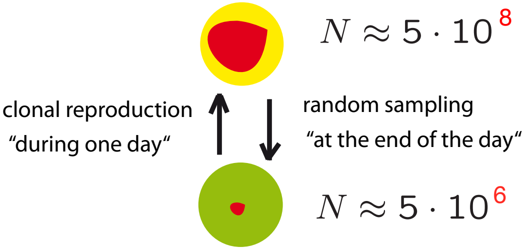
An individual-based model for the Lenski experiment, and the deceleration of the relative fitness

Adrián González Casanova

Joint work with Noemi Kurt, Anton Wakolbinger and Linglong Yuan

27-01-2016

The Lenski experiment (one day cycle)



$N \approx 5 \cdot 10^{8}$

clonal reproduction
"during one day"

random sampling
"at the end of the day"

$N \approx 5 \cdot 10^{6}$

Relative fitness

Measuring adaptation

- A population of size A_0 of the unevolved strain and a population of size B_0 of the evolved strain perform a direct competition.
- The respective population sizes at the end of the day are denoted by A_1 and B_1.
- The (empirical) *relative fitness* $F(B|A)$ of strain B with respect to strain A is

$$F(B|A) \quad = \quad \frac{\log(B_1/B_0)}{\log(A_1/A_0)}.$$

Lenski, Travisano, PNAS, 1994

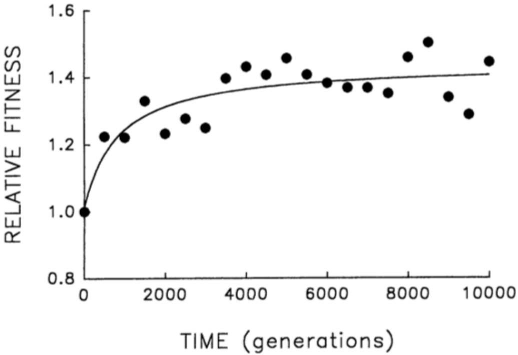
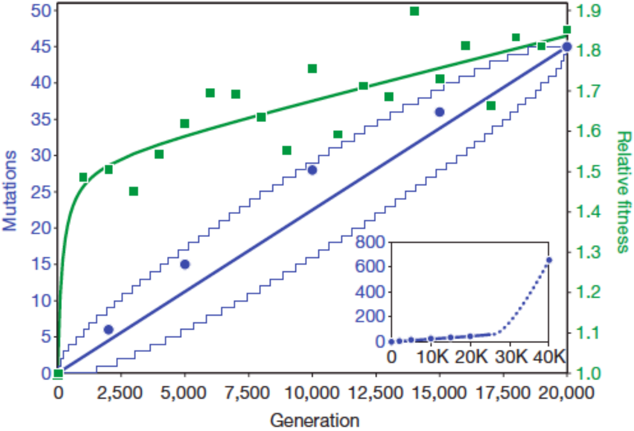


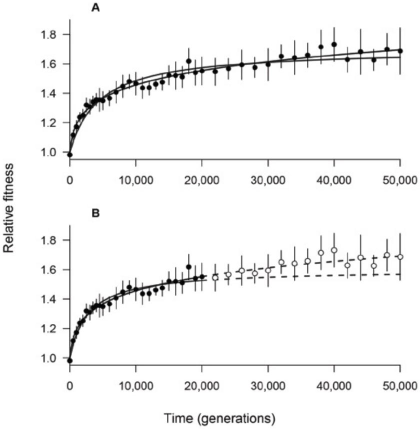
FIG. 4. Trajectory for mean fitness relative to the ancestor in one population of *E. coli* during 10,000 generations of experimental evolution. Each point is the mean of three assays. Curve is the best fit of a hyperbolic model.

Barrick, Yu, Yoon, Jeong, Oh, Schneider, Lenski, Kim, Nature 2009

$$\omega(t) = 1 + at/(t + b)$$



Wiser, Ribeck, Lenski, Science express 13-11-12



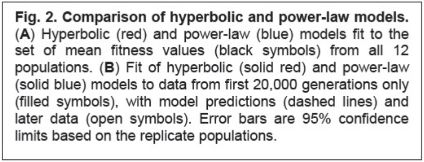


Fig. 2. Comparison of hyperbolic and power-law models. (**A**) Hyperbolic (red) and power-law (blue) models fit to the set of mean fitness values (black symbols) from all 12 populations. (**B**) Fit of hyperbolic (solid red) and power-law (solid blue) models to data from first 20,000 generations only (filled symbols), with model predictions (dashed lines) and later data (open symbols). Error bars are 95% confidence limits based on the replicate populations.

$$w(t) = (1 + ct)^{1/2g}$$

Questions

- Which curve describes better the trajectory of the relative fitness?
- Why is the relative fitness decelerating?

Possible explanations

- Clonal interference
- Epistasis
- The design of the experiment

The daily cycle model[1]

Information about the experiment

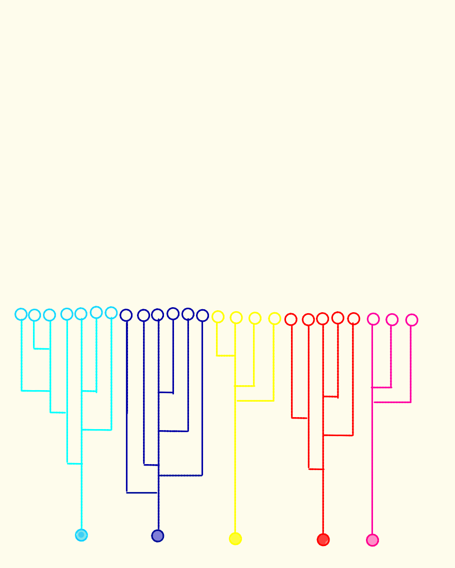
- At the beginning of each day there are N individuals.
- Within each day individuals split at constant rate.
- The reproduction process will stop when the glucose has been consumed. (This happens when there are around $100N$ individuals).
- N individuals out of the $\sim 100N$ are uniformly sampled without replacement, to be starting individuals at the next day.

---

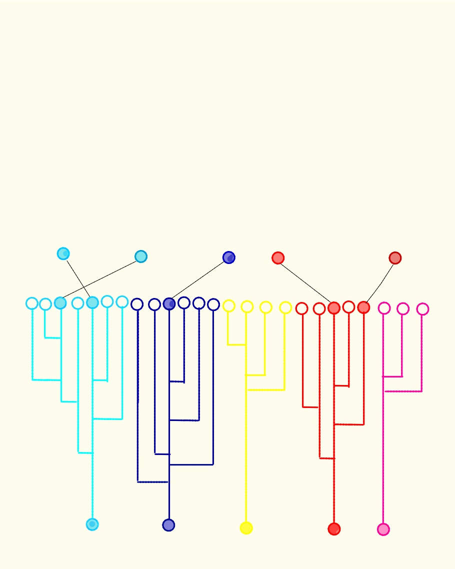
[1] An individual-based model for the Lenski experiment, and the deceleration of the relative fitness. Adrian Gonzalez Casanova, Noemi Kurt, Anton Wakolbinger and Linglong Yuan. (2015) arXiv 1505.0175

Pruned Yule trees

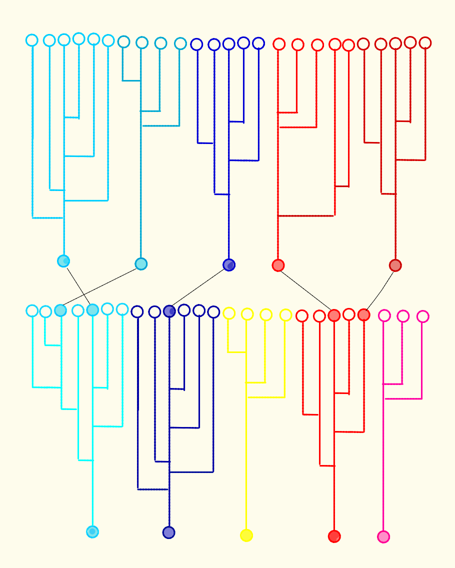
Pruned Yule trees



Pruned Yule trees



Pruned Yule trees



Inside a day

Let $Y_{i,j}(t)$ be a pure birth process with rate $r \in \mathbb{R}^+$, for every $i \in \mathbb{N}$ and $j \in \{1, 2, ..., N\}$. (Yule Processes with parameter r).
The total population size at time t of day i is

$$\sum_{j=1}^{N} Y_{i,j}(t)$$

Stopping rule

Each day, the reproduction stops at time σ , where

$$\sigma = \inf\{t : E[\sum_{j=1}^{N} Y_{i,j}(t)] = \gamma N\}.$$

Stopping rule

Each day, the reproduction stops at time σ , where

$$\sigma = \inf\{t : E[\sum_{j=1}^{N} Y_{i,j}(t)] = \gamma N\}.$$

Note that $E[\sum_{j=1}^{N} Y_{i,j}(t)] = Ne^{rt}$, so

$$\sigma = \frac{\ln(\gamma)}{r}.$$

Stopping rule

Each day, the reproduction stops at time σ , where

$$\sigma = \inf\{t : E[\sum_{j=1}^{N} Y_{i,j}(t)] = \gamma N\}.$$

Note that $E[\sum_{j=1}^{N} Y_{i,j}(t)] = Ne^{rt}$, so

$$\sigma = \frac{\ln(\gamma)}{r}.$$

The total population size at the end of the day is

$$\sum_{j=1}^{N} Y_{i,j}(\sigma) \sim \gamma N.$$

Sampling rule

To go from one day to the next, we sample uniformly at random N individuals (out of $\sim \gamma N$), and we say that each sampled individual is a root of a Yule tree in the next day.

Introducing selective advantage

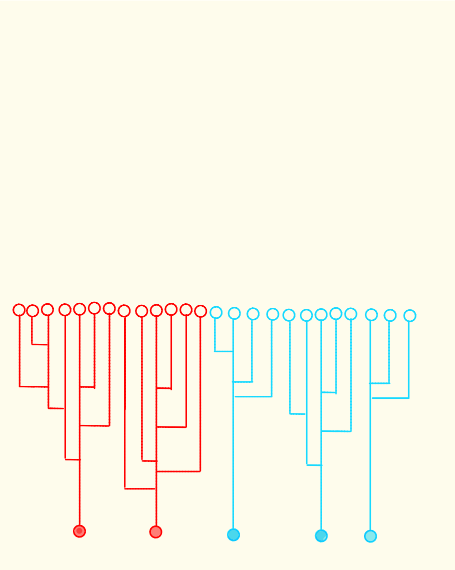
- Assume that some individuals reproduce at rate $r + \varrho_N$ (mutants), while other reproduce at rate r (basis population).

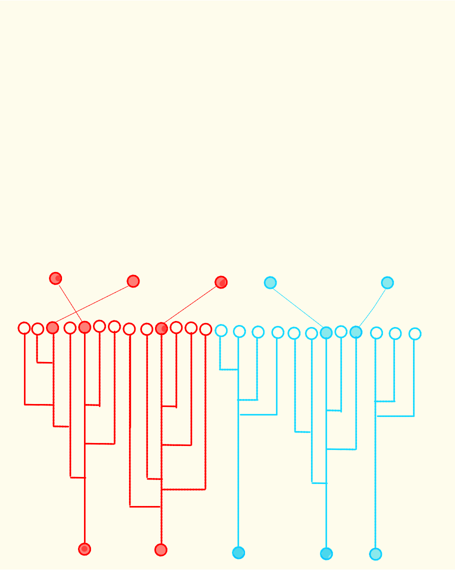
Introducing selective advantage

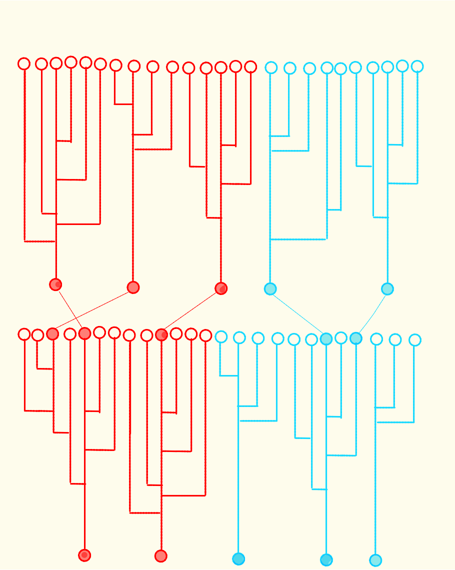
- Assume that some individuals reproduce at rate $r + \varrho_N$ (mutants), while other reproduce at rate r (basis population).
- Let $M_i(t)$ be the number of mutants at time t of day i.
- We are interested in the process

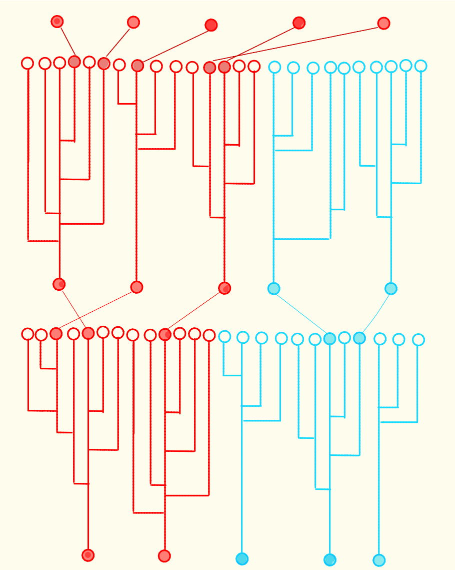
$$\{K_i\}_{i\in\mathbb{N}} := \{M_i(0)\}_{i\in\mathbb{N}},$$

which is constructed recursively using uniform sampling.







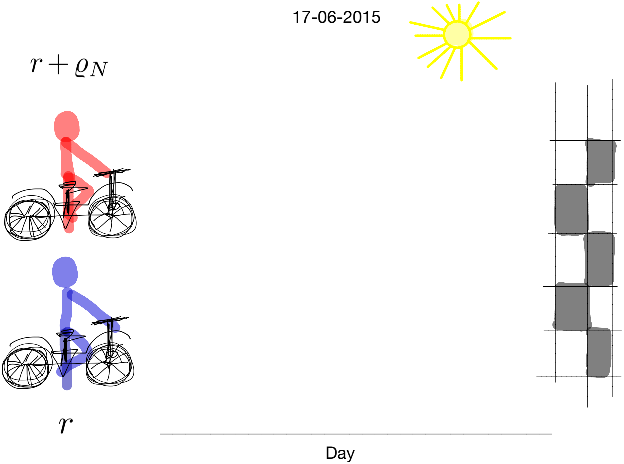


Selective advantage

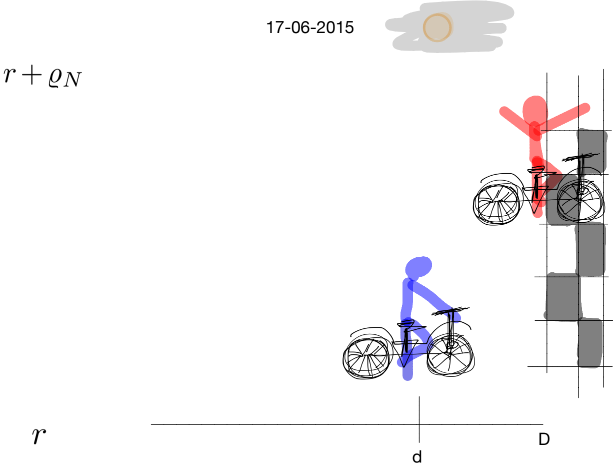
Basic population reproduces at rate r.
Mutants reproduce at rate $r + \varrho_N$.

$$\mathbb{E}[K_1 | K_0 = 1] = 1 + \varrho_N \frac{\log \gamma}{r} + o(\varrho_N).$$

Race until the sun is gone



Race until the sun is gone

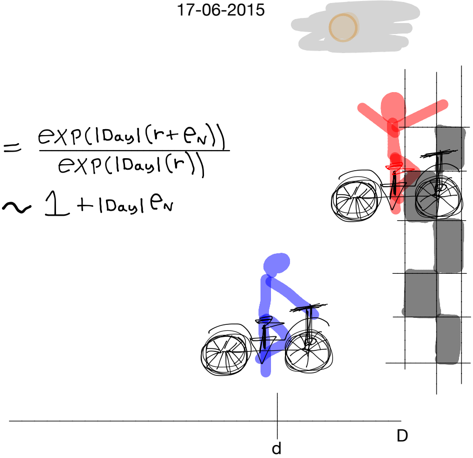


Race until the sun is gone

$r + \varrho_N$

$$\frac{\exp(D)}{\exp(d)} = \frac{\exp(|Day|(r+e_N))}{\exp(|Day|(r))}$$

$$\sim 1 + |Day|\, e_N$$

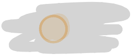


17-06-2015

r

Race until the sun is gone

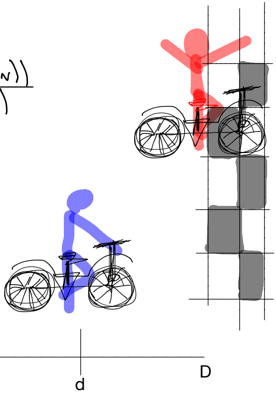
17-06-2015



$r + \varrho_N$

$$\frac{\exp(D)}{\exp(d)} = \frac{\exp(|Day|(r+e_N))}{\exp(|Day|(r))}$$

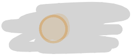
$$\sim 1 + 23 e_N$$



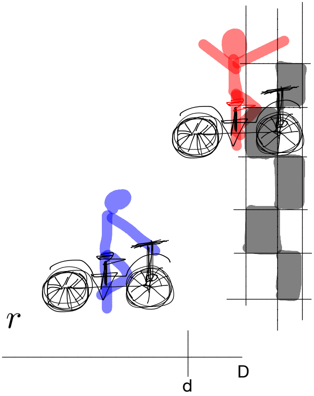
r

Race until the sun is gone

17-12-2015



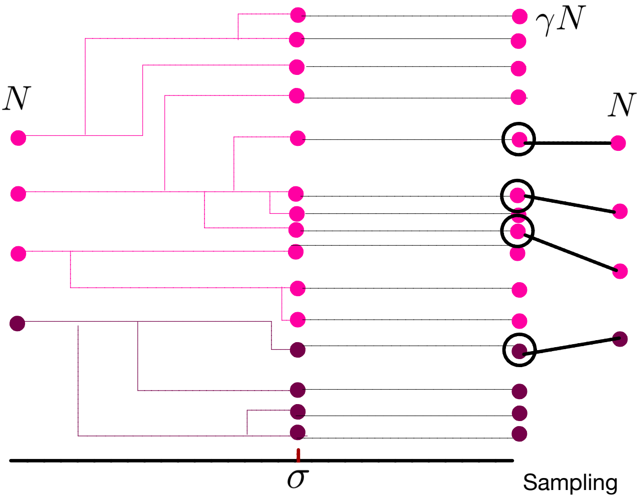
$r + \varrho_N$



$$\frac{\exp(D)}{\exp(d)} = \frac{\exp(|\mathrm{Day}|(r+\varrho_N))}{\exp(|\mathrm{Day}|(r))}$$

$$\sim 1 + 1\,\varrho_N$$

The *effective competition time*, and its dependence on *r*



Let $\pi_N := \mathbb{P}\big(\exists i \in \mathbb{N} : K_i = N \,|\, K_0 = 1\big)$, and $\tau^N := \tau^N_{\mathrm{fix}} \wedge \tau^N_{\mathrm{ext}}$.

Theorem (Probability and speed of fixation)

Under the assumptions of our model, as $N \to \infty$,

$$\pi_N \sim \frac{\gamma}{\gamma - 1} \frac{\varrho_N \log \gamma}{r}.$$

Moreover, for any $\delta > 0$ *there exists* $N_\delta \in \mathbb{N}$ *such that for all* $N \geq N_\delta$

$$\mathbb{P}(\tau^N > \varrho_N^{-1-3\delta}) \leq (7/8)^{\varrho_N^{-\delta}}.$$

The weak mutation - moderate selection model (Assumption A)

i) Beneficial mutations add ϱ_N to the reproduction rate of the individual that suffers the mutation.

ii) In each generation, with probability μ_N there occurs a beneficial mutation. The mutation affects only one (uniformly chosen) individual, and every offspring of this individual also carries the mutation.

iii) There exists $0 < b < 1/2$, and $a > 3b$, such that $\mu_N \sim N^{-a}$ and $\varrho_N \sim N^{-b}$ as $N \to \infty$.

$$\mu_N << \varrho_N$$

We define the fitness of the population at the beginning of day i with respect to that at the beginning of day 0 as

$$F_i := \frac{\log \frac{1}{N} \sum_{j=1}^{N} e^{R_{i,j} t}}{\log e^{r_0 t}}$$

where $R_{i,j}, j = 1, \ldots, N$ are the reproduction rates of the individuals present at the beginning of day i, and t is a given time for which the two populations are allowed to grow together.

If the whole population reproduces at the same rate (R_i), then

$$F_i = \frac{R_i}{r_0}$$

where $r_0 := R_0$.

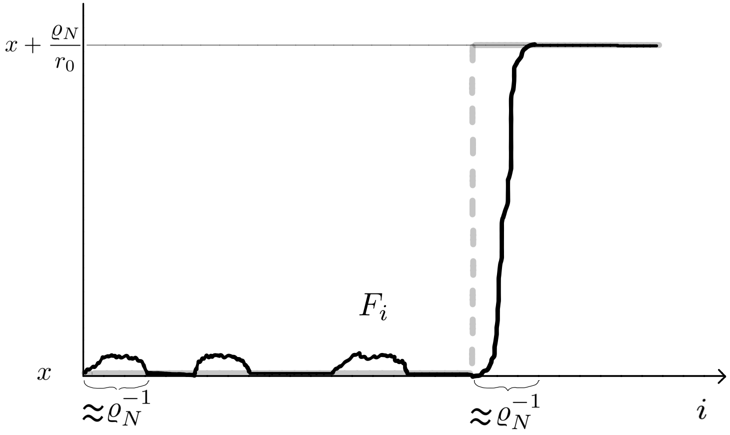


Figure: The number of attempts to go to fixation, when the reproduction rate of the basic poulation is x, is distributed Geometric with parameter $\pi_N \sim \varrho_N \frac{C(\gamma)}{x}$.

Theorem (Convergence of the relative fitness process)

Assume $R_{0,j} = r_0$ *for* $j = 1, \dots, N$*, and let* $(F_i)_{i \in \mathbb{N}_0}$ *be the process of relative fitness. Then under Assumption A, the sequence of processes* $(F_{\lfloor (\varrho_N^2 \mu_N)^{-1} t \rfloor})_{t \geq 0}$ *converges in distribution as* $N \to \infty$ *locally uniformly to the deterministic function*

$$f(t) = \sqrt{1 + \frac{\gamma \log \gamma}{\gamma - 1} \frac{2t}{r_0^2}}, \ \ t \geq 0\,.$$

Table: Our model compared with Wiser et al.

	Our model	Wiser et al
Clonal interference	No	Yes
Epistasis	No	Yes
Design of the experiment	Yes	No
	$f(t) = (1 + \frac{2C(\gamma)t}{r_0^2})^{1/2}$	$w(t) = (1 + ct)^{1/2g}$

Thank you!