BIO-463
Genomics and bioinformatics

Lecture 6:
Population genetics beyond selection & drift

Prof. Anne-Florence Bitbol

EPFL

Lemanic Life Sciences Hackathon

- Work on 12 exciting projects
- Develop new skills
- Collaborate with like-minded people
- Get expert mentorship
- Get inspired by talks from Alumni
- Pitch projects and win prizes
- Have fun during your Easter break

April 23 - 25, 2025

Hall SV and SV 1717

Reminder: Information about the first assignment

- **Assignment released on March 20, problem class on March 25, deadline on March 28**

This assignment is a graded problem set, and will count for **25% of your final grade**

You can discuss with TAs and with fellow students about the problem set, but in the end, you should hand in a **personal solution**
Detected plagiarism will result in a reduction of your grade

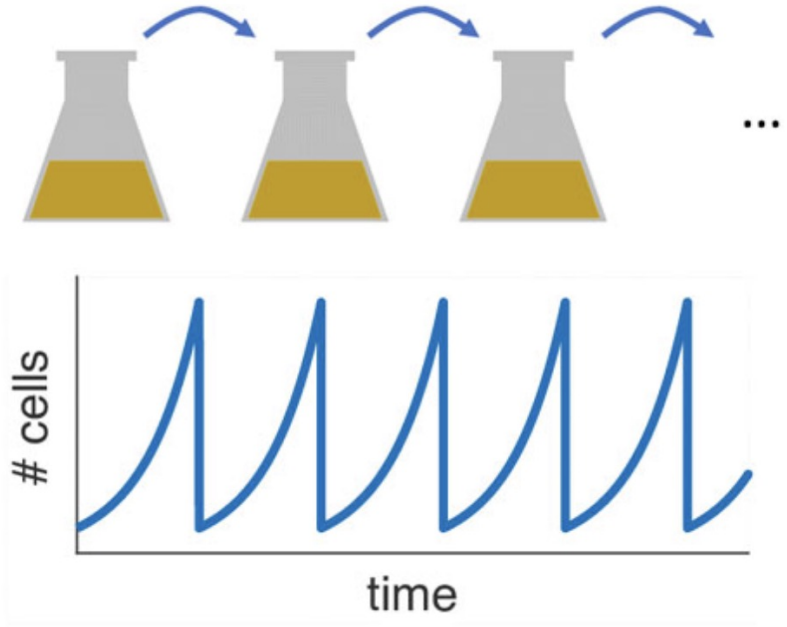
The expected language is R, as in all the class BIO-463 – some R functions and libraries may be recommended in the problems

Please hand in your solution in two versions of the same file:
• one should be your **source file**, with the following format: **.Rmd, .qmd**
• the other one should be the **html file** deriving from your source file

You will have to hand in your solution **via Moodle by Friday March 28**

Evolution experiments

- **Experimental protocol: serial transfer (or serial passage, or serial dilution)**



1- Cells are placed in culture medium and grow (exponentially – may or may not reach stationary phase)

2- Periodically, a small volume is sampled and placed in new medium – the rest is discarded → bottleneck

Phases 1&2 are repeated

Assume that the bottleneck has constant size K

Evolution experiments & population genetics

- **Modeling serial transfer (or serial passage, or serial dilution)**

1- Growth phase → deterministic exponential growth with no death, starting from K cells

Starting from mutant fraction $x_n=k_n/K$ at bottleneck n, the fraction after growth reads $x'_n = \frac{x_n e^{st}}{1+x_n(e^{st}-1)}$

Introducing $\sigma = e^{st}-1$, we can write $x'_n = \frac{(1+\sigma)x_n}{1+\sigma x_n}$ → as in the Wright-Fisher model, with σ instead of s

2- Transfer / bottleneck → binomial sampling of K individuals from the grown population

Binomial sampling: $P(k_{n+1}) = \binom{K}{k_{n+1}} (x'_n)^{k_{n+1}} (1-x'_n)^{K-k_{n+1}}$ → as in the Wright-Fisher model

where k_{n+1} is the number of mutants at bottleneck n+1

Mutant fixation probability: as in the Wright-Fisher model, the diffusion approximation gives

$\rho(1/K) = \frac{1-e^{-2\sigma}}{1-e^{-2K\sigma}} = \frac{1-e^{-2st}}{1-e^{-2Kst}}$ starting from one mutant at a bottleneck (fraction 1/K)

K>>1, |σ|<<1 K>>1, |s|t<<1

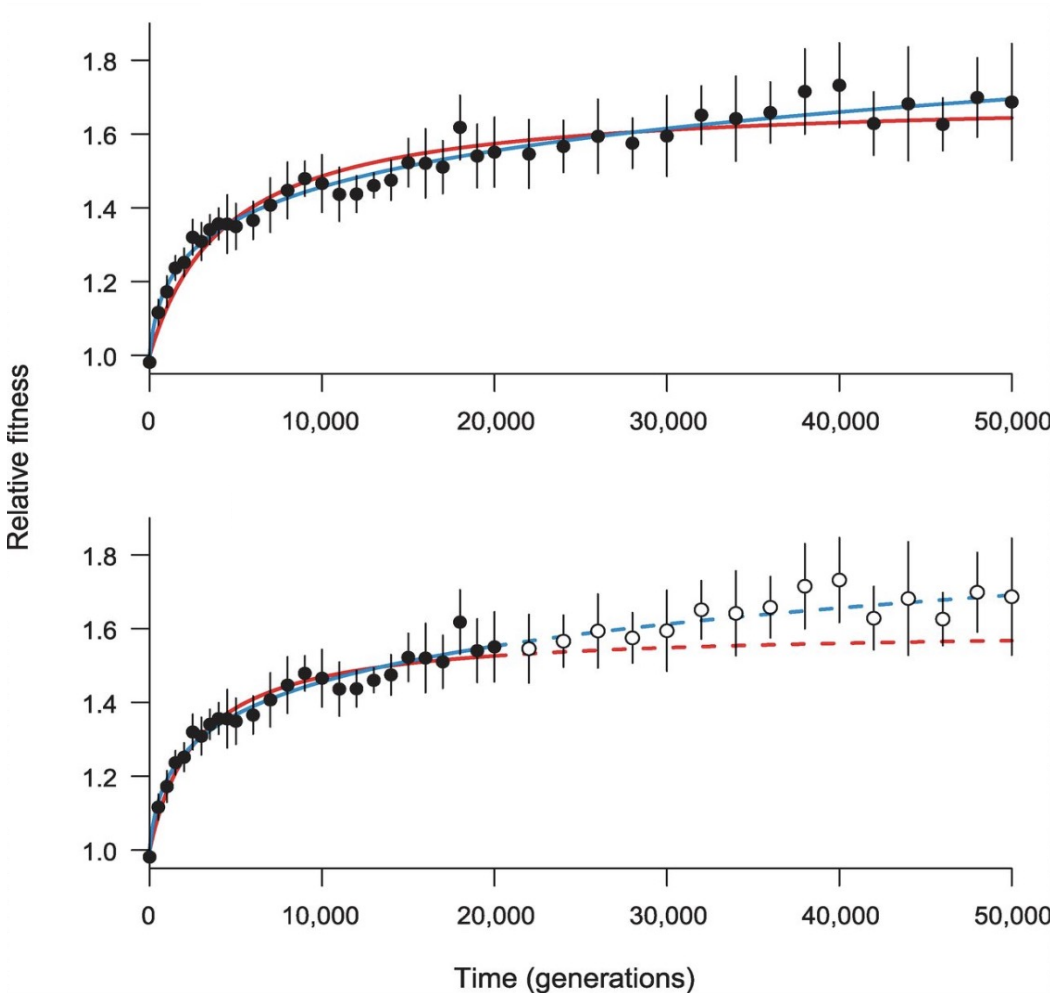
Evolution experiments

- **The Lenski long-term evolution experiment (LTEE), started in 1988**

12 initially identical populations of *E. coli*, serial transfer every day (1% is transferred)
6.64 generations (doublings) per day → >75,000 generations
Population A-3 has acquired the ability to metabolize citrate

Evolution experiments

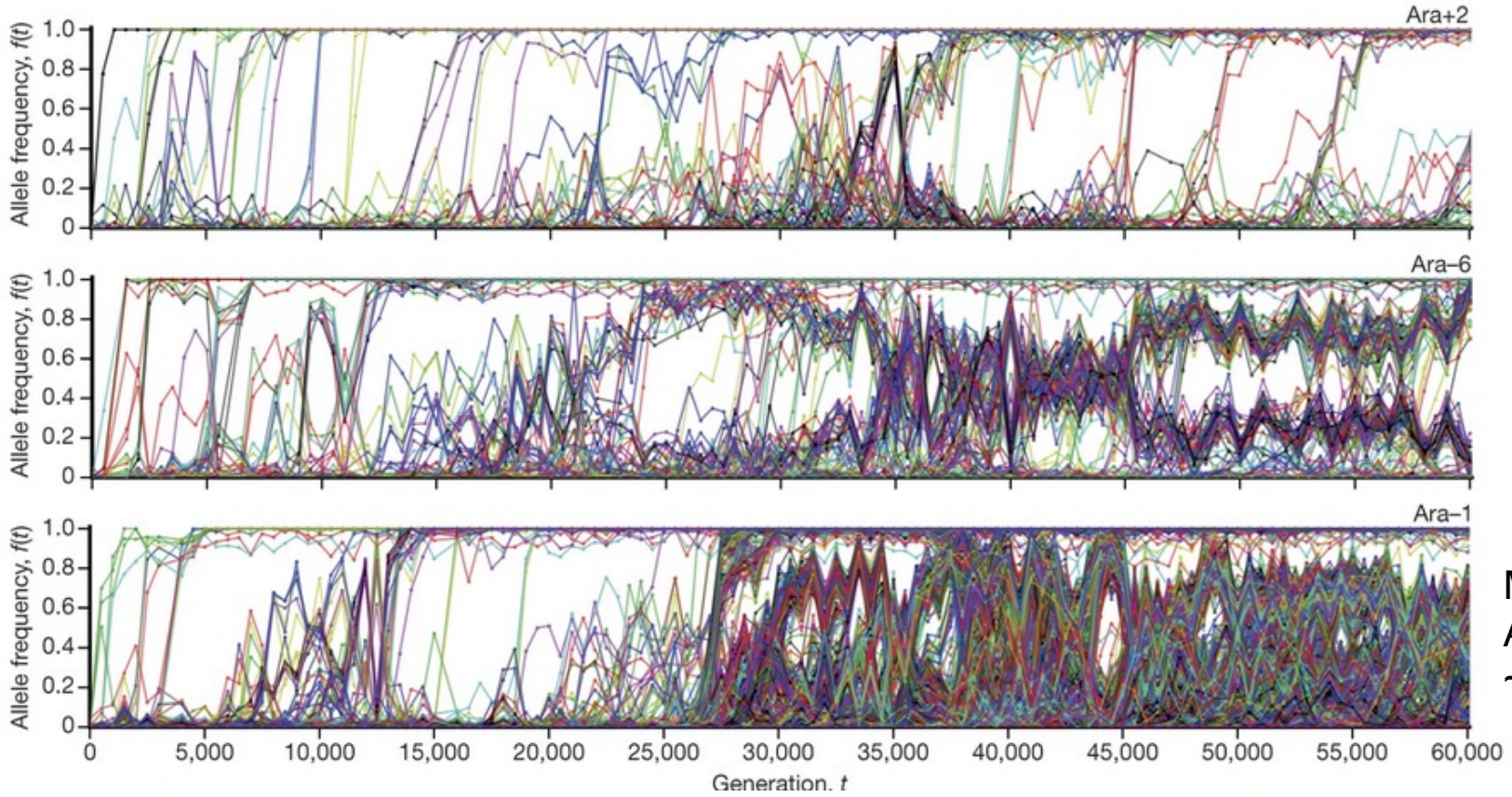
- **The Lenski long-term evolution experiment (LTEE), started in 1988 – Wiser et al, 2013**



- Fitness keeps increasing
- But the increase is getting slower

Evolution experiments

- **The Lenski long-term evolution experiment (LTEE), started in 1988 – Good et al, 2017**

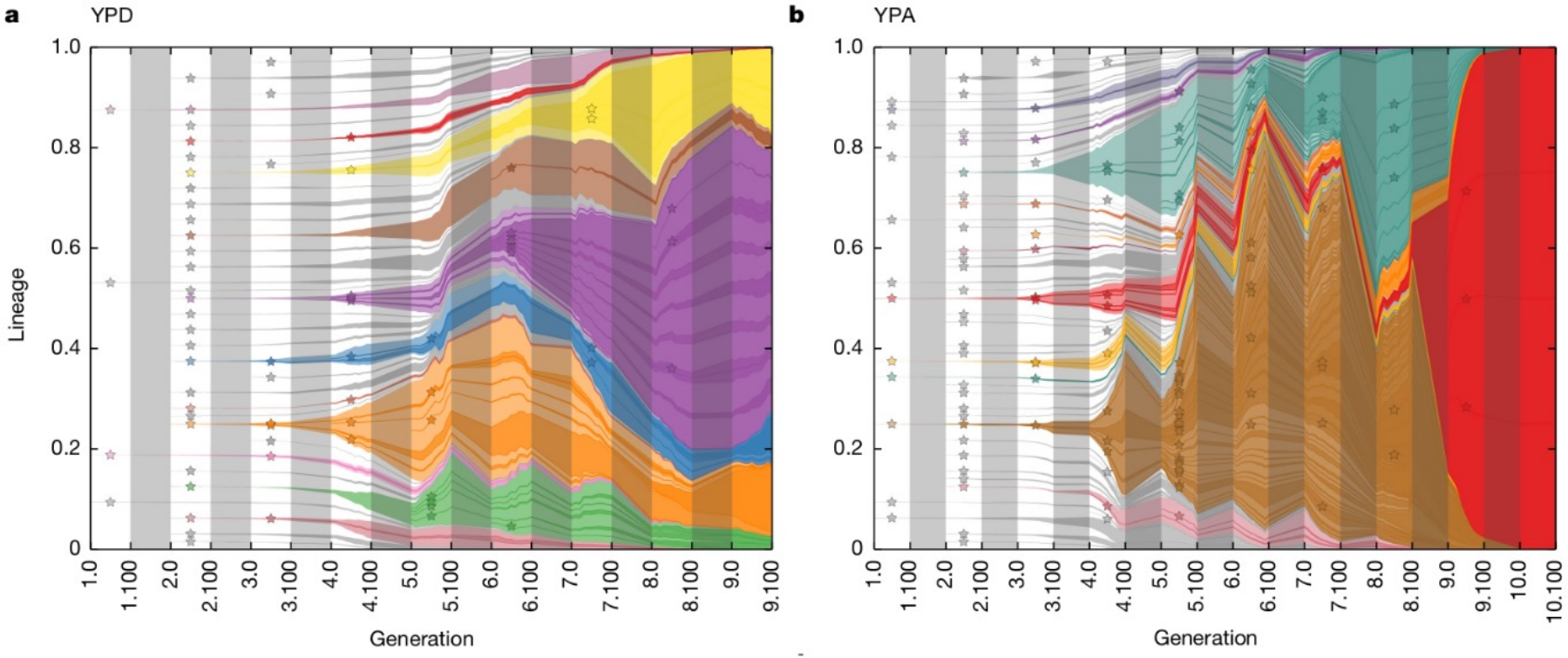


Mutator evolved in Ara-1 at generation ~28,000

Frequency trajectories of all mutations in 3 replicate populations
→ Different dynamics at the genetic level (while overall fitness increase is similar)

Evolution experiments

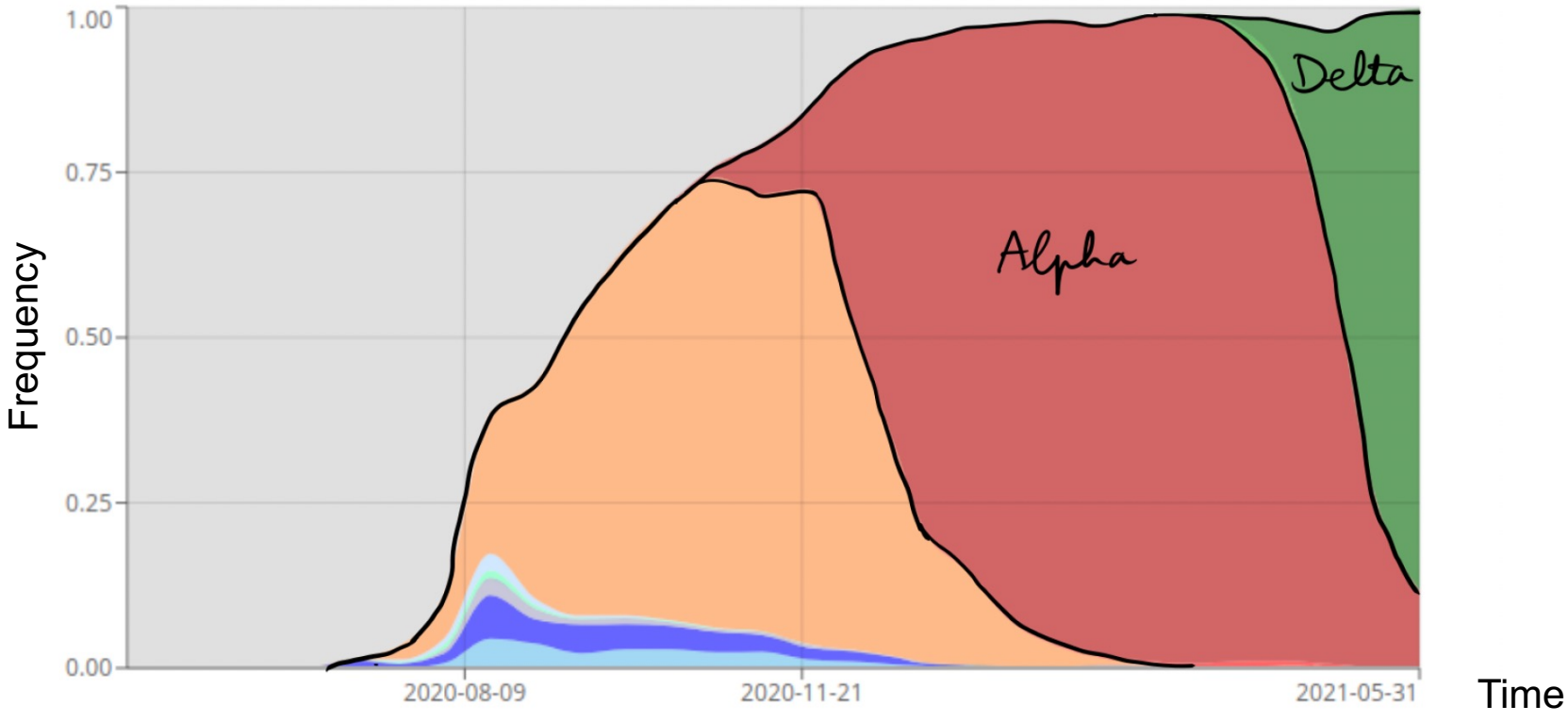
- **Evolving yeast population – Nguyen Ba et al 2019**



Stacked frequencies of barcoded lineages in a population of yeast versus time.
New barcodes are added in the gray phases.
(a): Yeast in a rich medium (YPD). (b): Same rich medium + added acetic acid (YPA).

Evolution in natural populations

- **Example: Covid**



SARS-CoV-2 variants in the UK
Composition of the population of SARS-CoV-2 viruses infecting patients in the UK versus time, from sequencing data. Each color represents a different variant strain.
https://covariants.org/

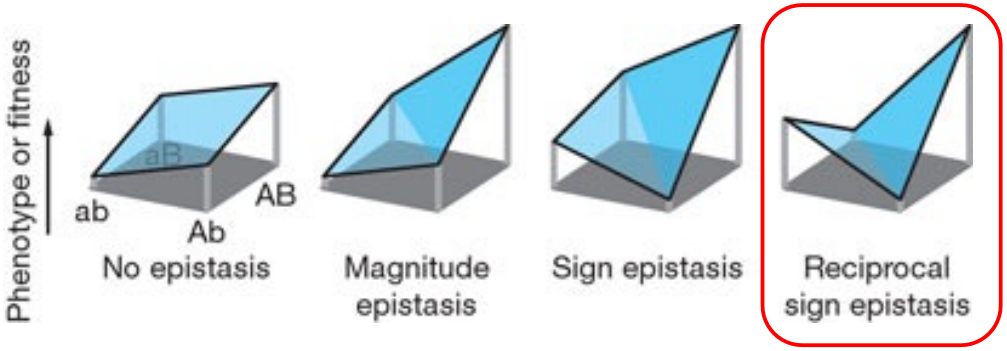
Overview: effects at play

- **Different effects so far**
  - **Mutations:**
    - generate diversity
    - most have small fitness effects, most are deleterious
  - **Natural selection:**
    - acts upon random mutations
    - because of it, beneficial mutations tend to take over and fix
  - **Genetic drift:**
    - corresponds to stochastic fluctuations
    - arises from finite population size (total population or mutant population)
    - means that moderately beneficial mutations often do not fix – new mutations start in a single individual
- **Additional important effects**
  - Recombination (horizontal gene transfer; sexual reproduction)
  - Interactions between mutations – not just additive fitness effects
  - Specific interactions between individuals (beyond mere competition) – cooperativity, attacks…
  - Spatial population structure, migrations and genetic flow
  - Environmental variability

Epistasis

- **The effect of a mutation depends on the context**

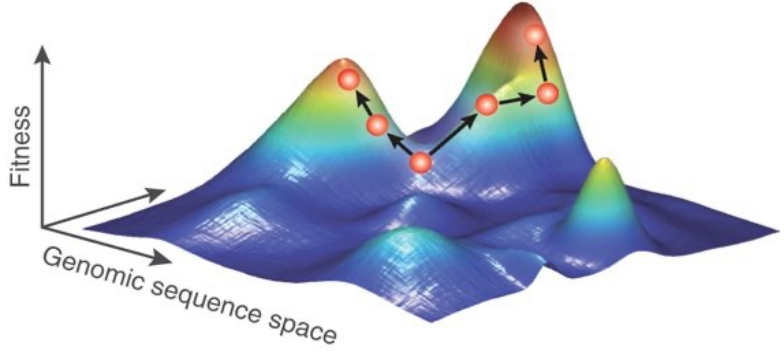
Consider 2 loci with 2 possible states each: A, a (first locus) and B, b (second one)
The fitness effect of A → a can depend on whether the organism possesses B or b on the second locus
This context-dependence (or interaction between mutations) is called epistasis



Can give rise to multiple fitness maxima

Poelwijk et al 2007

- **Fitness landscapes can be rugged**



Fitness landscape: fitness versus genotype (fixed environment) – Wright (1930s)

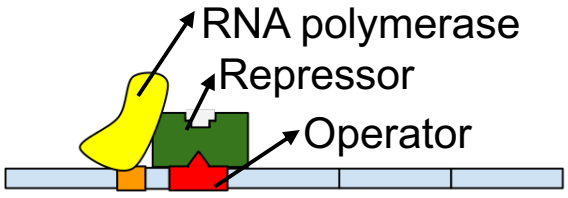
Reciprocal sign epistasis → ruggedness
Not obvious that the maximum of fitness will be reached
Implications for predictability of evolution

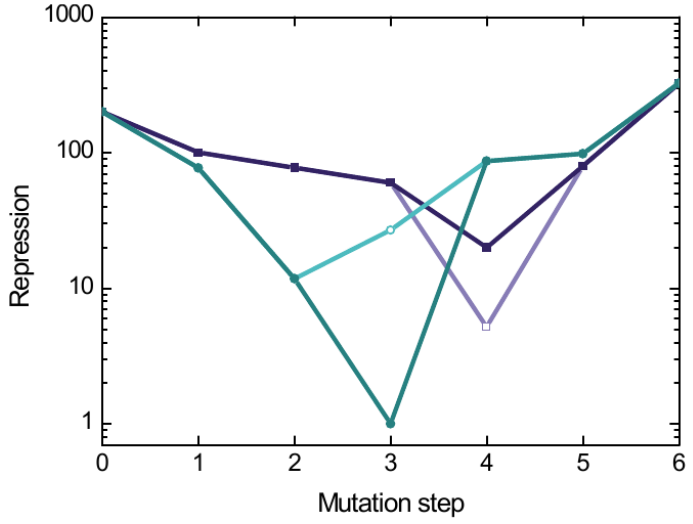
Epistasis

- **Molecular example**

  Co-evolving systems → fitness valleys

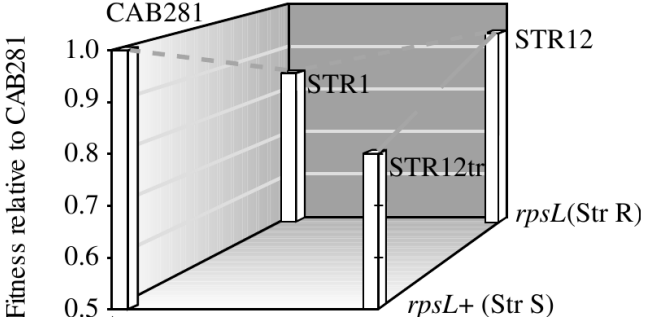
  The *lac* operon:





Dawid et al 2010

- **Fitness costs in the evolution of antibiotic resistance**

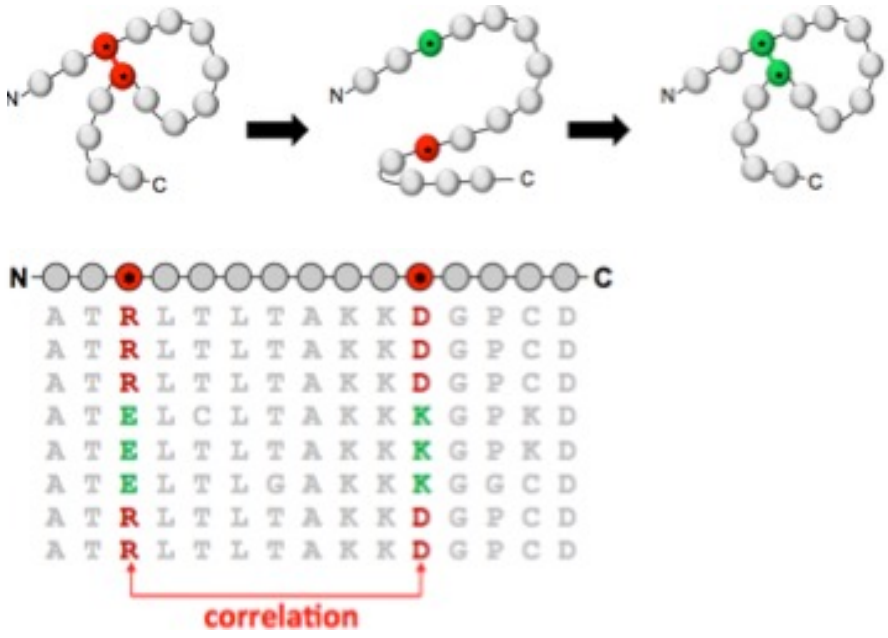


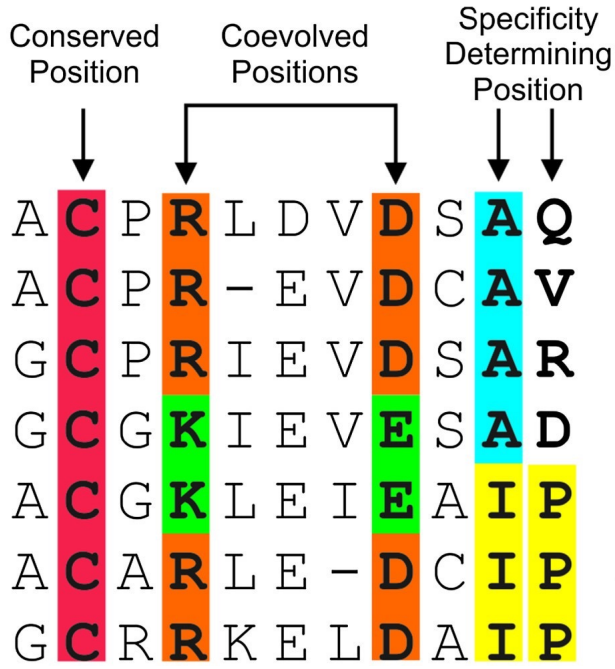
Evolution of streptomycin resistance in *E. coli*

Schrag et al 1997

Epistasis within a protein

- **Multiple sequence alignments and epistasis**





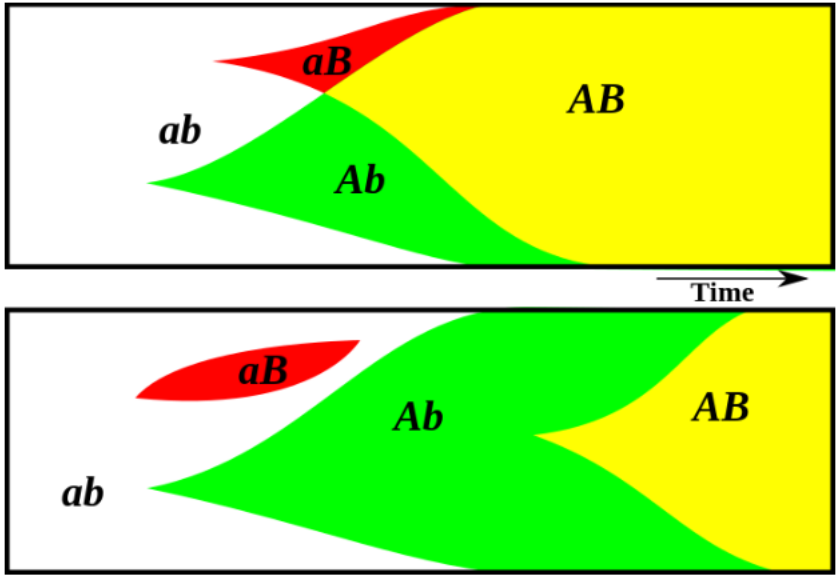
Teppa et al 2012

Epistasis within a protein gives rise to correlated sites (columns) in alignments
(Remark: the compensatory mutation picture is a simplification)

Recombination

- **Impact of recombination on evolution**

Recombination: exchange of genetic material between organisms, leading to offspring with combinations of traits that differ from parental ones
Sexual reproduction; horizontal gene transfer
Avoids clonal interference by allowing recombination of different beneficial mutations that appeared in different strains



a → A and b → B are beneficial

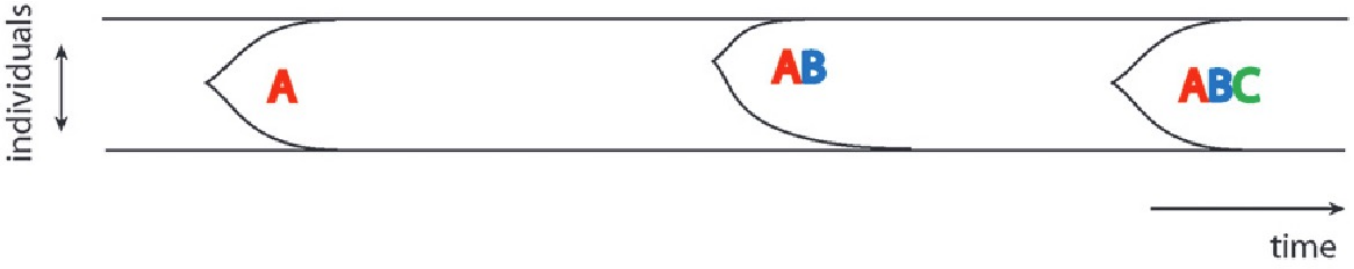
Recombination allows the double mutant AB to emerge from aB and Ab if they coexist for some time

Without recombination, AB tends to emerge later

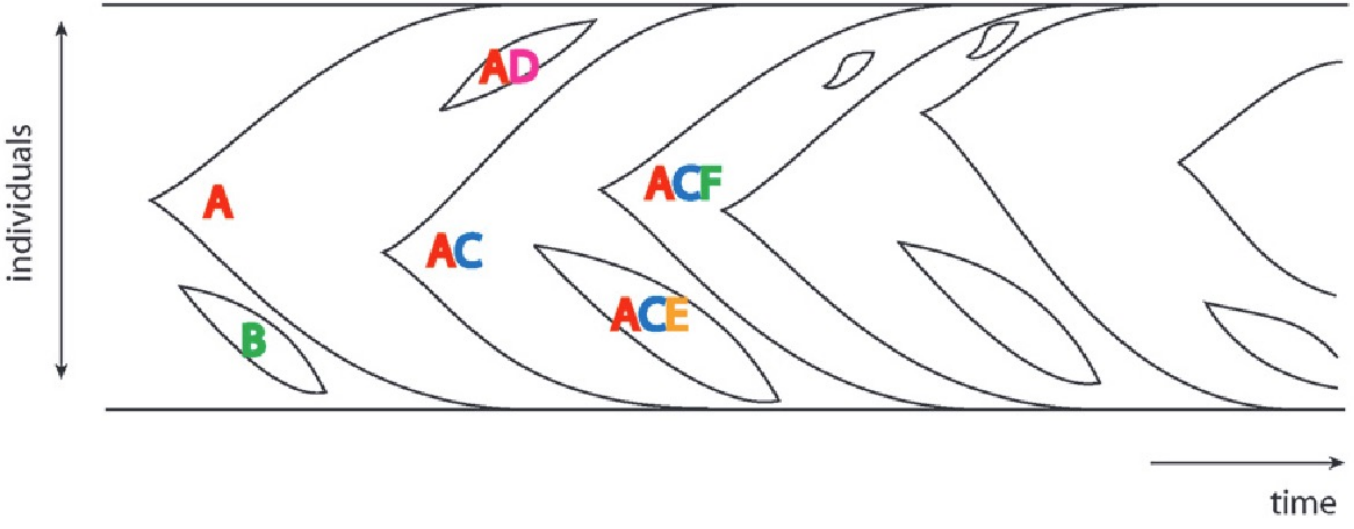
Reminder: clonal interference in asexual populations

- **Sequential fixation of mutations versus multiple mutations & clonal interference**

Two different regimes – Desai and Fisher 2007



Nμ<<1: sequential fixation of mutations



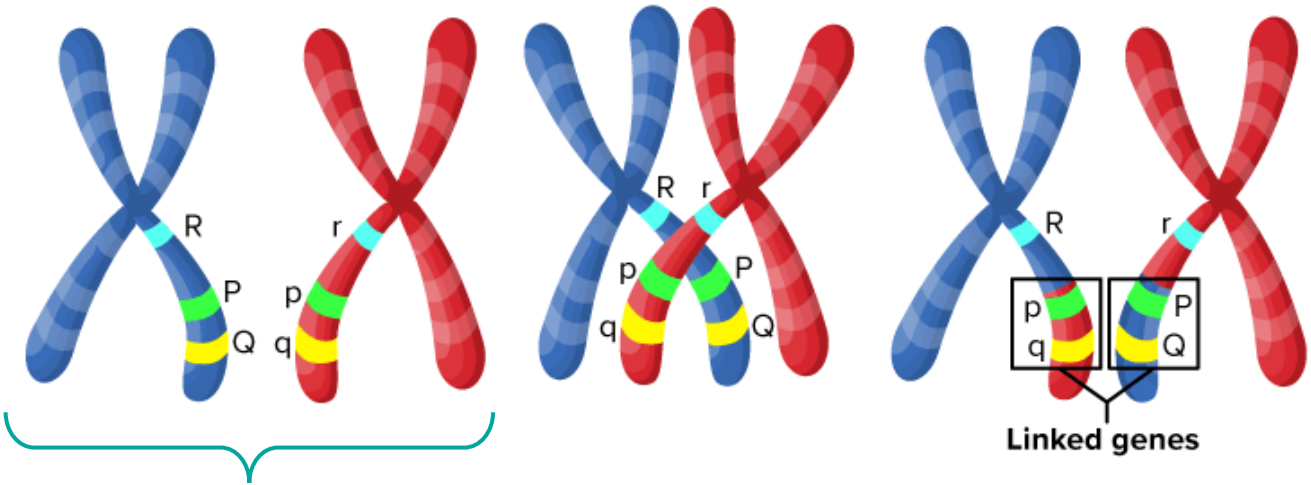
Nμ>1: multiple mutations growing in frequency at the same time
→ "clonal interference":
if two beneficial mutations that would fix if alone appear at similar times, they compete, and only one fixes

Recombination

- **Linkage**

Recombination between two loci is more or less likely depending on where they are located on the genome → linkage

Example: crossing over during meiosis: the closer genes are on a chromosome, the less likely their alleles will be mixed by the crossing over step

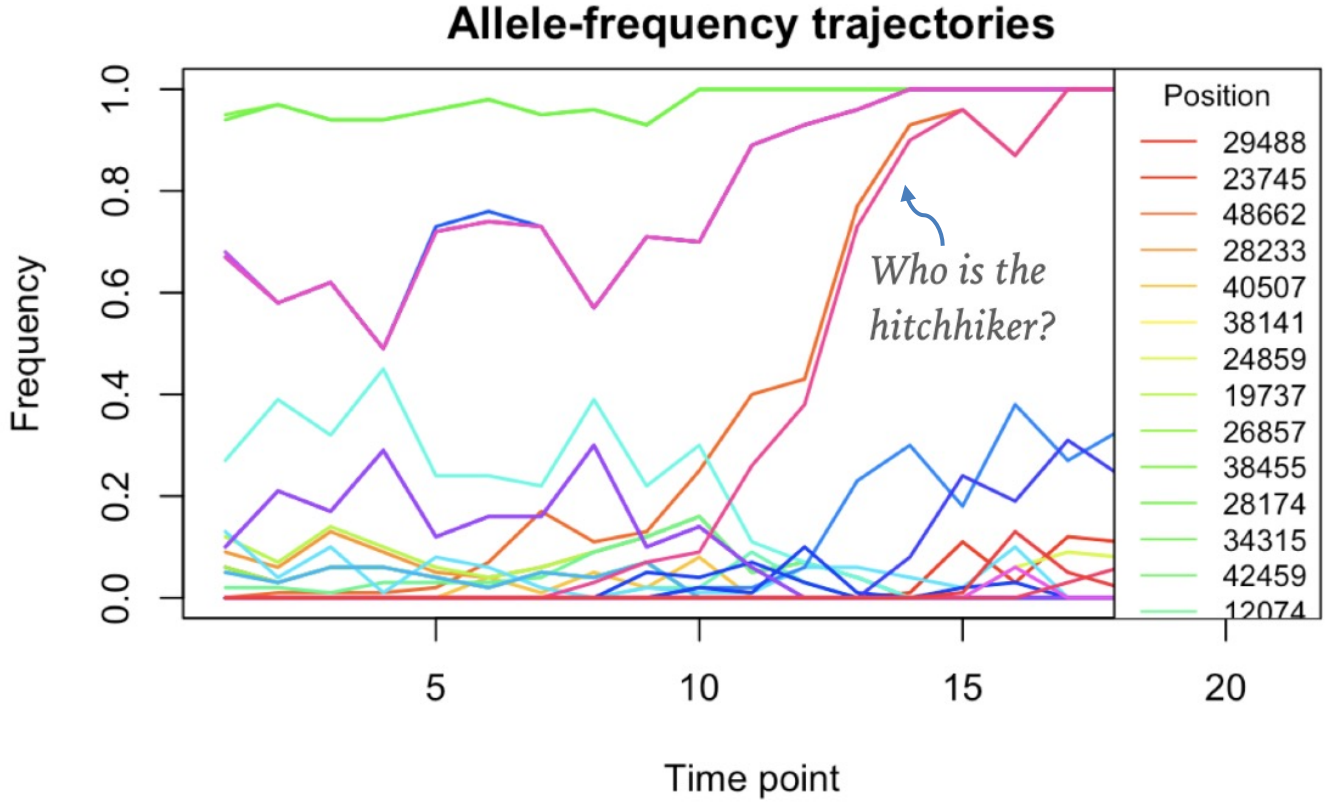


2 homologous chromosomes

Recombination

- **Linkage and genetic hitchhiking**

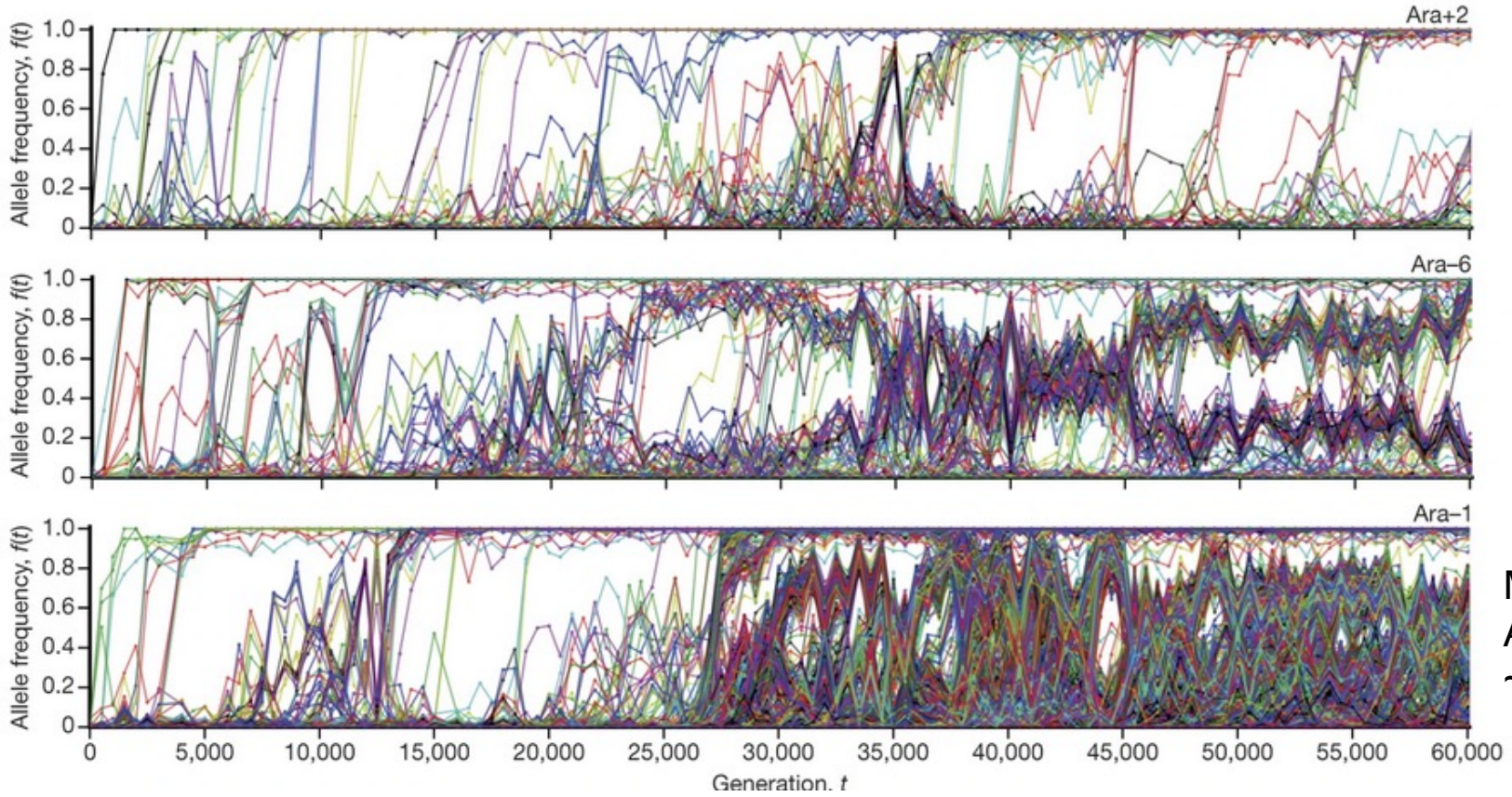
Genetic hitchhiking: a mutation that is not beneficial may grow in frequency because it is linked to a beneficial mutation



Claudia Bank,
U. of Bern

Reminder: evolution experiments

- **The Lenski long-term evolution experiment (LTEE), started in 1988 – Good et al, 2017**



Mutator evolved in Ara-1 at generation ~28,000

Frequency trajectories of all mutations in 3 replicate populations
→ Different dynamics at the genetic level (while overall fitness increase is similar)

Reminder: effects at play

- **Different effects**
  - **Mutations:**
    - generate diversity
    - most have small fitness effects, most are deleterious
  - **Natural selection:**
    - acts upon random mutations
    - because of it, beneficial mutations tend to take over and fix
  - **Genetic drift:**
    - corresponds to stochastic fluctuations
    - arises from finite population size (total population or mutant population)
    - means that moderately beneficial mutations often do not fix – new mutations start in a single individual
- **Additional important effects**
  - Interactions between mutations (epistasis)
  - Recombination (horizontal gene transfer; sexual reproduction)

  } genome-level

  - Specific interactions between individuals
  - Spatial population structure, migrations and genetic flow
  - Environmental variability

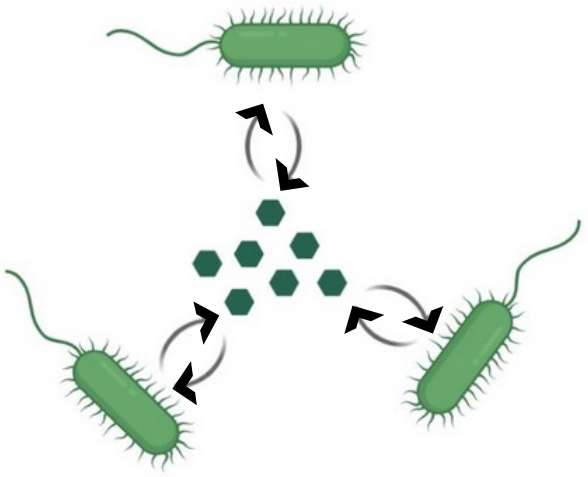
  } population or environment-level

Interactions between individuals or strains

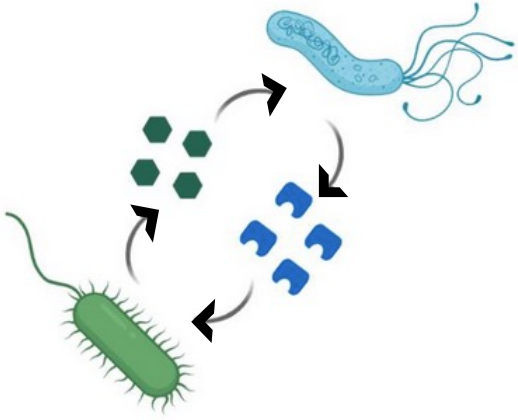
- **Competition**

So far, we considered no particular interaction between individuals
2 types (wild-type and mutant) with different constant fitnesses → natural selection
This is competition. In a finite-size population, without further mutations, one type fixes – no coexistence

- **Cooperation**



Public goods sharing



Cross-feeding

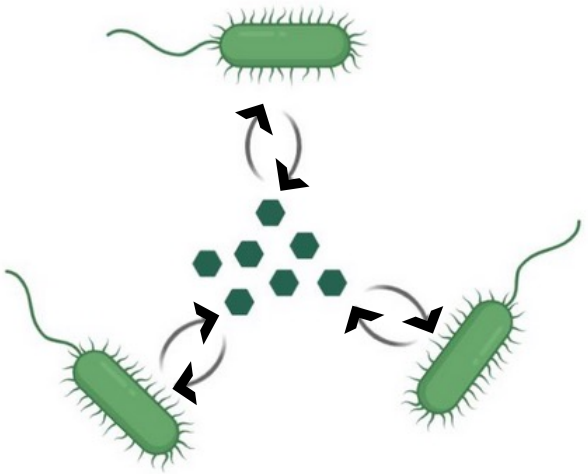
Figueiredo & Kramer 2020

Interactions between individuals or strains

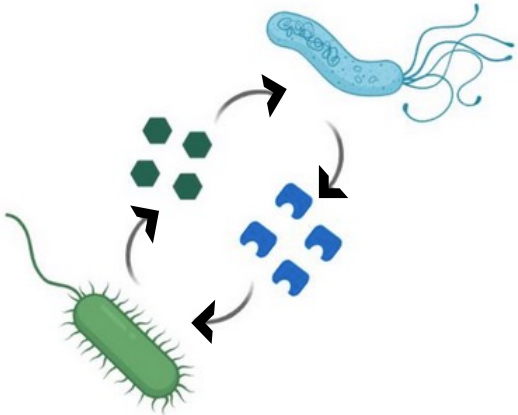
- **Competition**

So far, we considered no particular interaction between individuals
2 types (wild-type and mutant) with different constant fitnesses → natural selection
This is competition. In a finite-size population, without further mutations, one type fixes – no coexistence

- **Cooperation**

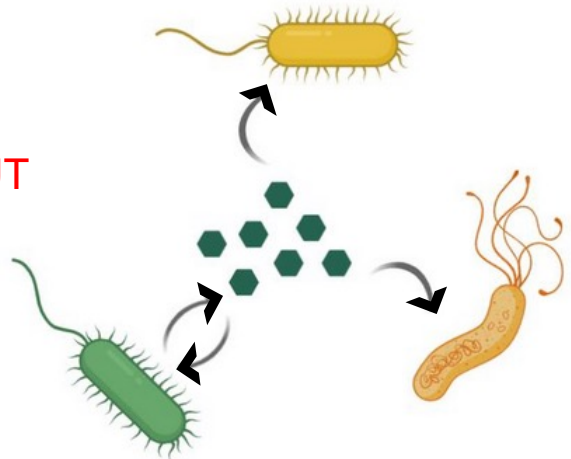


Public goods sharing



Cross-feeding

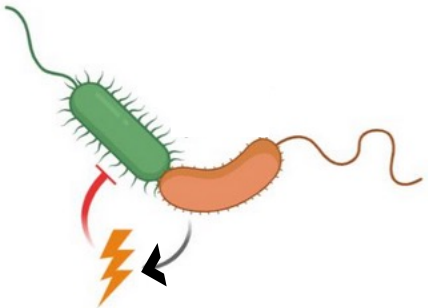
BUT



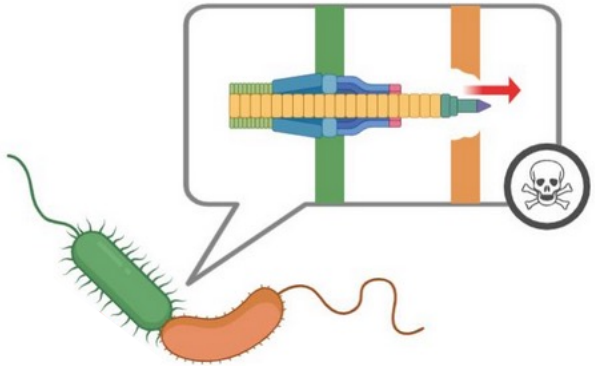
Cheating

Interactions between individuals or strains

- **Predation**
- **Interference**



Secretion of diffusible toxins



Contact-dependent mechanisms
(e.g. type VI secretion systems, can specifically kill cells of other types)

Figueiredo & Kramer 2020

- **Consequence of these interactions**

Fitness can depend on the composition of the population
Consider 2 strains. If x is the frequency (fraction) of mutants, mutant fitness may depend on x
This is called frequency-dependent selection

Interactions between individuals or strains

- **Deterministic description**

A: mutants with fitness f_A; B: wild-types with fitness f_B
Differential equation on the fraction x of mutants A: $\frac{dx}{dt} = (f_A - f_B)x(1-x)$

Evolutionary game theory model (payoffs are assumed to be nonnegative):

- If an A individual interacts with another A, it obtains a payoff a;
- If an A individual interacts with a B individual, it obtains a payoff b;
- If a B individual interacts with an A individual, it obtains a payoff c;
- If a B individual interacts with another B, it obtains a payoff d.

Assume no spatial structure, all individuals interact with all others
Assume that fitness of an individual is the average payoff it gets from interacting with all others

→ Frequency-dependent selection
→ Can get stable coexistence of A and B in this model

Spatially structured populations: models on graphs

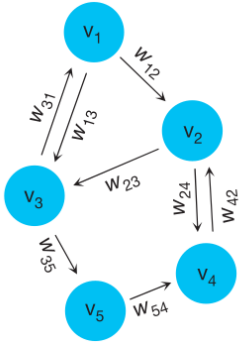
Importance of spatial structure in microbial populations

Natural microbial populations are not well-mixed:

- Infection → subdivision among organs within a patient
- Epidemic → subdivision among hosts
- Even a Petri dish is not well-mixed

Subdivided populations on graphs – Lieberman et al 2005

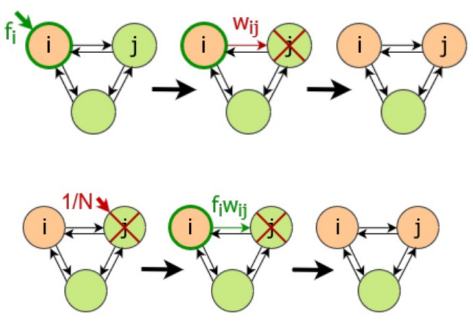
- General model that can represent complex population structures
- Constant population size, one individual per node of the graph
- Replacement probabilities specified on edges



$$W = \begin{bmatrix} 0 & w_{12} & w_{13} & 0 & 0 \\ 0 & 0 & w_{23} & w_{24} & 0 \\ w_{31} & 0 & 0 & 0 & w_{35} \\ 0 & w_{42} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_{54} & 0 \end{bmatrix}$$

Birth-death:

death-Birth:



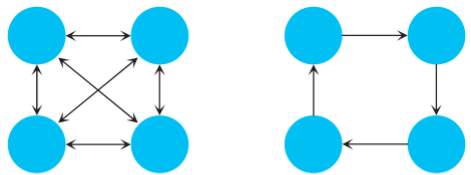
update rules

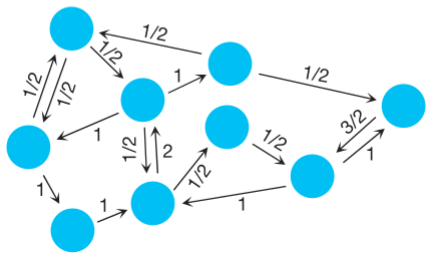
→ **Impact of graph structure on the probability that a mutant takes over?**

Spatially structured populations: models on graphs

- **Subdivided populations on graphs – Lieberman et al 2005**
  - Fixation probability of one M in a W population (assuming no further mutation):

  - **Circulation theorem:** if for each node, the sum of incoming migration probabilities is equal to the sum of outgoing ones, then the fixation probability of a mutant is the same as in a well-mixed population
  - Includes the case of **symmetric migrations** (cf. Maruyama 1970 & 1974)

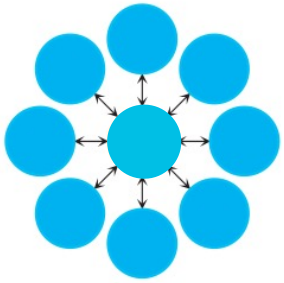




  - While spatial structure often has no impact on fixation probability, it does impact fixation time (usually, it makes it slower)

Spatially structured populations: models on graphs

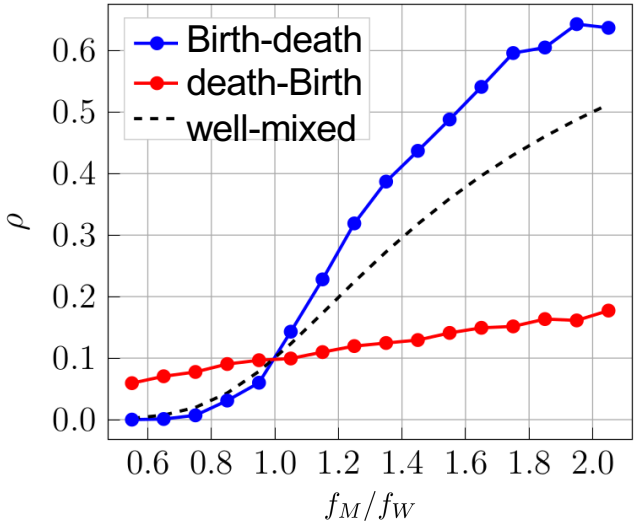
- **Subdivided populations on graphs – Lieberman et al 2005**
  - Some structures can amplify or suppress natural selection compared to the well-mixed case



The (Birth-death) star amplifies natural selection

… but this depends on the update rule (Kaveh et al 2015, Hindersin & Traulsen 2015, Pattni et al 2015)

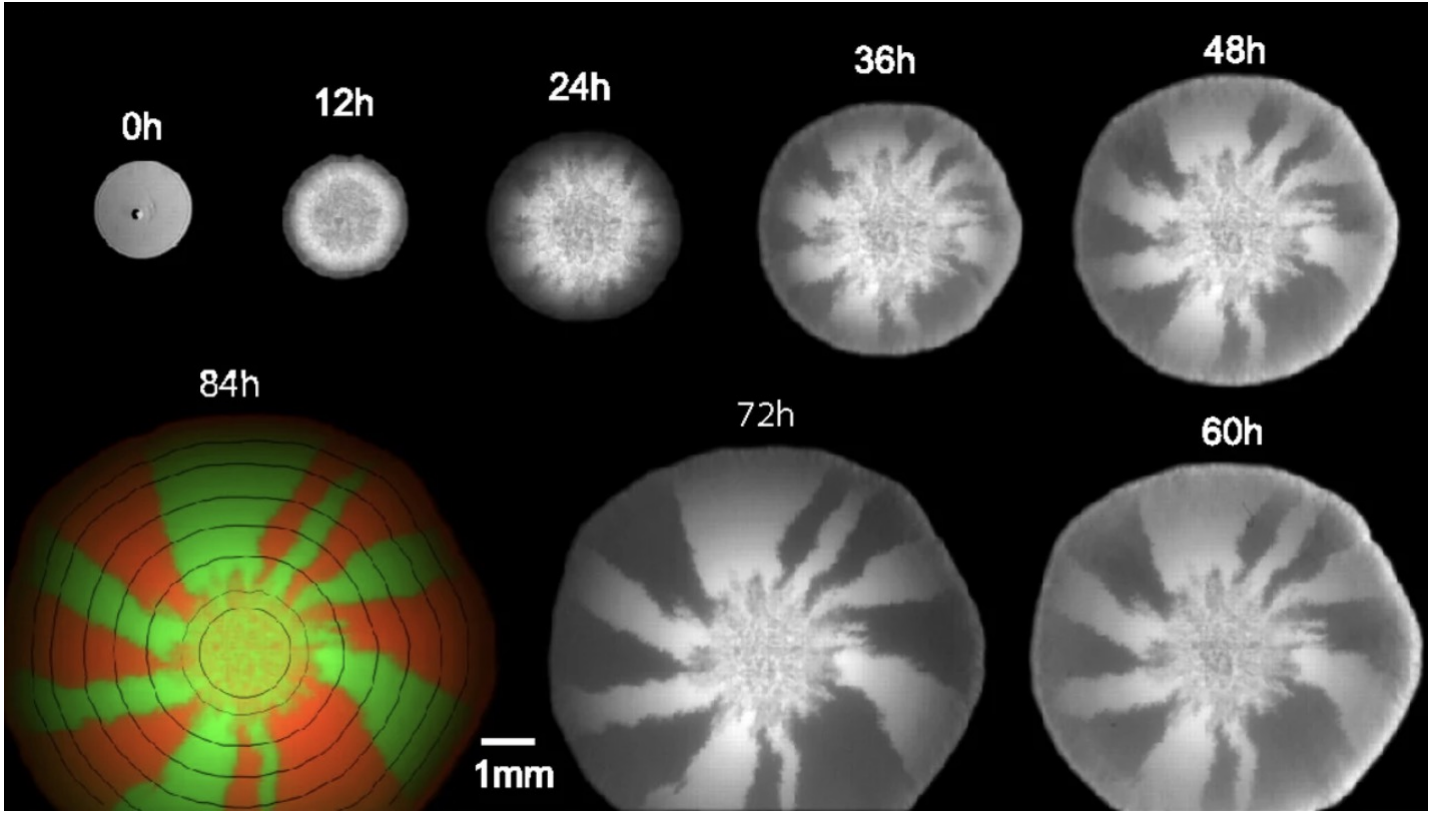
Fixation probability of one M in a W population with N=10 individuals, versus fitness ratio:



→ Star = amplifier in the Birth-death dynamics but suppressor in the death-Birth dynamics
Remark: The large impact of update rules is a limitation of these models

More spatial structure: expanding population

- **Droplet with 2 types of bacteria on Petri dish – Hallatschek et al 2007**



Images of a growing colony founded by a 50:50 mixture of bacteria (≈10^6 *E. coli* cells) carrying either CFP or YFP. Even though the bacteria were otherwise genetically identical, the growing colony shows complete segregation of the two neutral markers (CFP and YFP) over time.

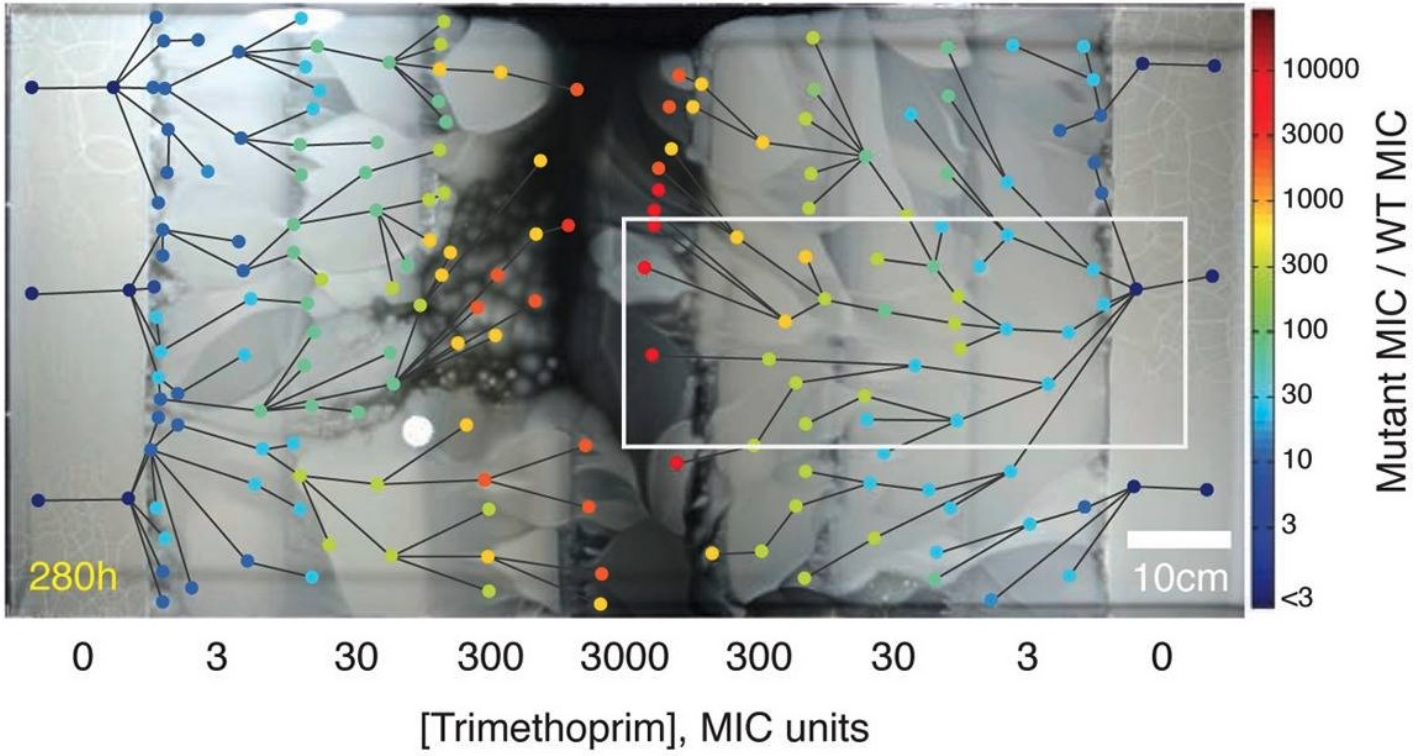
More spatial structure: environmental gradients

- **Discrete spatial variations of antibiotic concentrations – Baym et al 2016**

https://www.youtube.com/watch?v=plVk4NVIUh8

More spatial structure: environmental gradients

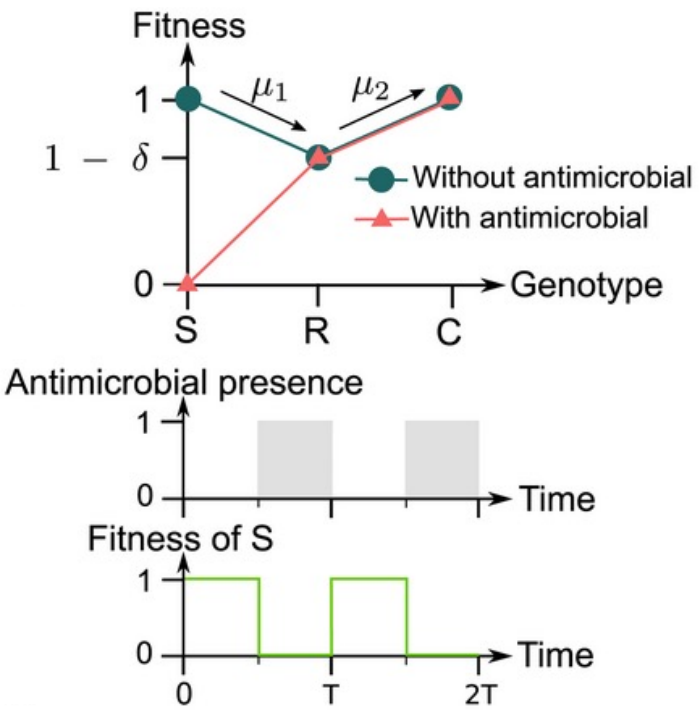
- **Discrete spatial variations of antibiotic concentrations – Baym et al 2016**

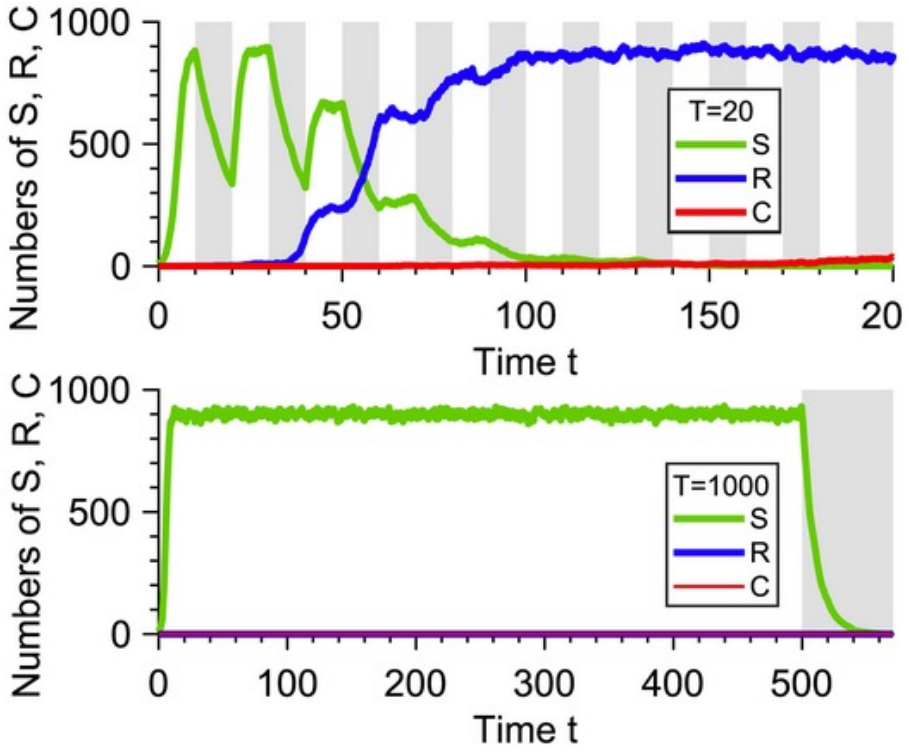


The four-step TMP MEGA-plate after 12 days. *E. coli* appear as white on the black background. The 182 sampled points of clones are indicated by circles, colored by their measured MIC. Lines indicate video-imputed ancestry.

Environment variability

- **Example: alternations of presence and absence of antibiotic– Marrec et al 2020**





Fitness of sensitive strain is assumed to be zero with antibiotic → huge impact
Fast alternations of presence and absence of antibiotic favour the spread of resistance
Long enough treatment yields eradication of bacteria (unless resistant ones are already present)

Reminder: effects at play

- **Different effects**
  - **Mutations:**
    - generate diversity
    - most have small fitness effects, most are deleterious
  - **Natural selection:**
    - acts upon random mutations
    - because of it, beneficial mutations tend to take over and fix
  - **Genetic drift:**
    - corresponds to stochastic fluctuations
    - arises from finite population size (total population or mutant population)
    - means that moderately beneficial mutations often do not fix – new mutations start in a single individual
- **Additional important effects**
  - Interactions between mutations (epistasis)
  - Recombination (horizontal gene transfer; sexual reproduction)

  } genome-level

  - Specific interactions between individuals
  - Spatial population structure, migrations and genetic flow
  - Environmental variability

  } population or environment-level