Universal constraints on selection strength in lineage trees

Arthur Genthon ⬡* and David Lacoste ⬡

Gulliver UMR CNRS 7083, ESPCI Paris, Université PSL, 75005 Paris, France

We obtain general inequalities constraining the difference between the average of an arbitrary function of a phenotypic trait, which includes the fitness landscape of the trait itself, in the presence or in the absence of natural selection. These inequalities imply bounds on the strength of selection, which can be measured from the statistics of trait values and divisions along lineages. The upper bound is related to recent generalizations of linear response relations in stochastic thermodynamics, and shares common features with Fisher's fundamental theorem of natural selection, and with its generalization by Price, although they define different measures of selection. The lower bound follows from recent improvements on Jensen's inequality, and both bounds depend on the variability of the fitness landscape. We illustrate our results using numerical simulations of growing cell colonies and with experimental data of time-lapse microscopy experiments of bacteria cell colonies.

I. INTRODUCTION

Quantifying the strength of selection in populations is an essential step in any description of evolution. With the development of single cell measurements, a large amount of data on cell lineages is becoming available both at the genotypic and phenotypic level. By analyzing the statistics of cell divisions in population trees, one can measure selection more accurately than using classical population growth rate measurements [1]. Similarly, by tracking phenotypes on cell lineages, one can obtain statistically reliable estimations of the fitness landscape of a given trait and of the selection strength of that trait [2]. In addition, an optimal lineage principle can be used to infer the population growth rate [3] or selective forces [4] from lineage statistics. All these methods contribute to bridging the gap between single-cell experiments at the population level and molecular mechanisms [5].

An alternate method to infer selection in evolution focuses on dynamical trajectories of frequency distributions [6,7]. In these works, Mustonen *et al.* introduced the notion of fitness flux to characterize the adaptation of a population by taking inspiration from stochastic thermodynamics. In fact, ideas from stochastic thermodynamics can be applied directly at the level of individual cell trajectories [8]. By following this kind of approach, we have derived general constraints on dynamical quantities characterizing the cell cycle such as the average number of divisions or the mean generation time [9,10]. These constraints are universal because they hold independently of the specific cell dynamic model and they are indeed verified in experimental data. Other examples of universality in the context of evolution includes the identification of universal families of distributions of selected values and the use of methods from extreme value statistics [11,12].

Here we derive universal constraints for the average value of a function of a trait, and for its selection strength, by exploiting a set of recent results known under the name of thermodynamic uncertainty relations (TUR). These relations take the form of inequalities, which generalize fluctuation-response relations far from equilibrium [13], and which capture important trade-offs for thermodynamic and non-thermodynamic systems [14] as recently reviewed in [15]. Although our results are framed in the context of cell population in lineage trees, they apply more broadly to general stochastic processes defined on any branched tree.

We start in Sec. II by laying the theoretical framework and the definitions of the forward and backward samplings of lineages within a tree, which are at the core of the notions of fitness landscape and selection strength. In Sec. III, we derive a general upper bound for the difference between the average values of an observable with respect to two different probability distributions, which we use in Sec. IV to obtain an upper bound for the strength of selection. That result goes beyond the Gaussian approximation. In Sec. V we study the case of small variability which leads to a simple expression of the strength of selection, reminiscent of the Gaussian case. Those expressions have mathematical similarities with Fisher's fundamental theorem of natural selection and Price's equation, although they correspond to different definitions of selection, as detailed in Sec. VI. To complement the upper bound on the strength of selection, we use a recent sharpened version of Jensen's inequality to derive in Sec. VII a lower bound for the strength of selection. Both bounds are tested with simulations and experimental data in Sec. VIII, showing a very good agreement with the theory. Finally, we conclude in Sec. IX. Several Appendices (Appendix A to G) present the details of the calculations, supplementary figures, and numerical comparisons between our results and results previously published.

---

*arthur.genthon@espci.fr

II. A FRAMEWORK FOR LINEAGE STATISTICS

A colony of cells can be represented as a branched tree, whose branches are called lineages and whose nodes correspond to cell divisions. We assume that each cell in the population divides after a stochastic time into exactly m daughter cells. In order to extract relevant statistics from such a tree, one needs to sample the lineages following a weighting scheme. The backward (or retrospective) and the forward (or chronological) samplings have been introduced in the context of populations of cells [2,5,16], and previously defined in the mathematical literature [17,18]. The backward sampling of lineages assigns a uniform weight $N(t)^{-1}$ to each of the $N(t)$ lineages, leading to an over-representation of cells coming from subpopulations that divided more than average. To compensate for this bias, the forward sampling takes into account the number of divisions K along a lineage and assigns to the lineage a weight $N_0^{-1} m^{-K}$, where N_0 is the size of the initial population. Intuitively, a lineage is followed forward in time from a cell in the initial population, by choosing with uniform probability $1/m$ which daughter to follow among the m daughters at each division. In this sense, the forward sampling cancels the effect of selection because the sister cells born from the same division have the same weight, regardless of their reproductive successes, i.e., the sizes of the subpopulations they generate. Thus, the statistics obtained with a forward sampling of the lineages within a tree reproduces the statistics obtained in single-lineage experiments, like in mother-machine configuration [19].

A general phenotypic trait $\mathcal{S}$ then admits a forward and a backward distributions, respectively defined by $p_{\text{for}}(s, t) = \sum_{K=0}^{\infty} n(s, K, t)/(N_0 m^K)$ and $p_{\text{back}}(s, t) = n(s, t)/N(t)$, where $n(x, t)$ is the number of lineages featuring a cell with trait value x at time t. Comparing $p_{\text{for}}(s, t)$ and $p_{\text{back}}(s, t)$ offers some insight on the effect of selection on trait $\mathcal{S}$. For this purpose we define the fitness landscape as [2]

$$h_t(s) = \Lambda_t + \frac{1}{t} \ln \left[\frac{p_{\text{back}}(s, t)}{p_{\text{for}}(s, t)} \right], \tag{1}$$

where $\Lambda_t = \ln[N(t)/N_0]/t$ is the population growth rate. Note that in the classical framework of evolutionary dynamics, the notion of fitness landscape finds its origin in Wright's seminal work [20], and is defined as a mapping between the values or versions of a phenotype or a genotype, with their associated fitnesses [21]. The notion of fitness in biology has multiple meanings, but is often understood in this context as the reproductive success, or growth rate. In contrast, this is not the case for the fitness landscape we defined, which therefore should not be confused with the growth rate of the subpopulation carrying the trait value s. Indeed, the reproductive success is defined by the comparison of the frequencies of a trait in a population over time, whereas the fitness landscape as we defined it compares the frequencies of a trait at the same time but in ensembles with and without selection. Therefore, $h_t(s_1) > h_t(s_2)$ means that the trait value s_1 benefits more from selection than s_2, but not necessarily that s_1 has a greater reproductive success than s_2. As a consequence, cells carrying the value s_1 could still be less represented in the population than those carrying trait s_2. The two points of view are linked by simple relations as detailed in Appendix A, and we argue in

Sec. VI that this subtle difference leads to different definitions of the effect of selection, and that the point of view which compares chronological and retrospective distributions could be more suitable to describe selection for certain applications.

When the statistics of trait $\mathcal{S}$ is unaffected by selection, that is when there is no correlation between the number of divisions undergone by a cell and the value s for this trait, then $p_{\text{back}}(s, t) = p_{\text{for}}(s, t)$ and the fitness landscape is flat, equal to the population growth rate. Instead, if the statistics of the trait is strongly perturbed by selection, then the fitness landscape is more rough and exhibits important deviations from its mean value.

Therefore, the variance of the fitness landscape appears as a natural candidate to quantify the roughness of this fitness landscape effect. However, this variance can be computed in both the forward or backward ensembles, giving related but different results, and it is therefore unclear which of the two should be used. To resolve this issue, we define the strength of selection $\Pi_{\mathcal{S}}$ acting on the trait $\mathcal{S}$ as the change in mean fitness landscape between the ensembles with and without selection [2]:

$$\Pi_{\mathcal{S}} = \langle h_t(s) \rangle_{\text{back}} - \langle h_t(s) \rangle_{\text{for}}. \tag{2}$$

This quantity indeed reflects the roughness of the fitness landscape, since it is null when the fitness landscape is flat and becomes larger as the difference between the backward and forward statistics for trait $\mathcal{S}$ increases. This behavior is well understood by writing the strength of selection as [2]

$$\Pi_{\mathcal{S}} = \frac{1}{t} \mathcal{J}[p_{\text{back}}(s, t) | p_{\text{for}}(s, t)], \tag{3}$$

where $\mathcal{J}$ is the Jeffrey's divergence, a non-negative and symmetric information-theoretic distance between the two distributions $p_{\text{back}}(s, t)$ and $p_{\text{for}}(s, t)$, defined as $\mathcal{J}[p(x)|q(x)] = \int [p(x) - q(x)] \ln[p(x)/q(x)]dx$.

Let us briefly comment on two points. First, the strength of selection defined here should not be confused with the coefficient of selection, usually defined as the relative difference in fitness associated with two values of a phenotypic trait [6]. Second, the strength of selection is a function of time, since fitness landscapes are time dependent by definition. Only if a steady state is reached in the long time limit, then $h_t(s)$ tends to a constant equal to the steady state population growth rate Λ, and the strength of selection tends to 0, as expected since selection no longer shifts trait frequencies.

In the particular case of Gaussian distributions, the strength of selection $\Pi_{\mathcal{S}}$ and the variance of the fitness landscape are in fact linked by a very simple relation. More precisely, when the forward distribution for the fitness landscape is Gaussian, and for a bijective function $h_t(s)$, then its backward distribution is also Gaussian, with the same standard deviation but a shifted average value, leading to (see Appendix B)

$$\Pi_{\mathcal{S}} = t \, \text{Var}(h_t), \tag{4}$$

where the variance can be indifferently taken over the forward or backward sampling. Note that we recover here a result known from [2], in a more direct way and with restricted assumptions, since in that reference the authors derived this relation assuming that the joint distribution of $h_t(s)$ and K was a bivariate Gaussian distribution.
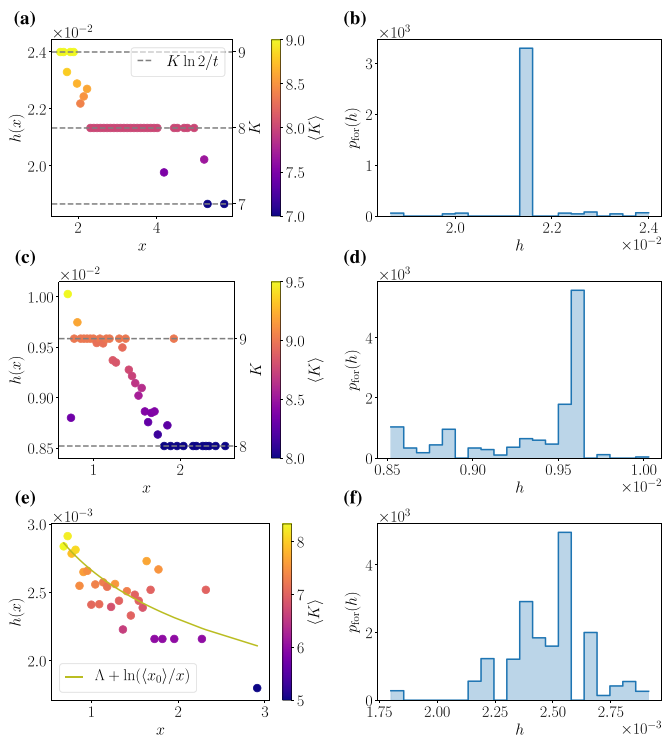
FIG. 1. Experimental fitness landscapes for size and their forward distributions, computed with data from [27]. Each row of the figure corresponds to a different experiment and the first column shows fitness landscapes $h_t(x)$ as functions of size x. In (a) and (c), the gray horizontal dashed lines correspond to theoretical plateaus, equal to $K \ln 2/t$, predicted when K is fully determined by the value s of the trait. The integers K corresponding to the plateaus are indicated on the right y axis. In (e), plateaus are blurred and replaced by a smoother scatter plot in good agreement with the general shape of the theoretical prediction, made in the case where there is no variability in individual growth rate nor volume partition at division [10]. Λ is the population growth rate and $\langle x_0 \rangle$ is the average size of initial cells. In (a), (c), and (e), each dot is made of all the cells having the same size, and the mean number of divisions among those cells is represented by the color of the dot. This shows that dots aligning on a plateau corresponding to a number K of divisions truly come from cells that underwent K divisions. In (b), (d), and (f), the three subfigures on the right represent the distribution $p_{\text{for}}(h)$ of the corresponding size fitness landscapes (i.e., on the same row) with the forward size distribution. For (b) and (d) the fitness landscapes are highly non-Gaussian, and the peaks in these distributions correspond to the value of one of the plateaus.

However, the Gaussian case only covers a small portion of realistic cases, and fitness landscapes can exhibit strong deviations from Gaussian distributions. For example, in the context of age-controlled divisions [10], it can be shown that the distribution of age fitness landscape is non-Gaussian and depends on the shape of the division rate as a function of the age. Moreover, we now show in Fig. 1 (and in Fig. 6 in Appendix G) experimental fitness landscapes that are non-Gaussian.

We use real data extracted from [27], which are made of 11 population trees corresponding to the growth of E. Coli in different nutrients. We focus on the number of divisions $\mathcal{K}$, the size $\mathcal{X}$, and the age $\mathcal{A}$, which are easily accessible and

for which we previously studied the theoretical fitness landscapes [10]. Only plots for the size are presented in the main text, and similar plots for the age can be found in Appendix G.

Fitness landscapes for three particular experimental conditions are shown in Fig. 1. Each row corresponds to a particular experiment, the first column displays the fitness landscapes as functions of the size x, and the second column shows the distributions of the corresponding fitness landscapes, computed with the forward size distributions. It is straightforward to demonstrate that in the case where the number of divisions K is completely determined by the value s of the trait, then $h_t(s) = h_t[K(s)] = K(s) \ln 2/t$ [2,10]. In this case, the fitness landscape is an ensemble of plateaus corresponding to the values of K featured in the population at time t, and cells on the same plateau have undergone the same number of divisions, even though they have a different value s. Those predicted plateaus are actually observed for several experiments, as shown in Figs. 1(a) and 1(c). Indeed, we evaluate the mean number of divisions for the set of cells used to compute each point of the fitness landscape, and represent it with a color code. We also plot the theoretical plateaus, with the discrete number of divisions K corresponding to the plateau on the right y axis of each plot. We see a very good agreement between the mean number of divisions of cells aligned on a particular plateau and the value K corresponding to this plateau on the y axis. This suggests a strong correlation between the value of the size and the number of divisions on the lineage. The dots between the plateaus correspond to sizes that have been reached by cells with different numbers of divisions (leading to noninteger mean values), as highlighted by the gradation from one color to another. By going from the top experiment to the bottom one, the plateaus gradually blur and are replaced in Fig. 1(e) by a smoother curve, in good agreement with the logarithmic prediction we made in [10]. This happens when lineages desynchronize because of the cumulative effect of various noises, leading to a weaker dependence of the number of divisions K on the final value of the trait s.

In the right column of the figure, we see that fitness landscapes strongly deviate from being normally distributed, which justifies the need to go beyond the results known in the Gaussian case. More precisely, in Figs. 1(b) and 1(d), fitness landscapes exhibit peaks at values of h corresponding to one of the plateaus appearing on the left-column plot. We notice that not all three plateaus of Fig. 1(a) [two plateaus of Fig. 1(b)] are mapped to a peak in the corresponding forward fitness landscape distribution in Fig. 1(b) [Fig. 1(d)]. This is because the cell size distribution tends to zero for extreme sizes (that is 0 and $+\infty$), thus cells of large sizes aligning on plateaus defined by small K and cells of small sizes aligning on plateaus defined by large K contribute very little to the cell size distribution and thus to the fitness landscape distribution.

In this article we derive universal relations going beyond the Gaussian assumption, and obtain a set of upper and lower bounds for the strength of selection, in terms of both the forward and backward variances for the fitness landscape. To do so, let us first derive a general inequality constraining the difference in average value for an observable between two probability distributions.

III. GENERAL FLUCTUATION-RESPONSE INEQUALITY

We consider a general system described by a reference probability distribution $p_a(s, t)$, where s is the value taken by a state variable $\mathcal{S}$. By perturbing the system, we change the distribution of the variable $\mathcal{S}$ from $p_a(s, t)$ to $p_b(s, t)$. We consider an observable depending on the variable $\mathcal{S}$, through a function $g_t(s)$, and ask the question of how the mean value of this observable is modified when the system is perturbed.

Assuming that $p_a(s, t)$ and $p_b(s, t)$ have the same support, we can define the ratio

$$q_t(s) = \frac{p_b(s, t)}{p_a(s, t)}. \tag{5}$$

Let us now compute the covariance between $g_t(s)$ and $q_t(s)$ with respect to $p_a(s, t)$:

$$\text{Cov}_a(g_t, q_t) = \langle g_t q_t \rangle_a - \langle g_t \rangle_a \langle q_t \rangle_a = \langle g_t \rangle_b - \langle g_t \rangle_a, \tag{6}$$

where we used $\langle q_t \rangle_a = 1$, due to the normalization of p_b, and $\langle q_t g_t \rangle_a = \langle g_t \rangle_b$.

Following the method used in [14] to derive mean-variance trade-off bounds in horse race gambling, we use the Cauchy-Schwarz inequality for the covariance:

$$\text{Cov}_a(g_t, q_t)^2 \leqslant \sigma_a^2(g_t) \sigma_a^2(q_t), \tag{7}$$

with σ_a^2 the variance with respect to $p_a(s, t)$. Finally, by combining Eqs. (6) and (7), we obtain a general bound for the difference in average values:

$$|\langle g_t \rangle_b - \langle g_t \rangle_a| \leqslant \sigma_a(g_t) \sigma_a(q_t). \tag{8}$$

The inequality can be understood as an out-of-equilibrium generalization of the fluctuation-dissipation theorem, because it involves a comparison between a reference unperturbed dynamics and a perturbed dynamics. The difference between the unperturbed and the perturbed averages of the function $g_t(s)$ is bounded by the unperturbed fluctuations of this function, measured by $\sigma_a(g_t)$, times $\sigma_a(q_t)$ which is a information-theoretic distance between the two probability distributions. Indeed, since $\langle q_t \rangle_a = 1$, the variance of q_t is given by $\sigma_a^2(q_t) = \int_s ds\, [p_b(s) - p_a(s)]^2 / p_a(s)$, and thus the larger $\sigma_a(q_t)$, the further away $p_b(s, t)$ and $p_a(s, t)$ are from each other.

To derive Eq. (8), we adopted the point of view of the unperturbed statistics $p_a(s, t)$ as reference, but a similar bound can be obtained in terms of standard deviations with respect to the perturbed dynamics $p_b(s, t)$. We consider the covariance between $g_t(s)$ and $r_t(s) = 1/q_t(s)$, with respect to $p_b(s, t)$:

$$\text{Cov}_b(g_t, r_t) = \langle g_t \rangle_b - \langle g_t \rangle_a, \tag{9}$$

Following the same steps, and using the Cauchy-Schwarz inequality for this covariance we finally obtain

$$|\langle g_t \rangle_b - \langle g_t \rangle_a| \leqslant \sigma_b(g_t) \sigma_b(r_t), \tag{10}$$

where the term $\sigma_b(r_t)$ is similarly interpreted as an information-theoretic distance measure between the two distributions. Thus, combining Eqs. (8) and (10), the change in mean value of the function g_t of the variable $\mathcal{S}$ between an unperturbed and a perturbed statistics is bounded by

$$|\langle g_t \rangle_b - \langle g_t \rangle_a| \leqslant \min\left[\sigma_a(g_t)\sigma_a(q_t),\, \sigma_b(g_t)\sigma_b(r_t)\right]. \tag{11}$$

A similar bound for $|\langle g_t \rangle_b - \langle g_t \rangle_a|$ was derived by Dechant *et al.* in [13], using Jensen's inequality. Their bound [Eq. (5) or (11) in their text] also involves a measure of the distance between the two probability distributions (Kullback-Leibler divergence) and the standard deviation of the observable considered in the unperturbed dynamics. We carry out a numerical comparison between the two bounds in Appendix C, to find which one is the tightest of the two. This shows that the relative performance of the two bounds depends on the shape of the perturbed and unperturbed distributions. In any case, our bound is easy to evaluate since it does not require an optimization over a free parameter, as it is the case in [13] [see Eq. (C2)].

IV. THE STRENGTH OF SELECTION IS BOUNDED BY THE VARIABILITY IN FITNESS LANDSCAPE

The results derived in the previous section for general distributions a and b are now used to obtain constraints on the strength of selection. Indeed, by setting the unperturbed distribution a to be the forward distribution of a phenotypic trait $\mathcal{S}$ and the perturbed distribution b to be the backward distribution of this trait (which is allowed since the forward and backward distributions have the same support), the difference $\langle g_t \rangle_\text{back} - \langle g_t \rangle_\text{for}$ is the change of mean value for g_t between an ensemble without selection (forward) and with selection (backward), while the perturbation is measuring the selection itself. In this context, the ratio $q_t(s)$ and the fitness landscape $h_t(s)$ are linked by the simple relation $q_t(s) = \exp\{t[h_t(s) - \Lambda_t]\}$.

An important application of the above results is when the arbitrary function $g_t(s)$ is the fitness landscape $h_t(s)$ itself. In this case, Eqs. (6) and (9) read

$$\Pi_\mathcal{S} = \text{Cov}_\text{for}(h_t, e^{th_t})\, e^{-t\Lambda_t} \tag{12}$$

$$= \text{Cov}_\text{back}(h_t, e^{-th_t})\, e^{t\Lambda_t}. \tag{13}$$

These equalities generalize the linear relation [Eq. (4)] between the strength of selection and the variance of the fitness landscape, valid in the Gaussian case. To better highlight the role of the variability of the fitness landscape, we write Eq. (11) in this context:

$$\Pi_\mathcal{S} \leqslant \min\left[\sigma_\text{for}(h_t)\sigma_\text{for}(q_t),\, \sigma_\text{back}(h_t)\sigma_\text{back}(r_t)\right], \tag{14}$$

where the absolute values on the left-hand side (l.h.s.) can be removed because the strength of selection is defined positive, as deduced from Eq. (3).

Note that the l.h.s. of Eq. (14) involves averages with respect to the two probability distributions, unlike what happens in the standard TUR where only one such average is present. The reason is that in stochastic thermodynamics, the two relevant probability distributions correspond to a forward and a time-reversed dynamics, and the quantity which replaces $g_t(s)$ is a current, which changes sign under time-reversal symmetry. Here there is no such symmetry present, hence the two averages are not the opposite of one another.

We obtained a universal upper bound for the strength of selection acting on trait $\mathcal{S}$, which involves the information-theoretic distances $\sigma_\text{for}(q_t)$ and $\sigma_\text{back}(r_t)$ between the backward and forward statistics, and the variances of the fitness

landscape in both ensembles, which are in general different from each other.

Even if the interpretation of $\sigma_{\text{for}}(q_t)$ as a distance in the framework of linear-response theory is general, $\sigma_{\text{for}}(q_t)$ can also be expressed in terms of measurable quantities for cell colonies:

$$\sigma_{\text{for}}(q_t) = \frac{\sigma_{\text{for}}(e^{th_t})}{\langle e^{th_t} \rangle_{\text{for}}}. \tag{15}$$

Thus, $\sigma_{\text{for}}(q_t)$ quantifies the relative fluctuation of the quantity $\exp[th_t(s)]$, which itself represents the ratio of the expected number of lineages ending with trait value s, rescaled by the number N_0 of initial cells, to the forward probability of this trait value (see Appendix D). A similar interpretation can be given for the term $\sigma_{\text{back}}(r_t)$.

V. A LINEAR RESPONSE EQUALITY

Let us now investigate precisely the conditions for which the previous inequalities become saturated. It is straightforward to show that when the forward and backward statistics are equal, inequalities Eqs. (14) and (11) are saturated. Indeed, the l.h.s terms are 0 and the right-hand side (r.h.s.) terms are null because they contain the standard deviation of the constant quantities $q_t(s) = r_t(s) = 1$.

We now study the case where the two probability distributions approach each other. One possible measure of the distance between the two distributions is $\sigma(q_t)$, or equivalently $\sigma(\ln q_t) = t\sigma(h_t)$. In the limit $t\sigma(h_t) \to 0$, referred to as the small variability limit, the l.h.s. of Eq. (11) reads (see Appendix E)

$$\langle g_t \rangle_{\text{back}} - \langle g_t \rangle_{\text{for}} \underset{t\sigma \to 0}{\sim} t \, \text{Cov}(h_t, g_t), \tag{16}$$

and the l.h.s. of Eq. (14) when the function g_t is the fitness landscape itself reads

$$\Pi_{\mathcal{S}} \underset{t\sigma \to 0}{\sim} t \, \text{Var}(h_t), \tag{17}$$

where the variance and the covariance can be equivalently taken over the forward or backward sampling. When computing the r.h.s. of Eqs. (14) and (11), we obtain that Eq. (14) is saturated in this limit, whereas Eq. (11) is not. The limit can also be written $t \ll \sigma(h_t)^{-1}$ which defines a characteristic timescale of the system. In practice, this limit can be reached either for short times or in the case of a strong control mechanism on the divisions, leading the lineages to stay synchronized even after a finite time. It is also possible to regard this limit as a regime of weak selection [22], since the strength of selection is small precisely because of Eq. (14).

VI. COMPARISON WITH FISHER'S FUNDAMENTAL THEOREM AND PRICE'S EQUATION

In this section we highlight the similarities between our results and the relations derived by Fisher and Price, in which the population growth rate, or fitness, associated with a trait value s plays a similar role to our fitness landscape $h_t(s)$. However, because these notions of fitness are distinct, as explained in Sec. II and further analyzed in Appendix A, the interpretations of selection contained in these equations are qualitatively different.

Fisher's fundamental theorem of natural selection states that the time derivative of the mean fitness of a population is equal to the variance of the fitness across the population [23,24]: $d\Lambda_t/dt = \text{Var}_{\text{back}}[\Lambda_i(s, t)]$, where $\Lambda_i(s, t) = [dn(s, t)/dt]/n(s, t)$ is the instantaneous growth rate, or instantaneous fitness, of the subpopulation of size $n(s, t)$ carrying the trait value s. The variance is computed with respect to the backward distribution, which puts equal weights on individuals, and therefore is the natural distribution to consider. The r.h.s. of both Eqs. (17) and (4) and Fisher's theorem involve the variance of a certain kind of fitness within the population. In contrast, the l.h.s. in Fisher's theorem is a measure of evolution of the population, while the l.h.s. in our result is a trait-dependent measure of selection. Moreover, some well-known limitations of Fisher's theorem lie in the implicit assumption that natural selection is the only possible phenomenon leading to a change in the gene frequencies [24]. This assumption neglects many important phenomena such as mutations and recombination events [22], random drift due to finite population size, and specific features of seascapes [7]. In contrast, our result does not suffer from any of these limitations, since it only requires the population to be represented as a branched tree, and is completely independent of the dynamics that generates the tree.

Price's equation [24] predicts the time evolution of the mean value of a trait, and involves two terms: a covariance term representing the selection effect, and the "environment change term", or dynamic effect, which accounts for all the other sources of variability leading to a change in the mean value of the trait. The part of the evolution of $\langle s \rangle_{\text{back}}$ over a time τ that is due to the separate effect of natural selection, in Price's sense which we denote with the superscript NS, can then be written as

$$\Delta \langle s \rangle_{\text{back}}^{\text{NS}} = \frac{1}{\Lambda} \, \text{Cov}_{\text{back}}[\text{s}, \Lambda(\text{s})], \tag{18}$$

where $\Delta \langle s \rangle_{\text{back}}^{\text{NS}} = [\langle s \rangle_{\text{back}}(t + \tau) - \langle s \rangle_{\text{back}}(t)]^{\text{NS}}$, $\Lambda(s) = n(s, t + \tau)/n(s, t)$, and $\Lambda = \langle \Lambda(s) \rangle_{\text{back}}$. We can draw a parallel between this equation and Eq. (16), as their r.h.s. both involve the covariance of the trait subjected to selection and a fitness associated with it. Note that there is no environment change term in Eq. (16) because the strength of selection is defined precisely in such a way as to isolate the effect of selection from other potential sources of variability.

Price's equation should be viewed as a way to separate the effect of selection from the effect of the environment, rather than as a predictive or quantitative formula to compute them, as remarked in [25]. The same can be said of all our results, where the strength of selection and the covariance between a trait value and its associated fitness landscape value can be viewed as two possible definitions of selection. These two notions of natural selection are different because of the distinction between growth rate and fitness landscape: one is concerned by the change in the frequencies of the trait values over time, and is computed by counting individuals, while the other one represents the shift in the frequencies of the trait values at snapshot time t between situations with and without selection, and is based on the comparison between chronological and retrospective samplings of the lineages.
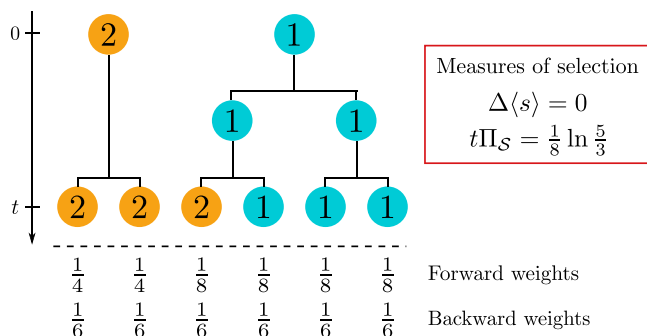
FIG. 2. Population tree where cells can have only two values $s = 1$ and $s = 2$ for the phenotypic trait $\mathcal{S}$. Cells with phenotype 1 divide more often than cells with phenotype 2, and phenotypic switching from 1 to 2 occurs randomly at division. These two phenomena balance so that the frequencies of $s = 1$ and $s = 2$ are the same at $t = 0$ and at t. However, the fitness landscapes for $s = 1$ and $s = 2$ are different, and are computed using the backward and forward weights of each lineage, leading to $t\Pi_{\mathcal{S}} = \ln(5/3)/8$. The strength of selection is nonzero while the measure of selection proposed by Price, namely $\Delta\langle s \rangle$, is null.

Let us give a minimal example for which the strength of selection is nonzero while the mean value of the trait $\mathcal{S}$ is unchanged, because of the balance between heterogeneity in reproductive success and phenotypic switching at division. This case is illustrated in Fig. 2 for a trait $\mathcal{S}$ taking only two values: $s = 1$ and $s = 2$. Individuals with trait value 1 reproduce typically twice as fast as those carrying trait value 2, but they can also switch to trait value 2 randomly at division. For simplicity, the values 1 and 2 of the trait cannot change themselves over time, in other words there is no environment effect here. The average value of trait $\mathcal{S}$ is the same at time $t = 0$ and at time t, the covariance term in Price's equation is zero, and there is no selection in Price's sense. Therefore, from this point of view, there is no difference between this situation and the situation where both values 1 and 2 reproduce at the same rate, without phenotypic switching at division. One the other hand, individuals with trait value 1 are over-represented in the backward statistics as compared to the forward statistics, while the opposite is true for trait value 2, meaning that the fitness landscapes for $s = 1$ and $s = 2$ are different. Indeed, $p_{\text{back}}(s = 1, t) = p_{\text{back}}(s = 2, t) = 1/2$, $p_{\text{for}}(s = 1, t) = 3/8$, $p_{\text{for}}(s = 2, t) = 5/8$, which leads to $t h_t(s = 1) = \ln(4/3) + t\Lambda_t$ and $t h_t(s = 2) = \ln(4/5) + t\Lambda_t$, with $t\Lambda_t = \ln 3$, using Eq. (1). This difference in fitness landscape results in a nonzero strength of selection $t\Pi_{\mathcal{S}} = \ln(5/3)/8$ using Eq. (2). We argue that the strength of selection $\Pi_{\mathcal{S}}$ may be a more appropriate way to define selection, since it gives a nonzero measure of selection for the example discussed above, and thus is more representative of the selection occurring in the population.

VII. A LOWER BOUND FOR THE STRENGTH OF SELECTION

We showed how the equality between the strength of selection and the variance of the fitness landscape distribution, which holds in the Gaussian case, becomes an inequality in

general. To complement the upper bound on the strength of selection given by Eq. (14), we now derive a nontrivial lower bound, which presents an interest to quantify the minimal effect of selection on a particular trait.

Using a property of the Jeffrey's divergence, the strength of selection can be decomposed as a sum of two Kullback-Leibler (KL) divergences: $\mathcal{J}(p_{\text{back}}|p_{\text{for}}) = \mathcal{D}_{\text{KL}}(p_{\text{back}}|p_{\text{for}}) + \mathcal{D}_{\text{KL}}(p_{\text{for}}|p_{\text{back}})$, where $\mathcal{D}_{\text{KL}}(p|q) = \int p(x)\ln[p(x)/q(x)]dx$. The positivity of both KL divergences, ensured by Jensen's inequality, gives $\Lambda_t - \langle h_t(s)\rangle_{\text{for}} \geqslant 0$ and $\langle h_t(s)\rangle_{\text{back}} - \Lambda_t \geqslant 0$. By combining these two inequalities, we recover that the strength of selection is a positive quantity.

We can therefore improve the trivial bound on the strength of selection, which is 0, by improving the two inequalities separately, using a sharpened version of Jensen's inequality, derived in [26]. Let us now detail how this works in our problem.

We define the convex functions $\varphi_{\text{for}}(x) = e^{tx}$, $\varphi_{\text{back}}(x) = e^{-tx}$ and the function

$$\psi(\varphi, x, \nu) = \frac{\varphi(x) - \varphi(\nu)}{(x - \nu)^2} - \frac{\varphi'(\nu)}{x - \nu}, \qquad (19)$$

where φ' stands for the derivative of φ. The sharpened version of Jensen's inequality reads

$$\langle e^{th_t}\rangle_{\text{for}} - e^{t\langle h_t\rangle_{\text{for}}} \geqslant \sigma^2_{\text{for}}(h_t) \inf_h \psi(\varphi_{\text{for}}, h, \langle h_t\rangle_{\text{for}}), \qquad (20)$$

and is used to improve upon the inequality $\Lambda_t - \langle h_t(s)\rangle_{\text{for}} \geqslant 0$. A similar improvement is obtained for $\langle h_t(s)\rangle_{\text{back}} - \Lambda_t \geqslant 0$ by considering φ_{back} instead of φ_{for}. Combining the two results gives (see Appendix F)

$$\Pi_{\mathcal{S}} \geqslant \frac{1}{t}\left[\frac{\sigma^2_{\text{for}}(h_t)}{\exp(t\Lambda_t)}\psi(\varphi_{\text{for}}, h_{\text{min}}, \langle h_t\rangle_{\text{for}}) \right.$$
$$\left. + \frac{\sigma^2_{\text{back}}(h_t)}{\exp(-t\Lambda_t)}\psi(\varphi_{\text{back}}, h_{\text{max}}, \langle h_t\rangle_{\text{back}})\right], \qquad (21)$$

which shows that the lower bound depends on the forward and backward variances of the fitness landscape, as well as on its average values and on the minimal (maximal) values of these distributions denoted h_{min} (h_{max}). When the fitness landscape is a monotonic function of the value of the trait, which is the case for cell age and size [10], or for the number of divisions, these extreme values are given by the extreme values of the trait itself.

Several weaker but simpler forms of this inequality, this time independent of the average fitness landscape values, or independent of the extreme values, or independent of the two, are derived in Appendix F. In any of those cases, the lower bound is a linear combination of the forward and backward variances of the fitness landscape.

VIII. TESTS OF THE LINEAR RESPONSE RELATIONS

We now illustrate the various bounds for growing cell populations, using both simulations and time-lapse video-microscopy experimental data [27].

First, we test Eq. (8) for the number of divisions $\mathcal{K}$, and for the linear function $g_t(K) = K$, so that the inequality bounds $\langle K \rangle_{\text{back}} - \langle K \rangle_{\text{for}}$. We simulate lineage trees starting from one cell, for a particular agent-based model in which cells are
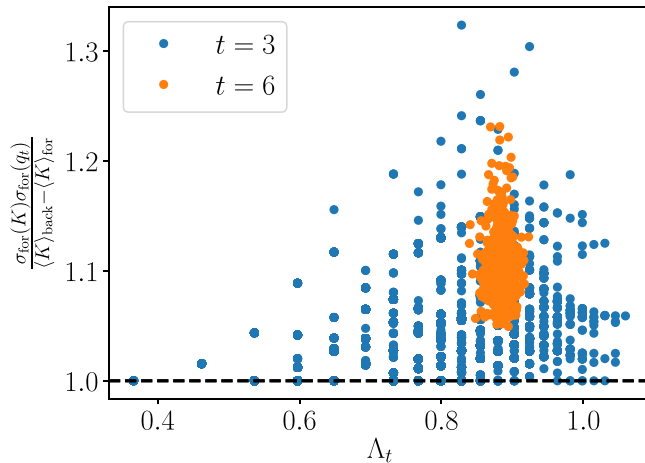
FIG. 3. Points of $\sigma_{for}(K)\sigma_{for}(q_t)/(\langle K \rangle_{back} - \langle K \rangle_{for})$ against Λ_t for many tree simulations using a size-controlled model. Each dot corresponds to a single tree, the two sets of data have the same parameters except for the final times of the simulation, which are $t = 3$ (blue) and $t = 6$ (orange). The black horizontal dashed line at $y = 1$ represents the point where the inequality of Eq. (8) is saturated.

described by their sizes. Cell sizes continuously increase at constant rate between divisions, and cells divide after a stochastic time only depending on their sizes. Each simulation of such a tree yields a single point on the scatter plot of Fig. 3, which shows the ratio of $\sigma_{for}(K)\sigma_{for}(q_t)$ to $\langle K \rangle_{back} - \langle K \rangle_{for}$ versus the population growth rate Λ_t. Two sets of points are presented, which only differ in the final time of the simulation. As expected from Eq. (8), all points in both sets are above 1. When the duration of the simulation is small ($t = 3$), the final population is small, around $N \sim 20$, therefore for a given tree the lineages do not have time to differentiate significantly and the variability in the number of divisions among the lineages is small. In that case, simulations points are approaching the horizontal dashed line at $y = 1$ corresponding to the saturation of the inequality. The final population N fluctuates significantly from one simulation to the next, because the simulation time is short and all simulations start with a single cell with random initial size. As a result, the dispersion of values of Λ_t is large.

Now, when doubling the duration of the simulation, the cloud of scattered points is considerably reduced in both directions. The horizontal dispersion reduces because as t increases, the state of the system at the final time becomes less and less affected by the initial condition. On the vertical axis there is a gap between the lower part of the scatter plot and the horizontal line at $y = 1$ due to the increase of heterogeneity in the number of divisions in the lineages with the simulation time.

Second, we test the upper and lower bounds on the strength of selection acting on cell size using experimental data from [27]. We show in Fig. 4 the upper bound $\mathcal{U}_{\mathcal{X}}$ given by Eq. (14) and the lower bound $\mathcal{L}_{\mathcal{X}}$ given by Eq. (21), normalized by the strength of selection $\Pi_{\mathcal{X}}$. The x axis labels in no particular order the colonies which have grown in different nutrient medium [27]. As expected, points representing the upper bound and those representing the lower bound are, respectively, above and below the horizontal dashed line at
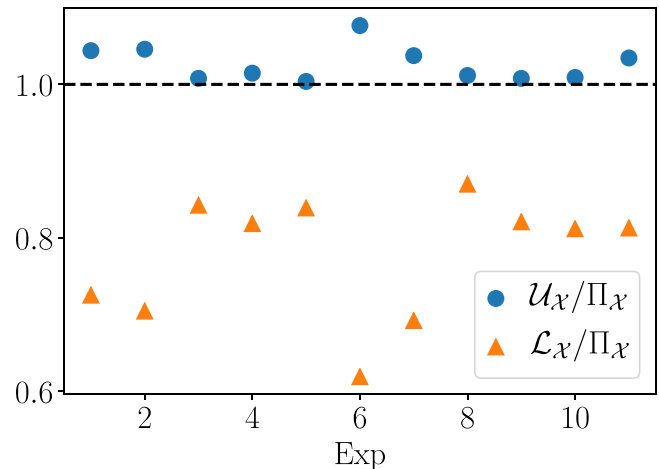

FIG. 4. Upper bound $\mathcal{U}_{\mathcal{X}}$ (blue dots) and lower bound $\mathcal{L}_{\mathcal{X}}$ (orange triangles) for the strength of selection acting on size $\Pi_{\mathcal{X}}$, normalized by the latter. The x axis represents the 11 colonies in different growth conditions from [27], in no particular order.

$y = 1$. Experiments for which the normalized upper bound approaches 1 indicate that cell cycles are almost synchronized and thus that there is small variability in terms of number of divisions among the lineages.

Note that Nozoe *et al.* also proved [2] that the strength of selection for the division bounds the strength of selection acting on any trait: $0 \leqslant \Pi_{\mathcal{S}} \leqslant \Pi_{\mathcal{K}}$. This bound is typically not as tight as Eq. (14) (see Appendix G for comparison). To improve upon it, one can use $\mathcal{S} = \mathcal{K}$ in Eq. (14) to obtain a bound for $\Pi_{\mathcal{K}}$ itself.

IX. DISCUSSION

The general idea of comparing the response of a system in the presence of a perturbation to its fluctuations in the absence of the perturbation lies at the heart of the fluctuation-dissipation theorem, which has a long history in physics, with some applications to evolution [1,28]. Remarkably, the present framework with forward (unperturbed) and backward (perturbed) dynamics can be conveniently applied to population dynamics without having to perform additional experiments, since both probabilities can be calculated with the same lineage tree. Our main result is a set of inequalities for the average of an arbitrary function of a trait or for its fitness landscape, valid beyond the Gaussian assumption, and which constrain the strength of selection in population dynamics even in the presence of time-dependent selection pressures.

These inequalities are universal because they only rely on the branching structure of the population tree and are completely independent of the dynamics of the tree, which is the ensemble of rules governing the division of the branches. In the context of cell populations, this means for instance that our results are valid for any control strategy (sizer, timer, adder, etc.) and in the presence of any source of noise (variability in single cell growth rate or size at division, asymmetry in resource partitioning between sister cells, etc.) [5], in the presence of possible mutations, and regardless of the nature of the cell (bacteria, yeast, stem cell, etc.). Although we illustrated our results with cell populations, they apply to

any stochastic process defined on a branched tree. In particular, they could be insightful in the context of ecology, where such trees are used to represent phylogeny [29]. In this case, each lineage could represent a species or a version of a gene, instead of an individual, and the divisions would represent speciations or mutations, respectively. The notions of fitness landscape and selection strength appear meaningful in this setting, as a quantification of the correlations between a feature of a species/genetic information and the number of speciations/mutations its phylogenetic lineage underwent.

For applications, we focused on phenotypic traits, like cell size or age, and in these cases, we found our upper bound on the strength of selection to be tight. In future work, it would be interesting to apply this framework to genotypic traits instead of phenotypic ones [22] and possibly exploit recent methods of lineage tracking [30]. This could open new perspectives to address a number of important problems like antibiotic resistance [4], the differentiation of stem cells [31], or virus evolution.

The search for universal principles in evolution is an active field of research [28,32]. An important step in this endeavor was made by Fisher, who boldly compared his theorem to the second law of thermodynamics. While the theorem turned out not to be as general as expected, Fisher had nevertheless the correct intuition about its importance for evolutionary biology, and he was also correct in expecting that such a general principle should be related to thermodynamics.

APPENDIX A: PHENOTYPIC FITNESS LANDSCAPE AND GROWTH RATE ASSOCIATED WITH THE VALUE s OF A TRAIT $\mathcal{S}$

The fitness landscape $h_t(s)$ of trait $\mathcal{S}$ defined in Eq. (1) should not be confused with the fitness associated with the value s of the trait, which can be identified with the growth rate of the subpopulation carrying that trait. However, they are linked by a simple relation. In line with the definition of the population growth rate, we define the growth rate of the subpopulation carrying the value s of the trait $\mathcal{S}$ as

$$\Lambda_t(s) = \frac{1}{t} \ln\left[\frac{n(s,t)}{n(s,0)}\right] = \Lambda_t + \frac{1}{t} \ln\left[\frac{p_{\text{back}}(s,t)}{p_{\text{back}}(s,0)}\right], \quad \text{(A1)}$$

where $n(s,t) = p_{\text{back}}(s,t)N(t)$ is the number of cells with trait value s at time t. Of course, defining this quantity only makes sense between two times $t = 0$ and t for which the value s is present in the population, otherwise the expression in undefined.

By comparing Eqs. (1) and (A1), we can link the growth rate associated with a trait value s to the fitness landscape of this trait value:

$$h_t(s) - \Lambda_t(s) = \frac{1}{t} \ln\left[\frac{p_{\text{back}}(s,0)}{p_{\text{for}}(s,t)}\right] = \frac{1}{t} \ln\left[\frac{p_{\text{for}}(s,0)}{p_{\text{for}}(s,t)}\right]. \quad \text{(A2)}$$

The last equality follows from the fact that, at $t = 0$, the cells have not divided yet and so the forward and backward samplings of the population are identical.

Finally, we see that the three quantities Λ_t, $\Lambda_t(s)$, and $h_t(s)$ are different but intimately linked. In the long time limit, for which equilibrium distributions do not depend on time anymore, they all become equal to the steady-state population growth rate Λ,

$$\lim_{t\to\infty} \Lambda_t = \lim_{t\to\infty} \Lambda_t(s) = \lim_{t\to\infty} h_t(s) = \Lambda. \quad \text{(A3)}$$

Moreover, as we already noted, Eq. (1) indicates that the sign of $h_t(s) - \Lambda_t$ informs us of the comparison between the backward and forward probabilities of trait value s in the population at time t, or in other words, if the trait value s is over-represented in the population as compared to a situation without selection. Thus, $h_t(s) - \Lambda_t$ quantifies the separate effect of selection.

Following the intuitive understanding of the growth rate associated with a trait value s, Eq. (A1) means that the sign of $\Lambda_t(s) - \Lambda_t$ informs us of the comparison between the backward statistics of the trait value s at time t and at time 0. A trait value is favored by the population dynamics, which includes all phenomena leading to changes in trait value frequencies, if its growth rate is larger than the population growth rate, which corresponds to an increase of the frequency of that trait value in the population as time grows.

The last relation, Eq. (A2), provides a new insight: the sign of $h_t(s) - \Lambda_t(s)$ is linked to the comparison between the forward probability of trait value s at time 0 and time t. The forward statistics is constructed to balance the effect of selection occurring in tree-structured data. However, it is affected by all the other sources of variability, as for example mutations. Therefore, the sign of $h_t(s) - \Lambda_t(s)$ indicates the evolution of the frequency of the value s of trait $\mathcal{S}$ as time grows, due to every phenomenon but selection.

Equation (1) can be written in the form of a fluctuation relation [2,10], with an exponential bias between the forward and backward probabilities, similar to Crooks fluctuation relation for work in stochastic thermodynamics. The same can be done with Eqs. (A1) and (A2), however, unlike Eq. (1), these new fluctuation relations both link two probability distributions that may not have the same support.

APPENDIX B: GAUSSIAN CASE

We show in this Appendix that a linear relation between the strength of selection and the variance of the fitness landscape holds in the case where the fitness landscape is normally distributed. To do so, let us first derive a very useful result: by isolating $p_{\text{back}}(s,t)$ in Eq. (1), and integrating over s using the normalization of p_{back}, the population growth rate is expressed as a forward average

$$e^{t\Lambda_t} = \langle e^{th_t} \rangle_{\text{for}}. \quad \text{(B1)}$$

We can do the same but isolating $p_{\text{for}}(s,t)$ this time, leading to

$$e^{-t\Lambda_t} = \langle e^{-th_t} \rangle_{\text{back}}. \quad \text{(B2)}$$

We now assume that $h_t(s)$ can be accounted for by a continuous probability distribution, even though the trait $\mathcal{S}$ may

be discrete, as in the case for the number of divisions. We set a Gaussian forward distribution with mean $\langle h_t \rangle_{\text{for}}$ and variance $\sigma_{\text{for}}(h)^2$ for the fitness landscape $h_t(s)$, then $\exp[th_t(s)]$ follows a log-normal distribution of mean

$$\langle e^{th_t} \rangle_{\text{for}} = e^{t\langle h_t \rangle_{\text{for}} + [t\sigma_{\text{for}}(h_t)]^2/2}. \tag{B3}$$

This relation shows that for a given forward average fitness landscape, the growth rate is positively affected by the variability between the lineages.

The backward average of the fitness landscape is given by the forward average of a biased fitness landscape:

$$\langle h_t \rangle_{\text{back}} = e^{-t\Lambda_t} \int h_t(s) e^{th_t(s)} p_{\text{for}}(s,t) ds. \tag{B4}$$

We make the hypothesis that the fitness landscape is a bijective function of the trait value and use the conservation of the probability: $p_{\text{for}}(s,t)ds = p_{\text{for}}(h)dh$, leading to a solvable Gaussian integral

$$\langle h_t \rangle_{\text{back}} = \frac{e^{-t\Lambda_t}}{\sqrt{2\pi\sigma_{\text{for}}(h_t)^2}} \int h e^{th} e^{-(h - \langle h_t \rangle_{\text{for}})^2/[2\sigma_{\text{for}}(h_t)^2]} dh$$

$$= e^{-t\Lambda_t} [\langle h_t \rangle_{\text{for}} + t\sigma_{\text{for}}(h_t)^2] e^{t\langle h_t \rangle_{\text{for}} + [t\sigma_{\text{for}}(h_t)]^2/2}. \tag{B5}$$

Finally, combining Eqs. (B1), (B3), and (B5), we obtain

$$\langle h_t \rangle_{\text{back}} = \langle h_t \rangle_{\text{for}} + t\sigma_{\text{for}}(h_t)^2, \tag{B6}$$

and thus

$$\Pi_{\mathcal{S}} = t\sigma_{\text{for}}(h_t)^2. \tag{B7}$$

Moreover, combining Eqs. (B3) and (B7) we deduce that $\langle h_t \rangle_{\text{for}}$ and $\langle h_t \rangle_{\text{back}}$ are not only respectively smaller and greater than Λ_t, as discussed in Sec. VII, but they are actually symmetrical around this value $\langle h_t \rangle_{\text{back}} - \Lambda_t = \Lambda_t - \langle h_t \rangle_{\text{for}} = t\,\sigma_{\text{for}}^2(h_t)/2$. In other words, in this particular case, the KL divergence is symmetrical: $\mathcal{D}_{\text{KL}}(p_{\text{for}}|p_{\text{back}}) = \mathcal{D}_{\text{KL}}(p_{\text{back}}|p_{\text{for}})$.

In the case where $h_t(s)$ follows a Gaussian distribution in the forward statistics, it also follows a Gaussian distribution in the backward statistics because the bias of the fluctuation relation between p_{back} and p_{for} is exponential in h_t. Since h_t follows a Gaussian distribution of mean $\langle h_t \rangle_{\text{back}}$ and standard deviation $\sigma_{\text{back}}(h_t)$ in the backward statistics, then $\exp[-th_t(s)]$ follows a log-normal distribution of mean

$$\langle e^{-th_t} \rangle_{\text{back}} = e^{-t\langle h_t \rangle_{\text{back}} + [t\sigma_{\text{back}}(h_t)]^2/2}. \tag{B8}$$

We now take the inverse of this formula and use Eqs. (B2) and (B6) to replace the backward average:

$$e^{t\Lambda_t} = e^{t[\langle h_t \rangle_{\text{for}} + t\sigma_{\text{for}}(h_t)^2] - [t\sigma_{\text{back}}(h_t)]^2/2}. \tag{B9}$$

By comparing Eqs. (B3) and (B9), it follows that $\sigma_{\text{back}}(h_t) = \sigma_{\text{for}}(h_t)$. Finally, the standard deviation in Eq. (B7) can be taken indifferently with respect to both statistics and we omit the index to write the general version of Eq. (B7):

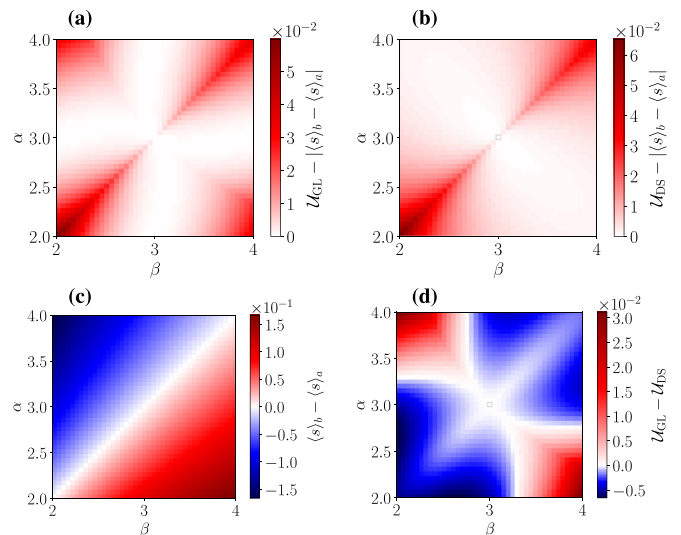$$\Pi_{\mathcal{S}} = t\,\text{Var}(h_t). \tag{B10}$$

FIG. 5. Comparison between the upper bounds $\mathcal{U}_{\text{GL}}$ [Eq. (C1)] and $\mathcal{U}_{\text{DS}}$ derived in [13] [Eq. (C2)], for beta distributions $p_a(s) = f(s, \alpha, \beta)$ and $p_b(s) = f(s, 3, 3)$. Parameters α and β are varied between 2 and 4. First row: Difference between the upper bounds and $|\langle s \rangle_b - \langle s \rangle_a|$, (a) for our bound, and (b) for Dechant-Sasa's bound, showing that all points are indeed above 0. (c) Exact difference $\langle s \rangle_b - \langle s \rangle_a$, in agreement with the theoretical value $\langle s \rangle_b - \langle s \rangle_a = 1/2 - \alpha/(\alpha + \beta)$. (d) Comparison between $\mathcal{U}_{\text{GL}}$ and $\mathcal{U}_{\text{DS}}$, blue regions indicate where our bound is tighter, i.e., smaller, and red regions indicate where Dechant-Sasa's bound is tighter. For all four plots, the grid is 41×41, and numerical values are rounded to 10^{-15} to avoid python floats precision errors.

APPENDIX C: UPPER BOUNDS NUMERICAL COMPARISON

In this Appendix we compare numerically the upper bound $\mathcal{U}_{\text{GL}}$ obtained in Eq. (11):

$$\mathcal{U}_{\text{GL}} = \min[\sigma_a(g_t)\sigma_a(q_t), \sigma_b(g_t)\sigma_b(r_t)], \tag{C1}$$

to the upper bound $\mathcal{U}_{\text{DS}}$ derived by Dechant and Sasa (Eq. (5) in [13]):

$$\mathcal{U}_{\text{DS}} = \inf_{\gamma > 0}[K^a_{g_t}(\gamma\sigma) - \gamma\sigma\langle g_t \rangle_a + \mathcal{D}_{\text{KL}}(p_b|p_a)], \tag{C2}$$

where $\sigma = \text{sgn}(\langle g_t \rangle_b - \langle g_t \rangle_a)$ and $K^a_{g_t}(\gamma) = \ln\langle\exp(\gamma g_t)\rangle_a$ the cumulant-generating function of g_t. Both quantities $\mathcal{U}_{\text{GL}}$ and $\mathcal{U}_{\text{DS}}$ bound the difference $|\langle g_t \rangle_b - \langle g_t \rangle_a|$ between the average values of a function g_t of a variable $\mathcal{S}$ with respect to probability distributions p_a and p_b.

To compare them, we took beta distributions for p_a and p_b, having the same support [0,1] so that both bounds are defined. Beta-distributed variables admit a probability density function (pdf) of the form $f(s, \alpha, \beta) = B(\alpha, \beta)s^{\alpha-1}(1-s)^{\beta-1}$, where $B(\alpha, \beta)$ is a normalization constant, and their mean is given by $\langle s \rangle = \alpha/(\alpha + \beta)$. We fix the pdf in the ensemble b to $p_b(s) = f(s, 3, 3)$, whose bell shape is reminiscent of a Gaussian distribution, on a finite interval. The pdf in the ensemble a is taken as $p_a(s) = f(s, \alpha, \beta)$, where α and β are varied in [2,4]. We choose the simple function $g_t(s) = s$.

Results are shown in Fig. 5. The first row of figures shows the difference between the upper bound and the actual

difference $|\langle s \rangle_b - \langle s \rangle_a|$, for our bound $\mathcal{U}_{\text{GL}}$ [Fig. 5(a)], and for Dechant-Sasa's bound $\mathcal{U}_{\text{DS}}$ [Fig. 5(b)]. As expected, all points on these two plots are positive. We plot in Fig. 5(c) the real difference $\langle s \rangle_b - \langle s \rangle_a$, which is in complete agreement with the theory: $\langle s \rangle_b - \langle s \rangle_a = 1/2 - \alpha/(\alpha + \beta)$. Finally, Fig. 5(d) shows a comparison between our bound and Dechant-Sasa's bound: blue regions represent sets of parameters (α, β) where our bound is numerically tighter, and the opposite is true in red regions. We note that, if the blue region is larger than the red region, on the other hand the advantage of one bound over the other $|\mathcal{U}_{\text{GL}} - \mathcal{U}_{\text{DS}}|$, is generally larger in the red region. Therefore, the answer to the question 'Which bound is tighter?' depends on the actual distributions $p_a(s)$ and $p_b(s)$. However, we note that our bound is easier to compute since it does not require the optimization over an external parameter, which is the case for $\mathcal{U}_{\text{DS}}$ in [13] [parameter γ in Eq. (C2)]. Note that this optimization can be bypassed by choosing a specific value γ in Eq. (C2), but then the corresponding bound is less tight than the version with the infimum.

APPENDIX D: INFORMATION-THEORETIC-DISTANCE $\sigma(q)$ BETWEEN PERTURBED AND UNPERTURBED DYNAMICS IN TERMS OF MEASURABLE QUANTITIES

We show in this Appendix how the information-theoretic-distance $\sigma_{\text{for}}(q_t)$ is linked to measurable quantities in cell colonies, for a general trait $\mathcal{S}$, and for the particular case of age models. Combining Eqs. (1) and (5) gives the ratio $q_t(s) = \exp[h_t(s) - \Lambda_t]$, which combined with Eq. (B1) leads to Eq. (15) in the main text. Thus, $\sigma_{\text{for}}(q_t)$ is the coefficient of variation in the forward statistics of the quantity

$$e^{th_t(s)} = e^{t\Lambda_t} \frac{p_{\text{back}}(s, t)}{p_{\text{for}}(s, t)} = \frac{N(t)}{N_0} \frac{p_{\text{back}}(s, t)}{p_{\text{for}}(s, t)}$$
$$= \frac{n(s, t)}{N_0} \frac{1}{p_{\text{for}}(s, t)}, \tag{D1}$$

where $n(s, t) = N(t)p_{\text{back}}(s, t)$ is the number of cells with the value s of trait $\mathcal{S}$ at time t.

In the case where $\mathcal{S}$ is the number of divisions, the fitness landscape is called the lineage fitness and is given by $h_t(K) = K \ln m/t$. Therefore, $\sigma_{\text{for}}(q_t) = \sigma_{\text{for}}(m^K)/\langle m^K \rangle_{\text{for}}$ is the relative fluctuation of quantity m^K, representing the expected number of lineages that underwent K divisions, normalized by the initial population N_0, divided by the forward probability of K.

For age models, where the division is only controlled by the age of the cell, we know a fluctuation relation linking the forward and backward distributions of generation times τ, defined as the time between two consecutive divisions on the same lineage [33]

$$f_{\text{back}}(\tau) = m f_{\text{for}}(\tau) \exp[-\Lambda \tau]. \tag{D2}$$

This relation can be understood as a version of the fluctuation relation on the number of divisions [2,10] at the scale of the cell cycle. However, let us note two differences: first, unlike the one for the number of divisions, Eq. (D2) is only true in the long time limit, when the population is growing at a constant steady state growth rate Λ; and second, the distributions f_{back} and f_{for} are not snapshot distributions at time t, but distributions of generation times computed along the weighted lineages.

We define $q(\tau) = f_{\text{back}}(\tau)/f_{\text{for}}(\tau)$ in the same way, and following the same steps for a general function $g(\tau)$ we derive

$$|\langle g(\tau) \rangle_{\text{back}} - \langle g(\tau) \rangle_{\text{for}}|$$
$$\leqslant \min\{\sigma_{\text{for}}[g(\tau)]\sigma_{\text{for}}[q(\tau)], \sigma_{\text{back}}[g(\tau)]\sigma_{\text{back}}[r(\tau)]\}. \tag{D3}$$

From Eq. (D2) we express the information-theoretic-distance term as

$$\sigma_{\text{for}}[q(\tau)] = \frac{\sigma_{\text{for}}(e^{-\Lambda \tau})}{\langle e^{-\Lambda \tau} \rangle_{\text{for}}}, \tag{D4}$$

which is the relative fluctuation in the forward sampling of the quantity $\exp[-\Lambda \tau] = f_{\text{back}}(\tau)/m f_{\text{for}}(\tau)$. We know from [34] that the backward distribution is also the generation time distribution of the direct ancestor cells. Therefore, $\exp[-\Lambda \tau]$ represents the ratio of the probability for the ancestor cell to divide at age τ to the expected number of daughter cells born from that division that divide at age τ.

APPENDIX E: SMALL VARIABILITY LIMIT

In this Appendix we study the two sides of the fluctuation-response inequality for an arbitrary function g_t of a phenotypic trait $\mathcal{S}$ [Eq. (11)] in the limit where the forward and backward distributions approach each other, and show that they are mathematically equivalent in this limit in the case where the function g_t is the fitness landscape h_t. The difference between the two distributions is captured by the "distance" term $\sigma[q_t(s)]$, or equivalently $\sigma[\ln q_t(s)] = \sigma\{\ln[p_{\text{back}}(s, t)/p_{\text{for}}(s, t)]\} = t\sigma[h_t(s)]$, where the standard deviations can be taken either in the backward or forward statistics. From now on, we refer to the limit where the forward and backward distributions are close to each other as the small variability limit, defined by $t\sigma[h_t(s)] \to 0$.

We first use this limit in the forward statistics: $t\sigma_{\text{for}}[h_t(s)] \to 0$. Starting from Eq. (1) we isolate $p_{\text{back}}(s, t)$, multiply both sides by $g_t(s)$ and integrate over s, leading to the expression of the backward average of function g_t as the forward average of a biased version of the same function:

$$\langle g_t \rangle_{\text{back}} = e^{t(\langle h_t \rangle_{\text{for}} - \Lambda_t)} \int g_t(s) e^{t[h_t(s) - \langle h_t \rangle_{\text{for}}]} p_{\text{for}}(s, t) ds. \tag{E1}$$

In order to expand the exponential, we assume that for any s, $t[h_t(s) - \langle h_t \rangle_{\text{for}}]$ is small, which corresponds to $t\sigma_{\text{for}}(h_t)$ small because $\sigma_{\text{for}}(h_t)$ is the characteristic distance to the mean. Therefore,

$$\langle g_t \rangle_{\text{back}} \underset{t\sigma \to 0}{\sim} e^{t(\langle h_t \rangle_{\text{for}} - \Lambda_t)} \int g_t(s)\{1 + t[h_t(s) - \langle h_t \rangle_{\text{for}}]\} p_{\text{for}}(s, t) ds \sim e^{t(\langle h_t \rangle_{\text{for}} - \Lambda_t)}[\langle g_t \rangle_{\text{for}} + t(\langle g_t h_t \rangle_{\text{for}} - \langle g_t \rangle_{\text{for}} \langle h_t \rangle_{\text{for}})], \tag{E2}$$

where $\langle g_t h_t \rangle_{\text{for}} - \langle g_t \rangle_{\text{for}} \langle h_t \rangle_{\text{for}} = \text{Cov}_{\text{for}}(h_t, g_t)$ is the covariance of h_t and g_t with respect to the forward probability. The term in the bracket is a first-order correction to $\langle g_t \rangle_{\text{for}}$ in $t\sigma_{\text{for}}(h_t)$. Now we need to compute the prefactor $\exp[-t\Lambda_t]$, starting with Eq. (B1) and using the same expansion

$$
e^{t\Lambda_t} = \langle e^{th_t} \rangle_{\text{for}} = \int e^{th_t(s)} p_{\text{for}}(s,t)ds \underset{t\sigma \to 0}{\sim} e^{t\langle h_t \rangle_{\text{for}}} \int \left[1 + t[h_t(s) - \langle h_t \rangle_{\text{for}}] + \frac{t^2}{2}[h_t(s) - \langle h_t \rangle_{\text{for}}]^2 \right] p_{\text{for}}(s,t)ds
$$

$$
\sim e^{t\langle h_t \rangle_{\text{for}}} \left[1 + \frac{[t\sigma_{\text{for}}(h_t)]^2}{2} \right], \tag{E3}
$$

which is a second-order correction to $\exp[t\langle h_t \rangle_{\text{for}}]$ in $t\sigma_{\text{for}}(h_t)$.

Combining Eqs. (E2) and (E3) we find at first order

$$
\langle g_t \rangle_{\text{back}} - \langle g_t \rangle_{\text{for}} \underset{t\sigma \to 0}{\sim} t\, \text{Cov}_{\text{for}}(h_t, g_t). \tag{E4}
$$

As mentioned at the beginning of this Appendix, we can use the backward point of view, with a first-order expansion in $t\sigma_{\text{back}}[h_t(s)]$ instead. In this case,

$$
\langle g_t \rangle_{\text{for}} = e^{t(\Lambda_t - \langle h_t \rangle_{\text{back}})} \int g_t(s) e^{t[\langle h_t \rangle_{\text{back}} - h_t(s)]} p_{\text{back}}(s,t)ds \underset{t\sigma \to 0}{\sim} e^{t(\Lambda_t - \langle h_t \rangle_{\text{back}})}[\langle g_t \rangle_{\text{back}} - t\text{Cov}_{\text{back}}(h_t, g_t)]. \tag{E5}
$$

The prefactor is computed with the same expansion starting from Eq. (B2):

$$
e^{-t\Lambda_t} = \langle e^{-th_t} \rangle_{\text{back}} \underset{t\sigma \to 0}{\sim} e^{-t\langle h_t \rangle_{\text{back}}} \int \left[1 + t[\langle h_t \rangle_{\text{back}} - h_t(s)] + \frac{t^2}{2}[\langle h_t \rangle_{\text{back}} - h_t(s)]^2 \right] p_{\text{back}}(s,t)ds
$$

$$
\sim e^{-t\langle h_t \rangle_{\text{back}}} \left[1 + \frac{[t\sigma_{\text{back}}(h_t)]^2}{2} \right], \tag{E6}
$$

which is a second order correction in $t\sigma_{\text{back}}(h_t)$. Combining Eqs. (E5) and (E6), we find

$$
\langle g_t \rangle_{\text{back}} - \langle g_t \rangle_{\text{for}} \underset{t\sigma \to 0}{\sim} t\, \text{Cov}_{\text{back}}(h_t, g_t). \tag{E7}
$$

When comparing Eqs. (E4) and (E7), we conclude that the covariance can be taken equivalently in the forward or backward statistics.

Let us now turn to the r.h.s. of inequality Eq. (11). Using the expression of $\sigma_{\text{for}}(q_t)$ as the forward coefficient of variation of the quantity $\exp[th_t(s)]$ [Eq. (15)] ($\sigma_{\text{back}}(r_t)$ as the backward coefficient of variation of the quantity $\exp[-th_t(s)]$), it is straightforward to show from the same kind of Taylor expansion that

$$
\sigma_{\text{for}}(g_t)\sigma_{\text{for}}(q_t) \underset{t\sigma \to 0}{\sim} t\sigma_{\text{for}}(h_t)\sigma_{\text{for}}(g_t), \tag{E8}
$$

$$
\sigma_{\text{back}}(g_t)\sigma_{\text{back}}(r_t) \underset{t\sigma \to 0}{\sim} t\sigma_{\text{back}}(h_t)\sigma_{\text{back}}(g_t). \tag{E9}
$$

Thus, the inequality Eq. (11) does not get necessarily saturated in this limit. However, in the particular case where $g_t(s)$ is the fitness landscape $h_t(s)$, then Eq. (E4) reads

$$
\Pi_{\mathcal{S}} \underset{t\sigma \to 0}{\sim} t\, \text{Var}(h_t), \tag{E10}
$$

and thus the inequality Eq. (14) is saturated in this limit.

APPENDIX F: A LOWER BOUND FOR THE STRENGTH OF SELECTION

The trivial lower bound on the strength of selection $\Pi_{\mathcal{S}} \geqslant 0$ comes from the positivity of KL divergences $\mathcal{D}_{\text{KL}}(p_{\text{for}}|p_{\text{back}})$ and $\mathcal{D}_{\text{KL}}(p_{\text{back}}|p_{\text{for}})$, giving, respectively, $\Lambda_t - \langle h_t \rangle_{\text{for}} \geqslant 0$ and $\langle h_t \rangle_{\text{back}} - \Lambda_t \geqslant 0$. The positivity of KL divergence itself relies on Jensen's inequality. In order to improve upon this bound, we proceed in two steps, using a sharpened version of Jensen's inequality [26] to both inequalities. First, seeking a positive lower bound for $\Lambda_t - \langle h_t \rangle_{\text{for}}$, the sharpened version of Jensen's inequality reads

$$
\langle e^{th_t} \rangle_{\text{for}} - e^{t\langle h_t \rangle_{\text{for}}} \geqslant \sigma_{\text{for}}^2(h_t) \inf_h \psi(\varphi_{\text{for}}, h, \langle h_t \rangle_{\text{for}}), \tag{F1}
$$

where functions ψ, φ_{for}, and φ_{back} have been defined in the main text. The r.h.s. involves the minimum of the function ψ when varying h on its support at time t. We then divide this expression by $\exp(t\Lambda_t)$:

$$
\langle e^{t(h_t - \Lambda_t)} \rangle_{\text{for}} - e^{t(\langle h_t \rangle_{\text{for}} - \Lambda_t)} \geqslant \frac{\sigma_{\text{for}}^2(h_t)}{\exp(t\Lambda_t)} \inf_h \psi(\varphi_{\text{for}}, h, \langle h_t \rangle_{\text{for}}), \tag{F2}
$$

The first term is 1 because of the normalization of the probability distribution p_{back}. Finally the enhanced bound reads

$$
\Lambda_t - \langle h_t \rangle_{\text{for}} \geqslant -\frac{1}{t} \ln \left(1 - \frac{\sigma_{\text{for}}^2(h_t)}{\exp(t\Lambda_t)} \inf_h \psi(\varphi_{\text{for}}, h, \langle h_t \rangle_{\text{for}}) \right). \tag{F3}
$$

Similarly, we find

$$\langle h_t \rangle_{\text{back}} - \Lambda_t$$
$$\geqslant -\frac{1}{t} \ln \left(1 - \frac{\sigma_{\text{back}}^2(h_t)}{\exp(-t\Lambda_t)} \inf_h \psi(\varphi_{\text{back}}, h, \langle h_t \rangle_{\text{back}}) \right). \quad \text{(F4)}$$

Liao *et al.* proved [26] that when $\varphi'(x)$ is a convex (concave) function, then $\psi(\varphi, x, \nu)$ is an increasing (decreasing) function of x, and thus the infimum of function $\psi(\varphi, x, \nu)$ on x is reached for $x = x_{\min}$ ($x = x_{\max}$).

Because of the convexity of $\varphi'_{\text{for}}(x) = t \exp[tx]$, the minimum of ψ is reached when evaluating ψ at the minimum value $h_{\min}$ of the fitness landscape $h_t(s)$. At finite time, the support of $h_t(s)$ is finite and so is its minimum value. Similarly, because of the concavity of $\varphi''_{\text{back}}(x) = -t \exp[-tx]$, the minimum of ψ is reached when evaluating ψ at the maximum value $h_{\max}$. Finally, we use the relation $-\ln(1-x) \geqslant x$ valid for any real number x and we combine the two inequalities to obtain

$$\Pi_{\mathcal{S}} \geqslant \frac{1}{t} \left[\frac{\sigma_{\text{for}}^2(h_t)}{\exp(t\Lambda_t)} \psi(\varphi_{\text{for}}, h_{\min}, \langle h_t \rangle_{\text{for}}) \right.$$
$$\left. + \frac{\sigma_{\text{back}}^2(h_t)}{\exp(-t\Lambda_t)} \psi(\varphi_{\text{back}}, h_{\max}, \langle h_t \rangle_{\text{back}}) \right]. \quad \text{(F5)}$$

Note that the r.h.s. of Eqs. (F3) and (F4) are both positive numbers due to the convexity and concavity of φ'_{for} and φ'_{back}, respectively. As a result, their sum is also positive and therefore the r.h.s. of Eq. (F5) does represent an improvement with respect to the trivial bound which would be 0.

Liao *et al.* also proposed another lower bound, looser but simpler than the one involving the function ψ. Indeed, one can replace $\inf_h \psi(\varphi_{\text{for}}, h, \langle h_t \rangle_{\text{for}})$ by $\inf_h \varphi''_{\text{for}}(h)/2$ in Eq. (F1). Moreover, $\inf_h \varphi''_{\text{for}}(h)/2 = \varphi''_{\text{for}}(h_{\min})/2$ since $\varphi''_{\text{for}}(h)$ is an increasing function of h. The same goes for the other inequality, and combining the two leads to

$$\Pi_{\mathcal{S}} \geqslant \frac{1}{2t} \left[\frac{\sigma_{\text{for}}^2(h_t)}{\exp(t\Lambda_t)} \varphi''_{\text{for}}(h_{\min}) + \frac{\sigma_{\text{back}}^2(h_t)}{\exp(-t\Lambda_t)} \varphi''_{\text{back}}(h_{\max}) \right]. \quad \text{(F6)}$$

We notice that this version of the bound does not depend on the average values of the fitness landscape, unlike Eq. (F5).

Let us mention that if no information is known on the support of the fitness landscape, $h_{\min}$ can still be taken equal to 0, in both Eqs. (F5) and (F6), because fitness landscapes are positive functions. Indeed, the fitness landscape can be expressed as [2] $th_t(s) = \ln[\sum_K m^K R_{\text{for}}(K|s)]$, in terms of the conditional forward probability $R_{\text{for}}(K|s)$ of the number of divisions. Since $m \geqslant 1$ and $K \geqslant 0$, this relation implies that the fitness landscape is a positive quantity. In this case, Eq. (F5) [Eq. (F6)] gives a nontrivial bound based on the first two moments (second moment) of the forward fitness landscape distribution.

The simplest bound is thus obtained when considering Eq. (F6) with $h_{\min} = 0$ and $h_{\max} = +\infty$, which cancels the
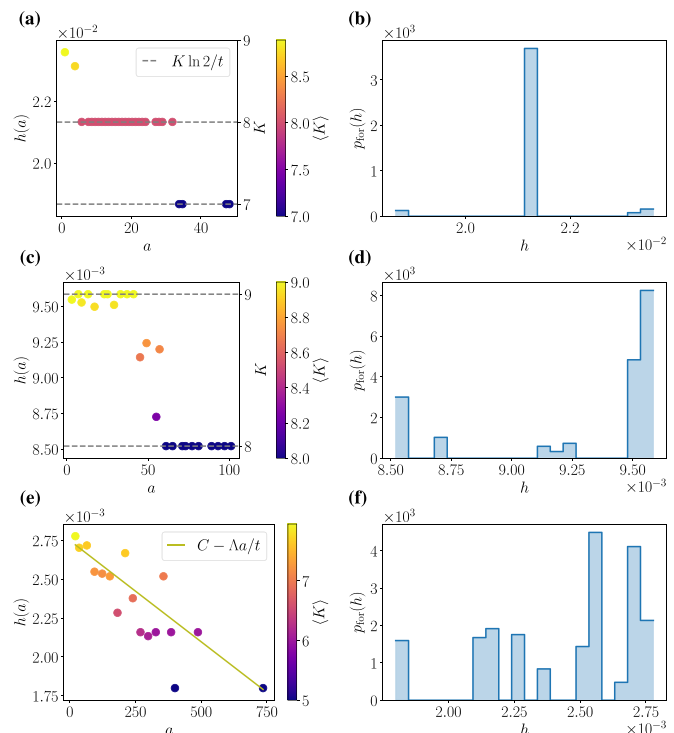

FIG. 6. Experimental fitness landscapes for age and their forward distributions, computed with data from [27]. Each line corresponds to a different experiment and the first column shows fitness landscapes $h_t(a)$ as functions of size a. In (a) and (c), the gray horizontal dashed lines correspond to theoretical plateaus, equal to $K \ln 2/t$, predicted when K is fully determined by the value s of the trait. The integers K corresponding to the plateaus are indicated on the right y axis. In (e), the plateaus are blurred and replaced by a smoother scatter plot in good agreement with the general shape of the theoretical prediction, made in the case of an age-controlled model in steady state [10]. Λ is the population growth rate and the constant C was adjusted to fit the scatter plot. In (a), (c), and (e), each dot is made of all the cells having the same age, and the mean number of divisions among those cells is represented by the color of the dot. This shows that dots aligning on a plateau corresponding to a number K of divisions truly come from cells that underwent K divisions. In (b), (d), and (f), the three subfigures on the right represent the distribution $p_{\text{for}}(h)$ of the corresponding age fitness landscapes (i.e., on the same line) with the forward age distribution. For all three rows, the fitness landscapes are highly non-Gaussian, and for (b) and (d) the peaks in these distributions correspond to the values of some of the plateaus.

second term in the bracket

$$\Pi_{\mathcal{S}} \geqslant \frac{t}{2} \frac{\sigma_{\text{for}}^2(h_t)}{\exp(t\Lambda_t)}. \quad \text{(F7)}$$

APPENDIX G: FURTHER TESTS ON EXPERIMENTAL DATA FROM [27]

In Fig. 1 in Sec. II we showed the size fitness landscapes $h_t(x)$ as a function of the cell size x, and the distribution $p_{\text{for}}(h)$ of size fitness landscape computed with the forward cell size distribution, for 3 of the 11 experiments from [27]. We now show the corresponding plots when choosing the age $\mathcal{A}$ of the cell as the phenotypic trait $\mathcal{S}$. The experiment on
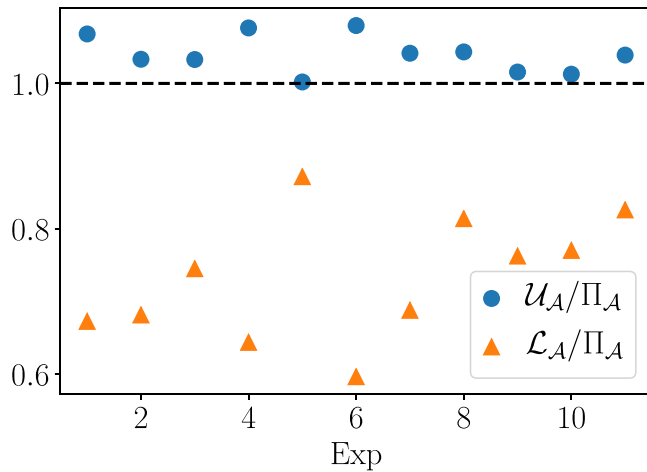
FIG. 7. Upper bound $\mathcal{U}_\mathcal{A}$ (blue dots) and lower bound $\mathcal{L}_\mathcal{A}$ (orange triangles) for the strength of selection acting on size $\Pi_\mathcal{A}$, normalized by the latter. The x axis represents the 11 colonies in different growth conditions from [27], in the same order as in Fig. 4.

line i in Fig. 6 is the same as the experiment on line i in Fig. 1. By comparing the two figures, we see that for the first two rows the theoretical plateaus at $h = K \ln 2/t$ are the same for $h_t(a)$ and for $h_t(x)$, which is logical since it is the same cells, and both the age and the size are highly correlated to the number of divisions. In Fig. 6(e), as for the size in Fig. 1(e), the plateaus start to blur to give rise to a smoother scatter plot, whose shape matches the linear prediction we made for age-controlled models in steady state [10]. Similarly to the case of the cell size, distributions of fitness landscapes shown in the second column are highly non-Gaussian.

Then we test the upper and lower bounds on the strength of selection acting on cell age using the same data. We show in Fig. 7 the upper bound $\mathcal{U}_\mathcal{A}$ given by Eq. (14) and the lower bound $\mathcal{L}_\mathcal{A}$ given by Eq. (21), normalized by the strength of selection $\Pi_\mathcal{A}$ acting on age. The x axis numbers the colonies which have grown in different nutrient media [27]. As expected, points representing the upper bound and those representing the lower bound are, respectively, above and below the horizontal dashed line at $y = 1$.

Nozoe *et al.* proved that the strength of selection acting on any trait $\mathcal{S}$ is bounded by the strength of selection acting on the number of divisions $\mathcal{K}$ [2]: $\forall \mathcal{S}$, $\Pi_\mathcal{S} \leqslant \Pi_\mathcal{K}$. We obtained in
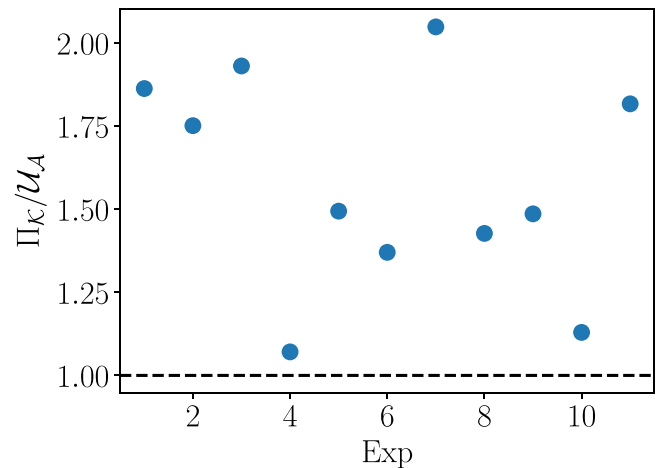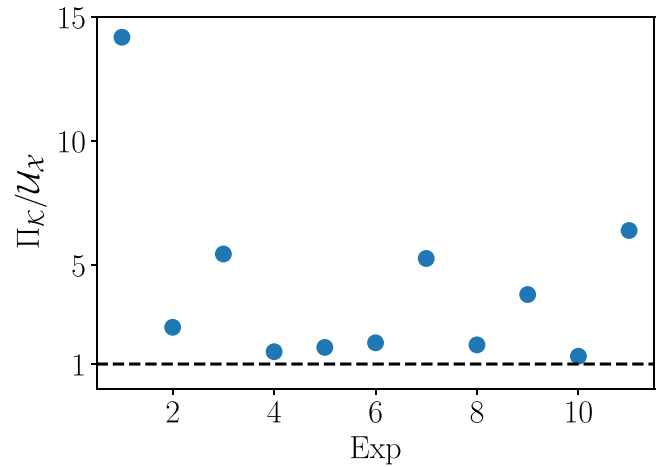

FIG. 8. Ratio of the general bound $\Pi_\mathcal{K}$ to our upper bound $\mathcal{U}_\mathcal{X}$ for size (top plot) and $\mathcal{U}_\mathcal{A}$ for age (bottom plot) for the 11 experiments from [27], in no particular order. All points are above the black horizontal dashed line at $y = 1$, which indicates that our upper bound is always smaller and thus better than $\Pi_\mathcal{K}$.

this paper a trait-dependent bound, which highlights the role of fluctuations of fitness landscape of that particular trait for the strength of selection, and which is often tighter than $\Pi_\mathcal{K}$. We show in Fig. 8 the ratio of $\Pi_\mathcal{K}$ to the upper bound $\mathcal{U}_\mathcal{X}$ for the size (top plot) and $\mathcal{U}_\mathcal{A}$ for the age (bottom plot). All points are indeed above 1.

[1] S. Leibler and E. Kussell, Individual histories and selection in heterogeneous populations, Proc. Natl. Acad. Sci. USA **107**, 13183 (2010).

[2] T. Nozoe, E. Kussell, and Y. Wakamoto, Inferring fitness landscapes and selection on phenotypic states from single-cell genealogical data, PLoS Genet. **13**, e1006653 (2017).

[3] Y. Wakamoto, A. Y. Grosberg, and E. Kussell, Optimal lineage principle for age-structured populations, Evolution **66**, 115 (2012).

[4] G. Lambert and E. Kussell, Quantifying Selective Pressures Driving Bacterial Evolution Using Lineage Analysis, Phys. Rev. X **5**, 011016 (2015).

[5] J. Lin and A. Amir, The effects of stochasticity at the single-cell level and cell size control on the population growth, Cell Syst. **5**, 358 (2017).

[6] V. Mustonen and M. Lassig, Fitness flux and ubiquity of adaptive evolution, Proc. Natl. Acad. Sci. USA **107**, 4248 (2010).

[7] V. Mustonen and M. Lässig, From fitness landscapes to seascapes: Non-equilibrium dynamics of selection and adaptation, Trends Genet. **25**, 111 (2009).

[8] T. J. Kobayashi and Y. Sughiyama, Fluctuation Relations of Fitness and Information in Population Dynamics, Phys. Rev. Lett. **115**, 238102 (2015).

[9] R. García-García, A. Genthon, and D. Lacoste, Linking lineage and population observables in biological branching processes, Phys. Rev. E **99**, 042413 (2019).

[10] A. Genthon and D. Lacoste, Fluctuation relations and fitness landscapes of growing cell populations, Sci. Rep. **10**, 11889 (2020).

[11] S. Boyer, D. Biswas, A. Kumar Soshee, N. Scaramozzino, C. Nizak, and O. Rivoire, Hierarchy and extremes in selections from pools of randomized proteins, Proc. Natl. Acad. Sci. USA **113**, 3482 (2016).

[12] M. Smerlak and A. Youssef, Limiting fitness distributions in evolutionary dynamics, J. Theor. Biol. **416**, 68 (2017).

[13] A. Dechant and S.-i. Sasa, Fluctuation–response inequality out of equilibrium, Proc. Natl. Acad. Sci. USA **117**, 6430 (2020).

[14] L. Dinis, J. Unterberger, and D. Lacoste, Phase transitions in optimal betting strategies, Europhys. Lett. **131**, 60005 (2020).

[15] J. M. Horowitz and T. R. Gingrich, Thermodynamic uncertainty relations constrain non-equilibrium fluctuations, Nat. Phys. **16**, 15 (2020).

[16] P. Thomas, Making sense of snapshot data: Ergodic principle for clonal cell populations, J. R. Soc. Interface **14**, 20170467 (2017).

[17] E. Baake and H.-O. Georgii, Mutation, selection, and ancestry in branching models: A variational approach, J. Math. Biol. **54**, 257 (2007).

[18] M. Smerlak, Quasi-species evolution maximizes genotypic reproductive value (not fitness or flatness), J. Theor. Biol. **522**, 110699 (2021).

[19] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Robust growth of Escherichia Coli, Curr. Biol. **20**, 1099 (2010).

[20] S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution, Proc. Sixth Intl. Congress Genetics **1**, 356 (1932).

[21] L. Peliti, Introduction to the statistical theory of Darwinian evolution, arXiv:cond-mat/9712027.

[22] R. A. Neher and B. I. Shraiman, Statistical genetics and evolution of quantitative traits, Rev. Mod. Phys. **83**, 1283 (2011).

[23] R. A. Fisher, *The Genetical Theory of Natural Selection: A Complete Variorum Edition*, edited by J. H. Bennet (Oxford University Press, Oxford, UK, 2000).

[24] G. R. Price, Fisher's 'fundamental theorem' made clear, Ann. Hum. Genet. **36**, 129 (1972).

[25] A. Gardner, Price's equation made clear, Philos. Trans. R. Soc. Sect. B **375**, 20190361 (2020).

[26] J. G. Liao and A. Berg, Sharpening Jensen's inequality, Am. Stat. **73**, 278 (2019).

[27] D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, and S. J. Tans, Stochasticity of metabolism and growth at the single-cell level, Nature (London) **514**, 376 (2014).

[28] K. Kaneko and C. Furusawa, Macroscopic theory for evolving biological systems akin to thermodynamics, Annu. Rev. Biophys. **47**, 273 (2018).

[29] R. T. Schuh and A. V. Z. Brower, *Biological Systematics: Principles and Applications*, 2nd ed. (Comstock/Cornell University Press, Ithaca, NY, 2009).

[30] A. N. Nguyen Ba, I. Cvijović, J. I. Rojas Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu, S. F. Levy, and M. M. Desai, High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast, Nature (London) **575**, 494 (2019).

[31] T. Tak, G. Prevedello, G. Simon, N. Paillon, K. R. Duffy, and L. Perié, Simultaneous tracking of division and differentiation from individual hematopoietic stem and progenitor cells reveals within-family homogeneity despite population heterogeneity, bioRxiv 586354 (2019).

[32] N. Goldenfeld and C. Woese, Life is physics: Evolution as a collective phenomenon far from equilibrium, Annu. Rev. Condens. Matter Phys. **2**, 375 (2011).

[33] E. O. Powell, Growth rate and generation time of bacteria, with special reference to continuous culture, J. Gen. Microbiol. **15**, 492 (1956).

[34] M. Hashimoto, T. Nozoe, H. Nakaoka, R. Okura, S. Akiyoshi, K. Kaneko, E. Kussell, and Y. Wakamoto, Noise-driven growth rate gain in clonal cellular populations, Proc. Natl. Acad. Sci. USA **113**, 3251 (2016).