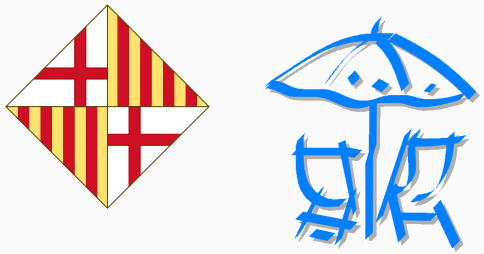
Fluctuation Relations and Fitness in Cell Populations

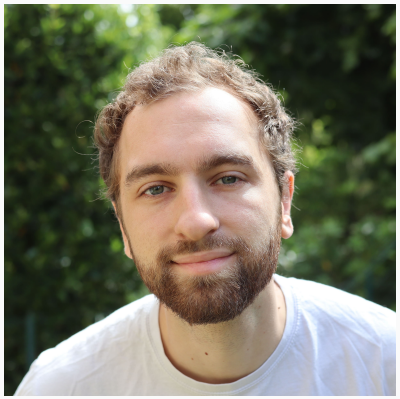
L. Peliti*

Barcelona, July 6, 2022

*SMRI, Santa Marinella (Italy)



Acknowlegments



A. Genthon

D. Lacoste

T. Nozoe

R. García-García

UMR 7083 uRiver

ESPCI PARIS

EDUCATION SCIENCE INNOVATION

Population growth rate

Starting from $N(0)$ cells at time 0:

$$N(t) \sim N(0)\, \mathrm{e}^{\Lambda t}$$

Population growth rate:

$$\Lambda_t = \frac{1}{t} \ln \frac{N(t)}{N(0)} \qquad \Lambda = \lim_{t \to \infty} \Lambda_t$$

Can we measure Λ?

- Infinite time limit
- Infinite population
- Dependence on phenotype distribution and environment
- Intrinsic stochasticity
- ...

Fitness

Fitness f_x of a phenotypic trait $\mathcal{X}$

Measured by the growth rate of a subpopulation:

$$\frac{\mathrm{d}N_x(t)}{\mathrm{d}t} \simeq f_x N(x,t), \qquad x \in \mathcal{X}$$

Fisher's fundamental theorem:

$$\frac{\partial}{\partial t}\overline{f_x} = \mathrm{var}\, f$$

Neglecting mutations, drift, phenotype change, ...

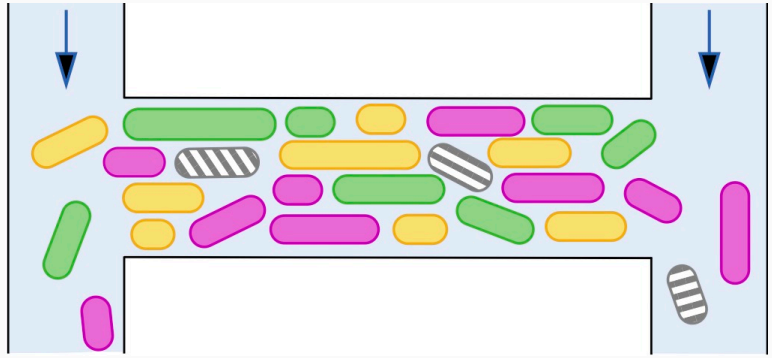
- Infinite population
- Dependence on phenotype distribution and environment
- Intrinsic stochasticity
- Epistasis, pleiotropy, ...
- ...

Fitness is central in model-building but elusive in experiment

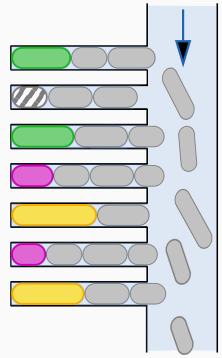
Monitoring single-cell dynamics

Experiments on single-cell dynamics:

Petri dish



Flow cytometer

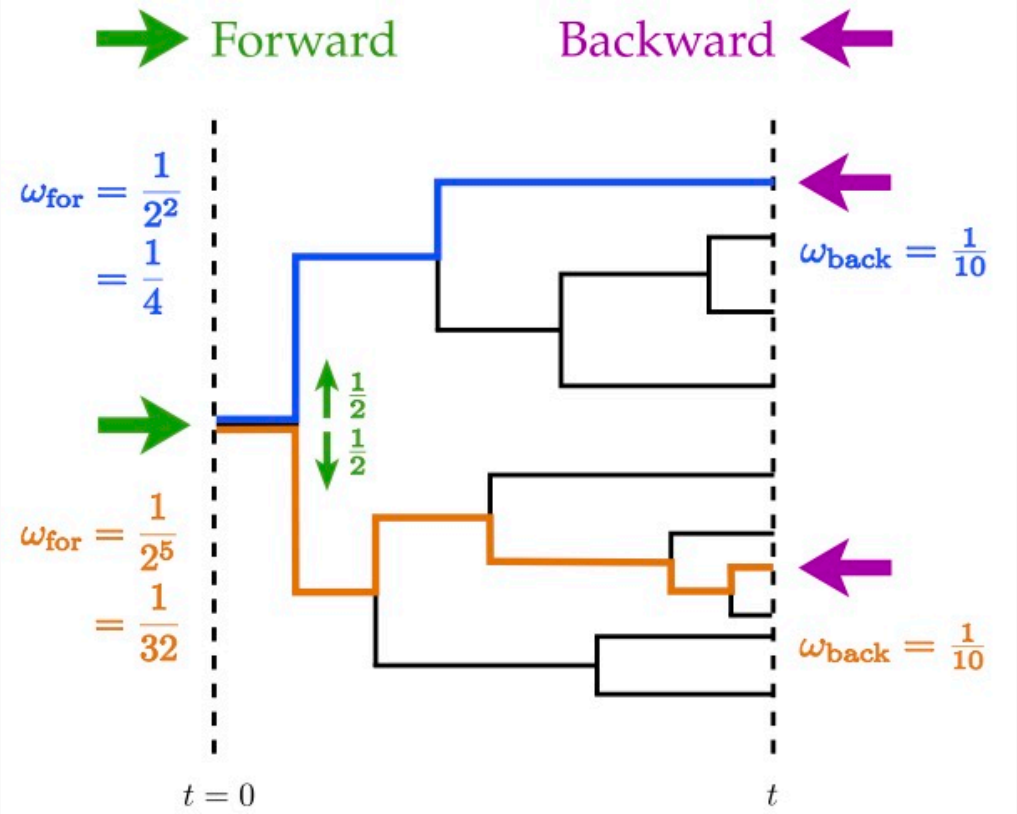


Mother machine

Phenotype (e.g., expression of some proteins) can be monitored by coupling to expression of fluorescent proteins

Can we harness genealogical information to evaluate fitness and population growth rate?

Forward and Backward sampling



$$\omega_{\text{back}} = \frac{1}{N(t)} \qquad \omega_{\text{for}} = \frac{2^{-K}}{N(0)}$$

$$\omega_{\text{back}}(\ell) = \mathrm{e}^{K(\ell)\ln 2 - t\Lambda_t}\omega_{\text{for}}(\ell)$$

Leibler and Kussell, 2010

Division and doubling times

Evaluate

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\omega_{\mathrm{back}}\|\omega_{\mathrm{for}}) &:= \sum_{\ell}\omega_{\mathrm{back}}(\ell)\ln\frac{\omega_{\mathrm{back}(\ell)}}{\omega_{\mathrm{for}}(\ell)}\\ &= \langle K\rangle_{\mathrm{back}}\ln 2 - t\Lambda_t \geq 0\\ \mathcal{D}_{\mathrm{KL}}(\omega_{\mathrm{for}}\|\omega_{\mathrm{back}}) &= t\Lambda_t - \langle K\rangle_{\mathrm{for}}\ln 2 \geq 0\end{aligned}$$

Thus

$$\frac{t}{\langle K\rangle_{\mathrm{back}}} \leq \frac{\ln 2}{\Lambda_t} \leq \frac{t}{\langle K\rangle_{\mathrm{for}}}$$

Define the inter-division time $\tau = \lim_{t\to\infty} t/\langle K\rangle$. Then

$$\langle\tau\rangle_{\mathrm{back}} \leq \mathcal{T}_{\mathrm{d}} \leq \langle\tau\rangle_{\mathrm{for}}$$

where $\mathcal{T}_{\mathrm{d}} = \ln 2/\Lambda$ is the population doubling time

García-García et al., 2019

Estimating the fitness landscape

Let $\mathcal{X}$ be a trait (phenotype): we then have, for each value x,

$$\boxed{p_{\mathrm{back}}(K,x) = \mathrm{e}^{K\ln 2 - t\Lambda_t} p_{\mathrm{for}}(K,x)}$$

and we can define the marginals

$$p_{\mathrm{back}}(x) = \sum_K p_{\mathrm{back}}(K,x) \qquad p_{\mathrm{for}}(x) = \sum_K p_{\mathrm{for}}(K,x)$$

Defining the **fitness landscape**

$$h_t(x) := \frac{1}{t}\ln\frac{N(t)p_{\mathrm{back}}(x)}{N(0)p_{\mathrm{for}}(x)} = \Lambda_t + \frac{1}{t}\ln\frac{p_{\mathrm{back}}(x)}{p_{\mathrm{for}}(x)}$$

we have

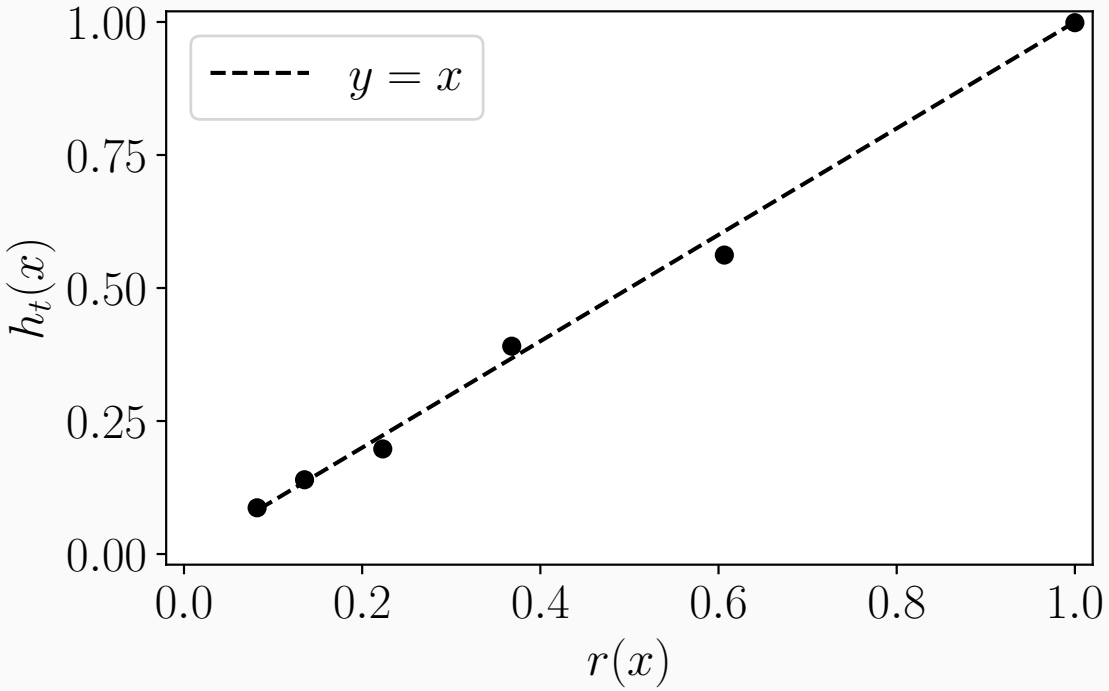
$$p_{\mathrm{back}}(x) = \mathrm{e}^{t(h(x)-\Lambda_t)} p_{\mathrm{for}}(x)$$

Defining the conditional distribution $p_{\mathrm{for}}(K|x) := p_{\mathrm{for}}(K,x)/p_{\mathrm{for}}(x)$ we obtain the estimator

$$h_t(x) = \frac{1}{t}\ln\sum_K 2^K p_{\mathrm{for}}(K|x)$$

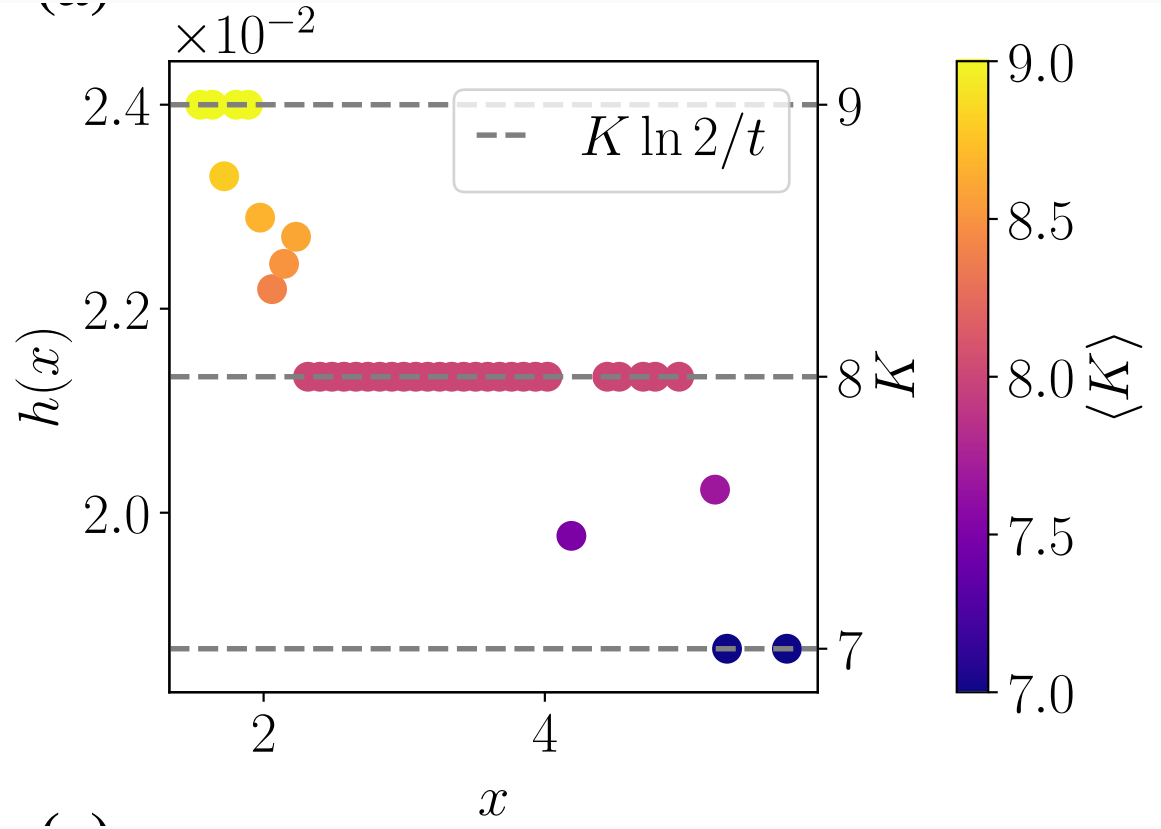
Nozoe et al. 2017

Estimating the fitness landscape



Estimated fitness landscape $h_t(x)$ for a Moran model of $N = 10\,000$ individuals, with division rate $r(x) = \mathrm{e}^{-x/2}$, $x \in \{0, \ldots, 5\}$, and $t = 5$. *The total weight of the forward sampling yields the number $N(0)$ of ancestors.* The effective population growth rate is given by $\Lambda_t = \ln(N(t)/N(0))/t$. Only lineages surviving at t are sampled.

Estimating the fitness landscape



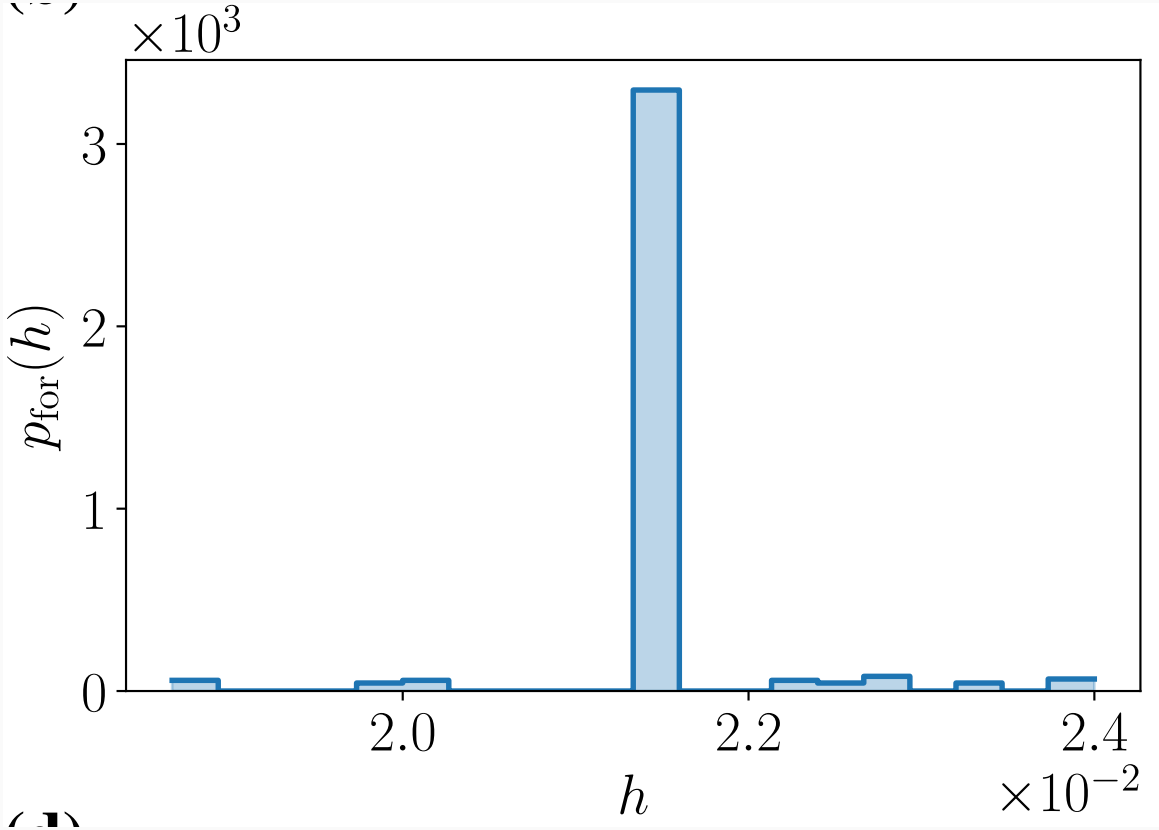
Trait: Cell size x

If $h(x)$ were fully determined by x we would have $h(x) = K \ln 2/t$

Genthon and Lacoste, 2021

Data by Kiviet et al., 2014 on *E. coli*

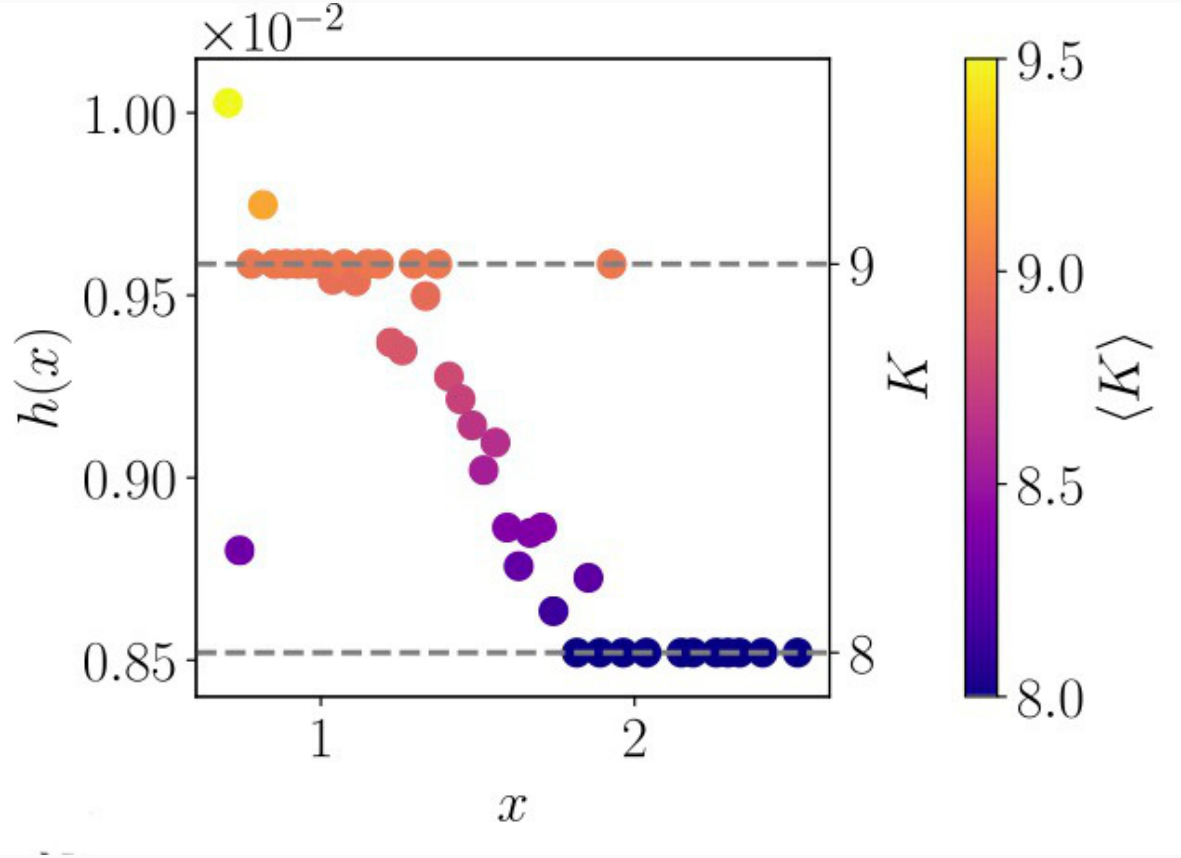
Estimating the fitness landscape



Genthon and Lacoste, 2021

Data by Kiviet et al., 2014 on *E. coli*

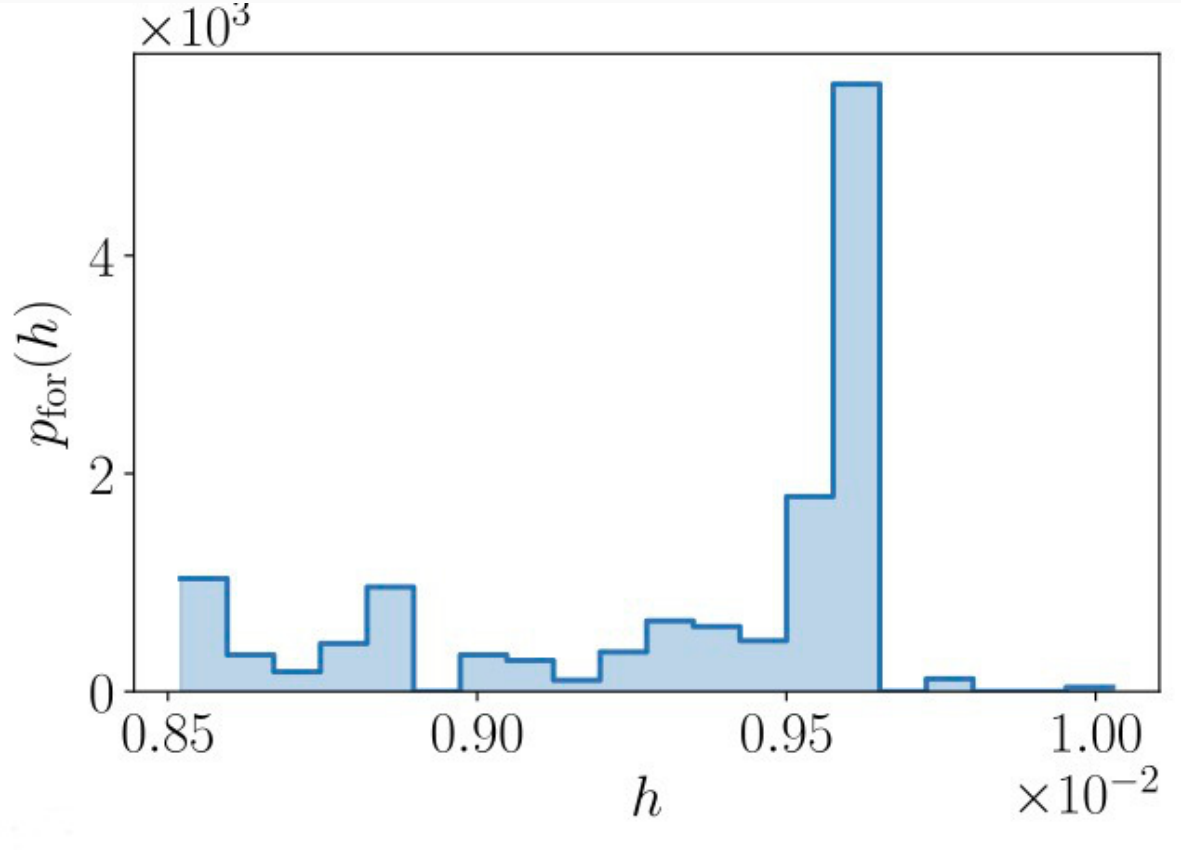
Estimating the fitness landscape



Genthon and Lacoste, 2021

Data by Kiviet et al., 2014 on *E. coli*

Estimating the fitness landscape



Genthon and Lacoste, 2021

Data by Kiviet et al., 2014 on *E. coli*

Fitness landscape vs. biological fitness

Estimating the biological fitness from the growth rate of a subpopulation:

$$f_x \simeq \Lambda_t(x) = \frac{1}{t} \ln \frac{N(x,t)}{N(x,0)} = \Lambda_t + \frac{1}{t} \ln \frac{p_{\mathrm{back}}(x,t)}{p_{\mathrm{back}}(x,0)}$$

Thus we obtain

$$\begin{aligned} h_t(x) - \Lambda_t(x) &= \frac{1}{t} \left[\ln \frac{p_{\mathrm{back}}(x,t)}{p_{\mathrm{for}}(x,t)} - \ln \frac{b_{\mathrm{back}}(x,t)}{p_{\mathrm{back}}(x,0)} \right] \\ &= \frac{1}{t} \ln \frac{p_{\mathrm{back}}(x,0)}{p_{\mathrm{for}}(x,t)} = \frac{1}{t} \ln \frac{p_{\mathrm{for}}(x,0)}{p_{\mathrm{for}}(x,t)} \end{aligned}$$

Strength of selection

Measure of the strength of selection for trait $\mathcal{X}$:

$$\begin{aligned}\Pi_{\mathcal{X}} &= \langle h_t(x)\rangle_{\text{back}} - \langle h_t(x)\rangle_{\text{for}} \\ &= \frac{1}{t}\sum_x \left(p_{\text{back}}(x) - p_{\text{for}}(x)\right) \ln \frac{p_{\text{back}}(x)}{p_{\text{for}}(x)} \geq 0 \qquad (*) \\ &= \frac{1}{t}\left[\mathcal{D}_{\text{KL}}(p_{\text{back}} \| p_{\text{for}}) + \mathcal{D}_{\text{KL}}(p_{\text{for}} \| p_{\text{back}})\right]\end{aligned}$$

Nozoe et al. 2017

Define

$$q_t(x) = \frac{p_{\text{back}}(x)}{p_{\text{for}}(x)} \qquad r_t(x) = \frac{p_{\text{for}}(x)}{p_{\text{back}}(x)} = \frac{1}{q_t(x)}$$

then, for an arbitrary function $g_t(x)$,

$$\text{cov}_{\text{back}}(g_t, q_t) = \langle g_t q_t\rangle_{\text{back}} - \langle g_t\rangle_{\text{back}} \langle q_t\rangle_{\text{back}} = \langle g_t\rangle_{\text{back}} - \langle g_t\rangle_{\text{for}}$$

$$\text{cov}_{\text{for}}(g_t, r_t) = \langle g_t\rangle_{\text{for}} - \langle g_t\rangle_{\text{back}}$$

and, by the Cauchy-Schwartz inequality,

$$\left|\langle g_t\rangle_{\text{for}} - \langle g_t\rangle_{\text{back}}\right| \leq \min\left(\sigma_{\text{back}}(g_t)\sigma_{\text{for}}(q_t), \sigma_{\text{for}}(g_t)\sigma_{\text{back}}(r_t)\right)$$

Strength of selection

Since

$$q_t(x) = \mathrm{e}^{t(h_t(x)-\Lambda_t)}$$

we obtain

$$\Pi_{\mathcal{X}} = \mathrm{cov}_{\mathrm{for}}(h_t, \mathrm{e}^{th_t})\,\mathrm{e}^{-t\Lambda_t} = \mathrm{cov}_{\mathrm{back}}(h_t, \mathrm{e}^{-th_t})\,\mathrm{e}^{t\Lambda_t}$$

and

$$0 \leq \Pi_{\mathcal{X}} \leq \min\Big(\sigma_{\mathrm{for}}(h_t)\sigma_{\mathrm{for}}(q_t), \sigma_{\mathrm{back}}(h_t)\sigma_{\mathrm{back}}(r_t)\Big)$$

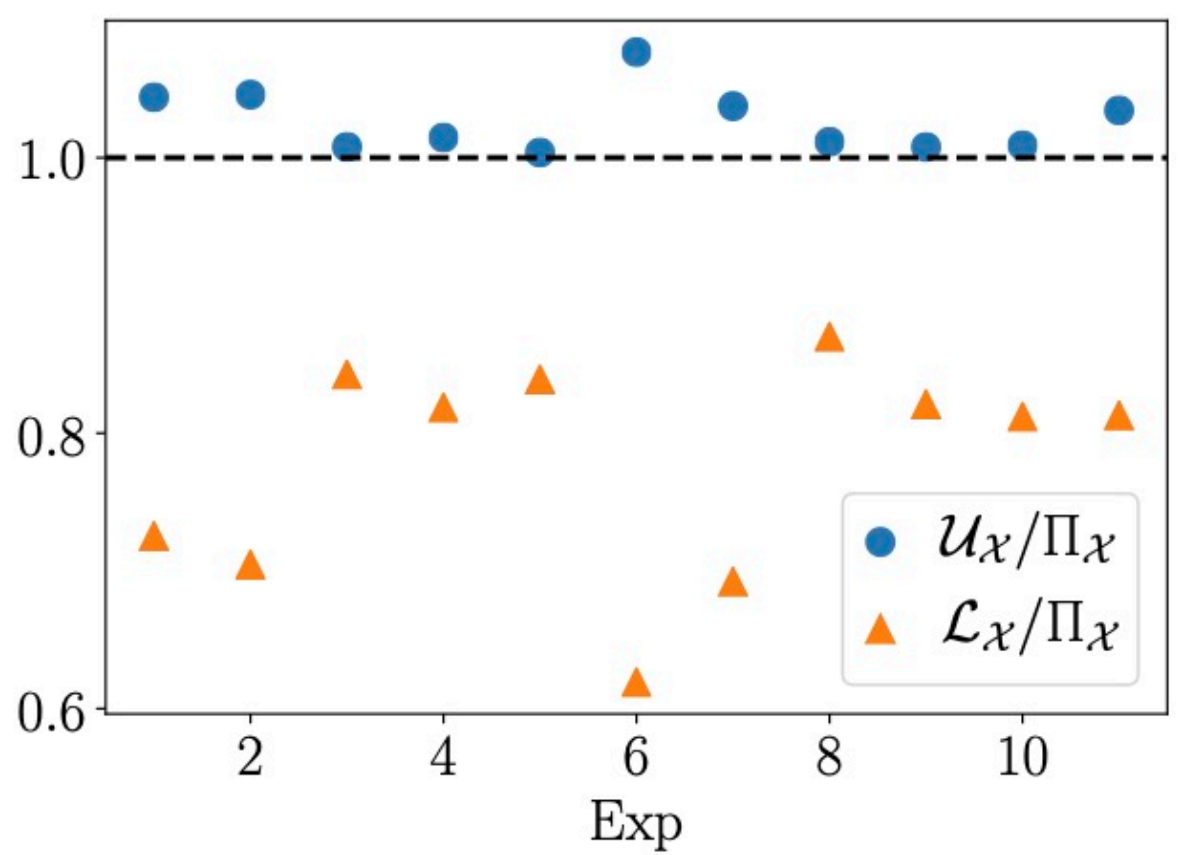
A tighter lower bound can also be obtained from Jensen's inequality applied on $(^*)$:

$$\Pi_{\mathcal{X}} \geq \frac{1}{t}\left[\frac{\sigma^2_{\mathrm{for}}(h_t)}{\exp(t\Lambda_t)}\psi(\varphi_{\mathrm{for}}, h_{\mathsf{min}}, \langle h_t\rangle_{\mathrm{for}}) + \frac{\sigma^2_{\mathrm{back}}(h_t)}{\exp(-t\Lambda_t)}\psi(\varphi_{\mathrm{back}}, h_{\mathsf{max}}, \langle h_t\rangle_{\mathrm{back}})\right]$$

$$\psi(\varphi, x, \nu) := \frac{\varphi(x) - \varphi(\nu)}{(x-\nu)^2} \qquad \varphi_{\mathrm{for}}(x) := \mathrm{e}^{tx} \qquad \varphi_{\mathrm{back}}(x) := \mathrm{e}^{-tx}$$

Genthon and Lacoste, 2021

Strength of selection

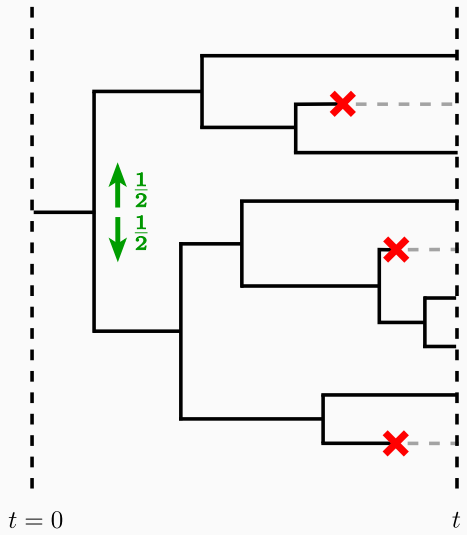


Genthon and Lacoste, 2021

Data by Kiviet et al., 2014 on *E. coli*

Role of cell death

Forward Backward



σ	ω_{back}	ω_{for}
1	1/6	1/4
0	0	1/8
1	1/6	1/8
1	1/6	1/8
0	0	1/16
1	1/6	1/32
1	1/6	1/32
1	1/6	1/8
0	0	1/8

$$p(K,\sigma,t) = \sum_{\ell\in\mathcal{L}(t)} \delta_{K,K(\ell)}\delta_{\sigma,\sigma(\ell)}\,\omega(\ell)$$

$$N(K,\sigma,t) = \sum_{\ell\in\mathcal{L}(t)} \delta_{K,K(\ell)}\delta_{\sigma,\sigma(\ell)}$$

$\mathcal{L}(t)$: set of *all* lineages present at time t (**DEAD or ALIVE!**)

Role of cell death

$$p_{\text{for}}(K,\sigma,t) = \frac{2^{-K} N(K,\sigma,t)}{N(0)} \qquad \sigma = 0,1$$

$$p_{\text{back}}(K,\sigma{=}0,t) = 0 \quad p_{\text{back}}(K,\sigma{=}1,t) = \frac{N(K,\sigma{=}1,t)}{N(t)} =: p_{\text{back}}(K,t)$$

$$p_{\text{surv}}(t) := p_{\text{for}}(\sigma{=}1,t) = \sum_K p_{\text{for}}(K,\sigma{=}1,t) = \frac{1}{N(0)} \sum_K 2^{-K} N(K,\sigma{=}1,t)$$

$$\Gamma_t = \frac{1}{t} \ln p_{\text{surv}}(t)$$

N.B. $p_{\text{surv}}(t) \neq N(\sigma{=}1,t)/\left|\mathcal{L}(t)\right|$ and $\Gamma_t \leq 0$, $\forall t$

$$\boxed{p_{\text{back}}(K,t) = \mathrm{e}^{K \ln 2 - t(\Lambda_t - \Gamma_t)}\, p_{\text{for}}(K,\sigma{=}1,t)}$$

$$\Lambda_t = \frac{1}{t} \ln \left\langle 2^K \right\rangle_{\text{for}|\sigma=1} + \Gamma_t$$

$$\left\langle \mathrm{e}^{t\Lambda_t - K \ln 2} \right\rangle_{\text{back}} = 1 - p_{\text{for}}(\sigma{=}0,t) = p_{\text{surv}}(t)$$

Genthon et al., 2022

Role of cell death

For the distribution $f(\tau|\sigma)$ of division times τ we have

$$f_{\mathrm{back}}(\tau) = 2 f_{\mathrm{for}}(\tau|\sigma{=}1)\, \mathrm{e}^{-\tau(\Lambda-\Gamma)}$$

and thus

$$\mathcal{D}_{\mathrm{KL}}\left(f_{\mathrm{back}}(\tau) \| f_{\mathrm{for}}(\tau|\sigma{=}1)\right) = -\left\langle \tau \right\rangle_{\mathrm{back}} (\Lambda - \Gamma) + \ln 2 \geq 0$$
$$\mathcal{D}_{\mathrm{KL}}\left(f_{\mathrm{for}}(\tau|\sigma{=}1) \| f_{\mathrm{back}}(\tau)\right) = \left\langle \tau \right\rangle_{\mathrm{for}} (\Lambda - \Gamma) - \ln 2 \geq 0$$

We thus have

$$\frac{1}{\langle \tau \rangle_{\mathrm{for}}} \leq \frac{1}{\mathcal{T}_{\mathrm{d}}} - \frac{\Gamma}{\ln 2} \leq \frac{1}{\langle \tau \rangle_{\mathrm{back}}}$$

and a generalized Euler-Lotka relation:

$$1 = 2 \int_0^\infty \mathrm{d}\tau\; f_{\mathrm{for}}(\tau|\sigma{=}1)\, \mathrm{e}^{-\tau(\Lambda-\Gamma)}$$

Genthon et al., 2022

Inferring population growth and selection

- Quantifying selection for a fixed trait x:

$$h(x) = \Lambda_t + \frac{1}{t} \ln \frac{p_{\text{back}}(x,t)}{p_{\text{for}}(x,t|\sigma{=}1)} = \frac{1}{t} \ln \left[\sum_k 2^K \, p_{\text{for}}(K,x,t|\sigma{=}1) \right] + \Gamma_t$$

- Fitness of trait x:

$$\Lambda_t(x) = \Lambda_t + \frac{1}{t} \ln \frac{p_{\text{back}}(x,t)}{p_{\text{back}}(x,0)}$$

Thus

$$h_t(x) - \Lambda_t(x) = \frac{1}{t} \ln \frac{p_{\text{for}}(x,t|\sigma=1)}{p_{\text{for}}(x,0)}$$

- Can be extended to the case of dilution (cytometer)
- In mother machines, only forward sampling is performed (but dead lineages contribute to averages)

Genthon et al., 2022

Inferring population growth and selection

Cytometer measurements:

- Dilution rate $\rho(x)$ (depending on trait x)
- Population size without dilution: $N^\circ(t)$, with dilution: $N(t)$
- Trait history $\boldsymbol{x} = (x(t))$

$$N^\circ(t) = N(t) \int \mathcal{D}\boldsymbol{x}\, p_{\mathrm{back}}(\boldsymbol{x}, \sigma{=}1)\, \exp\left[\int_0^t \mathrm{d}t'\, \rho(x(t'))\right]$$
$$= N(t) \left\langle \exp\left[\int_0^t \mathrm{d}t'\, \rho(x(t'))\right]\right\rangle_{\mathrm{back}}$$

Thus

$$\Lambda_t^\circ = \underbrace{\Lambda_t + \frac{1}{t} \ln \frac{N(0)}{N^\circ(0)}}_{\to 0 \text{ for } t\to\infty} + \frac{1}{t} \ln \left\langle \exp\left[\int_0^t \mathrm{d}t'\, \rho(x(t'))\right]\right\rangle_{\mathrm{back}}$$

- Sampling errors: requires sampling rare lineages
- Bias if dilution and trait are correlated

Inferring population growth and selection

- Empirical average:

$$\Lambda^\circ = \lim_{t\to\infty} \ln \left\{ \frac{1}{n(\sigma{=}1)} \sum_{\ell=1}^{n(\sigma=1)} \exp\left[\int_0^t \mathrm{d}t' \; \rho(x_\ell(t')) \right] \right\}$$

  Requires sampling rare lineages (with high dilution rate)

- Survivor bias: Probability of a trait history $\boldsymbol{x}$:

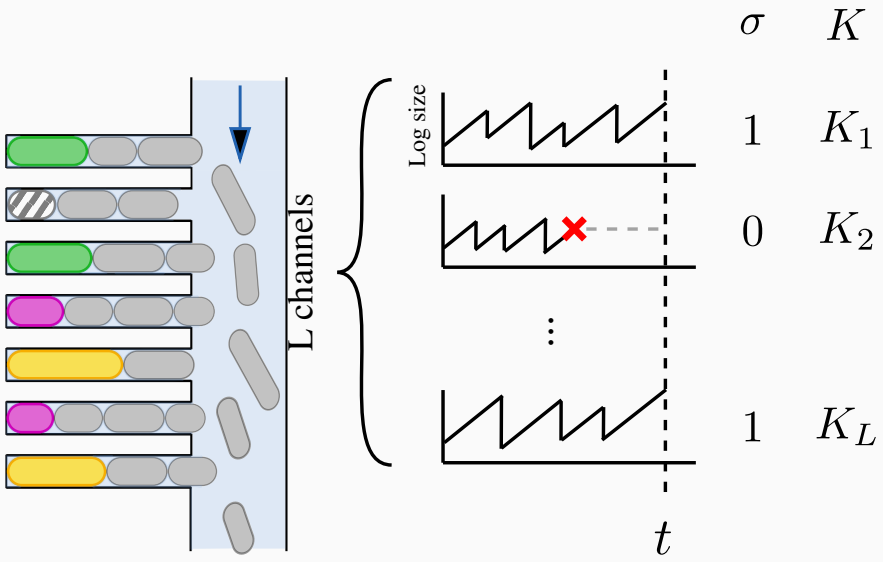
$$p^\circ_{\mathrm{for}}(\boldsymbol{x}) = p_{\mathrm{for}}(\boldsymbol{x}, \sigma{=}1) \exp\left[\int_0^t \mathrm{d}t' \; \rho(x(t')) \right]$$

$$p^\circ_{\mathrm{back}}(\boldsymbol{x}) = p_{\mathrm{back}}(\boldsymbol{x}) \frac{\exp\left[\int_0^t \mathrm{d}t' \; \rho(x(t')) \right]}{\left\langle \exp\left[\int_0^t \mathrm{d}t' \; \rho(x(t')) \right] \right\rangle_{\mathrm{back}}}$$

Inferring population growth and selection

Mother machines: A single lineage is followed in each channel

Only the forward sampling is available



$$p_{\mathrm{surv}}(t) = \frac{n_{\mathrm{lin}}(\sigma{=}1, t)}{L}$$

$$\Lambda_{\mathrm{lin}} = \frac{1}{t} \ln \left[\frac{1}{L} \sum_{j=1}^{L} 2^{K_j} \delta_{\sigma_j,1} \right]$$

Pitfalls

- Finite time: Average over many *independent* lineages to obtain $p_{\mathrm{for}}(K, t)$:

$$\Lambda_t = \frac{1}{t} \ln N(t) = \frac{1}{t} \ln \sum_K 2^K p_{\mathrm{for}}(K, t) = \frac{1}{t} \ln \left\langle 2^K \right\rangle_{p_{\mathrm{for}}}$$

$$= \Lambda + \mathrm{O}\left(\frac{1}{t}\right)$$

- Finite lineages number:
  - Averages are dominated by "exceptional" lineages, that are likely to be lost as time goes by
  - The mean of Λ_t approaches the most likely value of 2^K and eventually behaves as

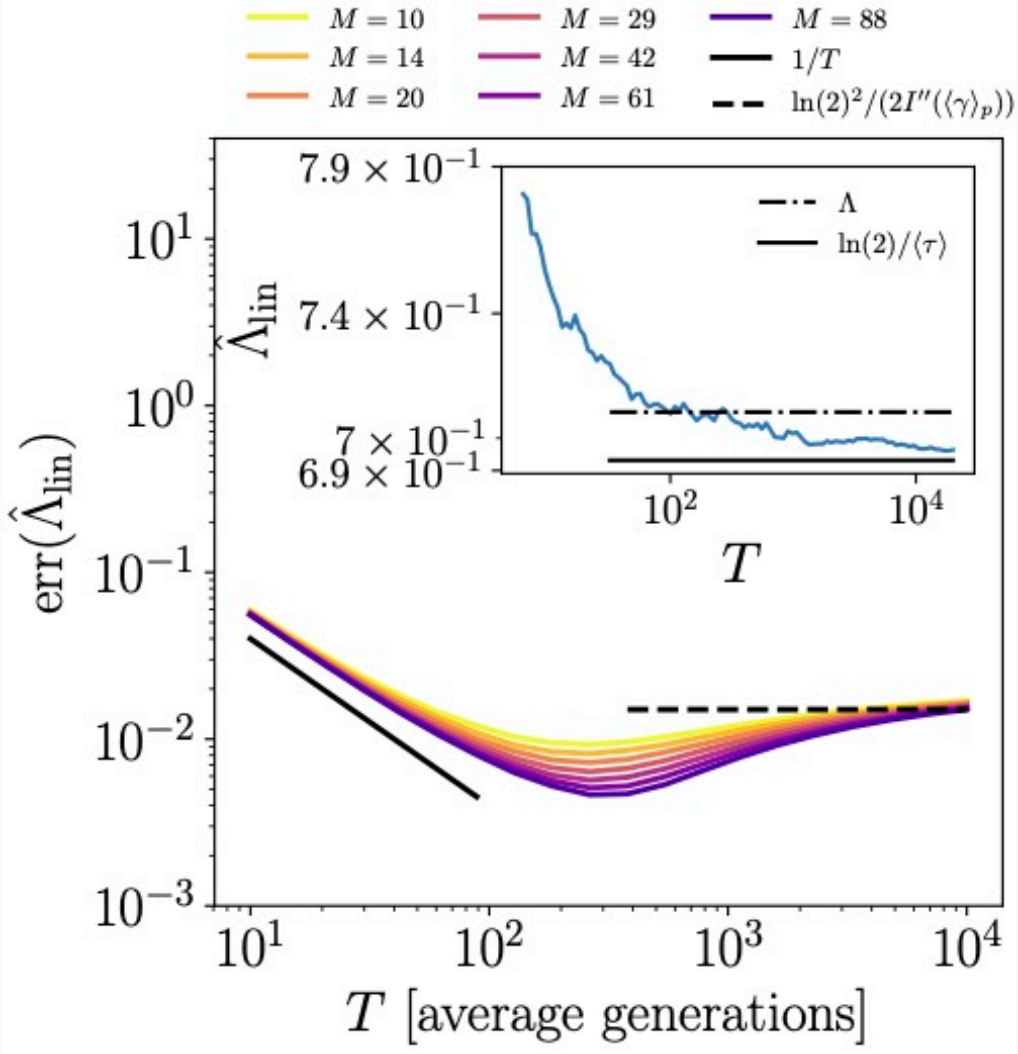
$$\lim_{t \to \infty} \overline{\Lambda_t} = \ln 2 \, r^*$$

    where r^* is the most likely division rate (in the forward ensemble)

For any number L of lineages there is a time window for the best results

Levien et al., 2020

Pitfalls



Levien et al., 2020

Summary

Conclusions

- Lineage statistics provide a useful tool to explore selection in microbial populations
- Comparison of forward and backward statistics provides bounds on the selection strength and other observables
- The method can encompass time-dependent phenotypes (**historical fitness**)
- One can take into account effects of dilution and cell death
- There is an error tradeoff between population size and runtime

Thank you!

References i


R. García-García, A. Genthon, and D. Lacoste.
Linking lineage and population observables in biological branching processes.
Physical Review E, 99:042413, 2019.


A. Genthon and D. Lacoste.
Fluctuation relations and fitness landscapes of growing cell populations.
Scientific Reports, 10:11889, 2020.


A. Genthon and D. Lacoste.
Universal constraints on selection strength in lineage trees.
Physical Review Research, 3:023187, 2021.

References ii

A. Genthon, L. Peliti, T. Nozoe, and D. Lacoste.
Sampling lineage trees with death.
In preparation, 2022.

J. Hermisson, O. Redner, H. Wagner, and E. Baake.
Mutation-selection balance: Ancestry, load, and maximum principle.
Theoretical Population Biology, 62:9–46, 2002.

D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova, and S. J. Tans.
Stochasticity of metabolism and growth at a single-cell level.
Nature, 514:376, 2014.

References iii


T. J. Kobayashi and Y. Sughiyama.
Fluctuation relations of fitness and information in population dynamics.
Physical Review Letters, 115:238102, 2015.


S. Leibler and E. Kussell.
Individual histories and selection in heterogeneous populations.

PNAS, 107:13183–13188, 2010.


E. Levien, T. GrandPre, and A. Amir.
Large deviation principle linking lineage statistics to fitness in microbial populations.
Physical Review Letters, 125:048102, 2020.

References iv


T. Nozoe, E. Kussell, and Y. Wakamoto.
Inferring fitness landscapes and selection on phenotypic states from single-cell genealogical data.
PLOS Genetics, 13:e1006653, 2017.