Cell lineage statistics and fitness with incomplete population trees

A. Genthon, L. Peliti, T. Nozoe and D. Lacoste
Laboratory Gulliver, ESPCI and Tokyo University

ESPCI PARIS
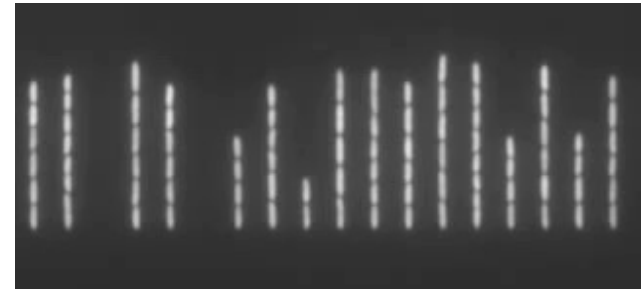
EDUCATION SCIENCE INNOVATION

Various types of single-cell experiments
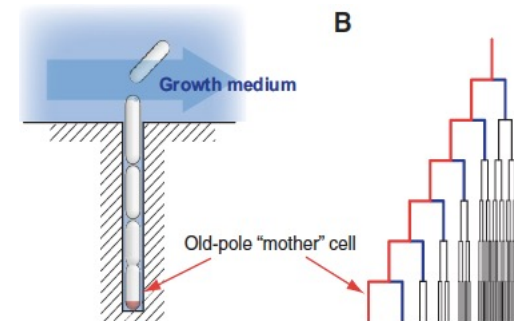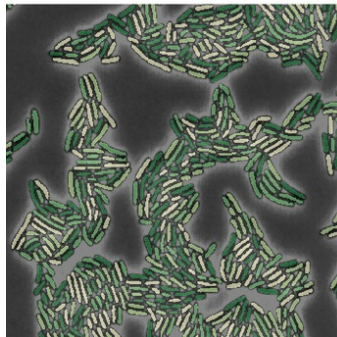
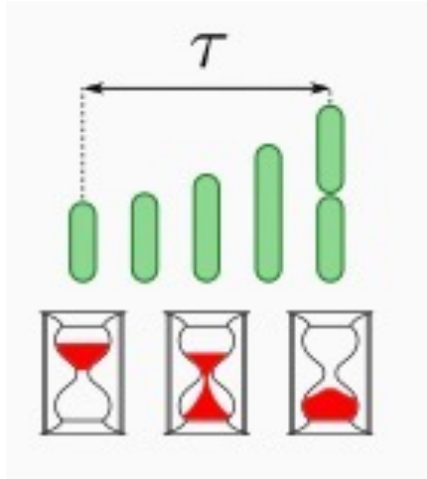**Time-lapse video-microscopy**

Mother machine :

Snapshot of live cells

P. Wang et al. 2010

Statistical biais in a population of cells

Powell relation (1956)

$$f_{pop}(\tau) = 2e^{-\Lambda\tau} f_{lin}(\tau)$$

Λ : population growth rate

Two kinds of averages:

pop : « snapshot » average in a population
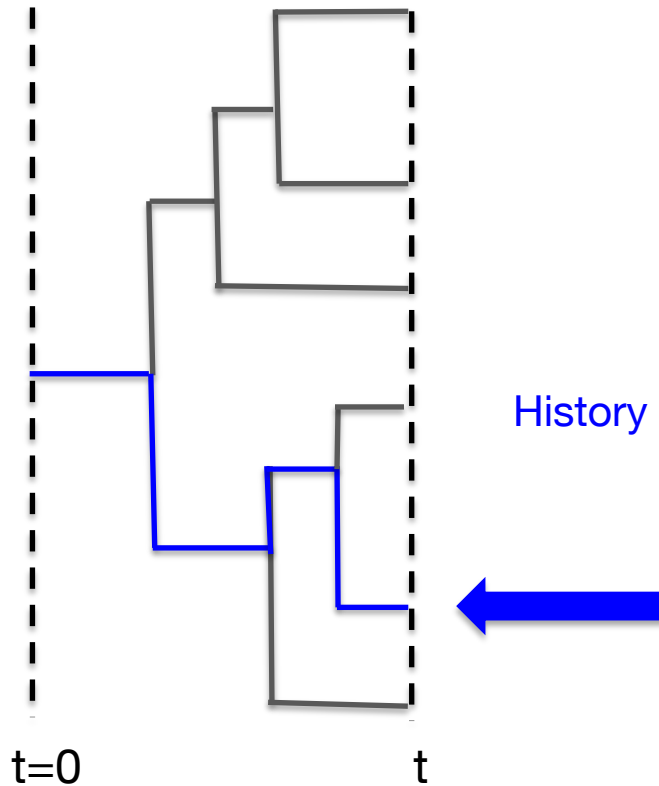*lin* : average along a lineage of cells

Powell inequalities :

$$\langle\tau\rangle_{pop} \leq T_d \leq \langle\tau\rangle_{lin}$$

$$T_d = \ln 2/\Lambda$$ doubling time of the population

Stochasticity of division time affects the population growth rate

Backward (retrospective) sampling

t=0   t

Weight :

$$\omega_{back} = \frac{1}{N(t)}$$

Probability to select a lineage :

$$\mathcal{P}_{back}(K) = \frac{n(K)}{N(t)}$$

Population growth rate :

$$\Lambda_t = \frac{1}{t} \ln \frac{N(t)}{N(0)}$$

History

Forward (chronological) sampling
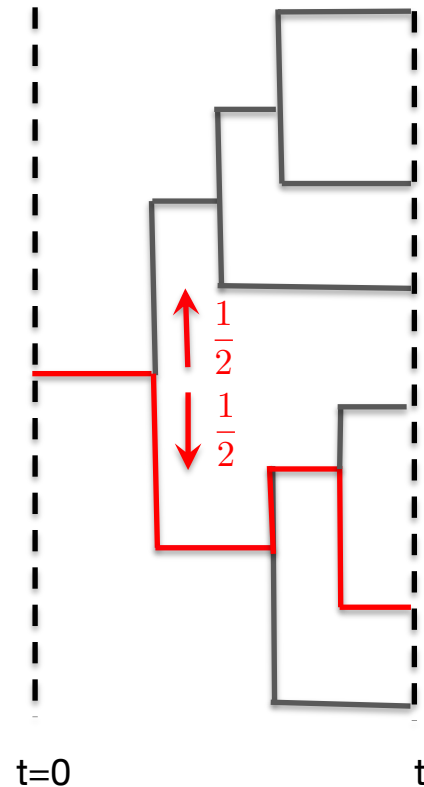
Weight on a lineage with K divisions :

$$\omega_{for}(K) = \frac{2^{-K}}{N(0)}$$

Forward lineage

Probability of a lineage with K divisions :

$$\mathcal{P}_{for}(K) = \omega_{for}(K) \times n(K)$$

number of cells at time t from lineages with K divisions

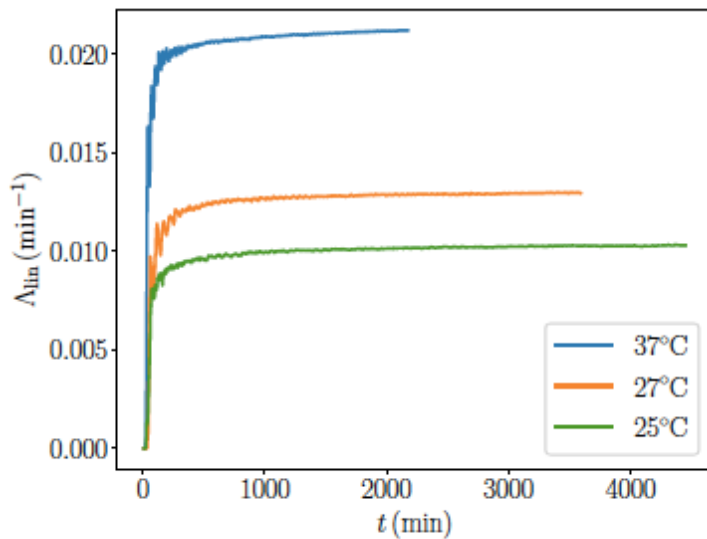t=0   t

Here $\quad n(3) = 4$ with $\omega_{for} = \dfrac{1}{2^3}$, and $n(2) = 2$ with $\omega_{for} = \dfrac{1}{2^2}$
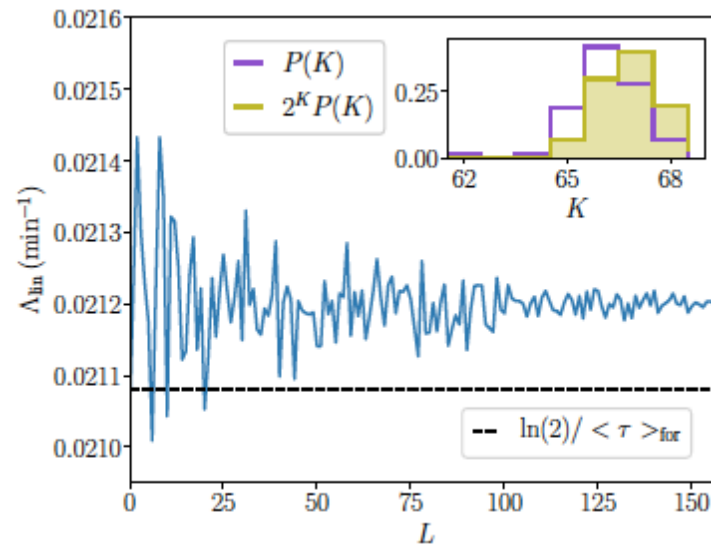
Inference of population growth rate

$$\Lambda_t = \frac{1}{t} \ln \langle 2^K \rangle_{\text{for}}$$

A. Genthon and D. L.,
Sci. Rep., **10**, 11889 (2020)

Mother machine data : Tanouchi et al. (2015)
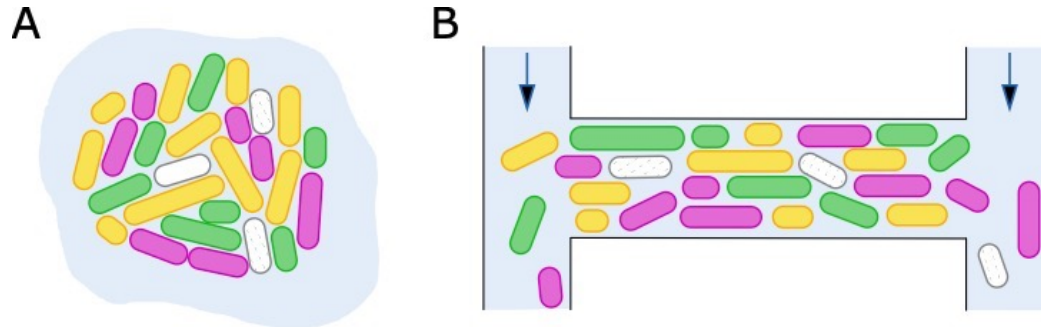

Convergence as function of time

Convergence as function of the number of samples
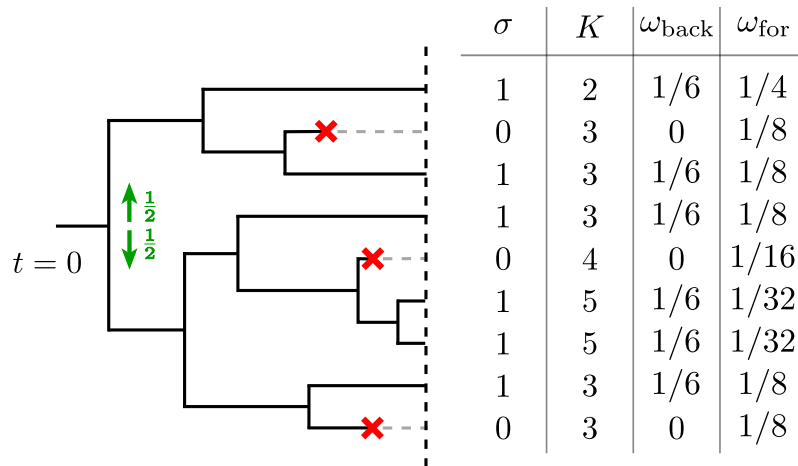
Incomplete cell lineages

- Cells can stop dividing or die because of changes in their environment

- Cells can be flushed away/diluted as in open microfluidic devices

- In either cases, the corresponding cell lineages are incomplete :

How should we treat dead lineages statistically ?

- Dead lineages should have no weight for backward sampling :

σ	K	ω_{back}	ω_{for}
1	2	1/6	1/4
0	3	0	1/8
1	3	1/6	1/8
1	3	1/6	1/8
0	4	0	1/16
1	5	1/6	1/32
1	5	1/6	1/32
1	3	1/6	1/8
0	3	0	1/8

$\sigma = 1$ lineage is alive at time t

$\sigma = 0$ lineage is dead at time t

- Forward distribution conditioned on survival :

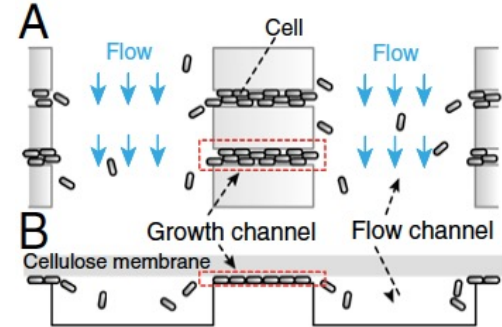$$p_{\text{for}}^{\star}(\cdot, t) = p_{\text{for}}(\cdot, t | \sigma = 1)$$

- Forward probability of survival and its rate Γ_t

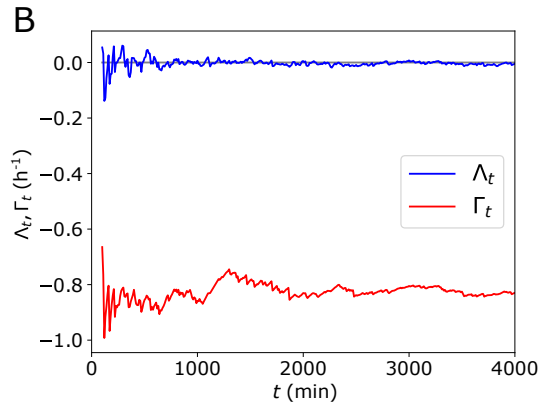$$p_{\text{for}}(\sigma = 1, t) = \sum_K p_{\text{for}}(K, \sigma = 1, t)$$

$$\Gamma_t = \frac{1}{t} \ln p_{\text{for}}(\sigma = 1, t)$$

Cell death induced by dilution

Data from M. Hashimoto et al. 2015 in an open channel that maintains a population of about 40 cells constant despite the loss due to dilution
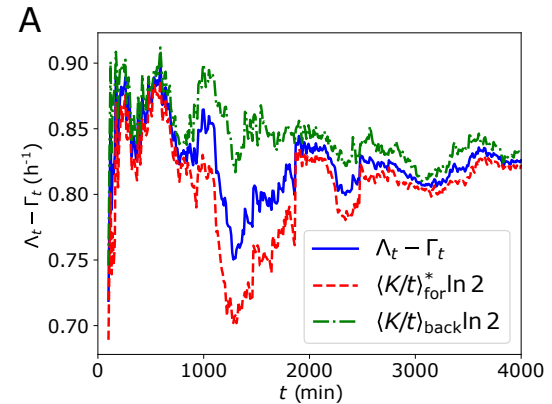
Death is equivalent to a reduction of fitness

$$\Lambda_t = \frac{1}{t} \ln \langle 2^K \rangle_{\text{for}}^\star + \Gamma_t$$

Universal inequalities bound the population growth rate

$$\frac{\langle K \rangle_{\text{for}}^\star \ln 2}{t} \le \Lambda_t - \Gamma_t \le \frac{\langle K \rangle_{\text{back}} \ln 2}{t}$$

The fitness landscape

Given a trait s and n(s,t) the number of lineages displaying this trait at time t :

$$p_{\text{for}}(s,t) = \sum_{K=0}^{\infty} n(s,K,t)/(N_0 2^K) \quad \text{and} \quad p_{\text{back}}(s,t) = n(s,t)/N(t)$$

Fitness landscape h(s) ignores the distinction between dead and surviving lineages :

$$h_t(s) = \Lambda_t + \frac{1}{t} \ln \left[\frac{p_{\text{back}}(s,t)}{p_{\text{for}}(s,t)} \right] \qquad \text{T. Nozoe et al. 2017}$$
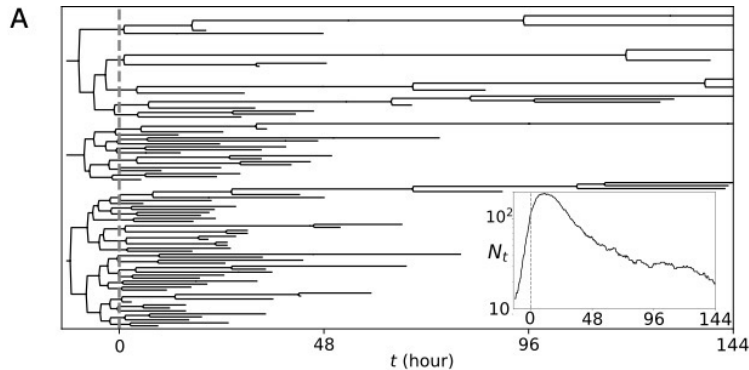
Proper fitness landscape :

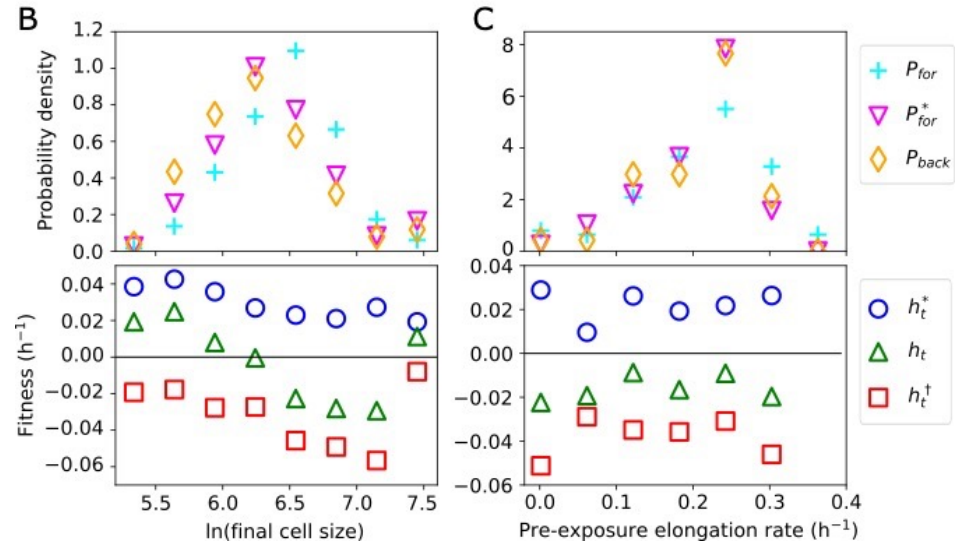$$h_t^*(s) = \frac{1}{t} \ln \left[\sum_K 2^K p_{\text{for}}^*(K,t|s) \right]$$

The survivor bias :

$$h_t^\dagger(s) = h_t(s) - h_t^*(s) = \Gamma_t + \frac{1}{t} \ln \left[\frac{p_{\text{for}}^\star(s,t)}{p_{\text{for}}(s,t)} \right]$$

Cell death induced by drug exposure

Drug exposure

Logarithmic cell-size measured at the end of the lineage (trait 2)

Pre-exposure single cell growth rate (trait 1)

Results of the analysis :

- Independance of trait 1 with survival and with the fitness of lineages

- The decrease of the landscape of trait 2 is not meaningful but is mainly due to the survivor bias

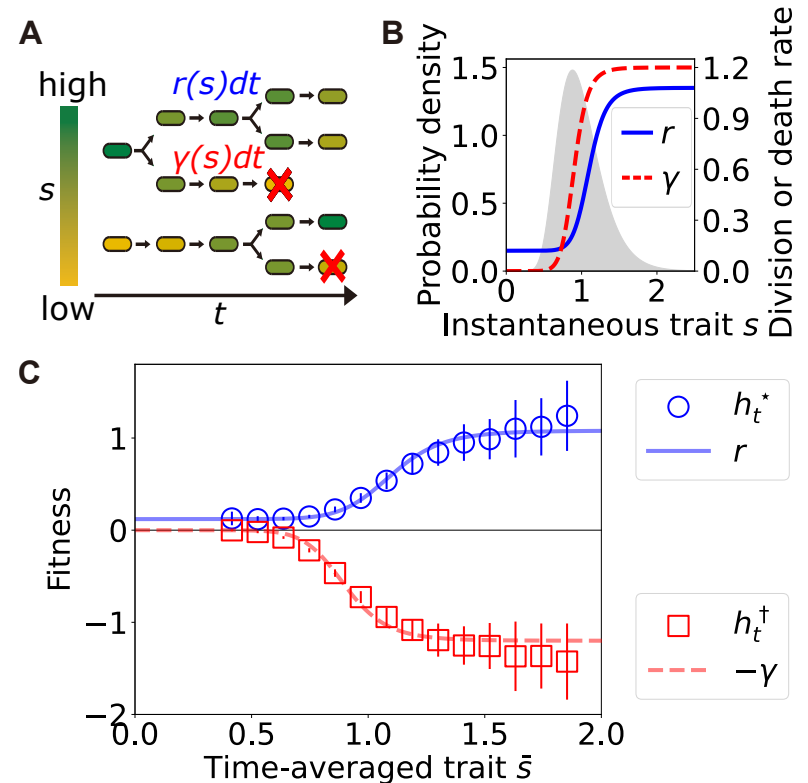# Inference of division and death rates from fitness landscapes

Using the time-averaged trait $\overline{s}$

$$h_t^{\star}(\overline{s}) = r(\overline{s})$$ division rate

$$h_t^{\dagger}(\overline{s}) = -\gamma(\overline{s})$$ death rate

Conditions:

i.   trait can fluctuate in time but should be unaltered at division

ii. Division rate is weakly non-linear or the autocorrelation time of the trait is large with respect to observation time

Strength of selection

- Strength of selection acting on trait s in the absence of death

$$\Pi(s) = \langle h(s) \rangle_{\text{back}} - \langle h(s) \rangle_{\text{for}} \geq 0 \qquad \text{T. Nozoe et al. 2017}$$

  becomes in the presence of death

$$\Pi_{\mathcal{S}} = \langle h_t^{\star} \rangle_{\text{back}} - \langle h_t^{\star} \rangle_{\text{for}}^{\star} \qquad \text{and} \qquad \Pi_{\mathcal{S}} = \Delta\Pi_{\mathcal{S}} + \Pi_{\mathcal{S}}^{\circ}$$

death induced part

Some observations :

- Selection can not be estimated only from growth rates, death rates matter too

- Selection is increased by death only when cells that divide faster also die faster
  but selection can be also decreased or be unaffected by death

- Variability of death rates among lineages leads to fitness gain for the population

Conclusion

- A general framework for the statistical lineage trees

- Inference of selection and fitness from lineage trees data

- Inference of division rate and death rate from lineage statistics

- Death matters for measuring the strength of selection

A. Genthon et al., Arxiv.org/abs/2305.05406 (under review)
R. Garcia-Garcia et al. Phys. Rev. E (2019)
A. Genthon et al., Sci. Rep. (2020)
A. Genthon et al., Phys. Rev. Res. (2021)
A. Genthon, J. Roy. Soc. Interface (2022)
A. Genthon et al., J. Phys. A (2022)