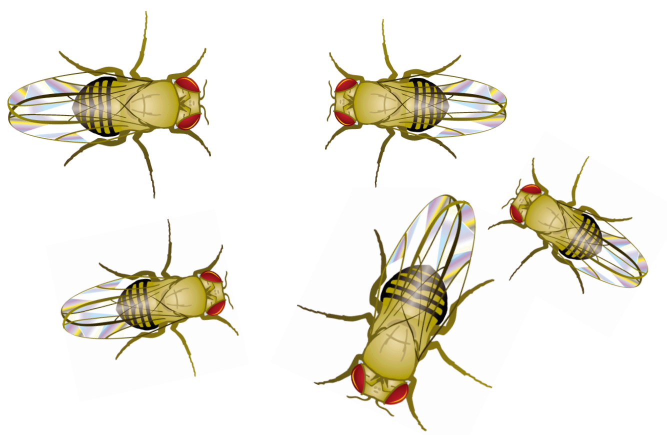
Intensive Course on Population Genetics

Ville Mustonen

- Population genetics: "The study of the distribution of inherited variation among a group of organisms of the same species" [Oxford Dictionary of Biology]



http://www.exploratorium.edu/exhibits/mutant_flies/normal.gif

http://www.emacswiki.org/emacs/EmacsIcons

Cross- and intra-species comparisons at the molecular level

m ingroup samples

```
AACTGTCCACGTTCTTCCGATGCAGCCTGA
AACTGTCCACGTTCTTCCGATGCAGCCTGA
AACTGTCCACGTTCTTCCGATGCAGCCTGA
AACTGTCCACGTTCTTCCGATGCAGCCTGA
AACTGTCCACGTTCTTCCGATGCAGCCTGA
AACTGTCCAGGTTCTTCCGATGCAGCCTGA
AACTGTCCAGGTTCTTCCGATGCAGCCTGA
AACTGTCCAGGTTCTTCCGATGCAGCCTGA
```

outgroup

```
AACTGTCCAGGTTCTTCCGACGCAGCCTGA
```

frequency counts

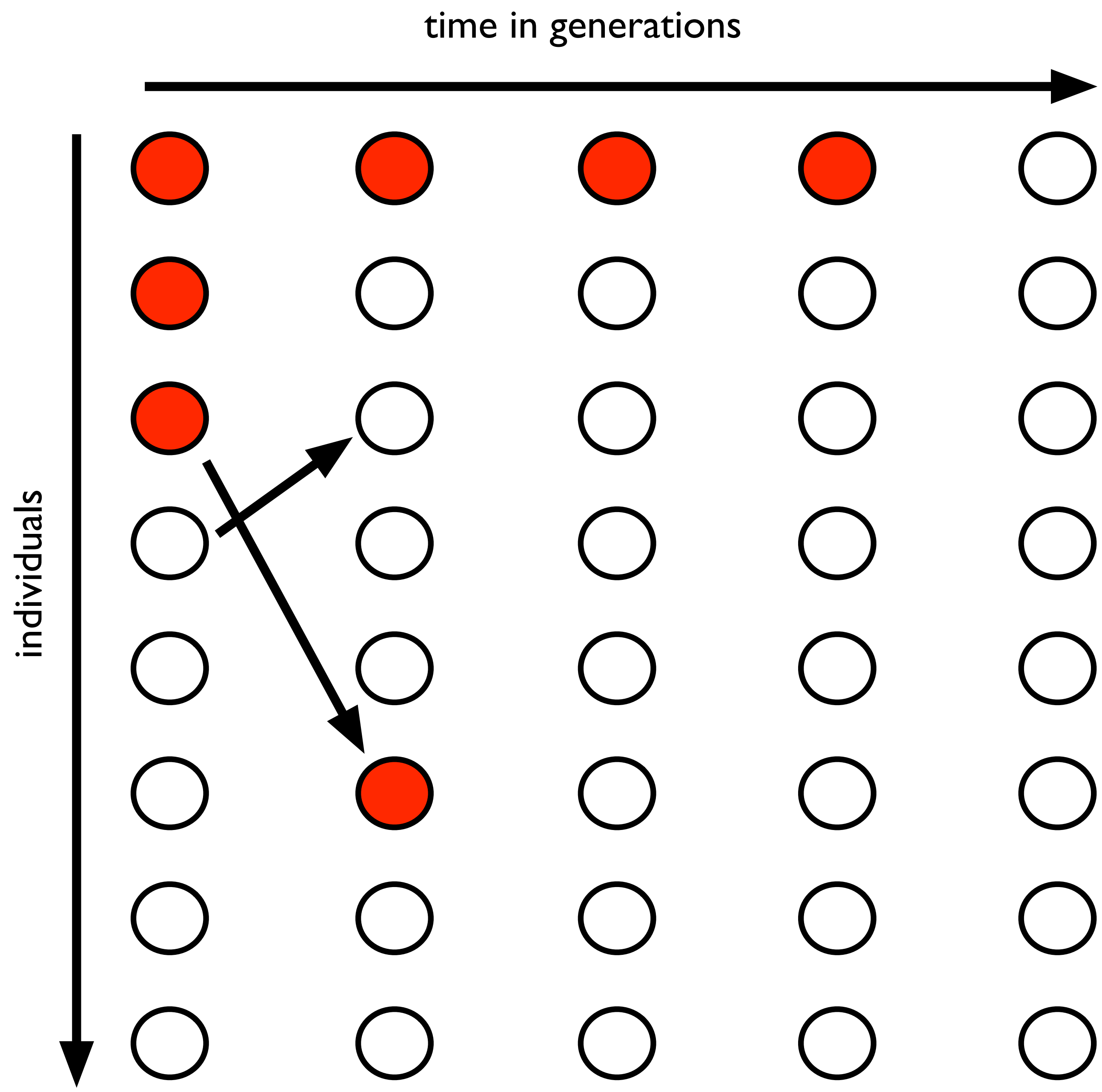
```
000000000500000000000800000000
```

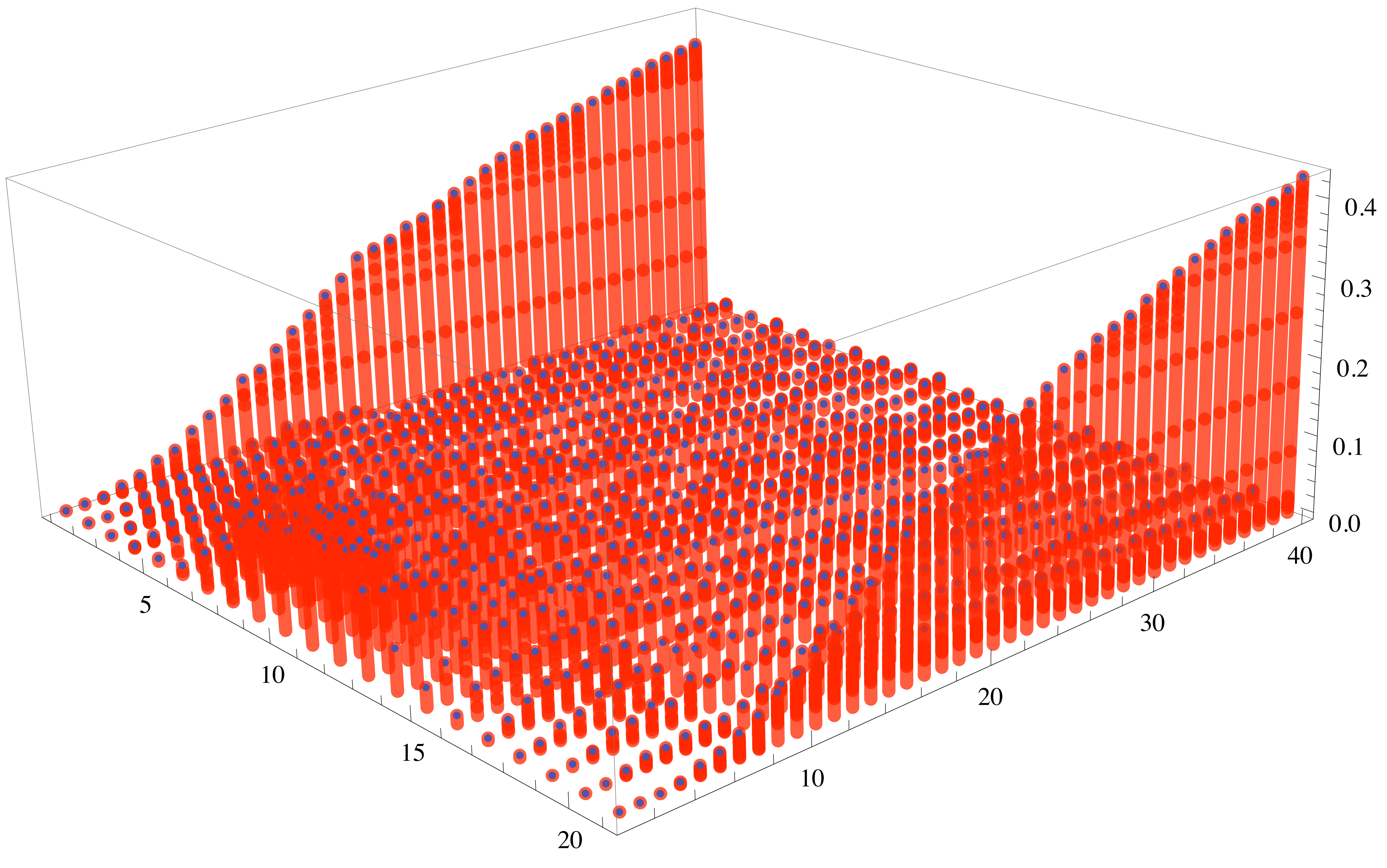
polymorphic locus (5) — substitution event (8)

- Let's study fundamental evolutionary forces:

  1. genetic drift (noise in reproduction)
  2. selection (differential reproductive success of individuals)
  3. mutation (process providing variation)
  4. ...

Genetic drift: noise in reproduction





0.4
0.3
0.2
0.1
0.0
5
10
15
20
10
20
30
40

Wright - Fisher process

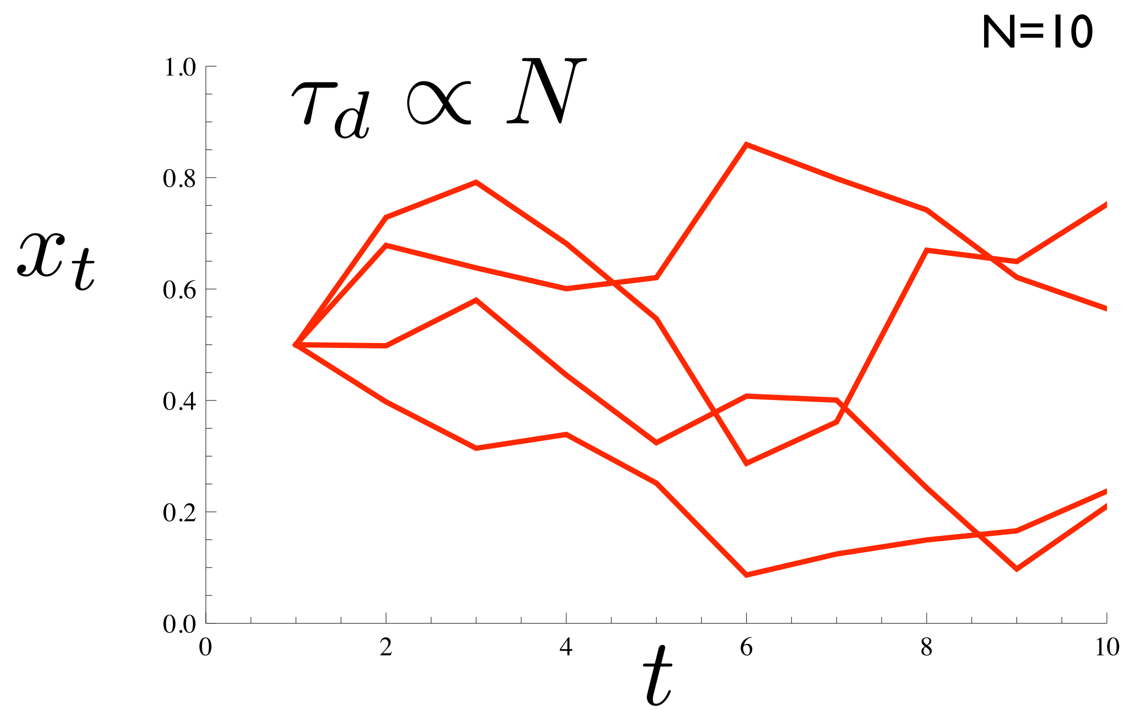
$$
\begin{aligned}
p_b &= N_b/N = x \\
p_a &= N_a/N = 1 - x
\end{aligned}
$$

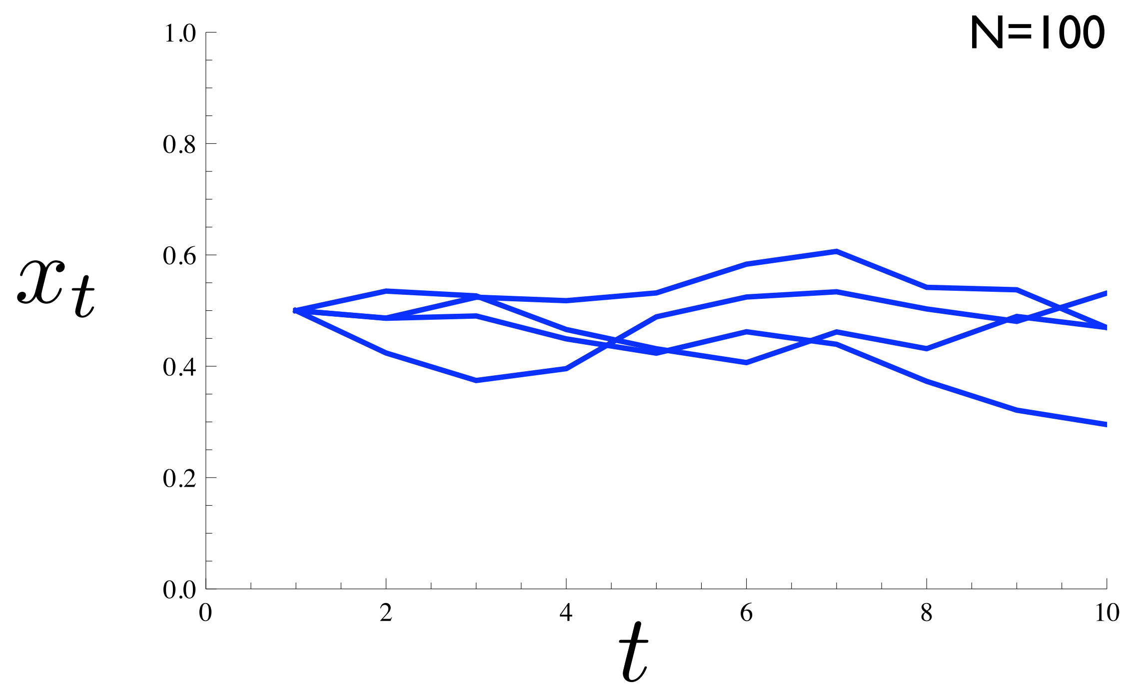
$$
P(m, N, t+1) = \binom{N}{m} x_t^m (1 - x_t)^{N-m}
$$

$$
\begin{aligned}
\langle m \rangle &= N x_t \\
\langle m^2 \rangle - \langle m \rangle^2 &= N x_t (1 - x_t)
\end{aligned}
$$

$$
\begin{aligned}
\langle x_{t+1} \rangle &= x_t \\
\langle x_{t+1}^2 \rangle - \langle x_{t+1} \rangle^2 &= \frac{x_t (1 - x_t)}{N}
\end{aligned}
$$

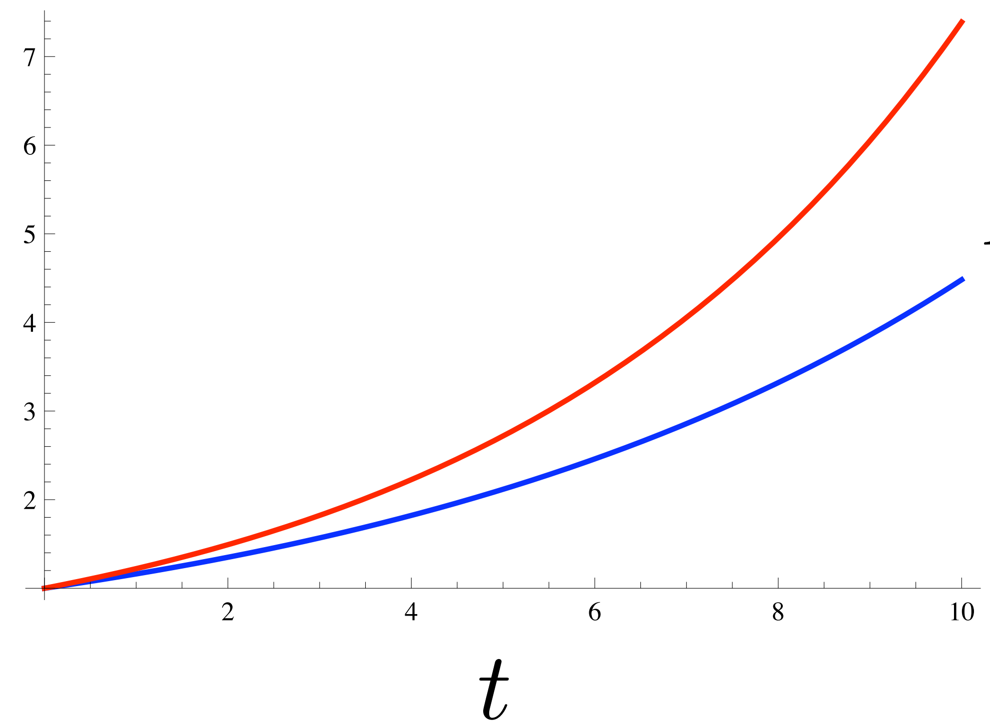
Wright - Fisher process





Selection: differential reproductive success of individuals

$$\dot{N}_a = F_a N_a$$



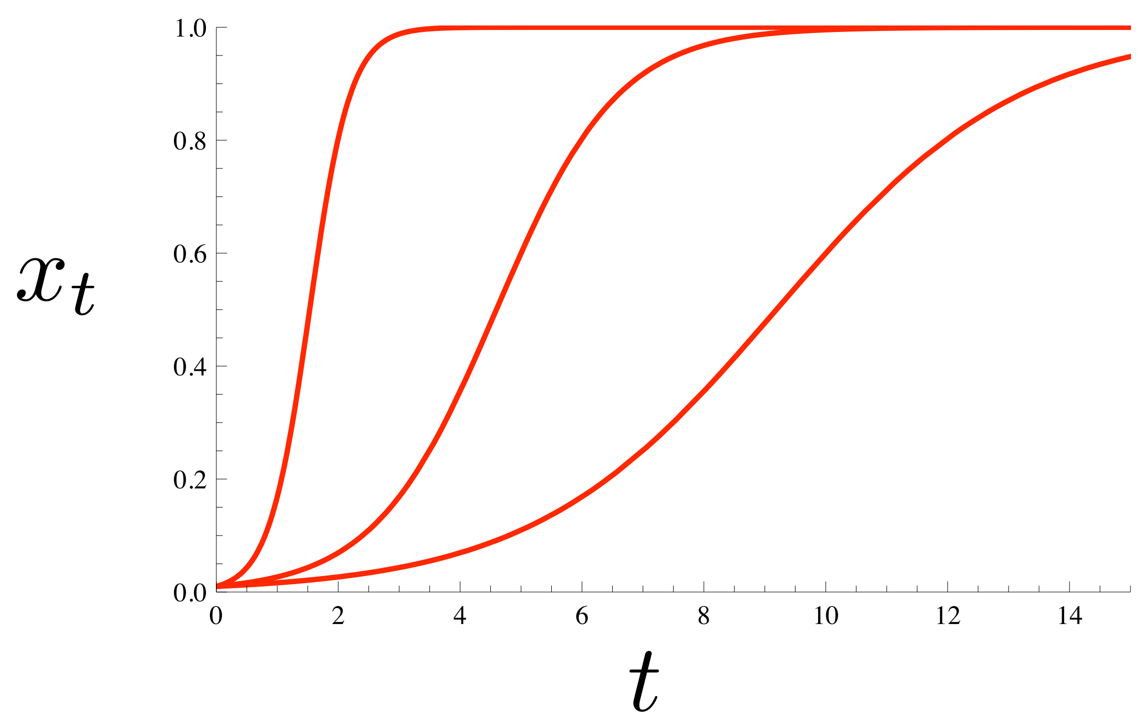
$$\dot{N}_b = F_b N_b$$

$$\dot{x} = \frac{d}{dt}\frac{N_b(t)}{N_a(t)+N_b(t)} = \frac{\dot{N}_b}{N_a+N_b} - \frac{N_b}{N_a+N_b}\frac{\dot{N}_b+\dot{N}_a}{N_a+N_b} = (F_b - F_a)x(1-x)$$

Selection

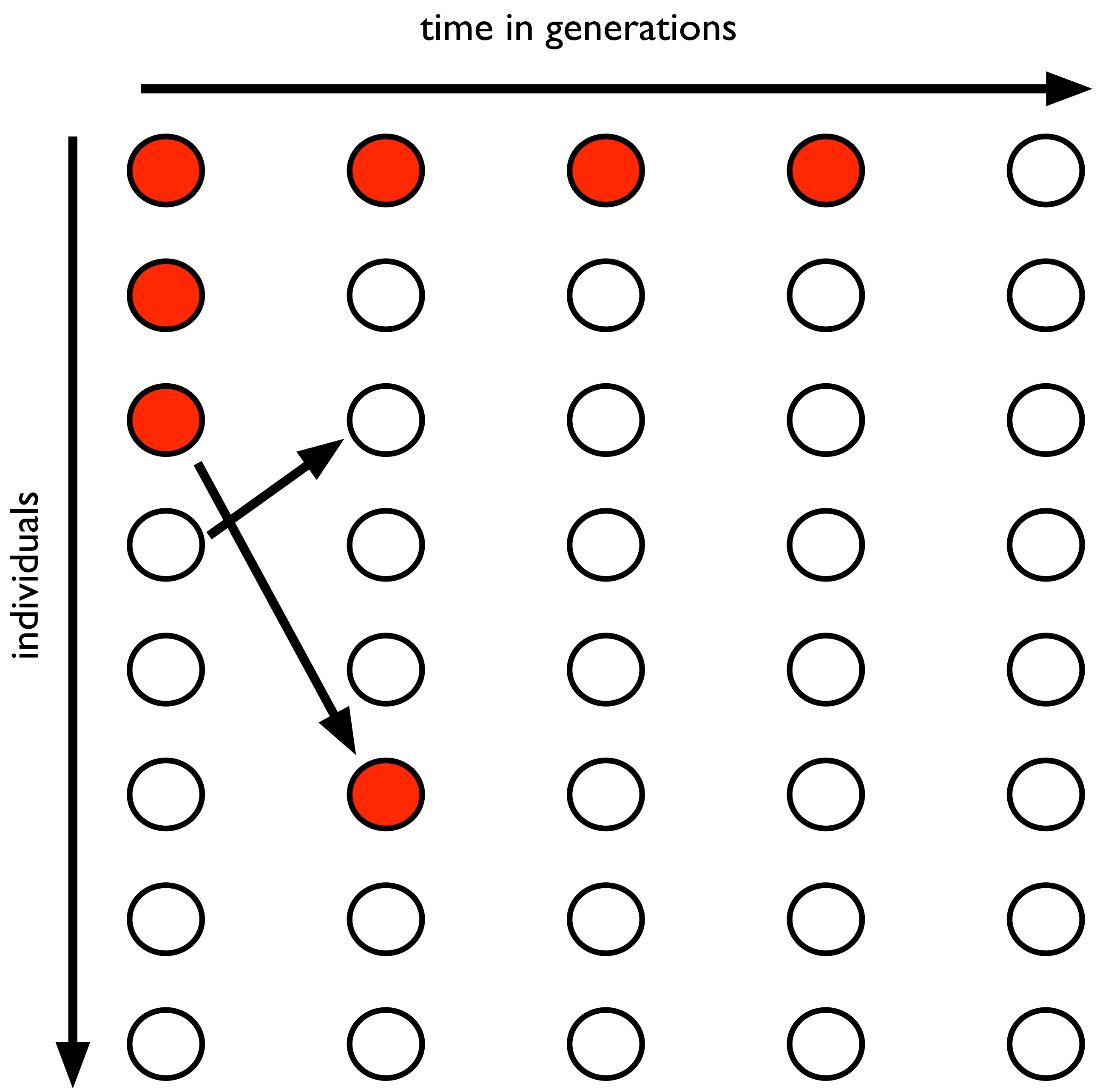
$$\dot{x} = F_{ba} x(1-x)$$

$$x(t) = \frac{x_0 \exp(F_{ba} t)}{1 + x_0(\exp(F_{ba} t) - 1)}$$



$$\tau_s \propto 1/F_{ba}$$

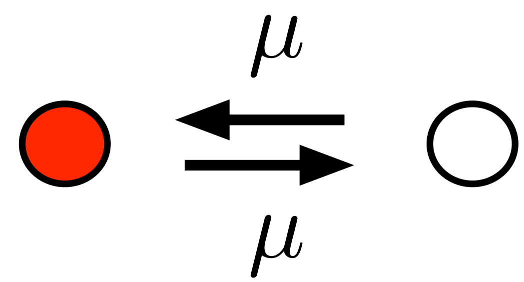
Mutation: source of variation



Mutation: source of variation

$$\dot{N}_a = \mu N_b - \mu N_a$$

$$\dot{N}_b = \mu N_a - \mu N_b$$

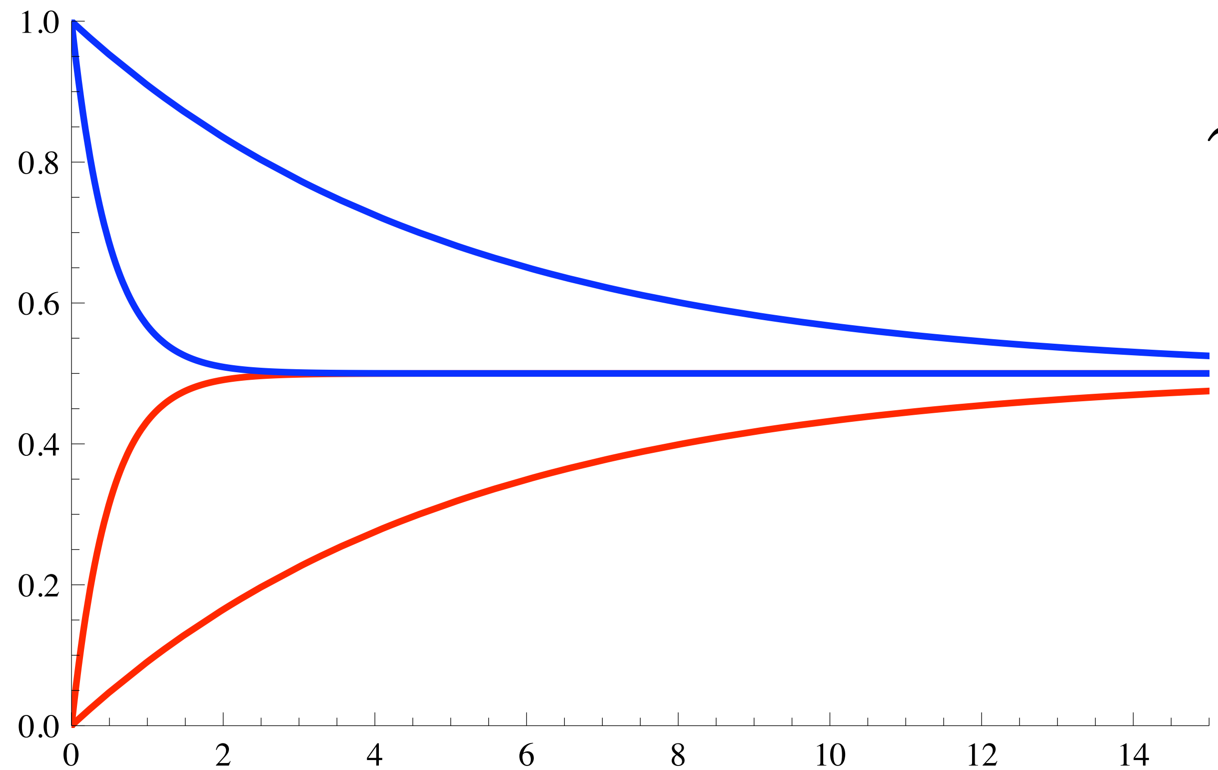


$$\dot{x} = \frac{d}{dt}\frac{N_b(t)}{N_a(t)+N_b(t)} = \frac{\dot{N}_b}{N_a+N_b} - \frac{N_b}{N_a+N_b}\frac{\dot{N}_b+\dot{N}_a}{N_a+N_b}$$
$$= \mu(1-2x)$$

Mutation

$$\dot{x} = \mu(1 - 2x)$$

$$x(t) = \frac{1}{2}(1 + \exp(-2\mu t)(2x_0 - 1))$$

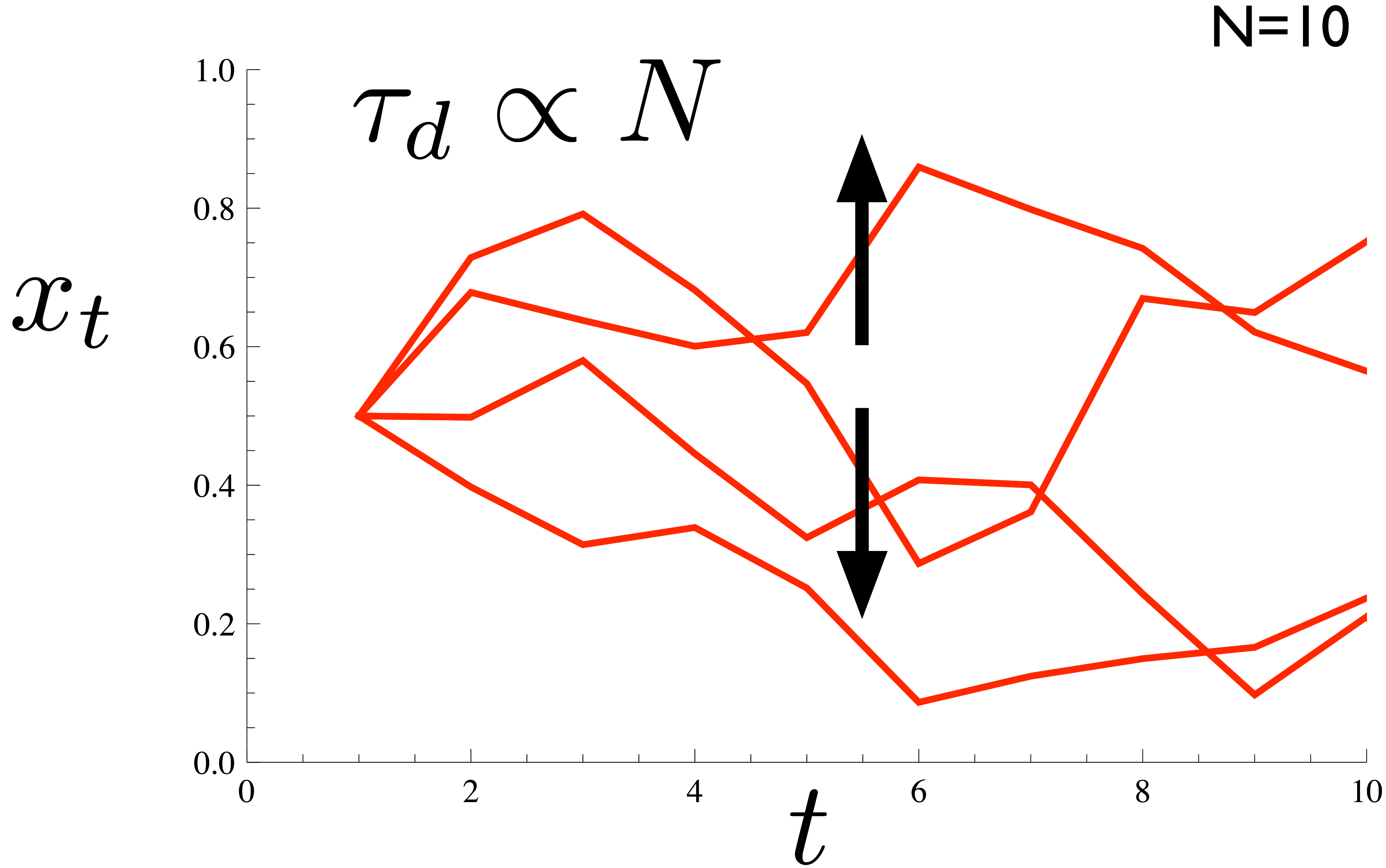


x_t

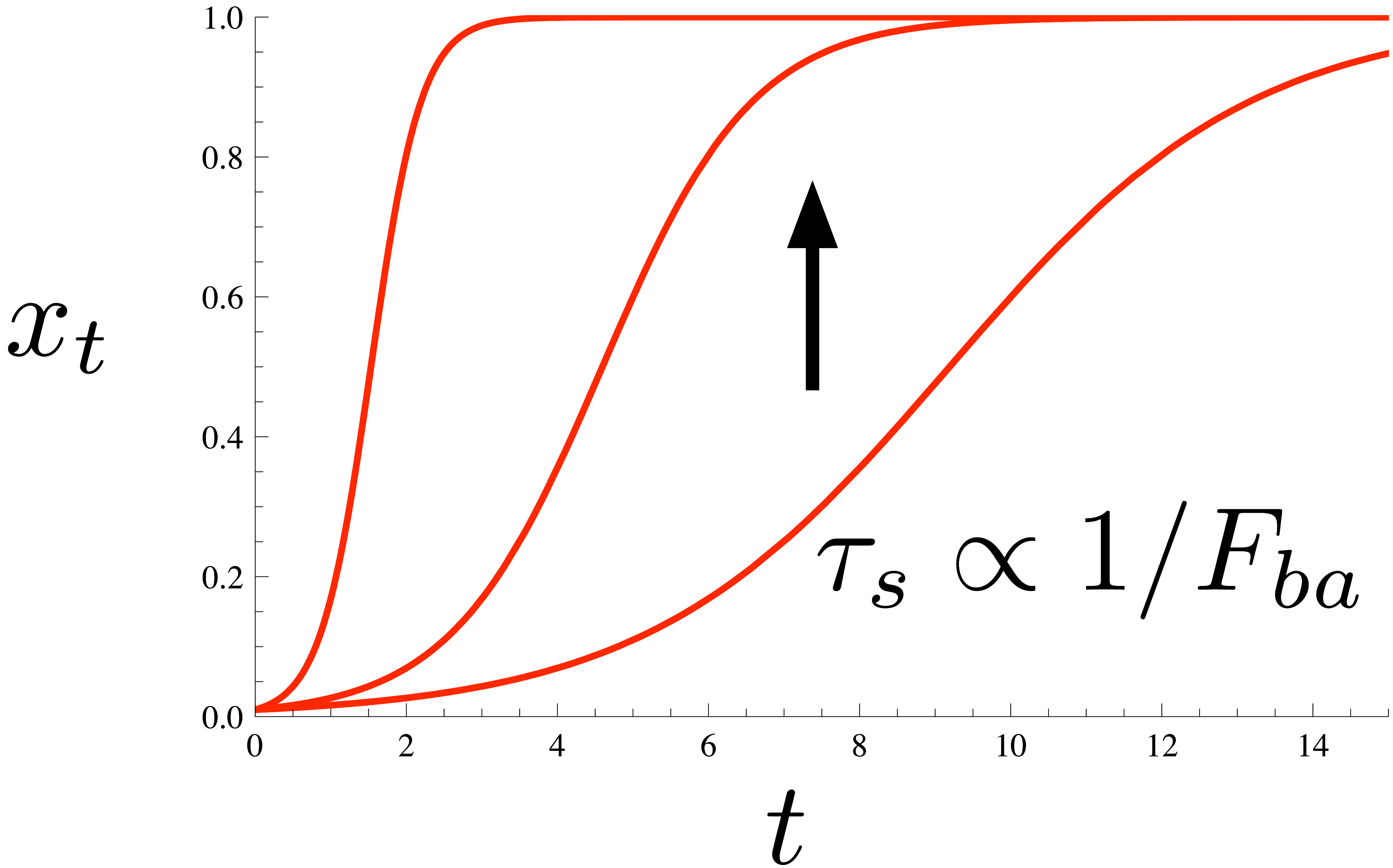
$\tau_m \propto 1/\mu$

t

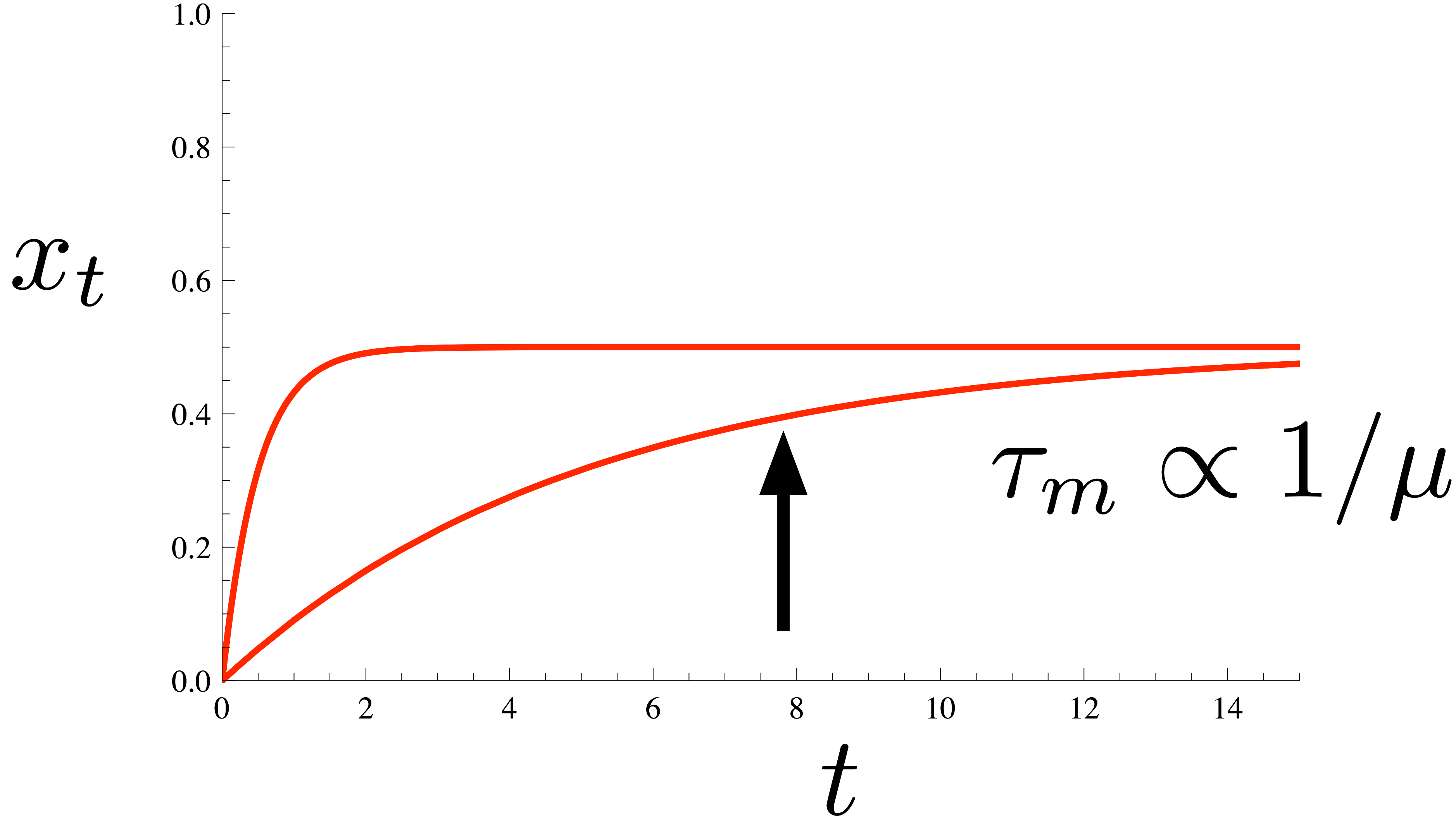
Drift, Selection and Mutation



Drift, Selection and Mutation



Drift, Selection and Mutation



Drift, Selection and Mutation

- Langevin equation:

$$\dot{x}(t) = F_{ba}\, x(t)[1 - x(t)] + \mu[1 - 2x(t)] + \chi_x(t)$$

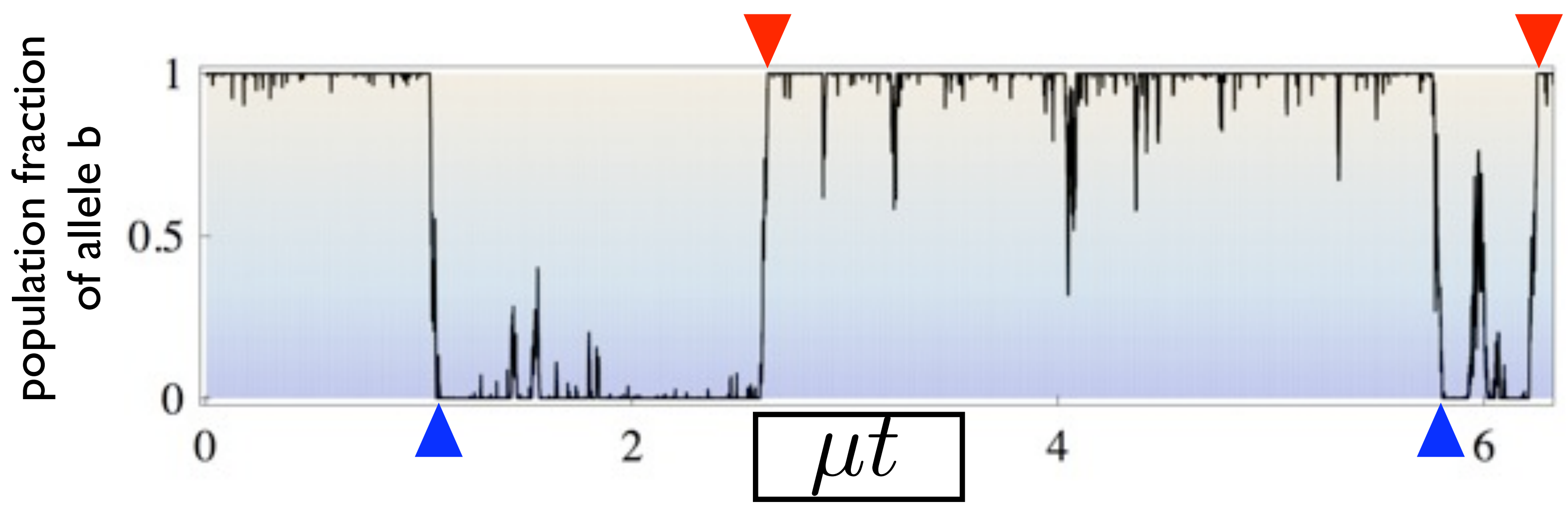
$$\langle \chi_x(t) \rangle = 0$$

$$\langle \chi_x(t)\chi_x(t') \rangle = (x(1-x)/N)\, \delta(t - t')$$

One locus two alleles model

- Wright-Fisher process with mutation and selection
- Ratio of time scales: $$\frac{N}{1/\mu} = \mu N \ll 1$$

lighter shading indicates the fitter state



Diffusion equation

$p(x,t)$ probability density of populations

$g(x,\epsilon,dt)$ probability of change $x \to x+\epsilon, dt$

$$
\begin{aligned}
p(x,t+dt) &= \int p(x-\epsilon,t)g(x-\epsilon,\epsilon,dt)d\epsilon \\
&= \int p(x,t)g(x,\epsilon,dt) - \epsilon\frac{\partial}{\partial x}pg + \frac{\epsilon^2}{2}\frac{\partial^2}{\partial x^2}pg + \ldots d\epsilon \\
&= p(x,t)\int g(x,\epsilon,dt)d\epsilon - \frac{\partial}{\partial x}p\int \epsilon g d\epsilon + \frac{1}{2}\frac{\partial^2}{\partial x^2}p\int \epsilon^2 g d\epsilon
\end{aligned}
$$

[SH Rice, Evolutionary Theory (2004)]

$$= \quad p(x,t)\int g(x,\epsilon,dt)d\epsilon - \frac{\partial}{\partial x}p\int \epsilon g d\epsilon + \frac{1}{2}\frac{\partial^2}{\partial x^2}p\int \epsilon^2 g d\epsilon$$

$$= \quad p(x,t) - \frac{\partial p(x,t)M(x)}{\partial x}dt + \frac{1}{2}\frac{\partial^2 p(x,t)V(x)}{\partial x^2}dt$$

$$\begin{aligned}
\int g(x,\epsilon,dt)d\epsilon &= 1 \\
\int \epsilon g(x,\epsilon,dt)d\epsilon &= \langle\epsilon\rangle \\
&= M(x)dt \\
\int \epsilon^2 g(x,\epsilon,dt)d\epsilon &= \langle\epsilon\rangle^2 + var(\epsilon) \\
&= V(x)dt
\end{aligned}$$

rate of directional change of allele frequency at point x

variance of allele frequency change due to nondirectional effects

$$\frac{p(x,t+dt)-p(x,t)}{dt} = -\frac{\partial p(x,t)M(x)}{\partial x} + \frac{1}{2}\frac{\partial^2 p(x,t)V(x)}{\partial x^2}$$

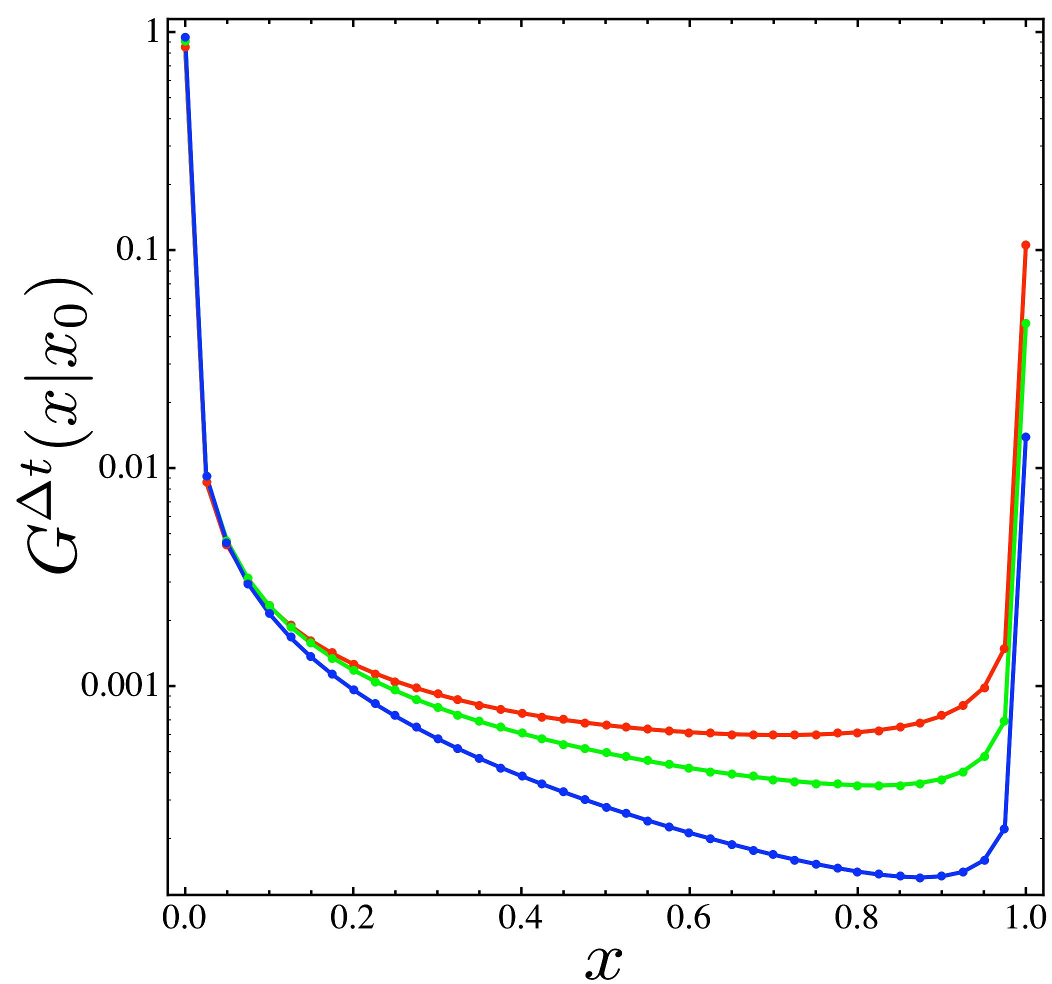
$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial p(x,t)M(x)}{\partial x} + \frac{1}{2}\frac{\partial^2 p(x,t)V(x)}{\partial x^2}$$

Now we just need to plug in our earlier results to get **Kimura's diffusion equation for one locus two alleles case:**

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\frac{x(1-x)}{N}p(x,t) - \frac{\partial}{\partial x}[F_{ba}x(1-x)+\mu(1-2x)]p(x,t)$$

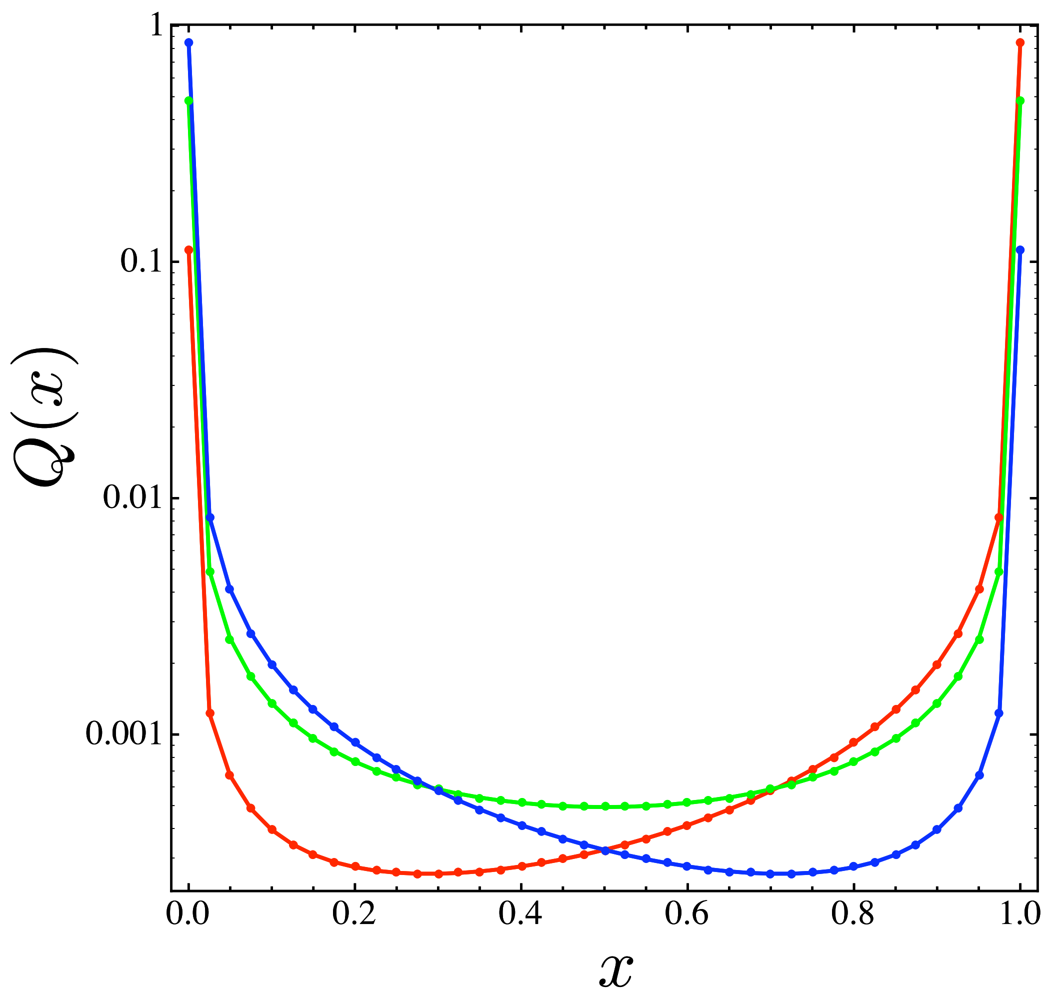
One locus two alleles model: looking at the averages

Equilibrium distributions



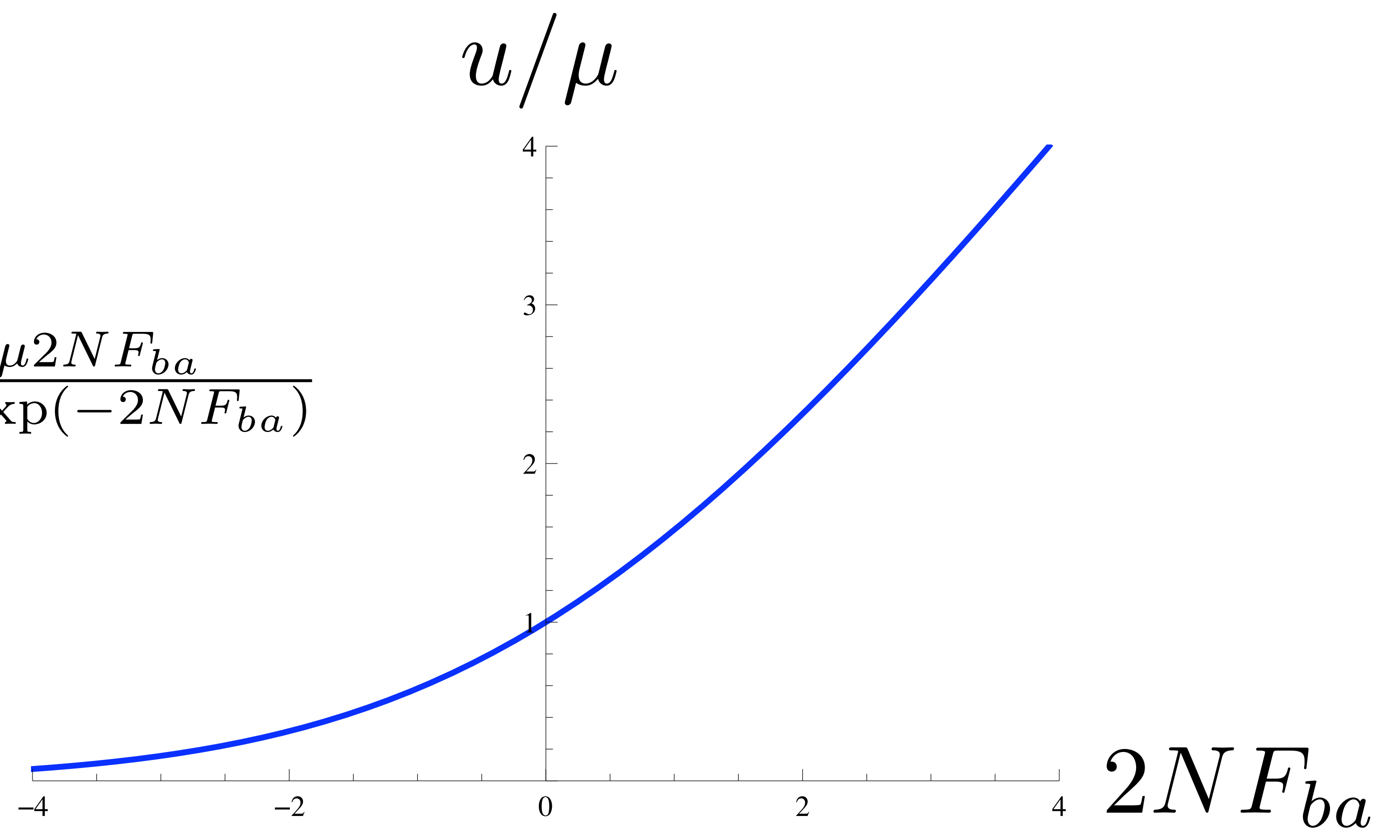
One locus two alleles model: looking at the averages

after infinite time



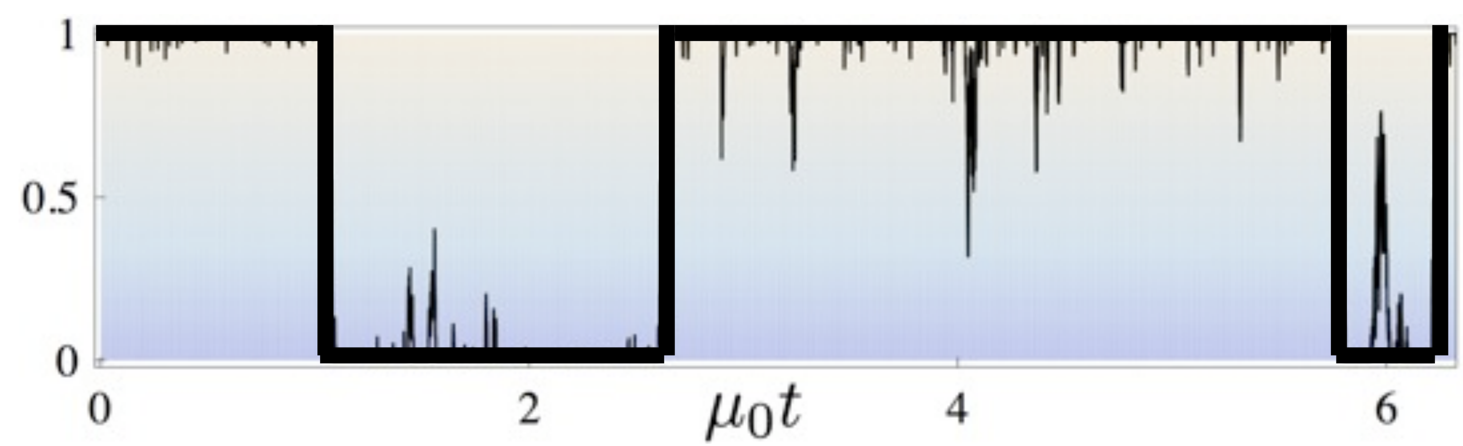
Substitution rate from Kimura's diffusion equation

$$u(2NF_{ba}) = \frac{\mu 2NF_{ba}}{1-\exp(-2NF_{ba})}$$



Substitution dynamics

$$\frac{N}{1/\mu} = \mu N \ll 1$$



$$\dot{p}_a = u_{ba} p_b - u_{ab} p_a$$

$$\dot{p}_b = u_{ab} p_a - u_{ba} p_b$$

Substitution dynamics

$$\langle p_a \rangle = \frac{u_{ba}}{u_{ba} + u_{ab}}$$

$$\langle p_b \rangle = \frac{u_{ab}}{u_{ba} + u_{ab}}$$

$$\boxed{\langle p_a \rangle = \langle p_b \rangle \exp(-2NF_{ba})}$$

$$\boxed{N_{deleterious} = N_{beneficial} \exp(2NF_{ba})}$$

Conclusions

- Fundamental evolutionary forces:
  1. genetic drift, diffusive force (!), effect strongest in small populations, changes allele frequencies at time-scale $2N$
  2. selection, directed force, increases the fitter phenotypes, time-scale $1/F$
  3. mutation, directed force, "restarts" the process from monomorphic states, time-scale $1/\mu$

- More complex scenarios, recombination, linkage, diploid genomes, demographics, frequency/time-dependent selection ...
- Further reading, Gillespie: *Population Genetics,* Hartl and Clark: *Principles of Population Genetics*