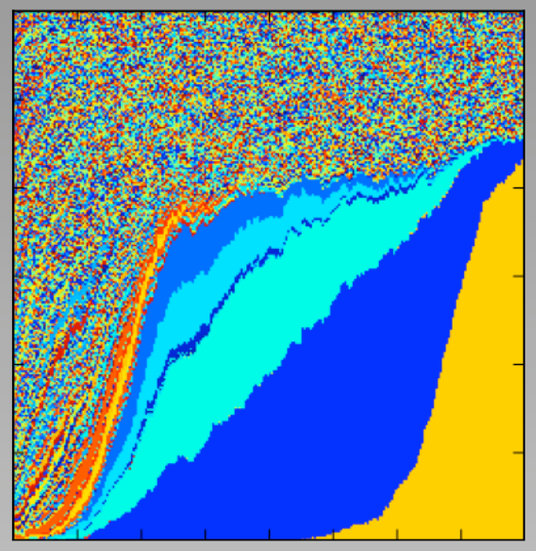
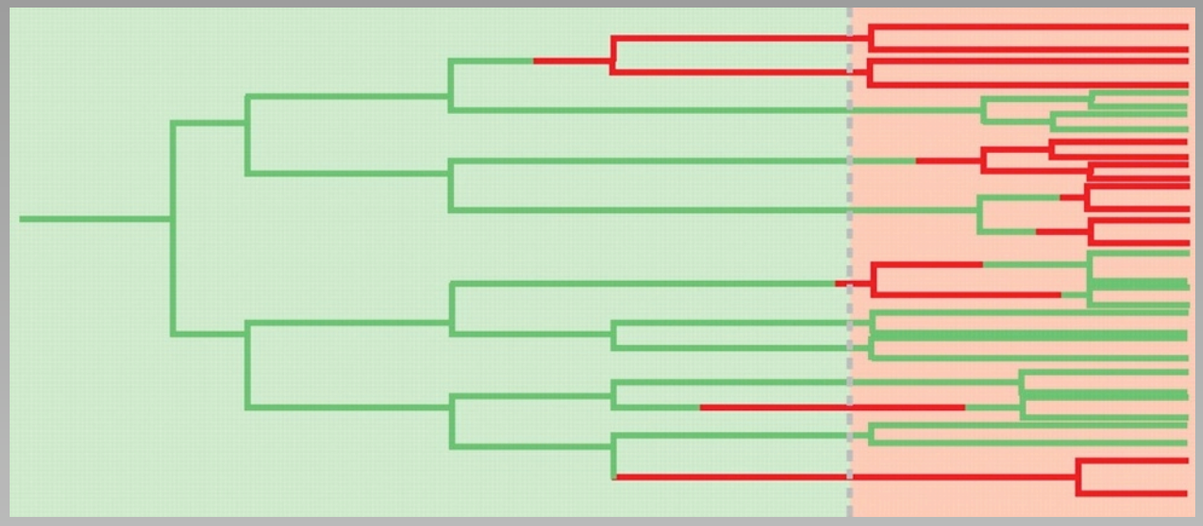




# *Non-equilibrium statistical physics, population genetics and evolution*

Marija Vucelja
The Rockefeller University

UVa Physics colloquium, 2013

# outline

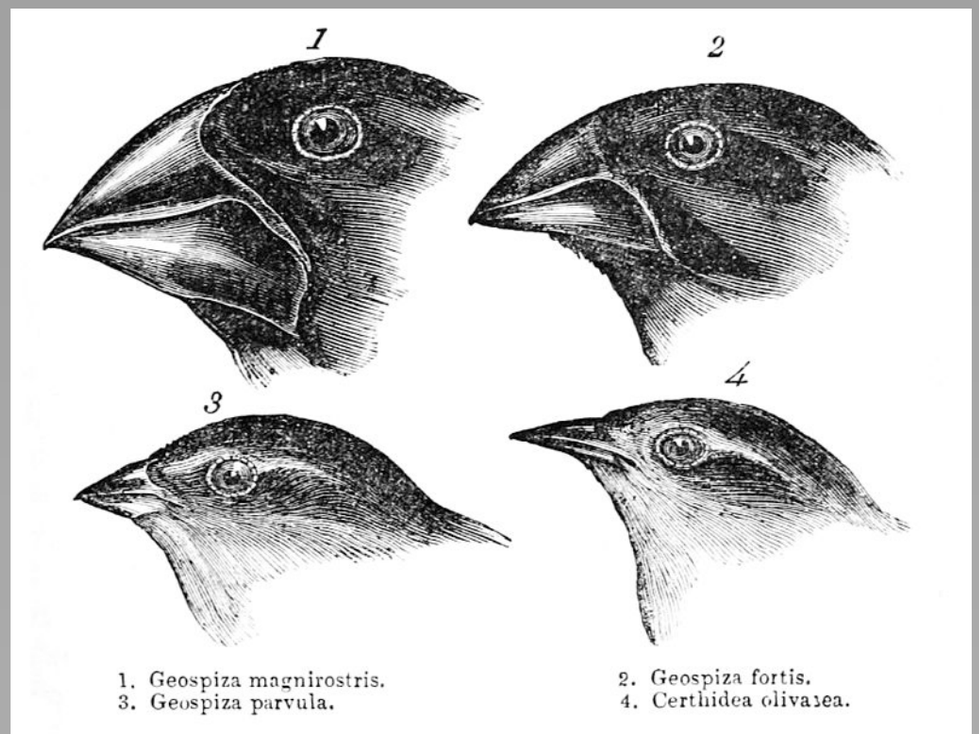
- traditional view of population genetics:
  mutations, recombination and selection

•questions of interest
•why all hasn't been solved yet?

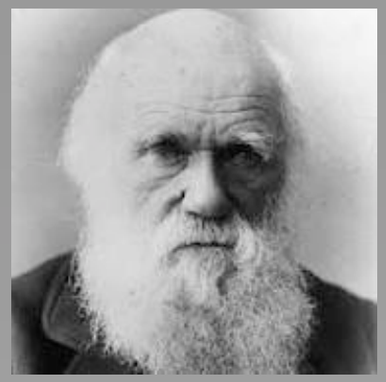
- relations to statistical physics
  - •spin glass (clonal interference)
  - •polymers, path integrals, localization phenomena (phenotype switching)

- unusual kind of non-equilibrium statistical physics:
  - new processes like recombination
  - effects of discreteness appearing even in the "thermodynamical limit"

Darwin's finches



# early ideas on evolution and the standard picture



Charles Darwin

Alfred Russel Wallace

Gregor Mendel

Jean-Baptiste Lamarck

$$g = \{s_1, ..., s_L\}$$

genotype = spin configuration

$$s_i \in \{+1, -1\}$$

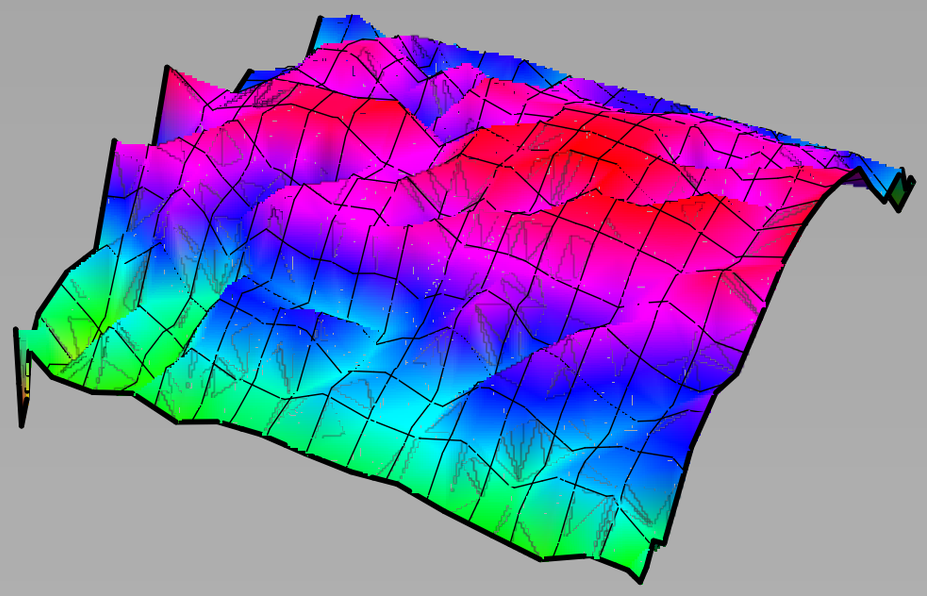
allele = spin state

example: eye color

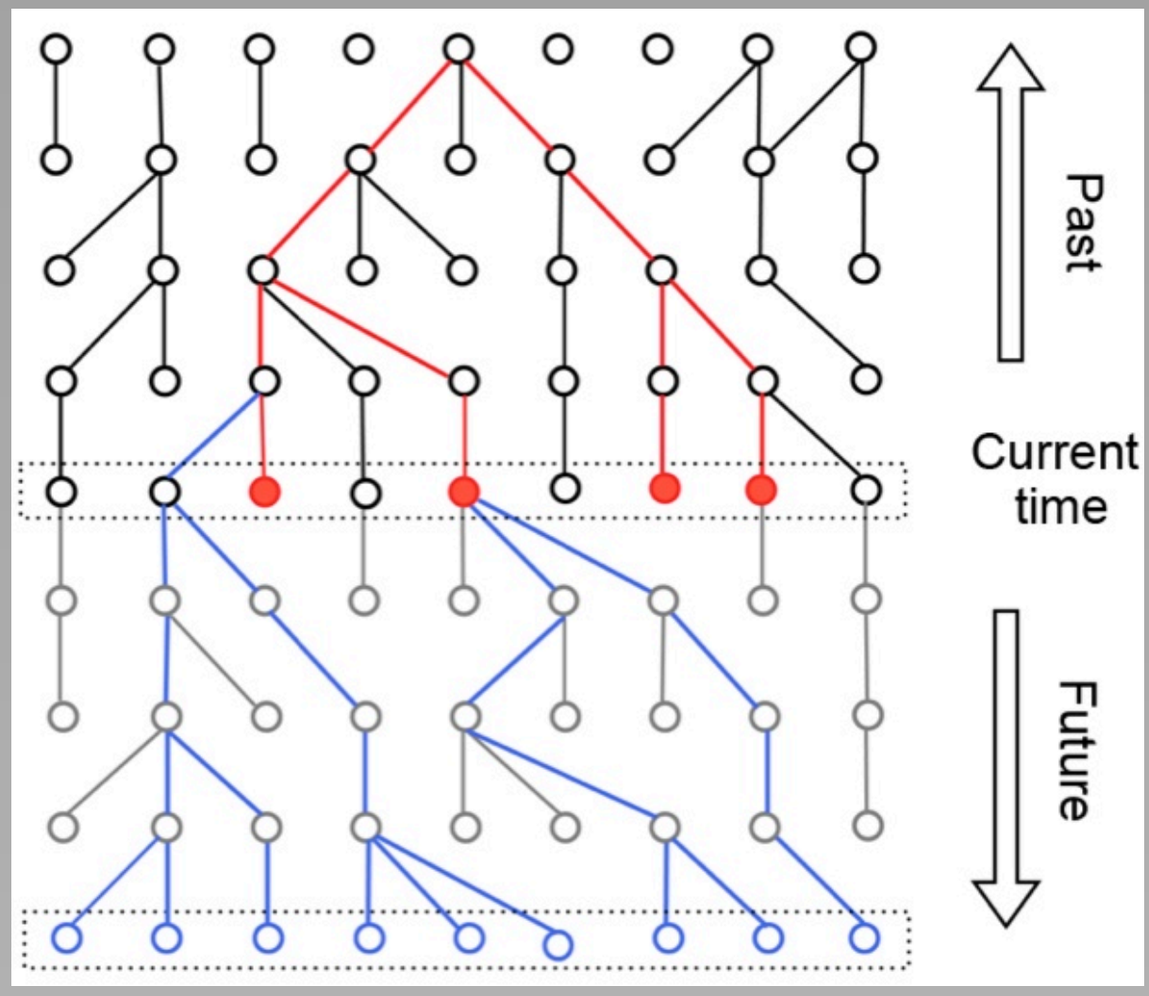
fitness landscape

$$F(g, \text{environment})$$

fitness = energy



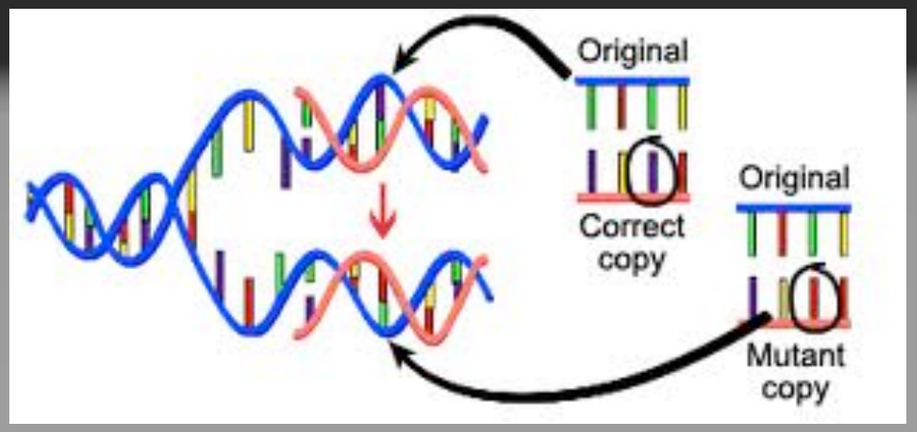
mutation creates variation: only favorable survive

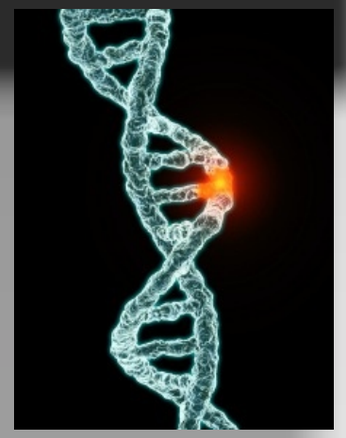


# mutation - spin flip

$$M : g \to Mg = \{s_1, \ldots, Ms_i, \ldots, s_L\}$$

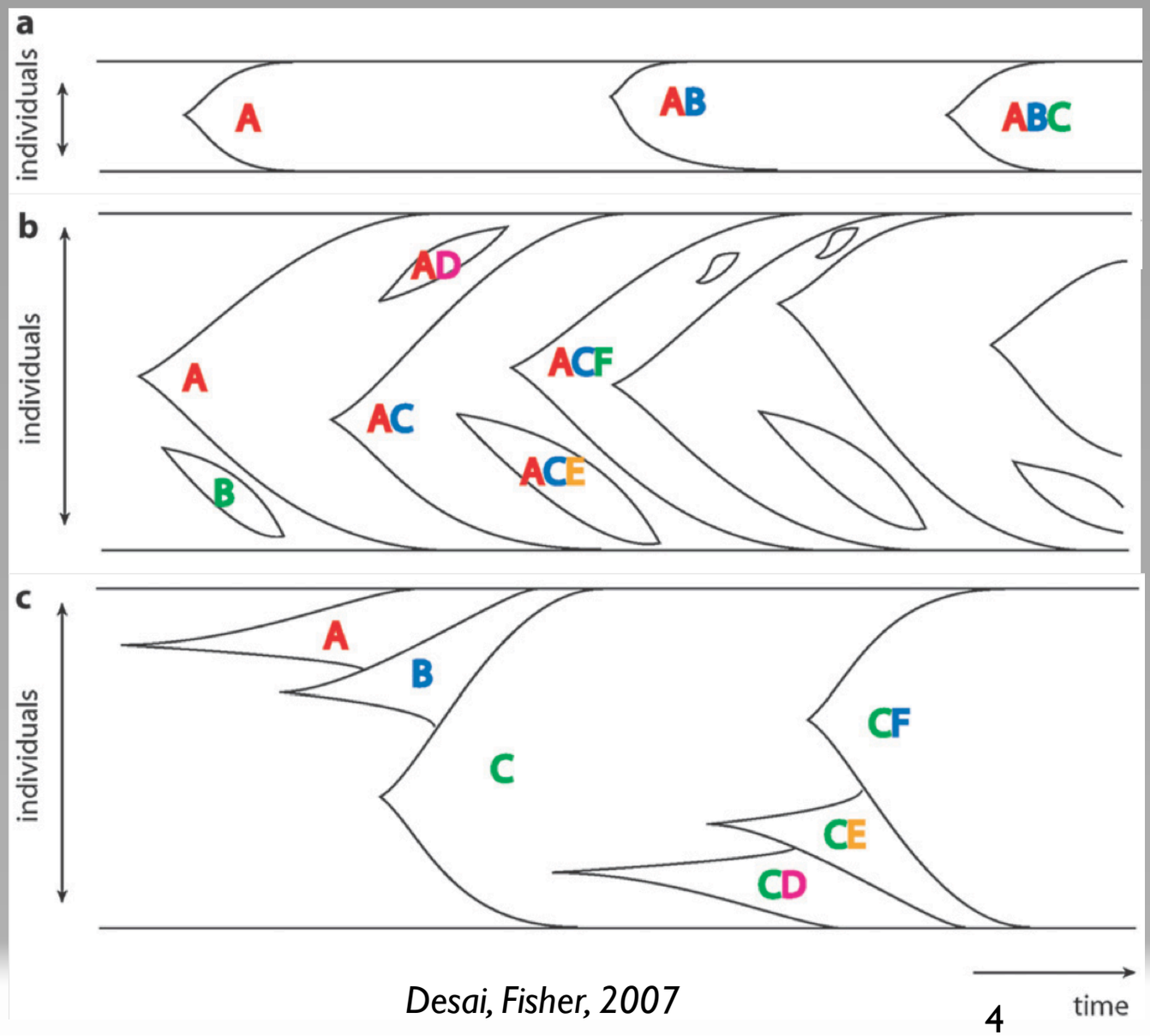
clones = organisms' with the same genotype



**successional mutations**
strong selection & weak mutations
present in small populations

**concurrent mutations**
strong selection & strong mutations
present in large populations
we see: clonal competition & multiple mutations




*Desai, Fisher, 2007*

**clonal competition**
weak beneficial mutation

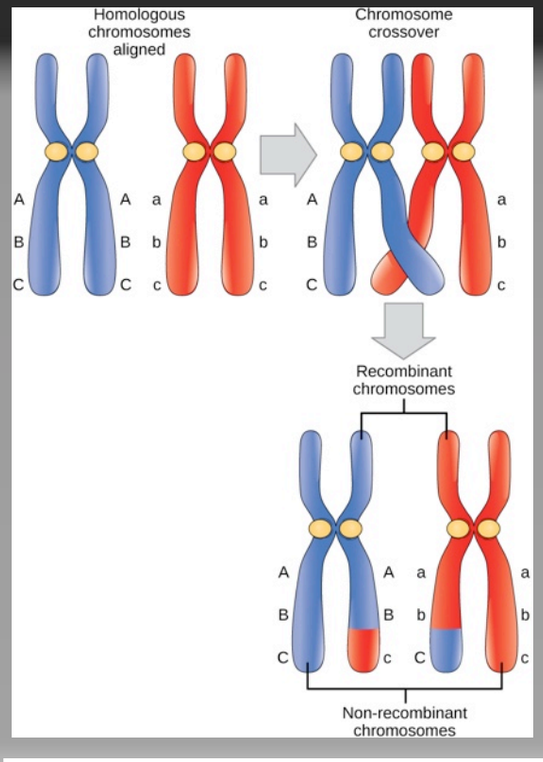
present in large populations

disregarding multiple mutations

# recombination - no physics analog

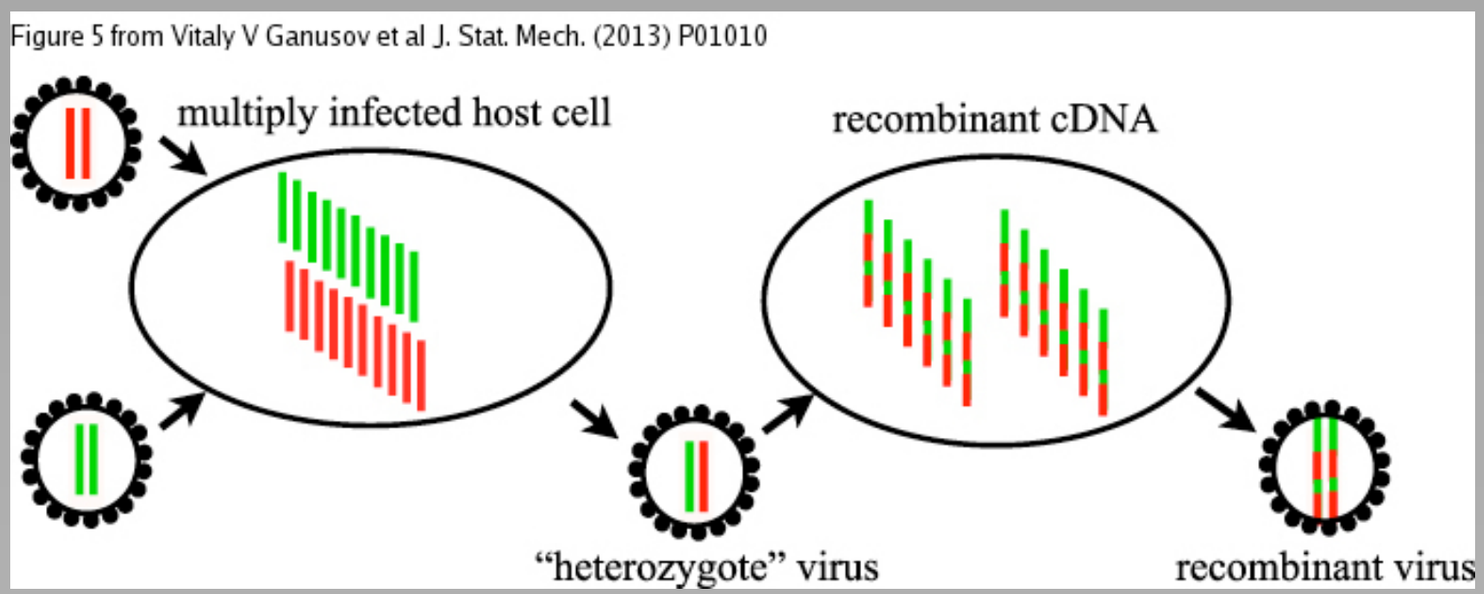
humans and other organisms
with two sets of chromosomes

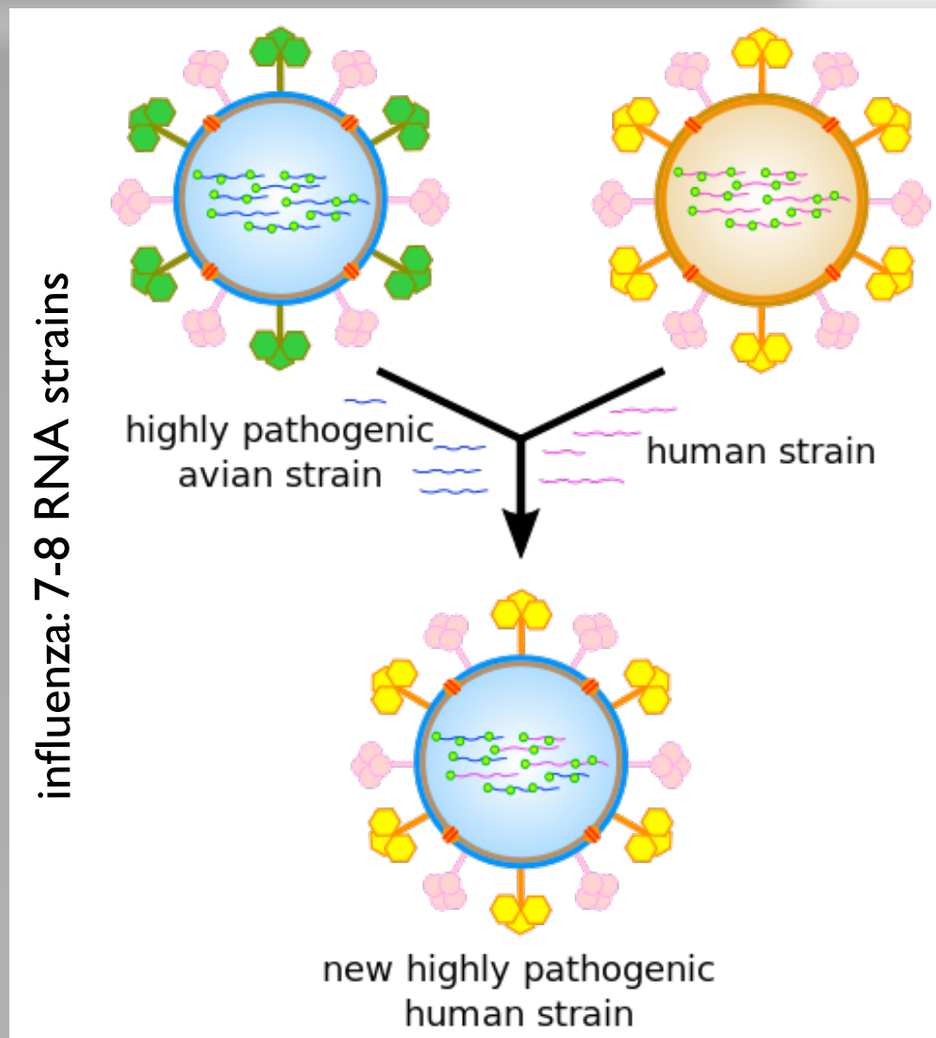
$$R : g^{(m)}, g^{(f)} \rightarrow g = \left\{ s_i | \xi^i s_i^{(m)} + (1 - \xi^i) s_i^{(f)} \right\}$$




viruses
recombine in the hosts' cell

HIV: 2 RNA strains



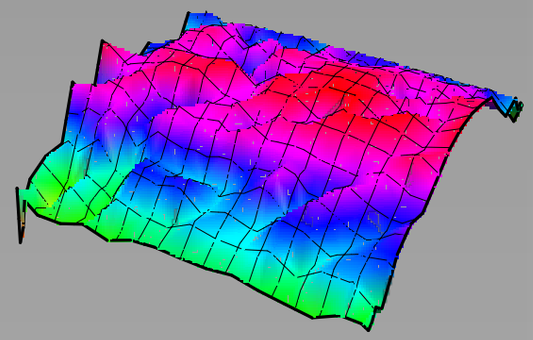



# selection

$F(g, \text{environment})$ proxy for fitness = the expected reproductive success

spin interactions

$$F = \bar{F} + \sum_i f_i s_i + \sum_{i<j} f_{ij} s_i s_j + \sum_{i<j<k} f_{ijk} s_i s_j s_k + \dots$$



$n_g$ number of individuals with genotype g

$$\dot{n}_g(t) = (F_g - \bar{F}) n_g(t) + \text{noise}$$

$$\bar{F} = \frac{1}{N} \sum_g^N n_g F_g$$ chemical potential

$$\frac{n_g(t)}{N} = \frac{e^{F_g t - \int \bar{F} dt}}{N} = \frac{e^{F_g t}}{Z}$$

**Boltzmann statistics**

time = inverse temperature

$1 \ll N \ll 2^L$ population = random sample from genotype space

# ρ(*F*,*t*) fitness distribution

$$\rho(F,t) = \sum_g \delta(F - F_g)$$

$$\langle F \rangle \equiv \int \mathrm{d}F p(F,t) F$$

R.A. Fisher

Fundamental theorem of natural selection

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle F \rangle = \langle (F - \langle F \rangle)^2 \rangle$$

variance = selection strength, since only $n_g$ with $F_g > F$ grow

$$n_g(t) = e^{F_g t - \int \bar{F} dt}$$

# Quasi-linkage equilibrium

*Kimura, 1965*

perturbative:

weak selection & weak interactions compared to mutation & recombination that act to decorrelate spins (bio: alelles)

$F_{int}$ interactions (bio: epistasis)

$$F = \bar{F} + \sum_i f_i s_i + \sum_{i<j} f_{ij} s_i s_j + \sum_{i<j<k} f_{ijk} s_i s_j s_k + \dots$$

$F_0$ non-interacting (bio: additive)

fitness distribution

$$\rho(F,t) = p(F_0,t)\omega(F_{\rm int},t)$$

separable solution:
distributions for non-interacting and interacting part

description with first two moments:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle s_i\rangle \approx \text{function of } \langle s_i\rangle, \langle s_i s_j\rangle$$
$$\langle s_i s_j\rangle \approx \text{function of } \langle s_i\rangle, \langle s_j\rangle$$

# Quasi-linkage equilibrium is allele competition

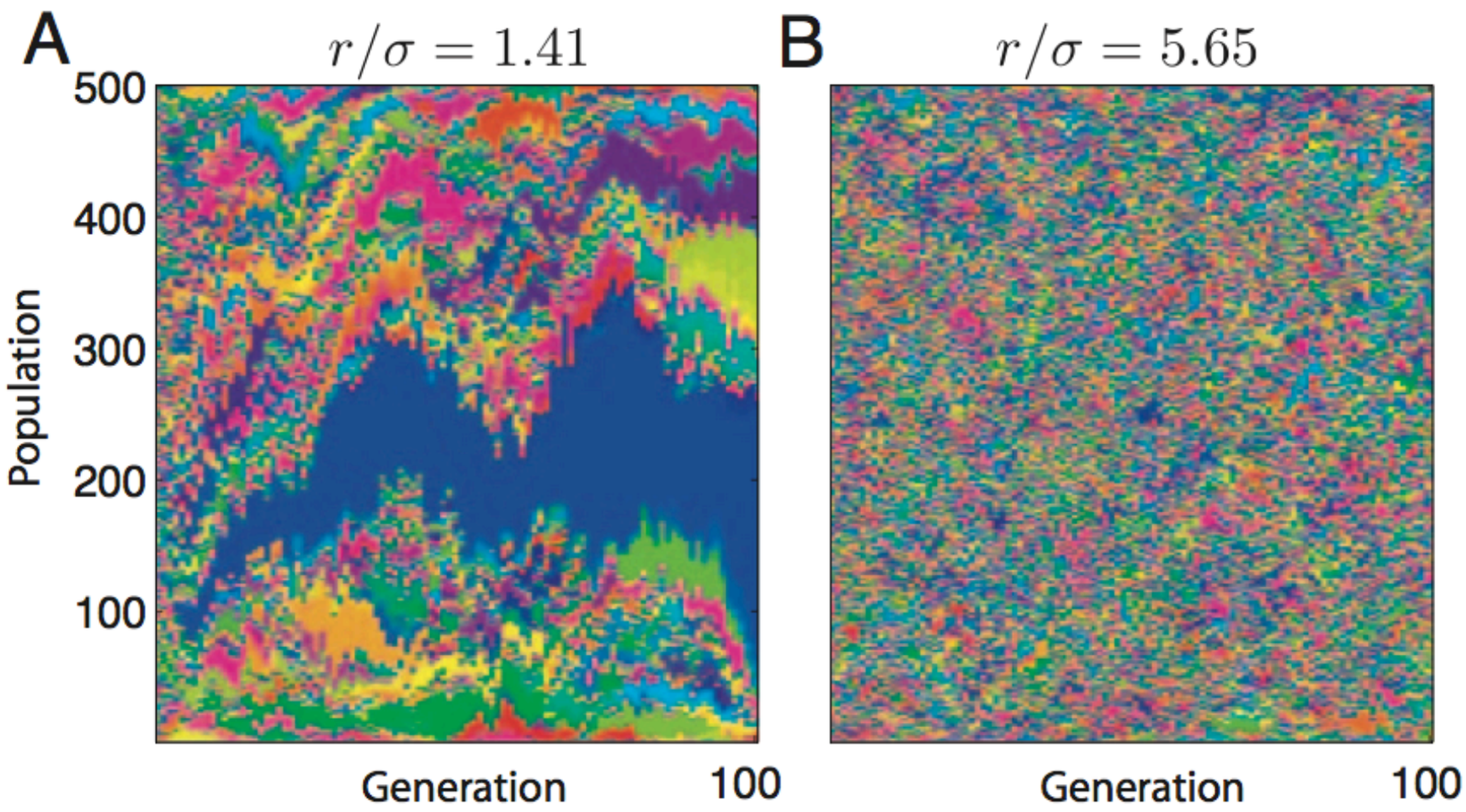
*Neher, Shraiman 2009*

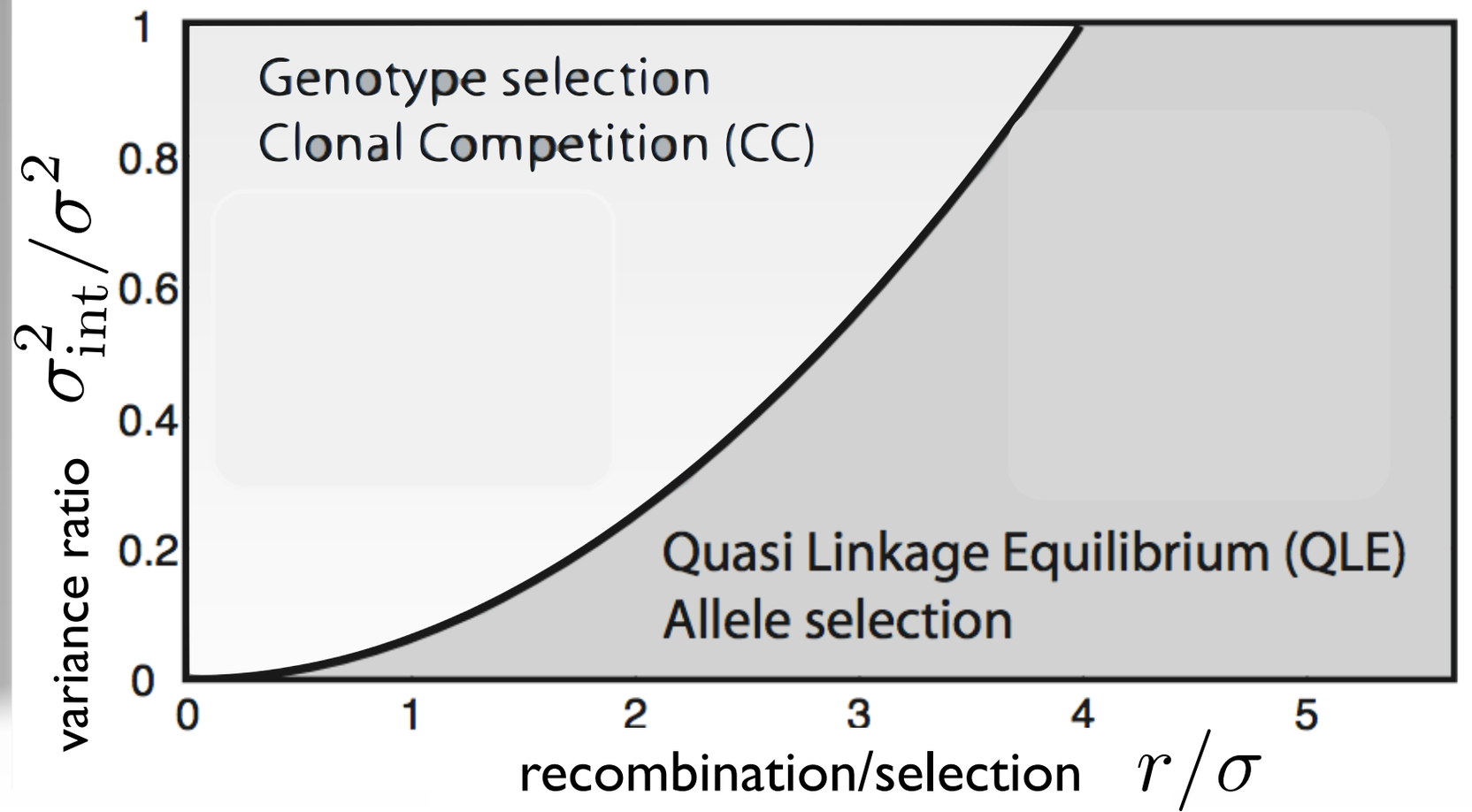
no mutations
recombination rate *r*
selection strength = variance of *F*
$= \sigma^2$
perturbation: $r/\sigma > 1$

color = different allele
same colors grouped together

$$F = F_0 + F_{\text{int}}$$

$\sigma_{int}^2$= variance $F_{int}$

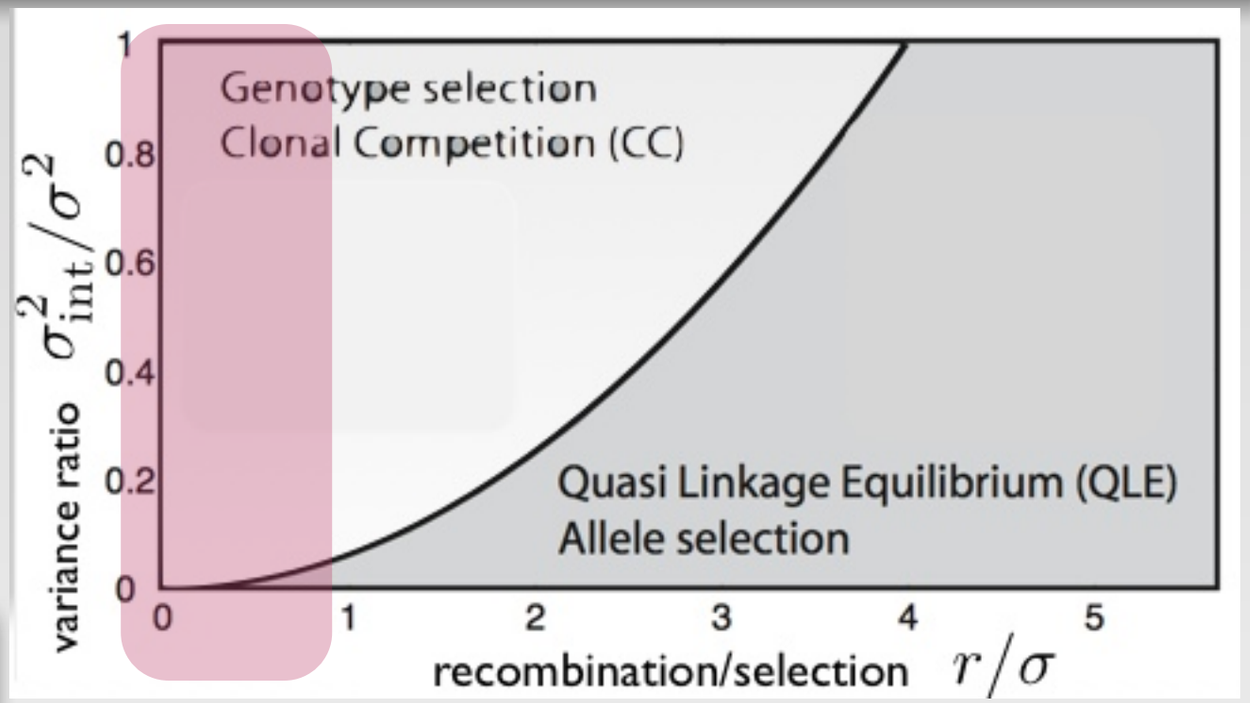


# Clonal Competition

*Neher, MV, Mezard, B. Shraiman 2009*

| Quasi-Linkage Equilibrium | Clonal Condensation |
|---|---|
| weakly nonlinear perturbative | strongly interacting non-perturbative |

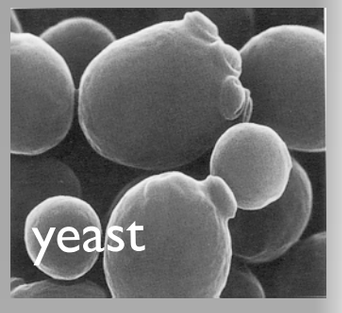



How do large clones form and persist?

**Relevant for:**

facultatively recombining populations like:

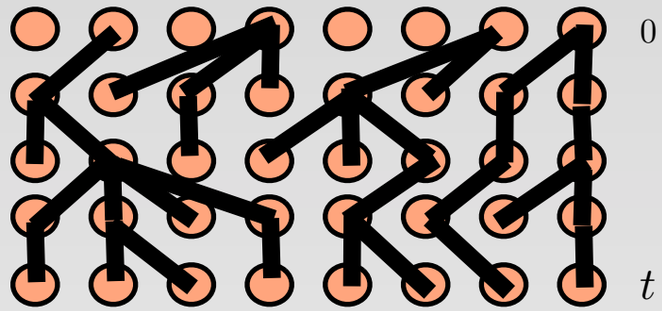
- fungi
- microorganisms
- plants
- contiguous parts of chromosome
- pathogens such as HIV and influenza
- nematodes



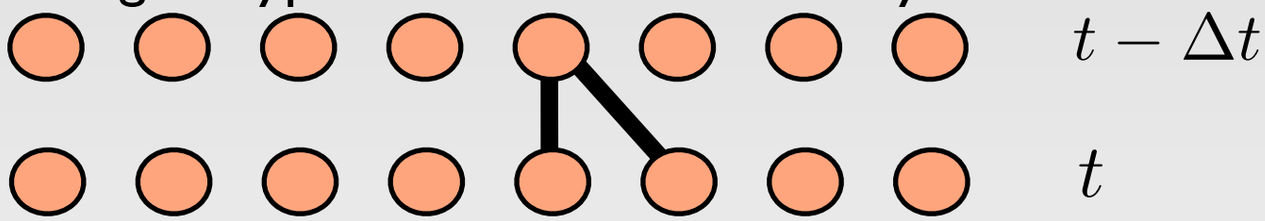

rapidly adapting populations,
hybridization zone, population bottlenecks, HIV, influenza

# Coalescence rate = good measure of distance in genotype space

Y(t) = probability that in a population of *N* individuals 2 are from the same clone (have a common ancestor)




if two genotypes are identical than they had a common ancestor in the past




| | |
|---|---|
| 1 | $\frac{1}{N}$ |

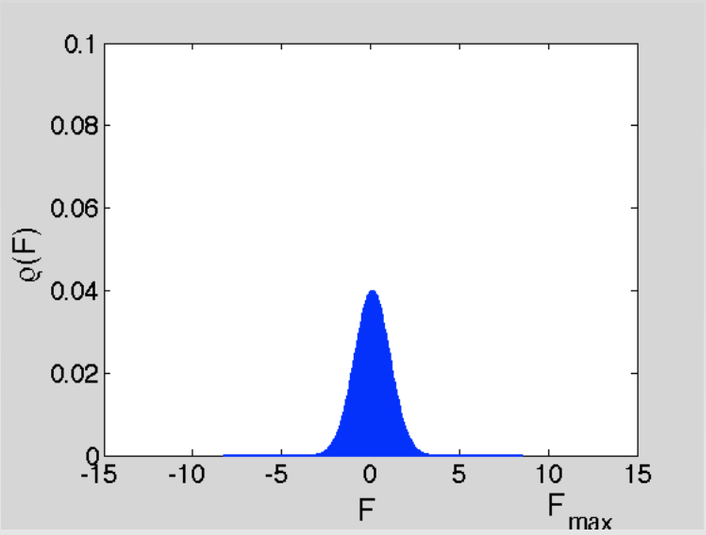
$\mathcal{O}(N^{-1})$ probability of two individuals in the present generation, having the same ancestor one generation in the past

actually:

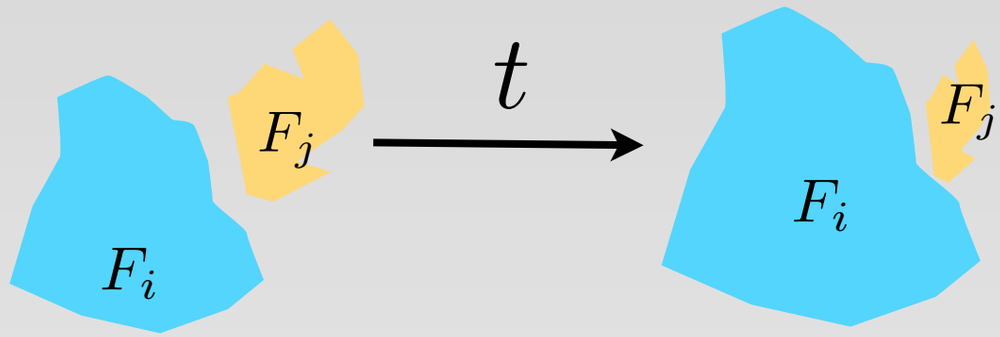
$$Y(t,...) = \begin{cases} \mathcal{O}(N^{-1}) \\ \mathcal{O}(1) \end{cases}$$

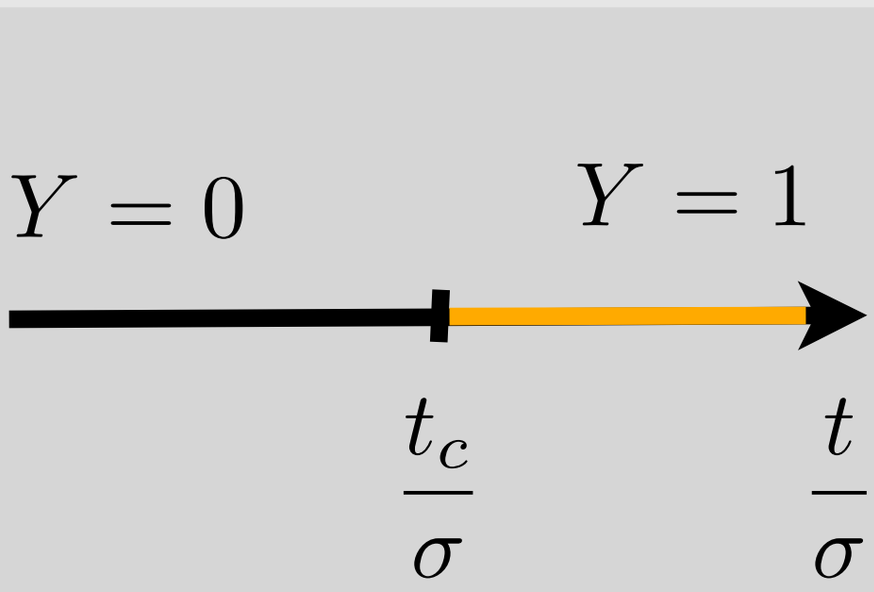
**clonal condensation** few clones grow to form a significant fraction of the population.

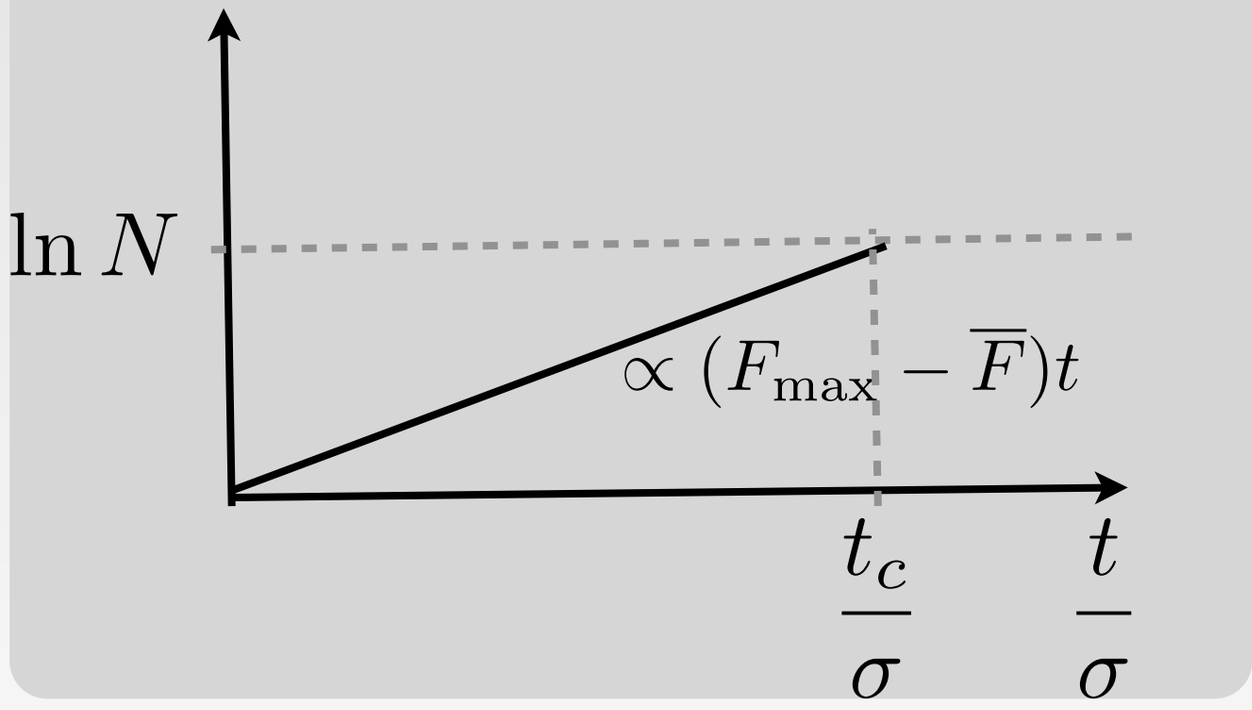
# Selection




the most fit genotype overtakes the whole population



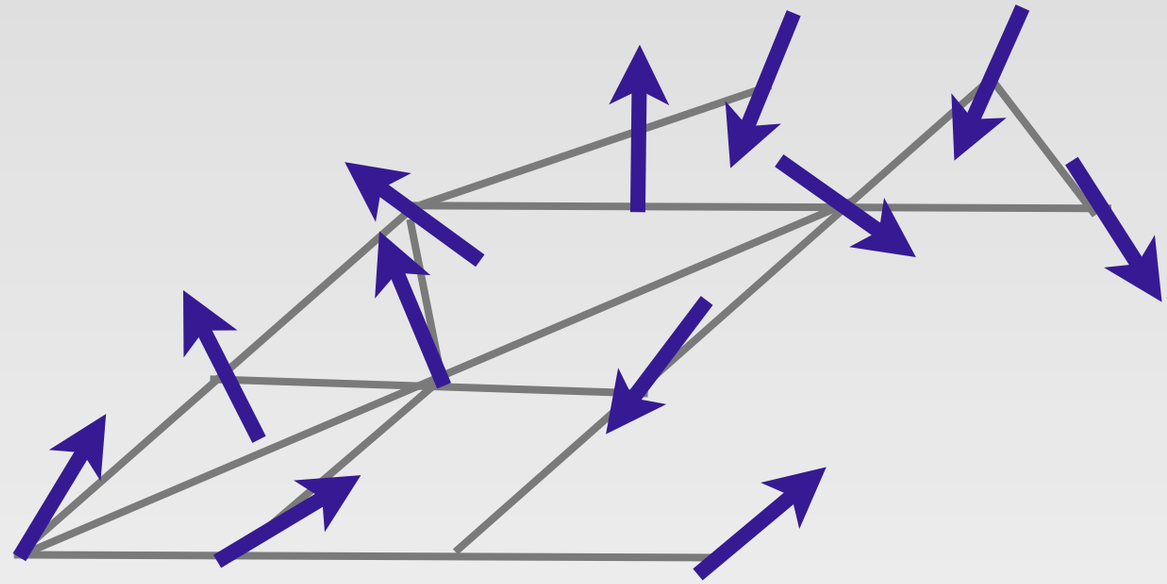

**“condensation time”**

time for the fittest clone to over takes the whole population

# the Random Energy Model and spin glasses

$$\rho(F) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{F^2}{2\sigma^2}}$$

many spins contribute independently; Central Limit Theorem



statistics of clones is identical to that of the Random Energy Model

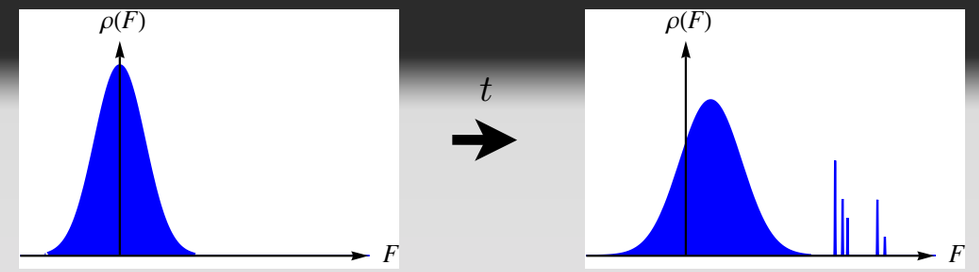
***spin glass:*** disordered magnet with frustrated interactions and stochastic positions of spins

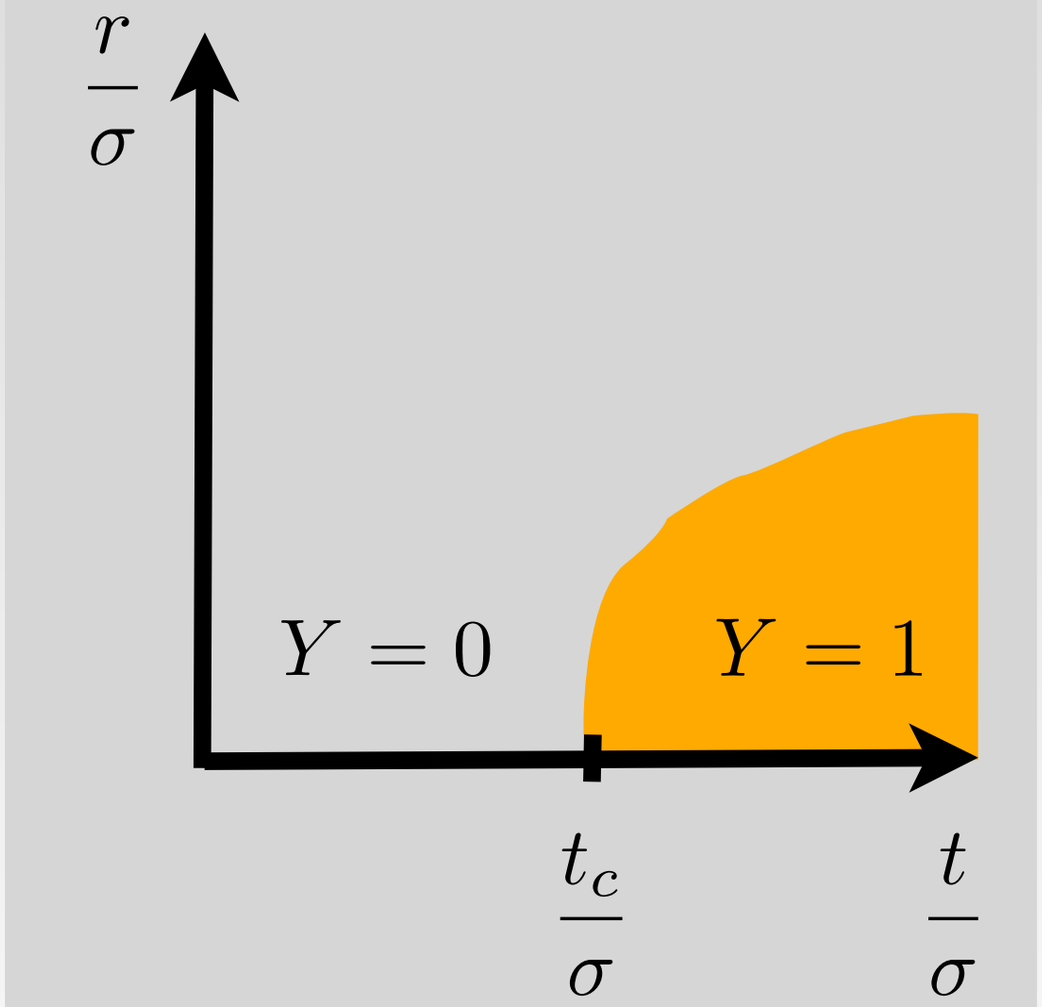
Random Energy Model = a set of configurations and an energy functional over those configurations.

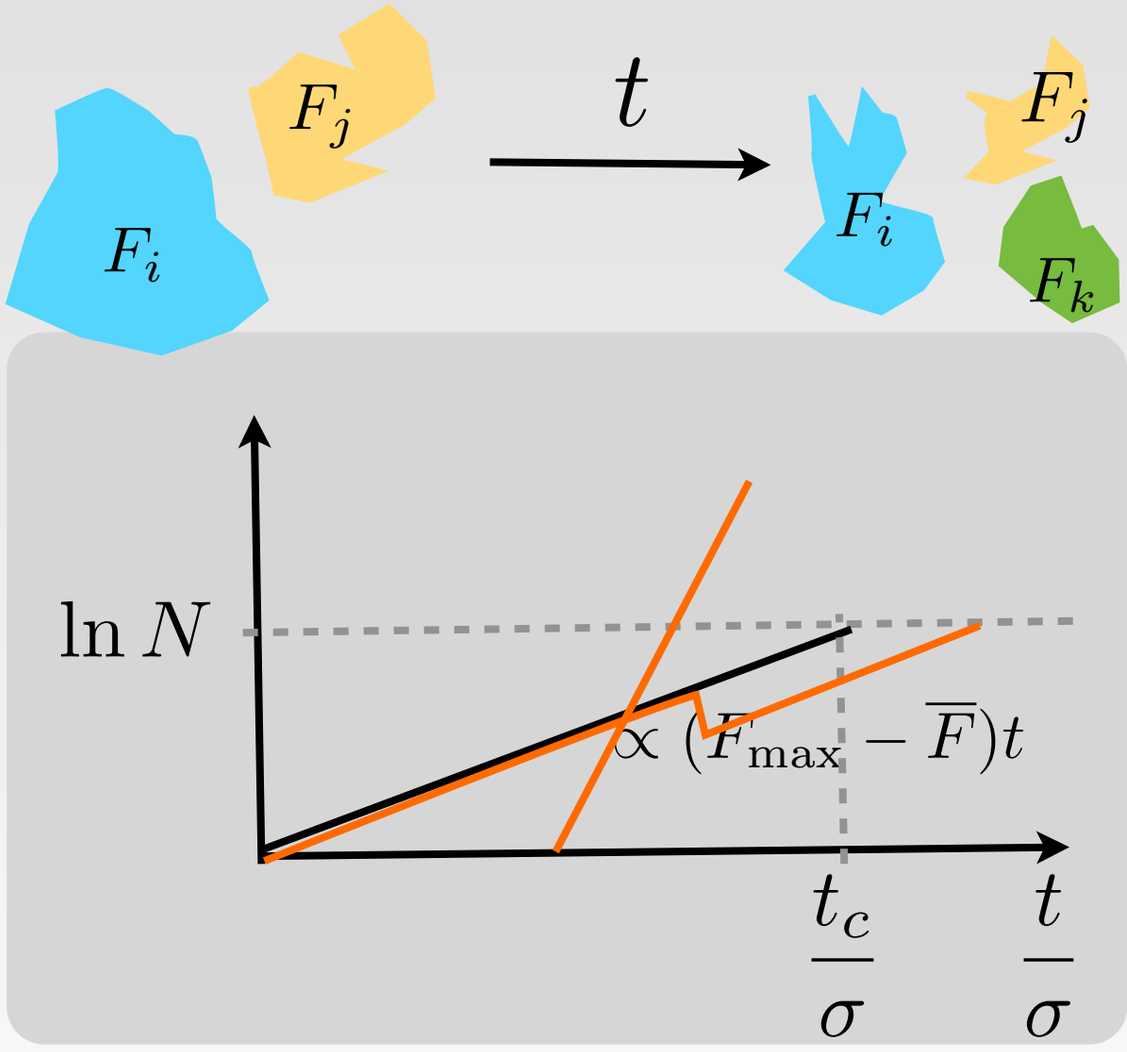
$\{E_i\}$ are i.i.d. random variables drawn from $\rho(E)$

"sample" = particular realization of such a process

# Recombination - reshuffles genetic variations and produces novel combinations of existing alleles



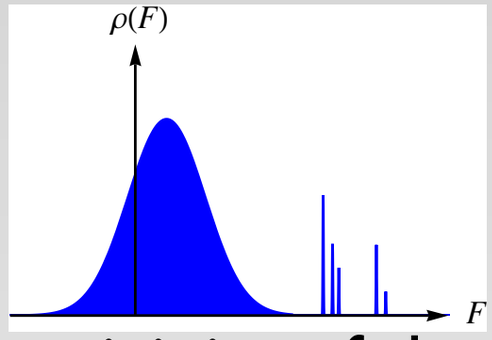




$$\dot{n}(\boldsymbol{g}) = \left(F(\{\boldsymbol{g}\}) - \overline{F}(t)\right) n(\boldsymbol{g}) + r\left[N^{-1}\sum_{\boldsymbol{g}',\boldsymbol{g}''} K(\boldsymbol{g}|\boldsymbol{g}',\boldsymbol{g}'')n(\boldsymbol{g}')n(\boldsymbol{g}'') - n(\boldsymbol{g})\right]$$

Smoluchowski equation

# Condensation r = 0




At short $t$ the averages are dominated by vicinity of the peak of $\rho(F)$

At $t > t_c$ the dominant contribution shifts to the leading edge of the distribution

With time the population shifts to fitter and fitter genotypes and eventually condenses.

$$\langle Y_t \rangle = \left\langle \sum_i \left( \frac{n_i(t)}{\sum_j n_j(t)} \right)^2 \right\rangle$$

probability of two individuals being identical

spin-glass order parameter

$$\langle Y_t \rangle = \left\langle \sum_{i=1}^{N} \int_0^\infty \mathrm{d}z z e^{2(F_i - \overline{F})t - z \sum_{i'=1}^{N} e^{(F_{i'} - r - \overline{F})t}} \right\rangle$$

$$\langle Y_t \rangle = N \int_0^\infty \mathrm{d}z z \int \mathrm{d}F_i \rho(F_i) n_i^2(t) e^{-z n_i(t)} \left[ \int \mathrm{d}F_j \rho(F_j) e^{-z n_j(t)} \right]^{N-1}$$

# Effective model

$$\sum_{i=1}^{N-M} e^{(F_i-r)t-\int_0^t \mathrm{d}t'\overline{F}(t')} + \sum_{j=1}^{M} e^{(F_j-r)(t-t_j)-\int_{t_j}^t \mathrm{d}t'\overline{F}(t')} = N$$

forefathers recombinants

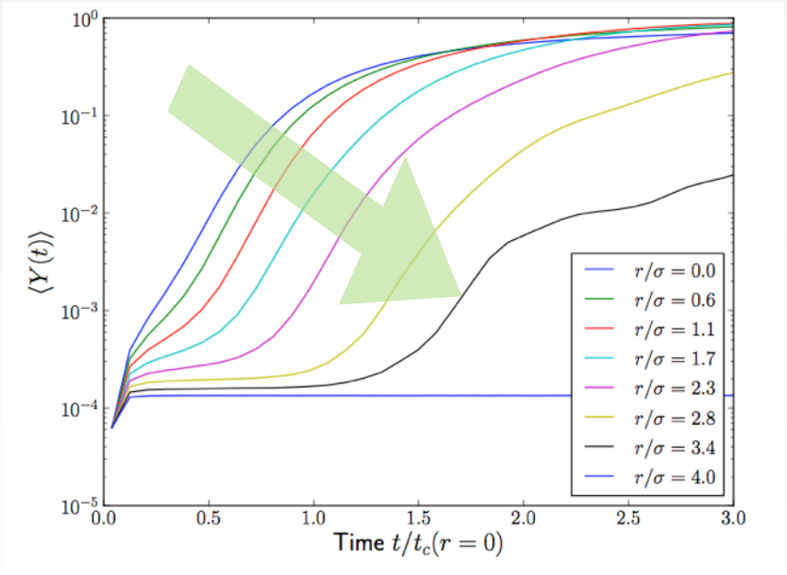
$$\langle Y_t \rangle = \left\langle \sum_i \left( \frac{n_i(t)}{\sum_j n_j(t)} \right)^2 \right\rangle$$

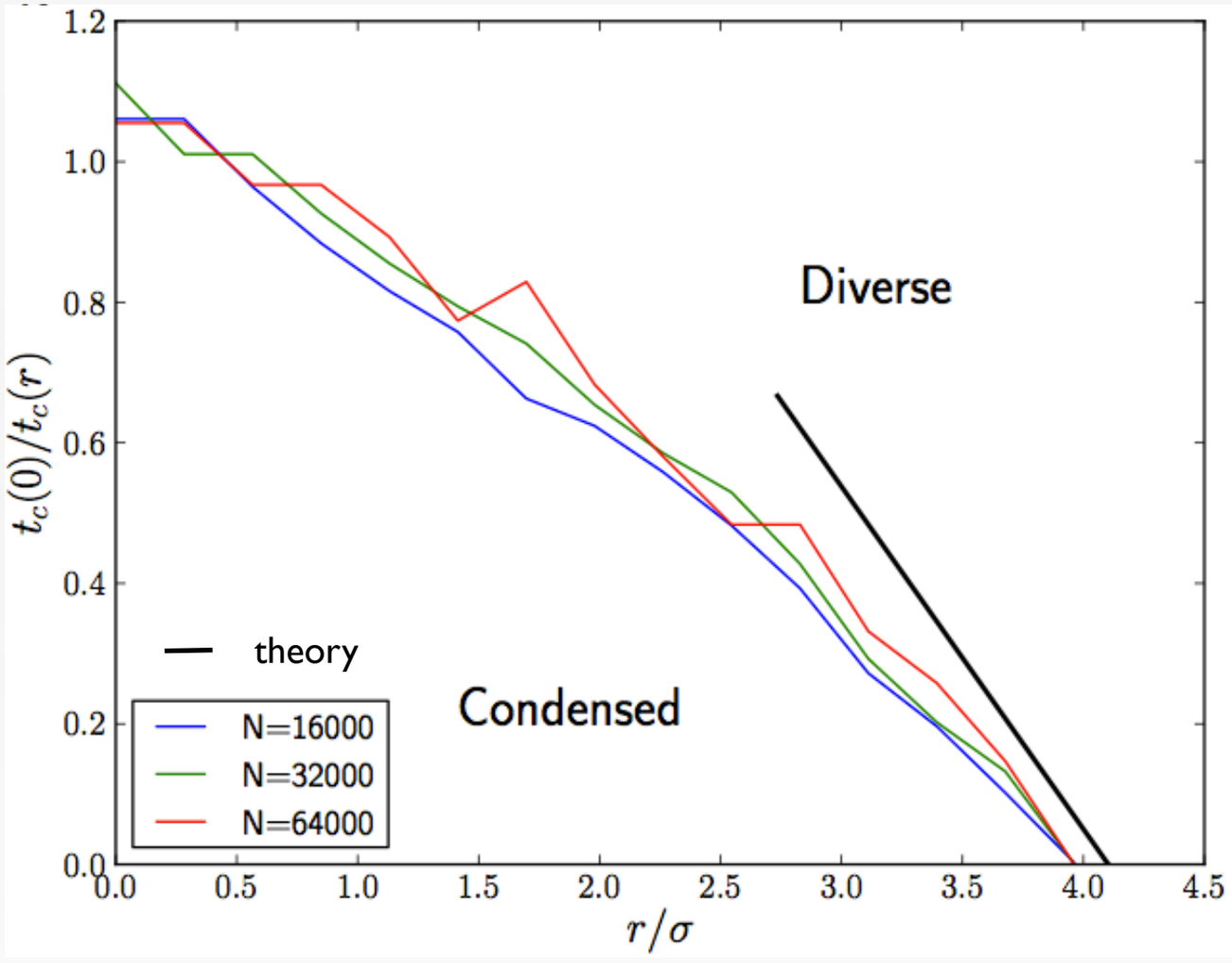
averaging:

- statistics of fitness
- Poisson process - arrival of $M$ recombinants at times $t_j$

$$\frac{t_c(r)}{t_c} = \frac{1}{1 - r/r_c}$$

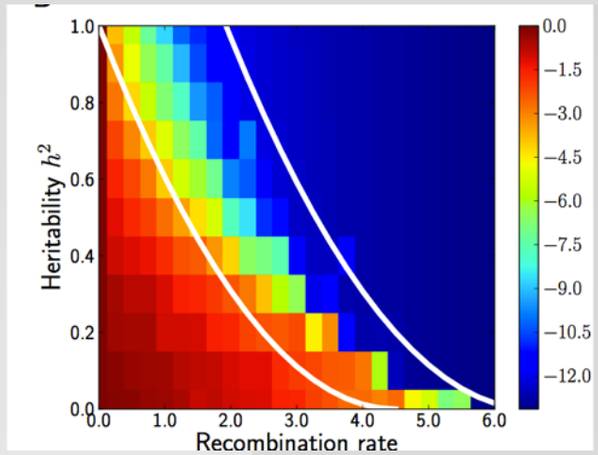
$$r_c \equiv \sigma\sqrt{2\ln(N)}$$

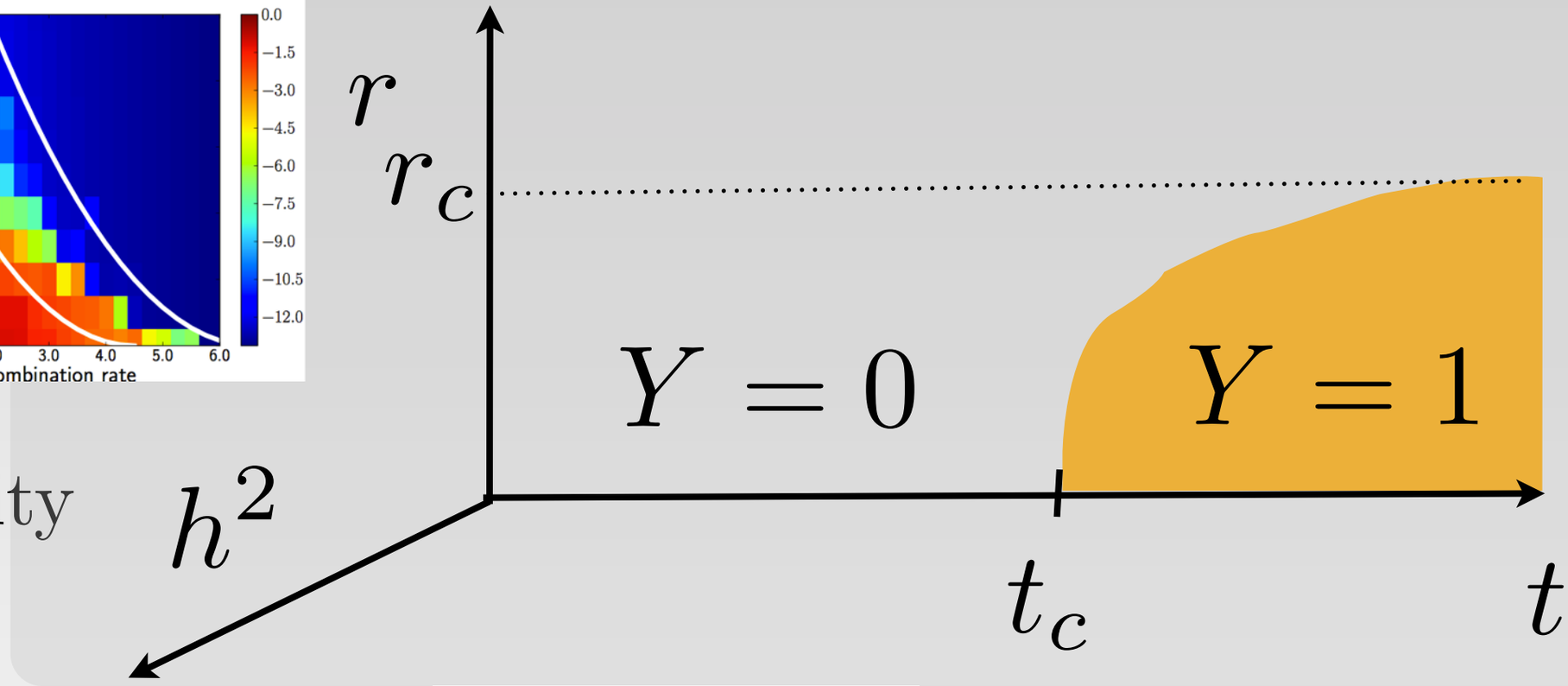



*Similar behavior observed in Barton 1983, Franklin and Levontin, 1970*

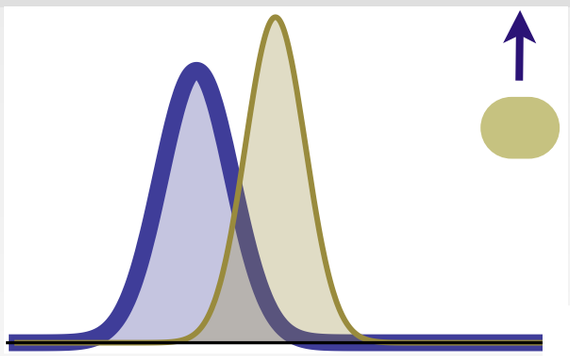
**Heritability -** how does the fitness of recombinants relate to that of the parents.



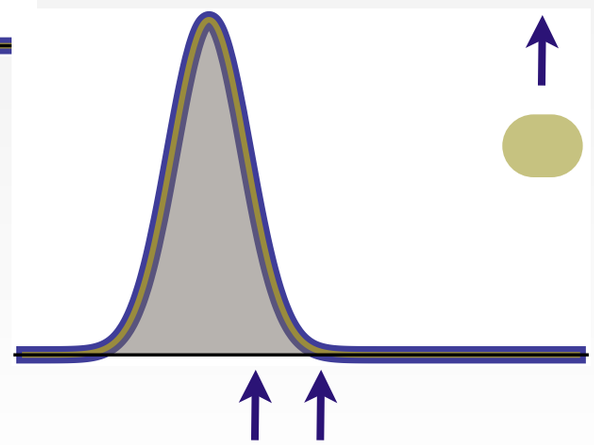




$h^2$ heritability

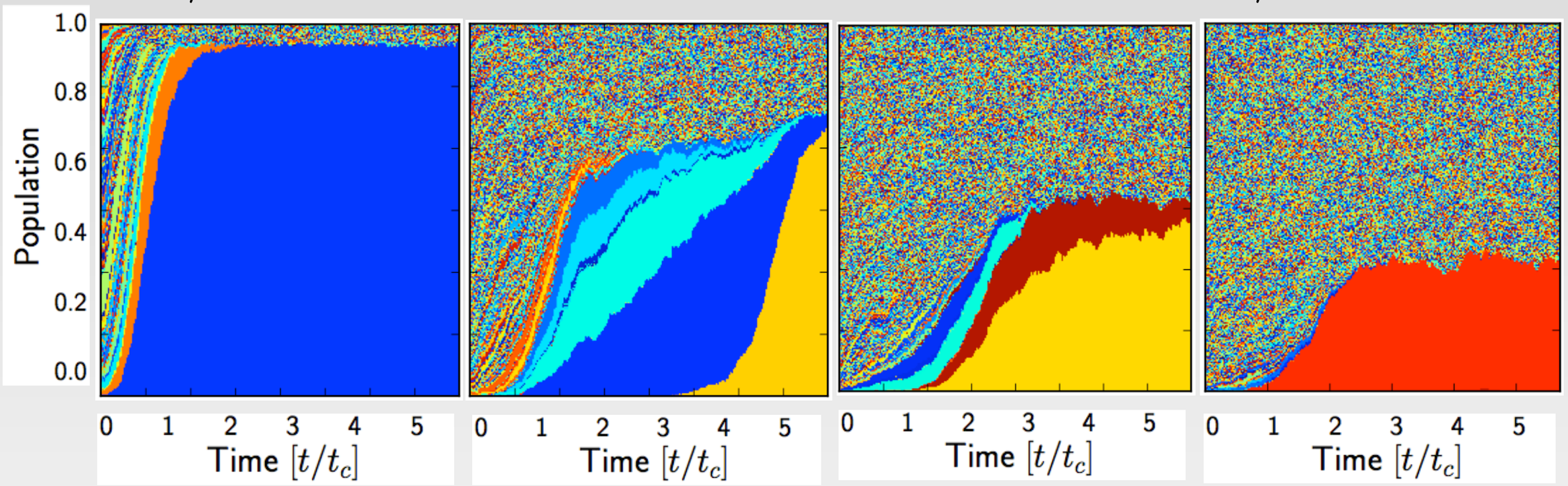
highly heritable trait




trait of low heritability



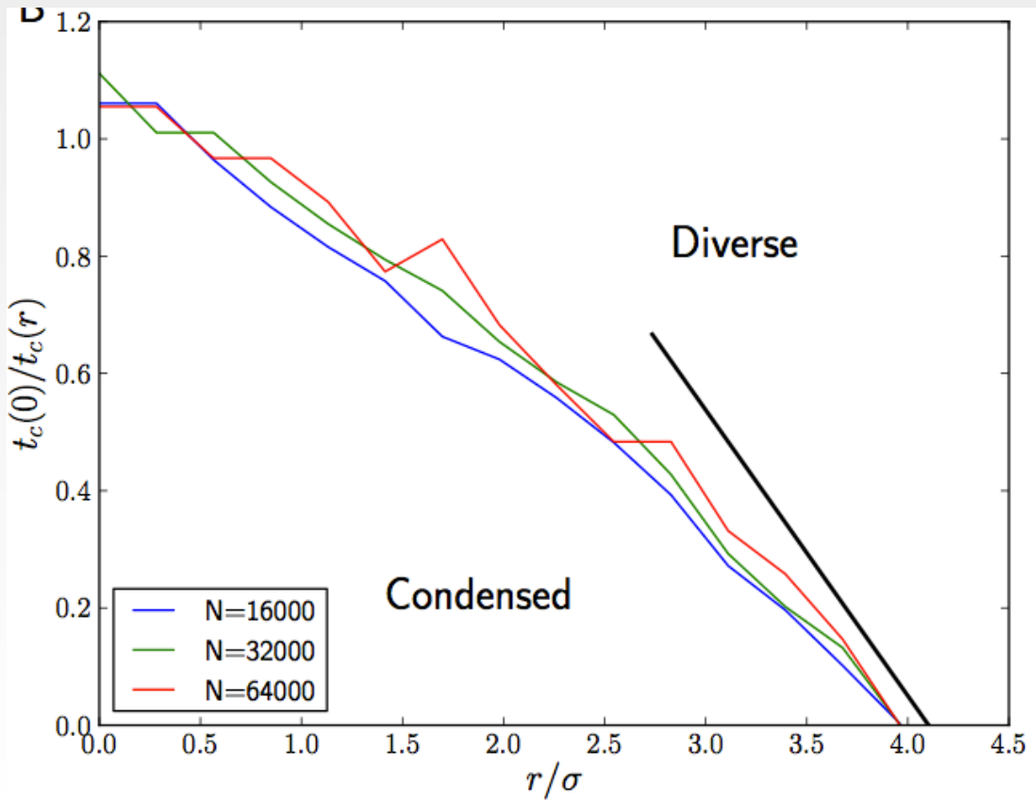
# without heritability



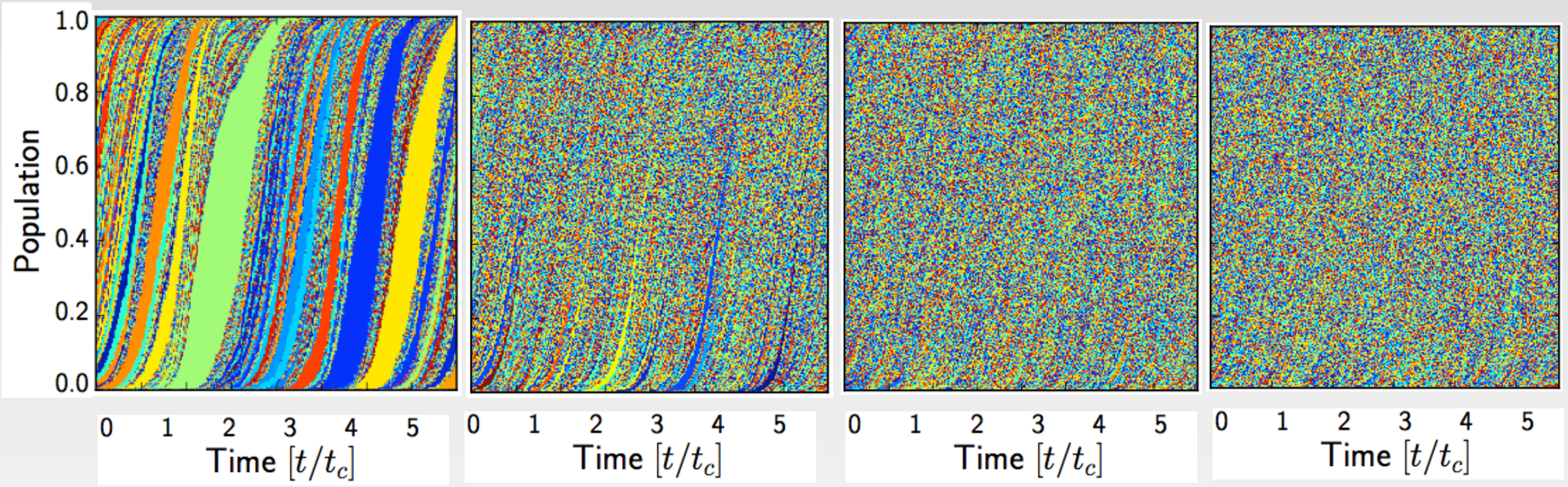

clone = color

same colors grouped

increasing recombination rate delays the transition to condensed state



# high heritability




Large clones cease to exist.
Most population is made out of short lived genotypes.

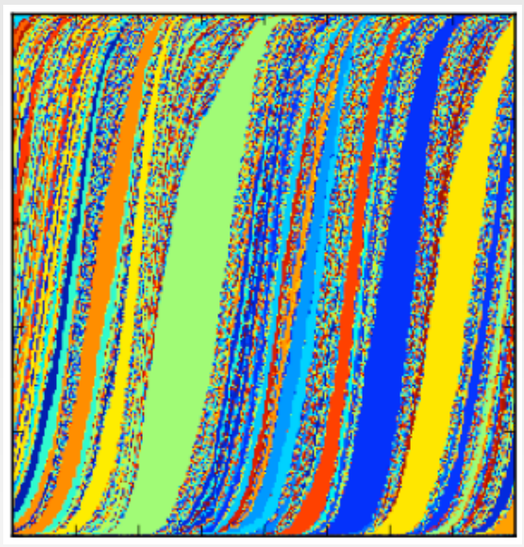
$$t_c = \sigma^{-1}\sqrt{2\ln(N)}$$

# Traveling solutions for additive fitness $h^2 = 1$

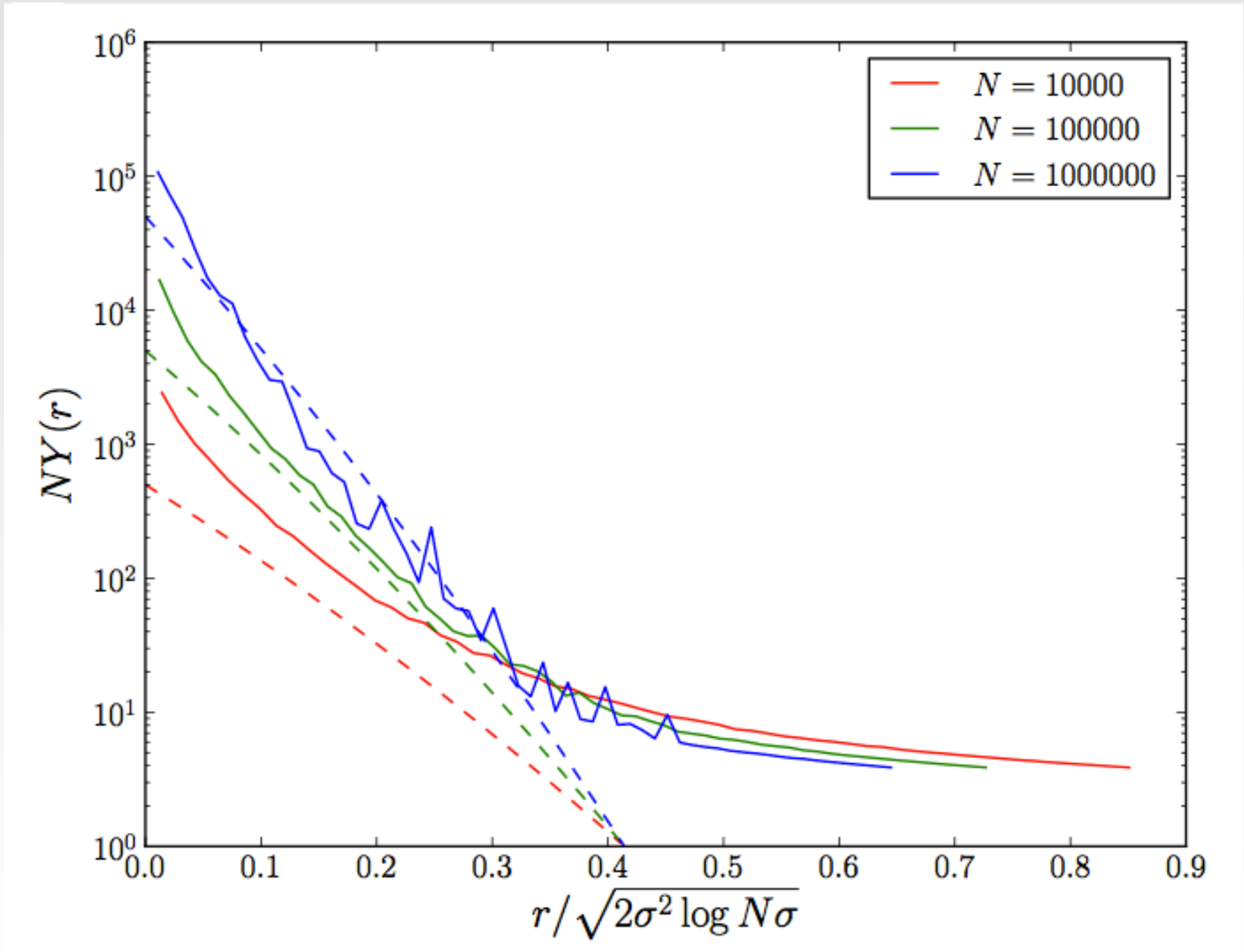
No aging = old genotypes are replaced; dominant genotypes have a finite characteristic age $\tau_j$

$$n_j = e^{(A_j - r)\tau_j - \frac{\tau_j^2 \sigma^2}{2}}$$

$$\langle Y(r) \rangle \approx 2N^{-1} + re^{-r\sigma^{-1}\sqrt{2\ln N} - \frac{r^2}{2\sigma^2}}$$



(Cohen et al., 2005a;
Desai and Fisher, 2007;
Hallatschek, 2011; Neher et al., 2010;
Rouzine et al., 2003;
Tsimring et al., 1996)



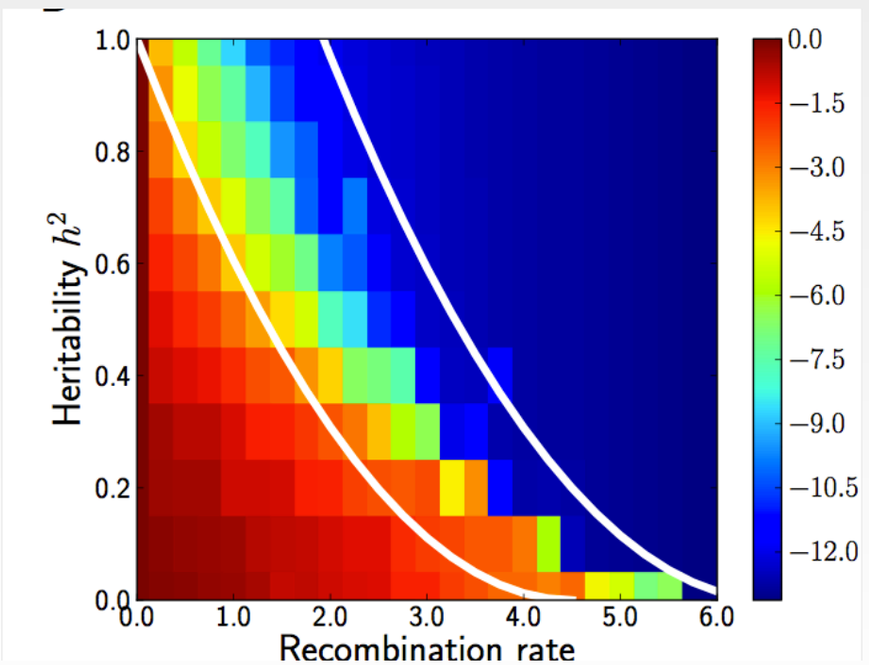
# Future directions: clonal condensation

Connection to real world populations: heritability (between 0 and 1), large number of loci.

- more complicated models would reveal more structure than this simple "dust/clone" dichotomy

- better understanding of a "mixed" phase $0 < h^2 < 1$
- adding mutations
- REM ⟶ Sherrington - Kirkpatrick
- dynamics population getting homogeneous and diverse

$$\log\langle Y_\infty(r,h)\rangle$$

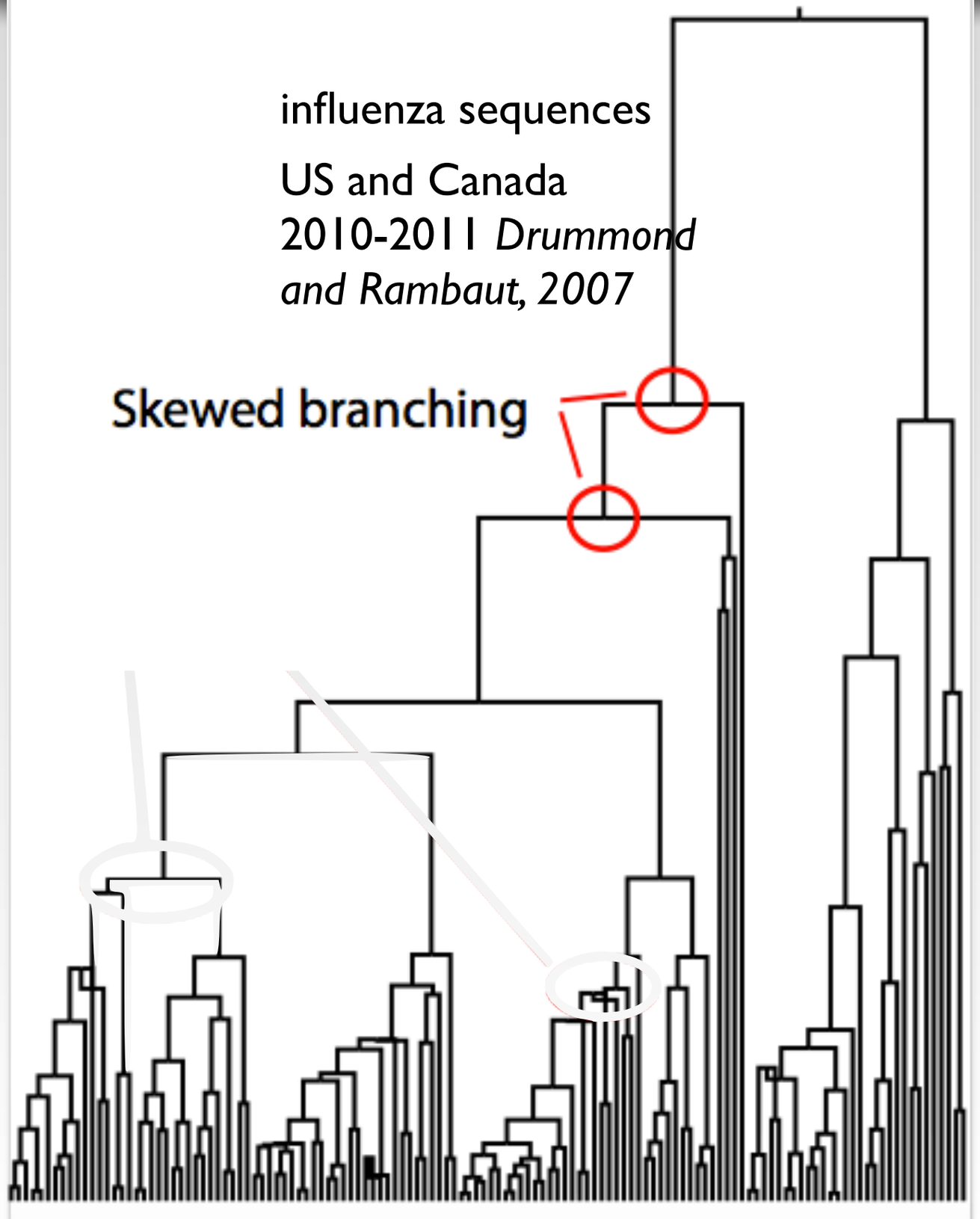


# Inference of fitness of the leaves from genealogical trees

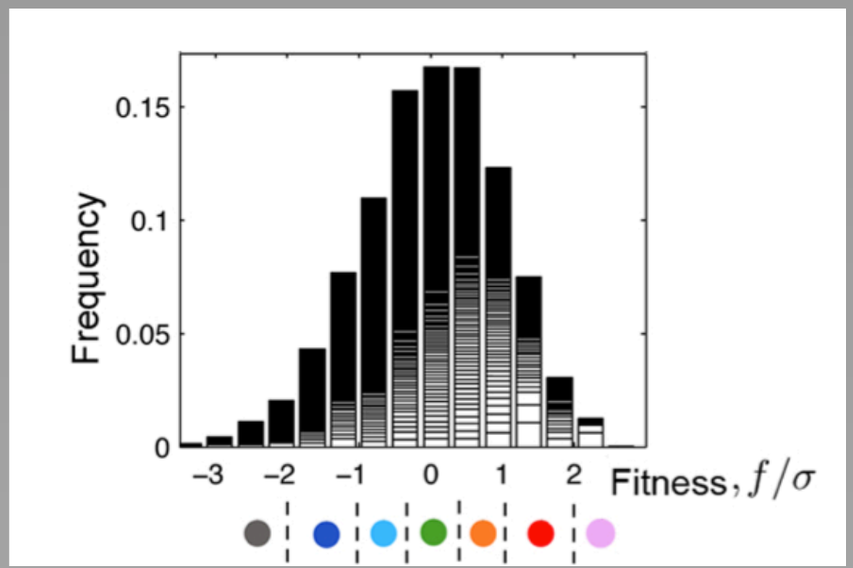
influenza sequences

US and Canada
2010-2011 *Drummond and Rambaut, 2007*

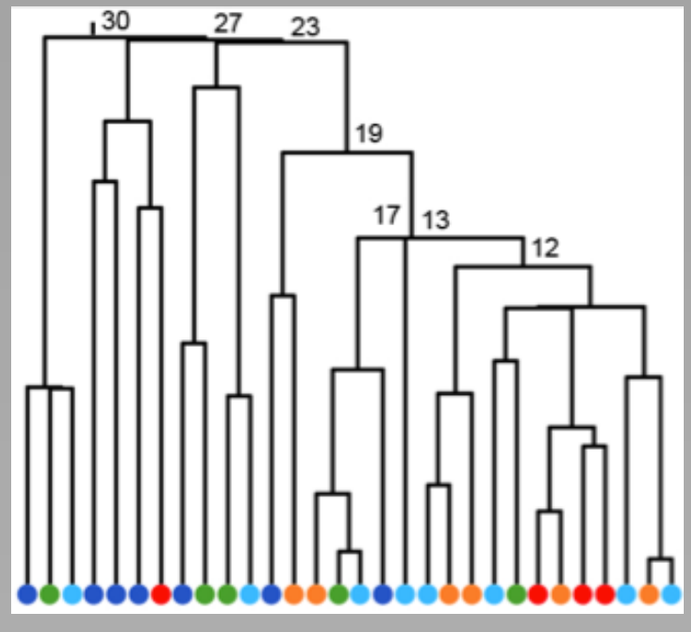
Skewed branching



*Dayarian, Shraiman 2012*




heuristic approach

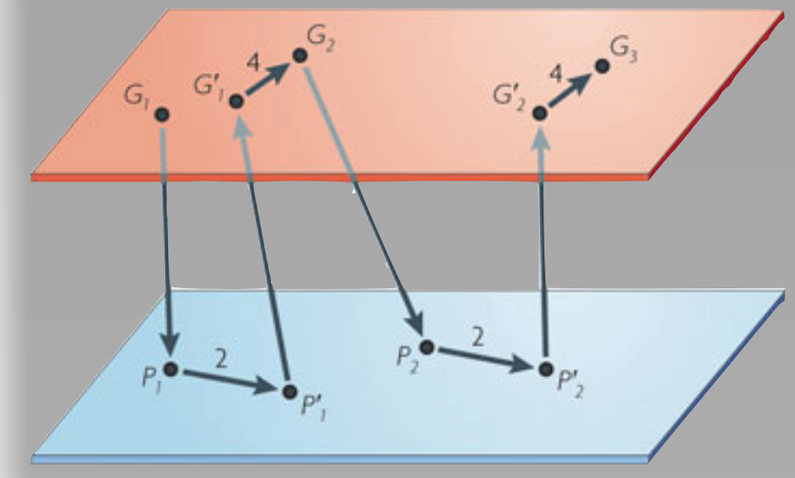


# Phenotype internal set of states of individual organisms

set of organisms' traits

example: stripes, color, biochemical or physiological properties, behavior...

phenotype-genotype map




+ environment

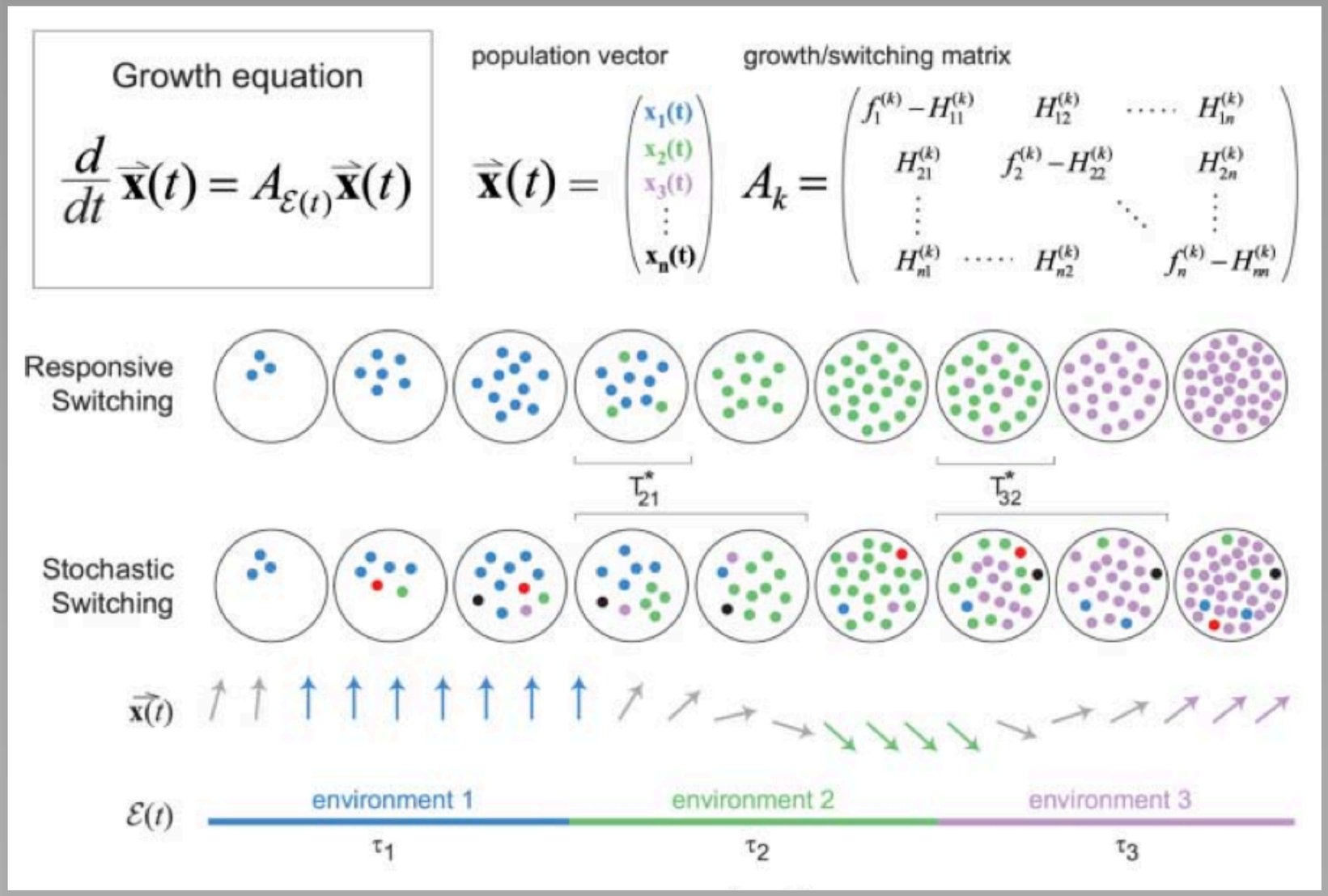
phenotypic switching

environmental fluctuations = external forcing

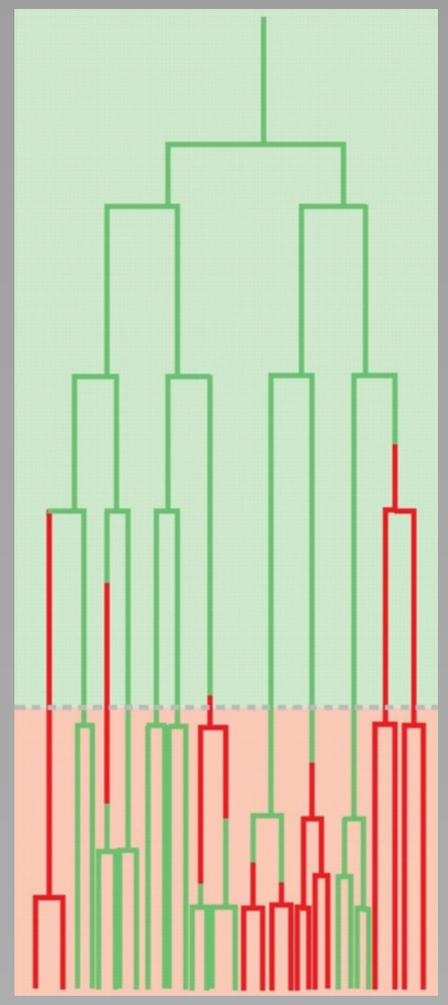
temporal fluctuations drive the system away from equilibrium

equilibrium perspective - individual histories (trajectories in the phenotypic space observed in the population )

# Phenotypic switching




Responsive and stochastic switching
*Kussell, Leibler 2005*

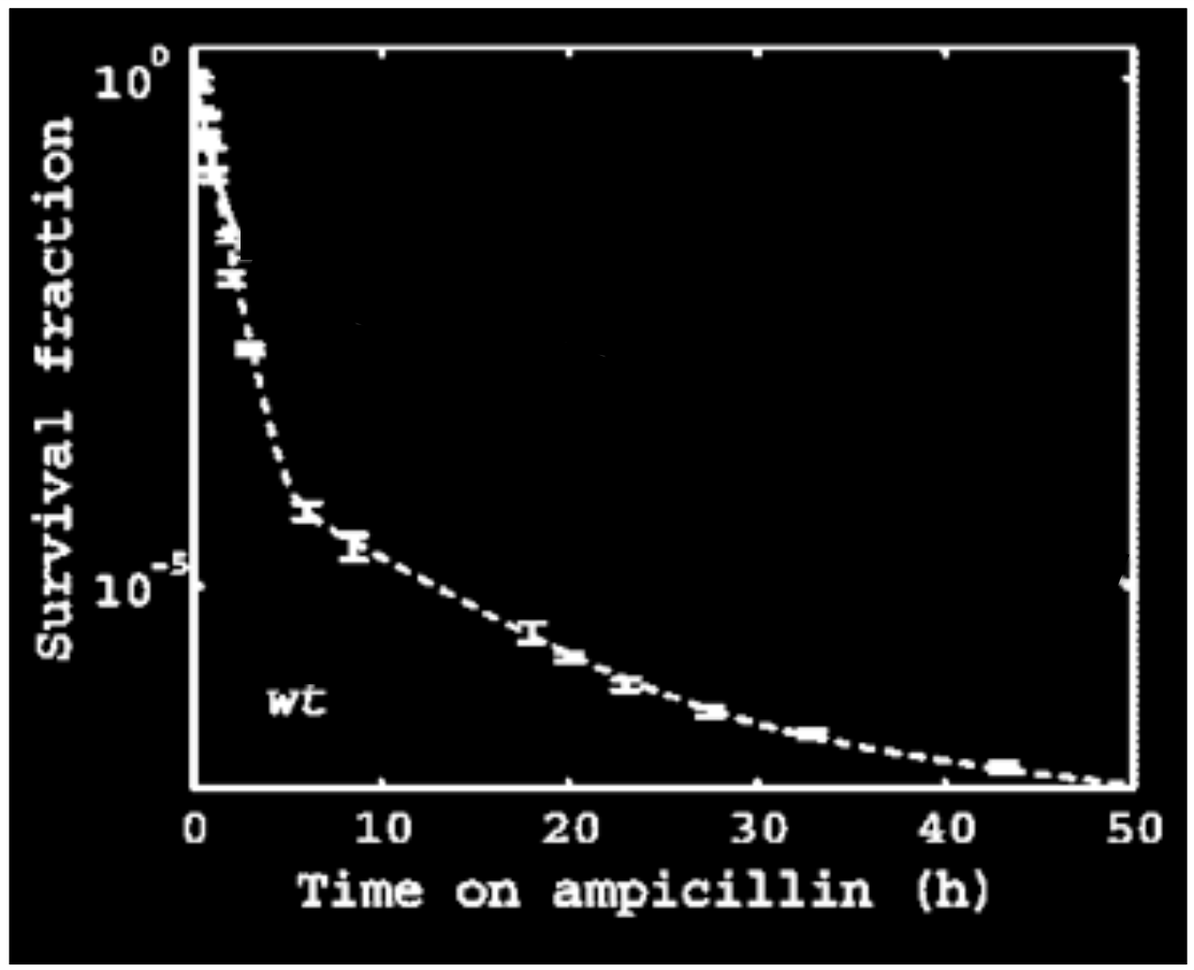


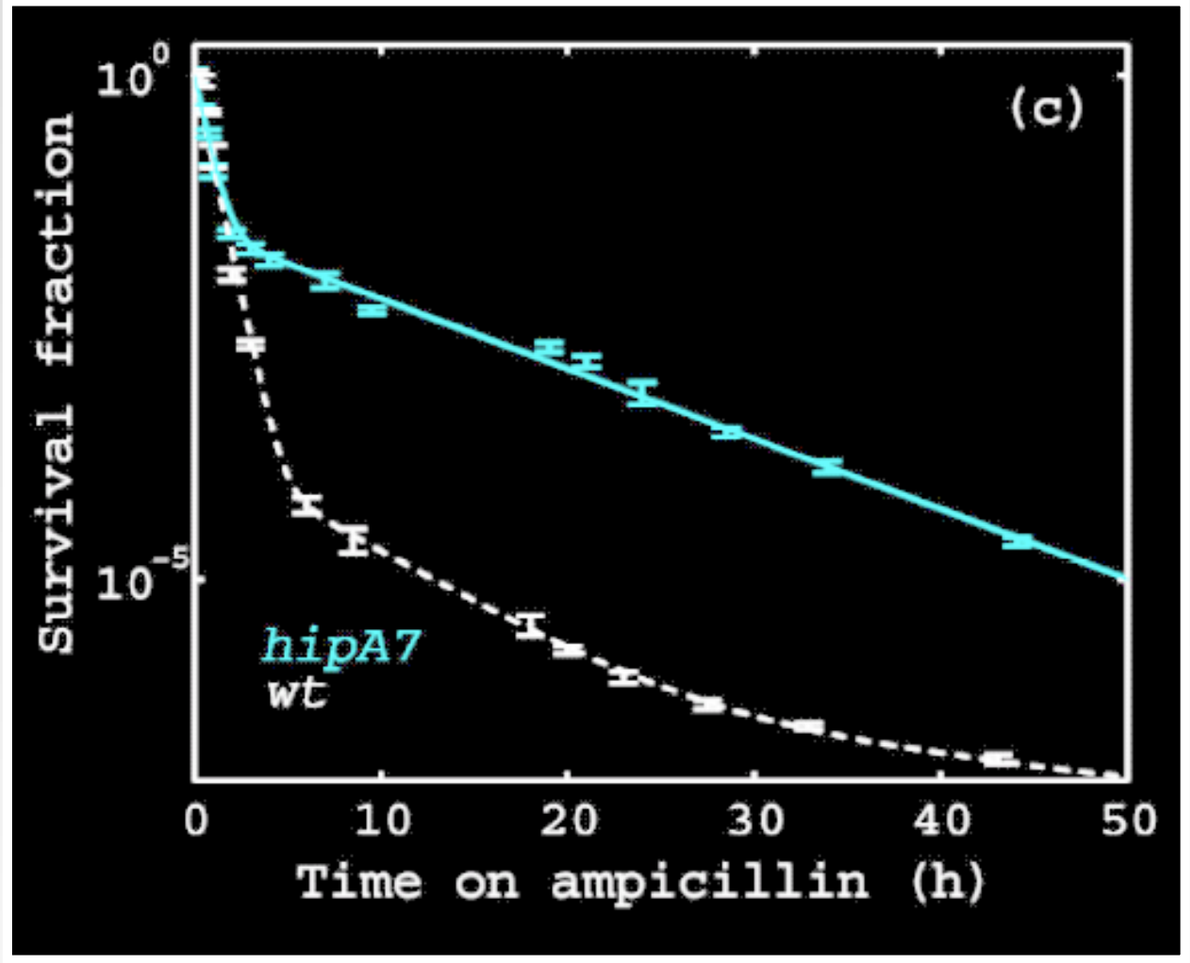
*Kussell, Leibler, 2010*

# Persistors

different behavior, dividing more slowly

## antibiotic persistence




*Balaban et al, 2004*

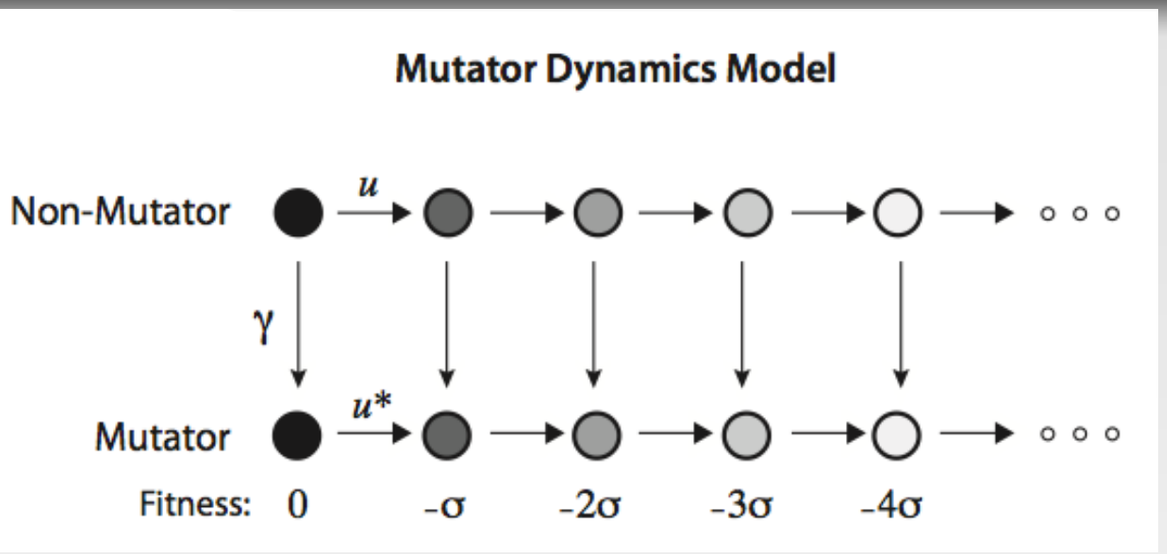
*hipA7 - high persistence mutants*

# Mutators

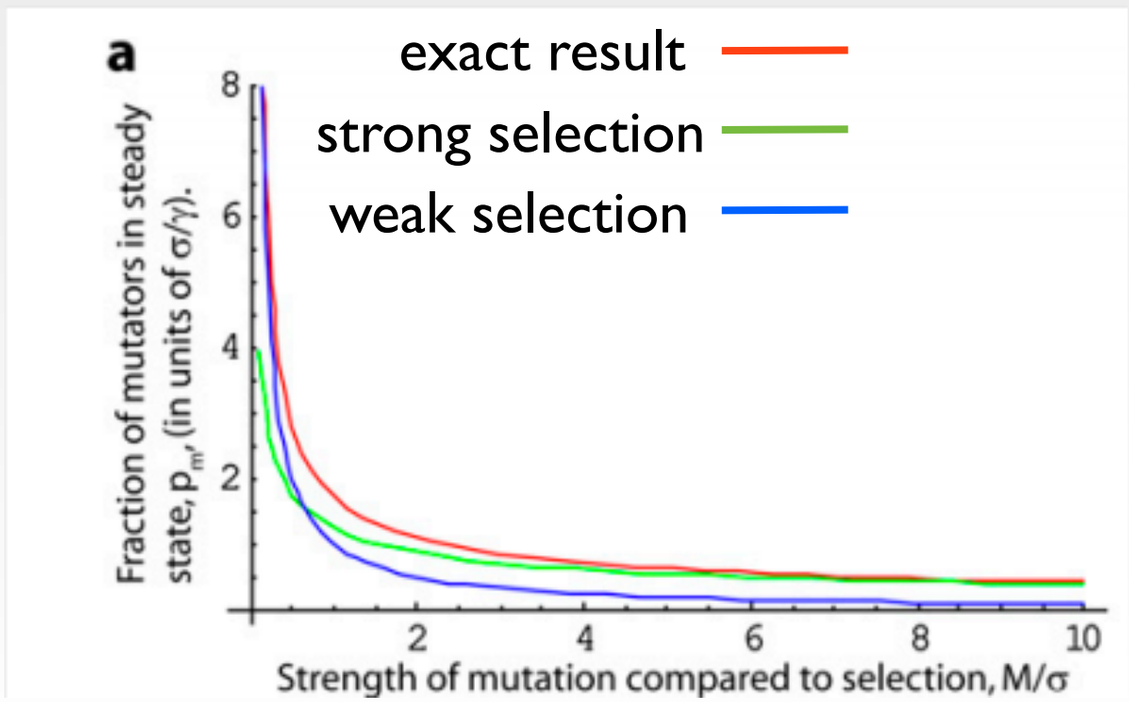
individuals with a much higher mutation rate in a population (up to $10^5$ higher)

*Desai, Fisher 2010*
*Travis, Travis 2002*
*Lenski experiments - evolving E. coli since 1988*

constant environment




*Desai, Fisher 2010:* Even in situations where selection on average acts against mutators, so they cannot stably invade, the mutators can still occasionally generate beneficial mutations and hence be important to the evolution of the population.

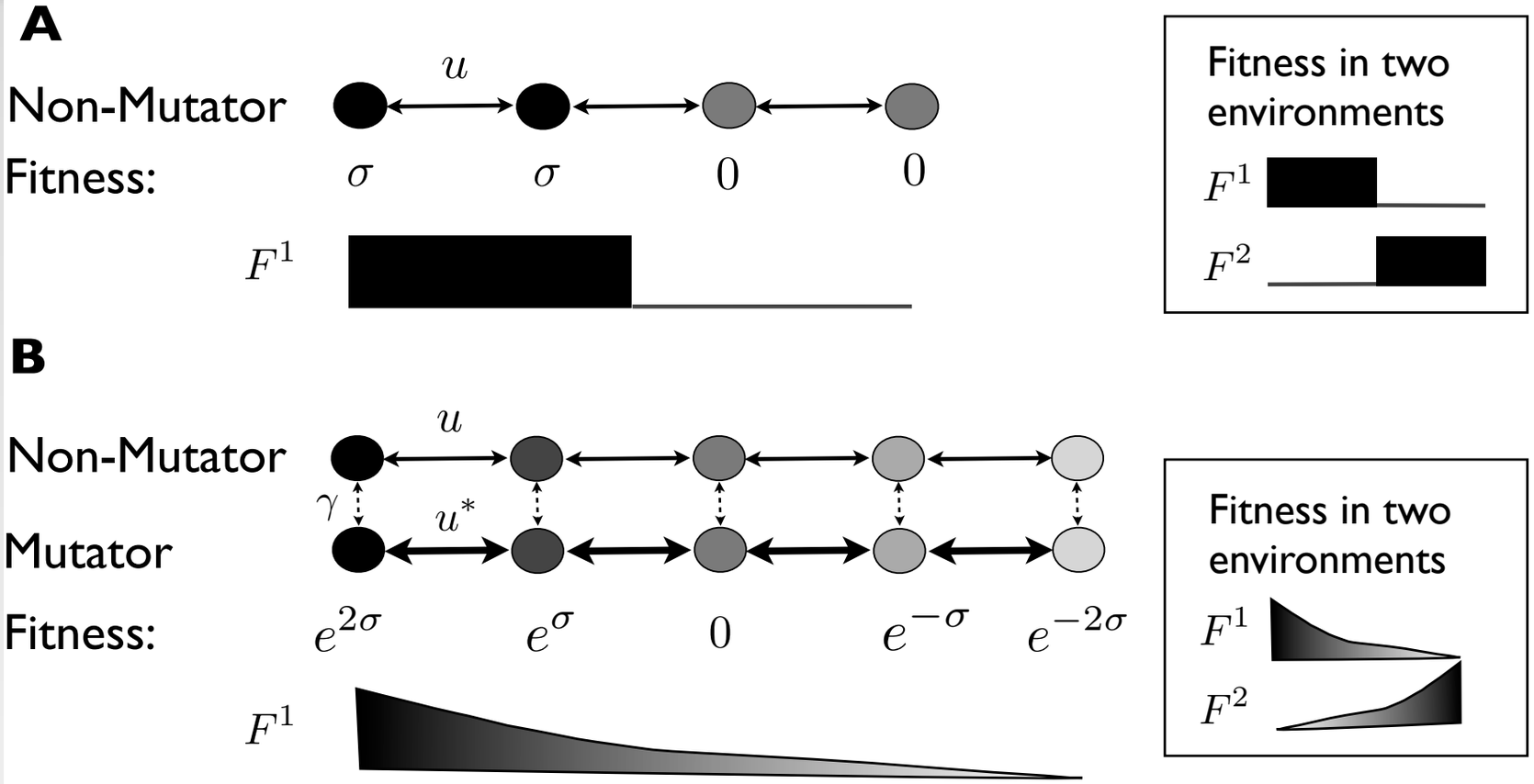



*Desai, Fisher 2010:* findings confirmed by R. Lenski experiments:

$$p_m \propto \gamma / M$$

# Mutators in fluctuating environments

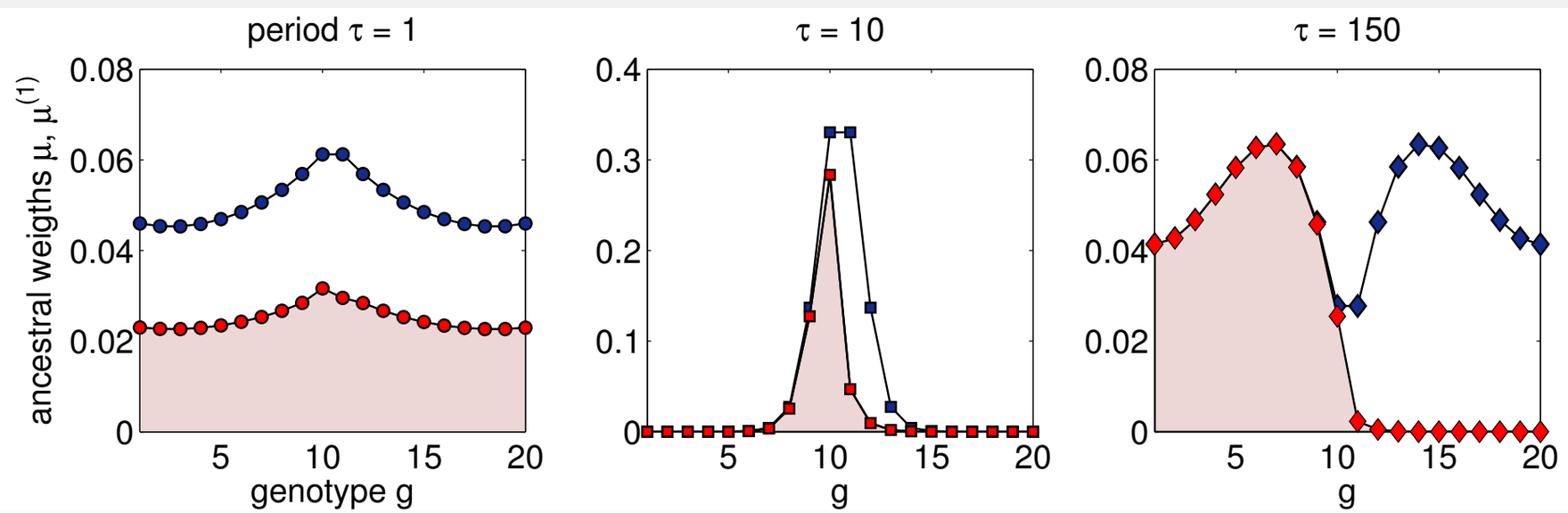
*Kussell, Leibler 2005; Kussell, Leibler, Grosberg, 2006; Travis Travis 2002*




Ancestral distributions of non-mutators in the periodic case of two environments

Analogous problem:

heteropolymer localization on the interface




localization for:

$$\frac{1}{f} < \tau < \frac{L^2}{4u}$$

# Summary

• **missing:** better theory and quantitative aspects of population genetics; relevant timescales?

• **present:** abundance of data.

•**difficulties**: numerous timescales, quenched disorder, incomplete statistics (we only see a single realization of the outcome of the evolution)

•**relevance:** drug resistance, disease evolution, origin of life

• **Population genetics - on statistical mechanics language** - relations to polymers, path integrals, localization phenomena naturally emerge, non-trivial "thermodynamic limit"

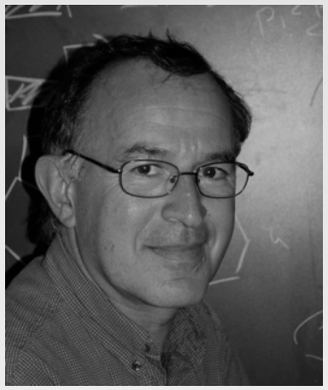
# Collaborators

Richard Neher

Max Planck, Tubingen

Marc Mézard

CNRS, Paris



Boris Shraiman

KITP

Adel Dayarian

KITP

Edo Kussell NYU