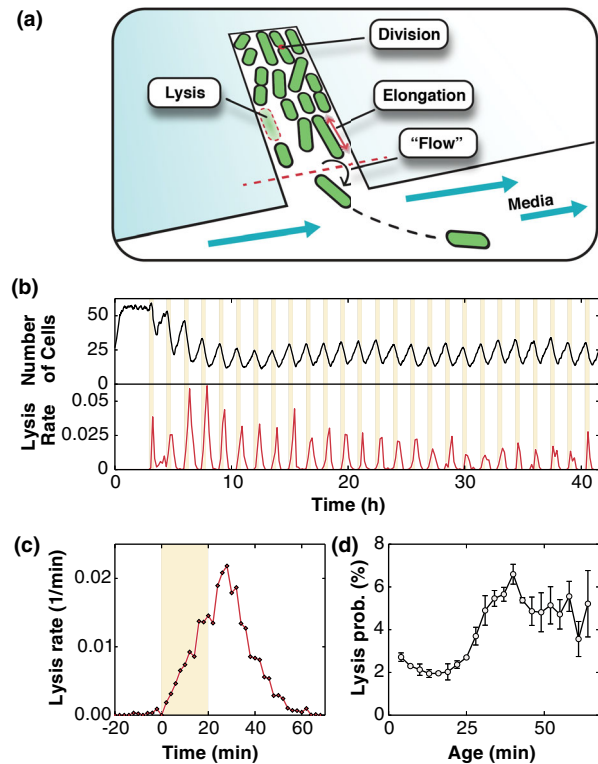

FIG. 2. Lytic response to carbenicillin antibiotics. (a) Schematic representation of the microfluidics device, where cells confined to growth chambers can be monitored for hundreds of generations. (b) The number of cells decreases sharply following each 100 μg/mL carbenicillin treatment. The measured lysis rate is also in phase with the antibiotic treatments. (c) Average lysis rate in response to each 20-min treatment of carbenicillin (number of treatments is 25). The lysis rate peaks approximately 10 min after antibiotic removal. (d) The rate of cell lysis increases with cell age (mean $\pm$ s.d., $n = 3$ min).

experimental conditions, a given cell has a cumulative probability of approximately 27% to undergo lysis following each treatment [31].

To determine whether cell lysis occurs in an age-dependent manner, we next calculate the age-dependent lysis probability for cells under cyclical carbenicillin treatments. We find in Fig. 2(d) that the lysis probability is strongly dependent on the time since the last division, and progressively increases by a factor of ~3 when the time since the last division exceeds 25 min. In fact, this threshold coincides with the average division age of cells growing in the absence of carbenicillin treatments (Fig. S2 in Ref. [30]), suggesting that cells nearing divisions may be more sensitive to carbenicillin treatments.

B. Life-history measurements of the reproductive function

In order to directly quantify the selective forces acting on the cells during the antibiotic treatments, we next measure the age-dependent division rate $q(x)$ and the survival

function $\ell(x)$. Note that since each cell division yields two cells with age zero, the reproductive rate is twice the division rate; i.e., $b(x) = 2q(x)$. We denote by $m(x)$ the age-specific antibiotic killing rate, and since both killing and cell divisions remove cells from any given age, we have $l'/l = -q - m$. Both $q(x)$ and $l(x)$ are measured for two different experiments in which cell populations are grown either under cyclical carbenicillin treatments [Figs. 3(a) and 3(b), blue line] or under constant conditions [Figs. 3(a) and 3(b), dashed line]. Under cyclical carbenicillin treatments, the large number of environmental periods (25 cycles) and the large number of independent populations ($n = 18$) ensures that the $q(x)$ and $l(x)$ functions converge to a stable form when averaged over the entire duration of the experiment.
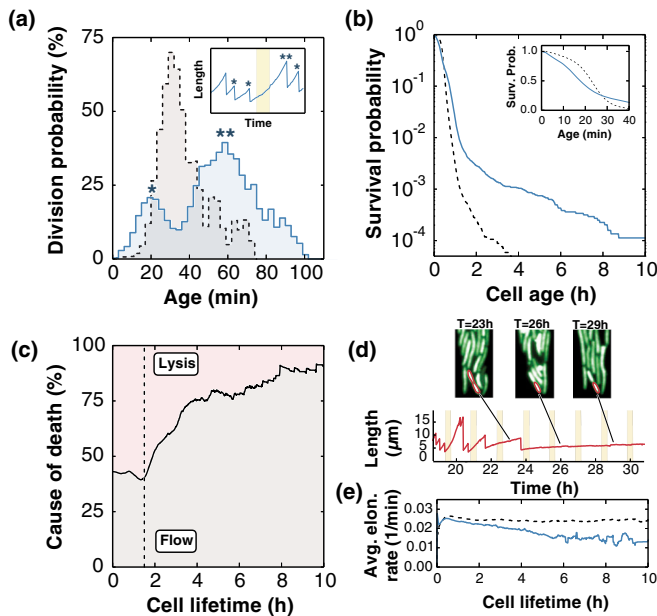

FIG. 3. Age-dependent growth and survival functions. (a) The age-dependent division probability obtained from life-history measurements of a population under cyclical carbenicillin treatments [$q(x)$, blue line] differs from a population grown under constant conditions (dashed line). Inset: The bimodal distribution of $q(x)$ is caused by the carbenicillin-dependent division rate, where cellular divisions are avoided during carbenicillin treatments. (b) Graph of the survival function $\ell(x)$ under carbenicillin treatments (blue line) as a function of the cell age. A non-negligible fraction of persister cells survive more than 2 h under carbenicillin treatments compared with a population grown in the absence of selection (dashed lines). Inset: Carbenicillin treatments decrease the survival probability for ages ≤ 28 min (blue line) compared to an untreated population (dashed line). (c) The fraction of cellular deaths caused by lysis decreases once cells survive a single carbenicillin treatment cycle (dashed line at $t = 90$ min). (d) An example of a persister cell that survives multiple antibiotic treatments but fails to revert to a proliferative state. (e) The average elongation rate decreases with increasing cell lifetime for cells under cyclical carbenicillin treatments (blue line) but remains constant for the untreated population (dashed line).

To measure $q(x)$, we first identify every division that occurs inside the growth chambers and compile the time since the cell's last division for each event. We next use this information to compute the age-dependent division probability for all ages $x \in [0, 100]$ min. Under carbenicillin treatments, we obtain a bimodal distribution, because cells tend not to divide under antibiotic treatments [Fig. 3(a), inset]. Cells avoid dividing during each treatment, and only resume divisions approximately 20 min following antibiotic removal (Fig. S3 in Ref. [30]).

The survival function $\ell(x)$ is shown as a function of cell age, the time since the last division [Fig. 3(b)]. The survival probability is lower at ages ≤ 28 min for cells subjected to carbenicillin treatments, confirming the age-dependent lysis rate measured above. The fraction of cells that reach older ages, on the other hand, is increased in the presence of carbenicillin treatments. This observation, as we show below, is consistent with the presence of persister cells within the population whose cell division rate is reduced during antibiotic exposures [32–34].

We show in Fig. 3(c) the cause of death for cells growing inside the growth chambers under carbenicillin treatments, which is either lysis or flow death. While the fraction of cells that die through cell lysis remains stable during the first cycle ($t < 90$ min), it progressively decreases with each antibiotic cycle. Figure S3 in Ref. [30] shows that carbenicillin treatments have no effect on cellular growth, implying that the increase in flow deaths is not attributed to a change in cell elongation rate. Instead, we observe that most of the cells that survive multiple treatments are slow-growing cells (i.e., persister cells [32–34]) that are unaffected by repeated antibiotic treatments [Fig. 3(d)]. Cells are more likely to be observed in a persister state with increasing cell lifetime (i.e., time since birth), which leads to a decrease in the average growth rate of the population [Fig. 3(e)]. Persistence, an example of a reversible stochastic switch, is a beneficial survival strategy when environments change infrequently [11,35]. In our experiments, the time scale of the carbenicillin treatments is fast, on the order of the cell division time; hence, reversal of persistence is not beneficial and does not contribute significantly to population growth [Fig. 3(d)].

C. Optimal lineages and selection: Experiment

Having directly measured the life-history parameters of individual cells in the population, we next attempt to extract the same information from lineage analysis of a single surviving cell within each population. Indeed, from the preceding analysis of lysis timing, we expect that the lineages that survive the 25 carbenicillin treatments should exhibit obvious differences from those that did not—e.g., by avoiding divisions just before, during, or just after carbenicillin treatments. Therefore, the lineage age distribution for surviving lineages should differ from the population-wide average (the complete pedigree of a

population is shown in Fig. 4(a) and the raw lineage datasets are included in the Supplemental Material [30]). We first compare in Fig. 4(b) the average age of the lineage and of the whole population over a single 90-min treatment, and observe that the average age of the surviving lineage decreases less rapidly upon antibiotic removal ($t = [20, 40]$ min), indicating that the surviving cells, on average, do not immediately divide after carbenicillin is removed from the environment. This observation indirectly confirms a strong age-dependent lysis rate caused by carbenicillin treatments, where the surviving lineage avoids division during sensitive ages. We stress that this observation does not imply active decision making in individual cells, and may arise as a direct consequence of age-dependent selection.

Conversely, the age-dependent structure of the surviving lineage differs from the population average. Figure 4(c) shows that the surviving lineage contains a larger fraction of cells whose age is between 25 and 60 min and a lower fraction of younger cells. This overrepresentation in the 25–60-min range suggests that cells in the surviving lineage are, on average, able to avoid cellular lysis during the carbenicillin treatments. The population, on the other hand, contains many lineages that terminate following a carbenicillin treatment, so a lower fraction of cells survive past

the age of 25 min. The overrepresentation of younger cells in the population distribution, a phenomenon also observed in the absence of selection [Fig. 4(c), inset], results from the continuous flow of cells out of the channel at a rate equal to the population growth rate Λ, whereas the surviving lineage consists precisely of those cells that escape this removal process.

The signatures of the age-dependent selection caused by carbenicillin treatments detected in Figs. 4(b) and 4(c) confirm that the surviving lineage differs from the population-wide distributions. We next use the surviving lineage distribution $\rho_s(x)$ to directly extract the reproductive function $k(x)$ and compare it with the life-history measurements obtained in Fig. 4. In Ref. [21], the following explicit expression for the optimal lineage age distribution was derived:

$$\rho_*(x) = \frac{1}{\bar{\tau}} \int_x^\infty k(\tau) e^{-\Lambda \tau} d\tau, \qquad (3)$$

where $\bar{\tau}$ is the mean age of the mothers at birth. Inverting this expression, we obtain

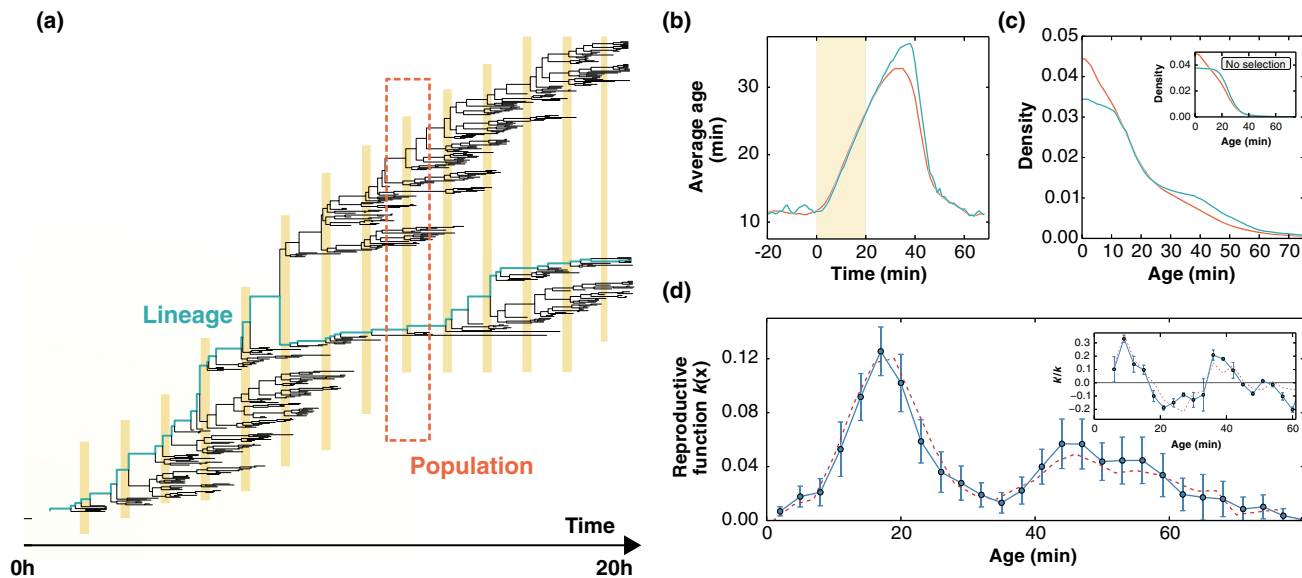$$k_s(x) = -\bar{\tau} e^{\Lambda x} \frac{d\rho_*(x)}{dx}, \qquad (4)$$

FIG. 4. Lineage distribution and selection. (a) Bacterial pedigree of a population for $t = [0, 20]$ h. The difference between lineage (teal) and population (orange) measurements is highlighted. The yellow bars denote antibiotic treatments. (b) Comparison between the average age of the population and surviving lineages during each treatment cycle. (c) The age distribution of population is underrepresented for ages between 25 and 50 min and enriched with younger cells. The different distributions reflect the age-dependent rate of lysis, whose effect is to remove older cells from the population but not from the surviving lineage. Inset: Lineage and population age distributions for cell populations grown in the absence of selection (i.e., constant conditions). (d) Comparison between the two measurements of the reproduction function, obtained using direct life-history measurements [$k(x) = b(x)\ell(x)$, dashed line] or extracted from the surviving lineage age distribution [$k_s(x)$ in Eq. (4), blue circles], is in very good agreement. Data are reported as the population average $\pm$ s.d. ($n = 18$ populations). Inset: Comparison between the rate of change of the reproduction function [$\kappa(x) \equiv k'(x)/k(x)$] obtained from the life history and surviving lineage measurements (mean $\pm$ s.d., $n = 18$ populations).

which enables the reproductive function, labeled here as $k_s(x)$, to be measured using the optimal lineage statistics ρ_*. In addition to ρ_*, this inversion requires Λ to be either independently measured or inferred directly from the lineage statistics. For example, in the absence of mortality ($m = 0$), all cells eventually divide, and the total expected number of offspring of each cell is $\int_0^\infty k(x) = 2$. Applying this constraint in Eq. (4) allows Λ to be determined from the lineage data alone. In our analysis of the data, which involves cell death, we avoid making this assumption and measure Λ directly in the microfluidics device. Interestingly, however, we find that the value of Λ measured (0.024 ± 0.003 min^{-1}) is in close agreement with the value inferred from lineage data (0.023 ± 0.004 min^{-1}). This observation indicates that cells are able to make up for reduced division rates at younger ages, when mortality is high [Fig. 3(b)], by increased division rates later in life [Fig. 3(a)]. In different biological contexts, different constraints on $k(x)$ may exist, which would similarly enable inference of the growth rate Λ from lineage analysis.

In Fig. 4(d), we plot the reproductive function $k_s(x)$ extracted from the age distribution of the surviving lineage shown in Fig. 4(c) using Eq. (4) (blue circles; see Sec. IV for details). The measurement of $k_s(x)$ is made using the lineage statistics of a single surviving cell in each of the 18 populations. We next compare this reproductive function with the experimentally determined $k(x)$ from the life-history traits $b(x)$ and $\ell(x)$ measured in Figs. 3(a) and 3(b), respectively [Fig. 4(d), dashed line]. The function $k(x)$ is obtained by combining every division and lysis event ($N_{\text{div}} = 15004$, $N_{\text{lysis}} = 6785$) that occurs in the 18 populations. Strikingly, both measurements use independent methods to find reproductive functions that are in very good agreement, especially when considering the shape of $\kappa(x)$, a more natural measure of an organism's fitness [21] [Fig. 4(d), inset].

Each method provides an accurate measurement of the reproductive function, and while both agree within experimental error, extracting $k(x)$ from the surviving lineage $\rho_s(x)$ requires a significantly lower amount of data. Indeed, instead of monitoring all members of a population and explicitly measuring the reproductive function $k(x)$, we show that it is enough to track a single cell back in time, record every division event to derive the age distribution $\rho_s(x)$, and use Eq. (4) to measure $k_s(x)$ with a similar level of accuracy. We note that, in contrast to lineage statistics, the population age distribution cannot, in general, be used to infer the reproductive function, as it is determined by $l(x)$ and Λ, and does not contain explicit dependence on $b(x)$ [21]. Our results demonstrate that (1) the surviving lineage measured in the growth chambers is a very good approximation to the optimal lineage and (2) lineage-based measurements provide an alternative, and more direct, method to characterize the selective forces acting on a population.

III. DISCUSSION

In this paper, we apply the optimal lineage formulation of population dynamics [21] to quantify the growth and survival of *E. coli* bacteria subjected to cyclical treatments of an antibiotic shown to affect cellular lysis in an age-dependent manner. We first measure the survival and reproductive functions directly from life-history data by monitoring and compiling every lysis and division event in 18 separate populations under 20-min treatments of a lethal dose of carbenicillin. We then confirm the optimal lineage formalism by inferring the age-dependent reproductive function using the lineage data alone. Strikingly, the properties of a few surviving lineages provide a highly accurate measurement of the reproductive function.

Our approach provides a simple yet powerful method to study selective pressures acting on populations. Using the optimal lineage relation [Eq. (2)], which links between the selective strength on life-history traits and surviving lineages' statistics, one can directly measure the strength of selection in a model-free way. Moreover, by applying this relation using a specific model of reproductive fitness, we show that signatures of age-dependent selection can be extracted from the statistical properties of a few surviving lineages.

The result outlined in Eq. (2) can be generalized to include other traits, for example, phenotypic states [20] or other physiological properties. Since a population's growth rate Λ depends exclusively on the form of the optimal lineage distribution ρ_*, knowledge of the optimal (or surviving) lineage distribution can be used to generate relations similar to Eq. (2). Broadly speaking, any observed difference in trait frequency along a surviving lineage versus in individual cells within a population suggests that the trait experiences selection. For example, we measure that the elongation rate along the surviving lineage is up to 10% faster than the population average (Fig. S4 in Ref. [30]), suggesting that elongation rate under constant conditions is under positive selection within our device. Similarly, we can also infer from lineage statistics that birth location inside the growth chamber is the strongest selective pressure acting on the population, and the optimal lineage relation suggests that the population would benefit the most by increasing the fitness of cells away from the main flow channel (Fig. S4 in Ref. [30]).

From an epidemiological perspective, studying the ancestry of a single individual provides a better estimate of the selection pressures than by looking at a cross section of the population at a fixed time [36]. By measuring the properties of the surviving lineages within a population, it may be possible to directly infer the traits, behaviors, or survival strategies that affect species' survival. Trait values that are most often observed along lineages will maximally affect the growth rate of the population. Recognizing the specific traits of surviving members of a population can provide great insight into the evolutionary

pressures acting on individuals. Generalizations of this path-integral approach to other biological systems can thus be used to detect and characterize the selective pressures acting on a population, and we hope they may provide significant insight into the emergence of survival strategies—such as antibiotic resistance—and the initial steps of speciation.

IV. METHODS AND MATERIALS

A. Strain construction

The nonmotile F3 strain of *E. coli* used in these experiments is from the Wakamoto lab (University of Tokyo). The strain is based on the W3110 background and has the *fliC*, *fimA*, and *flu* gene knocked out to decrease cell aggregation and surface adhesion [37]. The strains carry a plasmid conferring kanamycin resistance and expressed GFPmut2 under the control of the *rpsL* promoter [38].

B. Device description and fabrication

E. coli are grown inside growth chambers 7.5 μm wide by 50 μm long by 1 μm high growth chambers. The device was made using standard soft lithography techniques on a SU-8 master and details concerning the device fabrication have been described elsewhere [39]. A constant rate of liquid flow is maintained inside the main flow channel. Each reservoir is pressured to 6 psi and flow is controlled using in-line solenoid valves (Cole Parmer) connected to a computer-controlled Arduino microcontroller.

C. Experimental conditions and cyclical antibiotic treatments

Cells from a $-80\,°$C storage are plated on an agar plate containing Luria Bertani (LB) broth and 50 μg/mL of kanamycin. A single colony is suspended in fresh LB + kanamycin and grown until it reaches an OD_{600} of 0.2. Then, 4 mL of the cell culture is spun down and resuspended in 10 μL of fresh LB. The concentrated cell culture is then inoculated inside the output port of the microfluidic device and the cells are flushed out of the main chamber 5 min after inoculation with fresh LB broth. Tubes connected to either a LB + 25 μg/mL kanamycin reservoir or LB + 25 μg/mL kanamycin + 100 μg/mL carbenicillin are attached to each input port. For experiments performed under constant conditions, only a single input port connected to a LB + 25 μg/mL kanamycin reservoir is used. Micro-Manager [40] is used to control the microscope, record a fluorescent image (Andor Luca, Andor) of each growth chamber every minutes, and control the state of each solenoid valve. While 20 growth chambers are monitored during the experiment, only 18 are used in the analysis: two chambers are excluded because the number of cells decreased to zero less than 10 h after the start of the experiment.

D. Image processing and path reconstruction

A Gaussian blur filter is applied to each image to remove features smaller than 0.75 pixels and a boxcar filter is used to remove features greater than 7 pixels. The filtered image is then converted into a binary image using an adaptive threshold method (box size is 15 pixels) and a custom routine is applied to the resulting image to segment touching cells using a watershed algorithm [41]. Each cell within the binary image is then labeled and a blob detection algorithm is used to extract the position, length, width, orientation, and fluorescence intensity of each cell. The area that overlaps between all cells in different frames is recorded, and the generated assignment problem is solved using the Hungarian algorithm [42] to construct each cell path. Then, a second round of processing is applied to each individual cell path to correct mistakes in tracking [41] and generate the complete life-history data.

E. Divisions, age, elongation rates, and lysis events

Division events are found by locating sudden decreases in cell length that are either greater than 1.5 μm or 30% of the total length of a cell. Then, the algorithm assigns a mother-daughter relationship based on pole inheritance (i.e., the mother inherits the older pole) following each division event. This information is then used to assign to each cell an age and a lifetime variable: the age of both the mother and the daughter is reset to zero at every division event while the lifetime continues to increase over time through cell divisions. With the exception of Figs. 3(c) and 3(e), which use lifetimes, all of our analysis uses cell ages. The elongation rate is found by fitting an exponential function to the length of each cell between two division events [$L(t) = L(0)e^{rt}$]. Lysis events are found by locating cells that cease to exist a distance larger than 5 μm away from the end of the growth chamber; cells that die closer than 5 μm to the main flow channel are assigned a flow death.

F. Age distributions and the reproductive function

The age distribution for all cells present inside the growth chamber at each time is computed and averaged over time to find the age distribution in Figs. 4(b) and 4(c). The reproductive rate $b(x)$ is found directly by computing $q(x)$, the probability that a cell undergoes a division event between age x and $x + 1$ for all ages $x \in [0, 100]$ min [Fig. 3(a)], and using the relation $b(x) = 2q(x)$. The survival function $\ell(x)$ [Fig. 3(b)] is found by computing the probability that a cell survives to age x using every life history that ends in either cell division or lysis; those that end in a flow death are excluded since these events are due to population growth rather than intrinsic physiology of individual cells. The reproductive function is then computed using $k(x) = l(x)b(x)$, and $\kappa(x) \equiv k'(x)/k(x)$ is computed by taking the derivative of $k(x)$. A three-point moving average filter is applied to $k(x)$ before

differentiation to find $\kappa(x)$. Since we average over all periods with and without carbenicillin, our results pertain to effective, time-averaged parameters.

G. Measurements of the growth rate Λ

Two independent growth-rate measurements are used: one based on population-wide division rates (Λ_p) and a second that uses only the lineage data (Λ_ℓ). First, we determine Λ_p from the unique solution [21] to the expression

$$\int_0^\infty k(x)e^{-\Lambda_p x}dx = 1, \tag{5}$$

which can be obtained from Eq. (4) by multiplying both sides by $e^{-\Lambda x}$ and integrating, noting that $\rho_*(0) = 1/\bar{\tau}$. Using the reproductive function shown in Fig. 4(d) (dashed line), this method yields $\Lambda_p = 0.024 \pm 0.003$ min^{-1} (mean $\pm$ s.d., $n = 18$). The second method uses a constraint on $k(x)$, assuming the expected total reproductive output of each cell is 2; i.e., $\int_0^\infty k(x) = 2$. Applying this constraint, integrating both sides of Eq. (4) over x (by parts on the right-hand side) yields an equation for Λ_ℓ that depends on $\rho_*(x)$:

$$(\Lambda_\ell \bar{\tau}) \int_0^\infty e^{\Lambda_\ell x}\rho_*(x)dx = 1. \tag{6}$$

Equation (6) is used to find Λ_ℓ exclusively with the lineage data. The value of $\bar{\tau} = 28$ min (s.d. $= 2$ min, $n = 18$) is found by calculating the average cell division time along each lineage, and we numerically solve Eq. (6) using the measured ρ_* to find $\Lambda_\ell = 0.023 \pm 0.004$ min^{-1} (mean $\pm$ s.d., $n = 18$).

H. Lineage analysis

In each of the 18 growth chambers, a single cell present at $t = 40$ h is randomly selected and tracked backward in time until time $t = 0$ h to generate a lineage history (2 out of the 20 monitored populations are rejected because no cell survives past $T = 10$ h). The growth rate Λ_ℓ and the age distribution $\rho_s(x)$ along each surviving lineage are then used in combination with Eq. (4) to infer the reproductive function $k_s(x)$ from lineage data. The reproductive functions obtained from each of the 18 lineage histories are then averaged together to generate $k_s(x)$ shown in Fig. 4(d) (blue line).

ACKNOWLEDGMENTS

[1] C. Darwin, *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life* (J. Murray, London, 1859).

[2] H. Allen Orr, *The Genetic Theory of Adaptation: A Brief History*, Nat. Rev. Genet. **6**, 119 (2005).

[3] D. Schluter, *Estimating the Form of Natural Selection on a Quantitative Trait*, Evolution **42**, 849 (1988).

[4] R. G. Shaw and C. J. Geyer, *Inferring Fitness Landscapes*, Evolution (Lawrence, Kans.) **64**, 2510 (2010).

[5] S. F. Elena and R. E. Lenski, *Evolution Experiments with Microorganisms: The Dynamics and Genetic Bases of Adaptation*, Nat. Rev. Genet. **4**, 457 (2003).

[6] R. Kassen and P. B. Rainey, *The Ecology and Genetics of Microbial Diversity*, Annu. Rev. Microbiol. **58**, 207 (2004).

[7] K. M. Taute, S. Gude, P. Nghe, and S. J. Tans, *Evolutionary Constraints in Variable Environments, from Proteins to Networks*, Trends Genet. **30**, 192 (2014).

[8] S. S. Justice, D. A. Hunstad, L. Cegelski, and S. J. Hultgren, *Morphological Plasticity as a Bacterial Survival Strategy*, Nat. Rev. Microbiol. **6**, 162 (2008).

[9] T. Lenormand, D. Roze, and F. Rousset, *Stochasticity in Evolution*, Trends Ecol. Evol. **24**, 157 (2009).

[10] H. N. Lim and A. van Oudenaarden, *A Multistep Epigenetic Switch Enables the Stable Inheritance of DNA Methylation States*, Nat. Genet. **39**, 269 (2007).

[11] E. Kussell and S. Leibler, *Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments*, Science **309**, 2075 (2005).

[12] N. Dhar and J. D. McKinney, *Microbial Phenotypic Heterogeneity and Antibiotic Tolerance*, Curr. Opin. Microbiol. **10**, 30 (2007).

[13] R. D. H. Barrett and D. Schluter, *Adaptation from Standing Genetic Variation*, Trends Ecol. Evol. **23**, 38 (2008).

[14] R. R. Hudson, *Gene Genealogies and the Coalescent Process*, Oxford Surveys in Evolutionary Biology, Vol. 7 (Oxford University Press, Oxford 1990), pp. 44.

[15] G. R. Price, *Selection and Covariance*, Nature (London) **227**, 520 (1970).

[16] Z. Patwa and L. M. Wahl, *The Fixation Probability of Beneficial Mutations*, J. R. Soc. Interface **5**, 1279 (2008).

[17] J. E. Brommer, *The Evolution of Fitness in Life-History Theory*, Biol. Rev. Camb. Philos. Soc. **75**, 377 (2000).

[18] J. B. McGraw and H. Caswell, *Estimation of Individual Fitness from Life-History Data*, Am. Nat. **147**, 47 (1996).

[19] R. G. Shaw, C. J. Geyer, S. Wagenius, H. H. Hangelbroek, and J. R. Etterson, *Unifying Life-History Analyses for Inference of Fitness and Population Growth*, Am. Nat. **172**, E35 (2008).

[20] S. Leibler and E. Kussell, *Individual Histories and Selection in Heterogeneous Populations*, Proc. Natl. Acad. Sci. U.S.A. **107**, 13183 (2010).

[21] Y. Wakamoto, A. Y. Grosberg, and E. Kussell, *Optimal Lineage Principle for Age-Structured Populations*, Evolution (Lawrence, Kans.) **66**, 115 (2012).

[22] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).

[23] B. Charlesworth, *Evolution in Age-Structured Populations* (Cambridge University Press, Cambridge, England, 1994).

[24] W. D. Hamilton, *The Moulding of Senescence by Natural Selection*, J. Theor. Biol. **12**, 12 (1966).

[25] A. Tomasz, *The Mechanism of the Irreversible Antimicrobial Effects of Penicillins: How the Beta-Lactam Antibiotics Kill and Lyse Bacteria*, Annu. Rev. Microbiol. **33**, 113 (1979).

[26] M. A. Kohanski, D. J. Dwyer, and J. J. Collins, *How Antibiotics Kill Bacteria: From Targets to Networks*, Nat. Rev. Microbiol. **8**, 423 (2010).

[27] C. Miller, L. Elnif Thomsen, C. Gaggero, R. Mosseri, H. Ingmer, and S. N. Cohen, *SOS Response Induction by β-Lactams and Bacterial Defense against Antibiotic Lethality*, Science **305**, 1629 (2004).

[28] H. Suk Chung, Z. Yao, N. W. Goehring, R. Kishony, J. Beckwith, and D. Kahne, *Rapid β-Lactam-Induced Lysis Requires Successful Assembly of the Cell Division Machinery*, Proc. Natl. Acad. Sci. U.S.A. **106**, 21872 (2009).

[29] Z. Yao, D. Kahne, and R. Kishony, *Distinct Single-Cell Morphological Dynamics under Beta-Lactam Antibiotics*, Mol. Cell **48**, 705 (2012).

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.011016 for supporting figures and raw lineage datasets.

[31] While the number of cells per chamber decreases by ~50% following each antiiobic treatment, only half of this decrease is due to cell lysis. The rest of the decrease in cell density is due to filamentation, which artificially decreases the absolute number of cells inside the growth chambers.

[32] I. Keren, D. Shah, A. Spoering, N. Kaldalu, and K. Lewis, *Specialized Persister Cells and the Mechanism of Multidrug Tolerance in Escherichia coli*, J. Bacteriol. **186**, 8172 (2004).

[33] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, *Bacterial Persistence as a Phenotypic Switch*, Science **305**, 1622 (2004).

[34] B. R. Levin and D. E. Rozen, *Non-inherited Antibiotic Resistance*, Nat. Rev. Microbiol. **4**, 556 (2006).

[35] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler, *Bacterial Persistence: A Model of Survival in Changing Environments*, Genetics **169**, 1807 (2005).

[36] S. C. Stearns, S. G. Byars, D. R. Govindaraju, and D. Ewbank, *Measuring Selection in Contemporary Human Populations*, Nat. Rev. Genet. **11**, 611 (2010).

[37] J. Merrin, S. Leibler, and J. S. Chuang, *Printing Multistrain Bacterial Patterns with a Piezoelectric Inkjet Printer*, PLoS One **2**, e663 (2007).

[38] A. Zaslaver, A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon, *A Comprehensive Library of Fluorescent Transcriptional Reporters for Escherichia coli*, Nat. Methods **3**, 623 (2006).

[39] G. Lambert and E. Kussell, *Memory and Fitness Optimization of Bacteria under Fluctuating Evnvironments*, PLoS Genet. **10**, e1004556 (2014).

[40] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, in *Current Protocols in Molecular Biology*, edited by Frederick M. Ausubel, Roger Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (John Wiley & Sons, Hoboken, NJ, 2010).

[41] G. Lambert, http://github.com/guillambert/celltracker.

[42] J. Munkres, *Algorithms for the Assignment and Transportation Problems*, J. Soc. Ind. Appl. Math. **5**, 32 (1957).