SMEEG 2013, Stochastic Models in Ecology, Evolution and Genetics
Angers, December 9th-13th 2013

Balancing Selection in Subdivided Populations

Luca Dall'Asta - Politecnico di Torino (Italy)

POLITECNICO
DI TORINO

Evolutionary dynamics of a population of Ω identical individuals:

- selective pressure (fitness)
- random reproduction (death/reproduction, binomial sampling, ...)
- mutations

Microscopic dynamics: Wright-Fisher and Moran models.

Simplifying assumptions:

- haploid individuals
- two competing alleles {A,B}
- no mutation

Balancing selection is an umbrella concept:

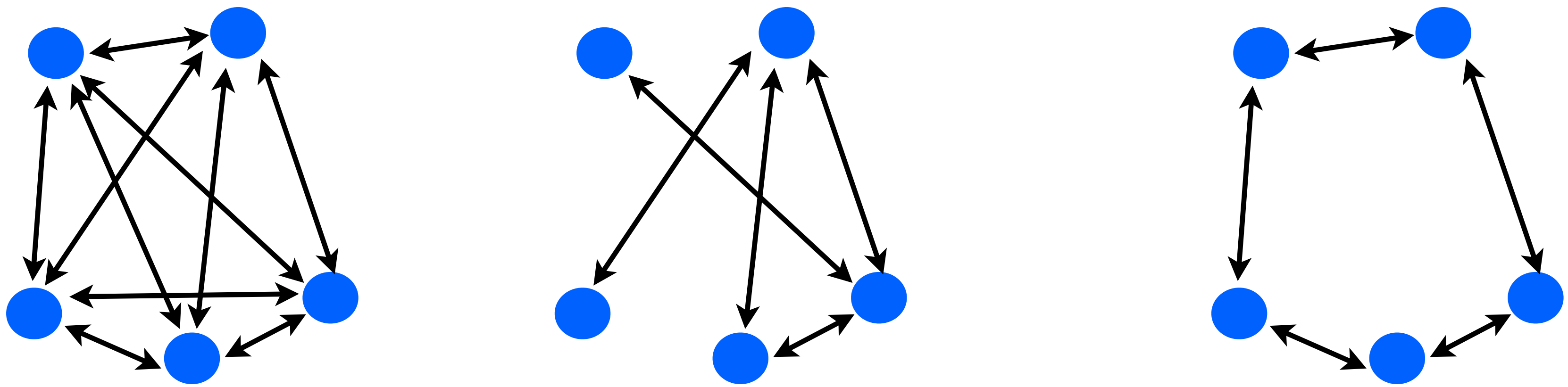
a class of selective processes by means of which multiple alleles are maintained at significant frequencies in a population.

Examples: self-incompatibility systems in plants, Major Histocompatibility Complex in mammalian, certain genetic diseases in humans (sickle-cell anemia, thalassemia, cystic fibrosis)

In the two alleles model (Robertson, 1962):

- favors an internal mixed equilibrium
- induces retardation in fixation in well-mixed populations.

Subdivision: N subpopulations (demes/colonies) of Ω individuals each, interacting by means of ***migration.***



no structure ("mean-field") | complex structures | spatial structures (diffusion effects)

Migration will be homogeneous ($m_{ij} = m$, $\forall(i,j)$) and conservative.

Study: 1) fixation properties ($\Omega < \infty$, $N < \infty$),

2) phase transitions ($\Omega < \infty$, $N = \infty$)

1. Mean-Fixation Time in unstructured subdivided populations with balancing selection.

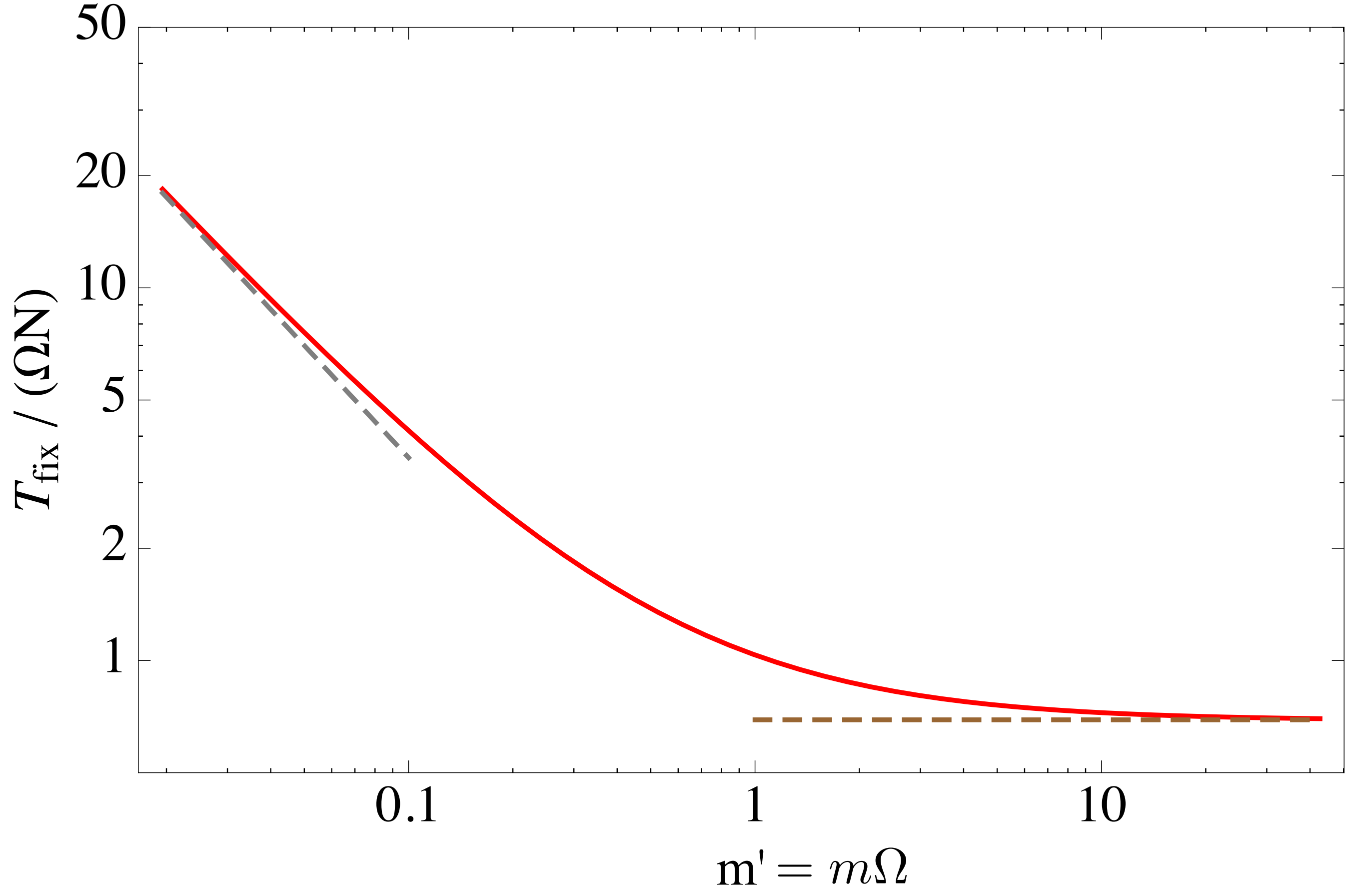
In collaboration with: P. Lombardo, SISSA-Trieste (Italy)
A. Gambassi, SISSA-Trieste (Italy)

See: P. Lombardo, A. Gambassi, and L. Dall'Asta, **arXiv:1310.5072** (2013)

Known results on the mean-fixation time (MFT):

1) High-migration limit: $T_{\text{fix}}(m \to \infty) \propto N\Omega$

2) Low-migration limit (Slatkin, 1981): $T_{\text{fix}}(m \to 0) \approx \dfrac{N}{\Omega m u_1(1/\Omega)}$



Diffusion approximation for the Island Model with balancing selection

For $\Omega \gg 1$, the frequency $x_i = \Omega_A^{(i)}/\Omega$ of allele A in deme *i* satisfies the Langevin equation:

$$\dot{x}_i = s(x_* - x_i)x_i(1 - x_i) + m\left(\frac{1}{N}\sum_{i=1}^{N} x_i - x_i\right) + \sqrt{\frac{x_i(1-x_i)}{\Omega}}\ \eta_i$$

Itô gaussian noise

$$\langle \eta_i(t)\eta_j(t')\rangle = \delta_{i,j}\delta(t - t')$$

Diffusion approximation for the Island Model with balancing selection

For $\Omega \gg 1$, the frequency $x_i = \Omega_A^{(i)}/\Omega$ of allele A in deme i satisfies the <u>Langevin equation</u>:

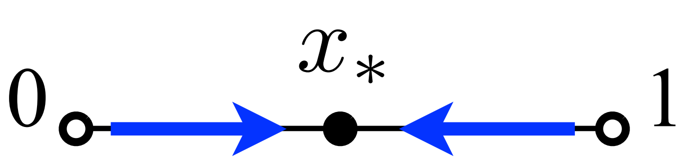
$$\dot{x}_i = s(x_* - x_i)x_i(1-x_i) + m\left(\frac{1}{N}\sum_{i=1}^{N} x_i - x_i\right) + \sqrt{\frac{x_i(1-x_i)}{\Omega}}\,\eta_i$$

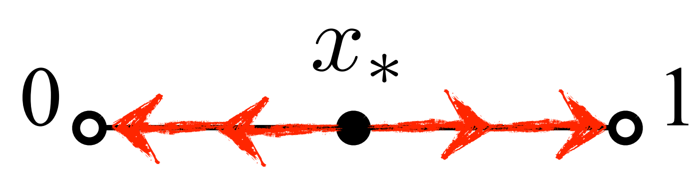
mean selective force $\mu(x_i)$

migration

genetic drift $v(x_i)$

?

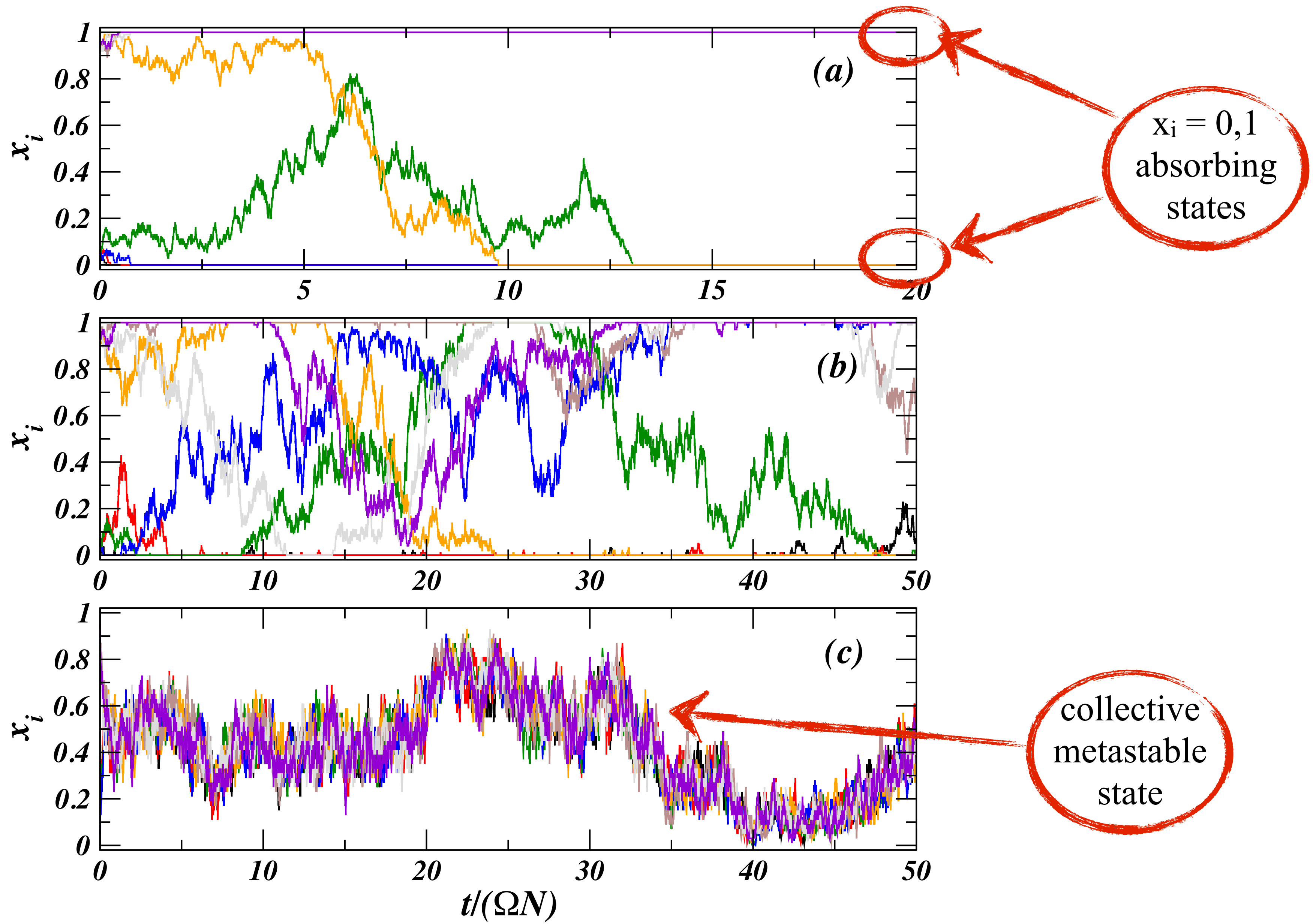




Collective dynamics: simulations

m = 0

m > 0



Approximate descrition of the Collective dynamics

Consider the average quantities $$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\begin{aligned}\dot{\overline{x}} &= s[x_*\overline{x} - (1+x_*)\overline{x^2} + \overline{x^3}] + \sqrt{(\overline{x} - \overline{x^2})/(\Omega N)}\,\eta, \\ &= M[\overline{x}, \overline{x^2}, \overline{x^3}] + \sqrt{V[\overline{x}, \overline{x^2}]}\eta\end{aligned}$$

$$\langle \eta(t)\eta(t')\rangle = \delta(t-t')$$

We need a **moment-closure** scheme!

$$\Rightarrow \quad \boxed{\dot{\overline{x}} = M(\overline{x}) + \sqrt{V(\overline{x})}\,\eta}$$

Cherry and Wakeley (2003) and Cherry (2003):

For $s = 0$ and large N, $\overline{x}$ is a much slower variable than local frequencies x_i :

1) the $\{x_i\}$ can be considered as almost independent random variables, each one described by the same conditional "quasi-stationary" distribution

$$P_{\mathrm{qs}}(x_i|\overline{x}) \propto x^{2m\Omega\bar{x}-1}(1-x)^{2m\Omega(1-\bar{x})-1}$$

that satisfies $\overline{x} = \int_0^1 \mathrm{d}x\, x\, P_{\mathrm{qs}}(x|\overline{x})$

2) we can approximate $\overline{x^k}$ with $\langle x^k \rangle_{\mathrm{qs}}$ as function of $\overline{x}$

$$M(\bar{x}) = s_{\mathrm{e}}\,\bar{x}(1-\bar{x})(x_* - \bar{x}), \qquad V(\bar{x}) = \bar{x}(1-\bar{x})/N_{\mathrm{e}}$$

The subdivided population behaves like a well-mixed one with

- *effective population size* $N_{\mathrm{e}} = N\Omega\left(1 + \frac{1}{2m\Omega}\right)$
- *effective selection coefficient* $s_{\mathrm{e}} = s/\left[\left(1 + \frac{1}{m\Omega}\right)\left(1 + \frac{1}{2m\Omega}\right)\right]$

Issues: When does the approximation hold? Can be improved?

Timescales: - migration $T_{\mathrm{migr}} \simeq 1/m$

- relaxation $T_{\mathrm{rel}} \simeq 1/M[\bar{x}] \simeq 1/s_{\mathrm{e}}$

- fluctuations $T_{\mathrm{fluct}} \simeq 1/V[\bar{x}] \simeq N_{\mathrm{e}}$

The approximation is correct when $T_{\mathrm{migr}} \ll \min\{T_{\mathrm{fluct}}, T_{\mathrm{rel}}\}.$

- when $N\Omega s > 1 + 1/m\Omega$, it requires $s_{\mathrm{e}}/m \ll 1$
- otherwise it only requires $N \gg 1$

Generalization to small but non-vanishing values of s_{e}/m

1) Parametric ansatz:

$$P_{\mathrm{qs}}(x|y) \propto x^{2m\Omega y-1}(1-x)^{2m\Omega(1-y)-1}\mathrm{e}^{s\Omega x(2x_*-x)}$$

2) Find y as function of $\bar{x}$ using the consistency condition

$$\bar{x} = \int_0^1 \mathrm{d}x\, x\, P_{\mathrm{qs}}(x|y)$$

$$\Rightarrow \quad y = \bar{x} - (s_{\mathrm{e}}/m)\bar{x}(1-\bar{x})(x_* - \bar{x}) + O((s_{\mathrm{e}}/m)^2)$$

3) Calculate corrections to $M[\bar{x}], V[\bar{x}]$

Mean Fixation Time

Using the backward Fokker-Planck approach (Kimura and Ohta, 1969)

$$V(\bar{x})T''_{\mathrm{fix}}(\bar{x})/2 + M(\bar{x})T'_{\mathrm{fix}}(\bar{x}) = -1$$

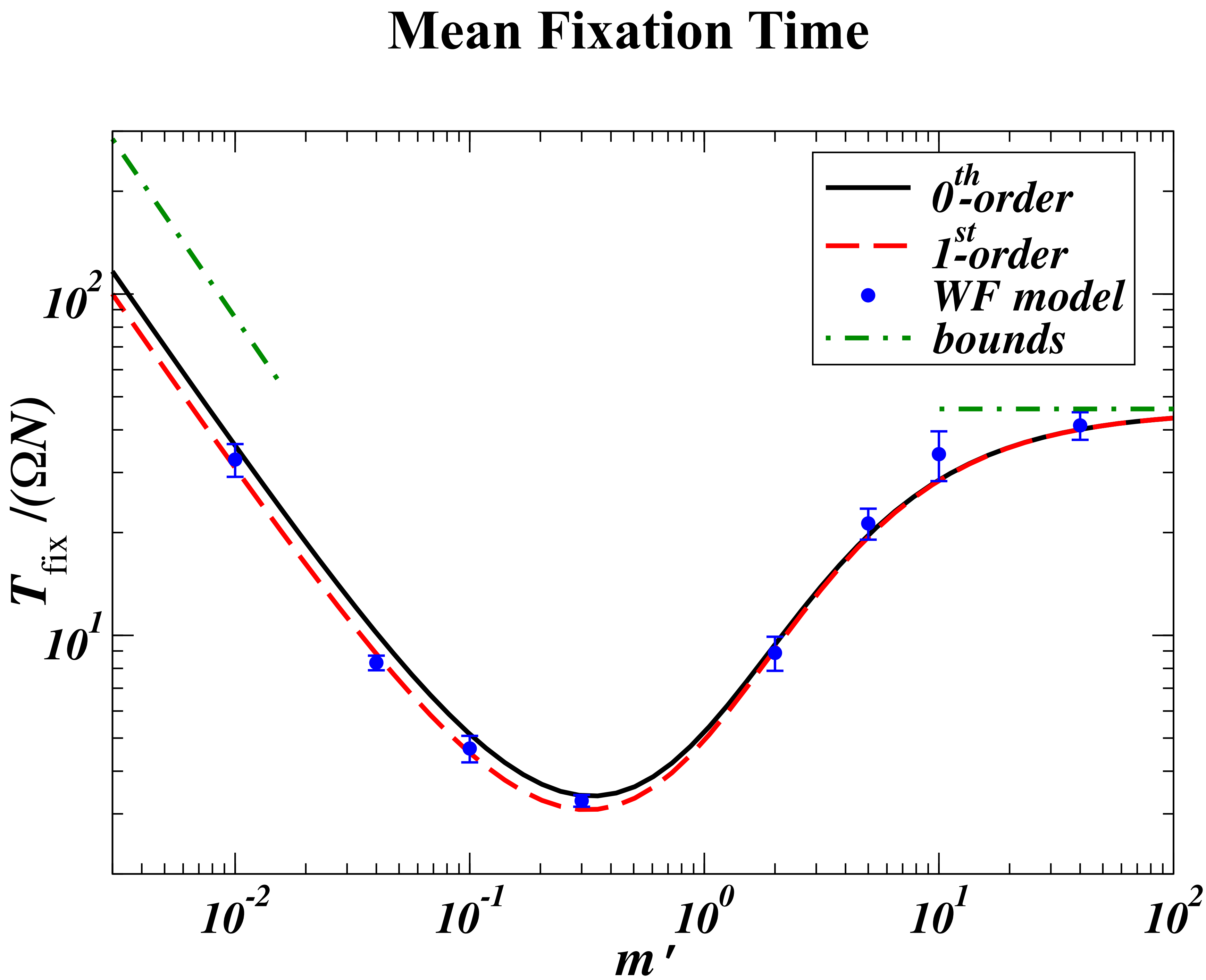
with initial condition $\bar{x} = \hat{x}$ (the metastable state)

At order zero in s_{e}/m,

$$T_{\mathrm{fix}}^{(0)} = \frac{N_{\mathrm{e}}}{2} \int_0^1 \mathrm{d}y \int_0^1 \mathrm{d}z \frac{\mathrm{e}^{s_{\mathrm{e}} N_{\mathrm{e}} y (1-z^2)/4}}{1 - yz^2}$$

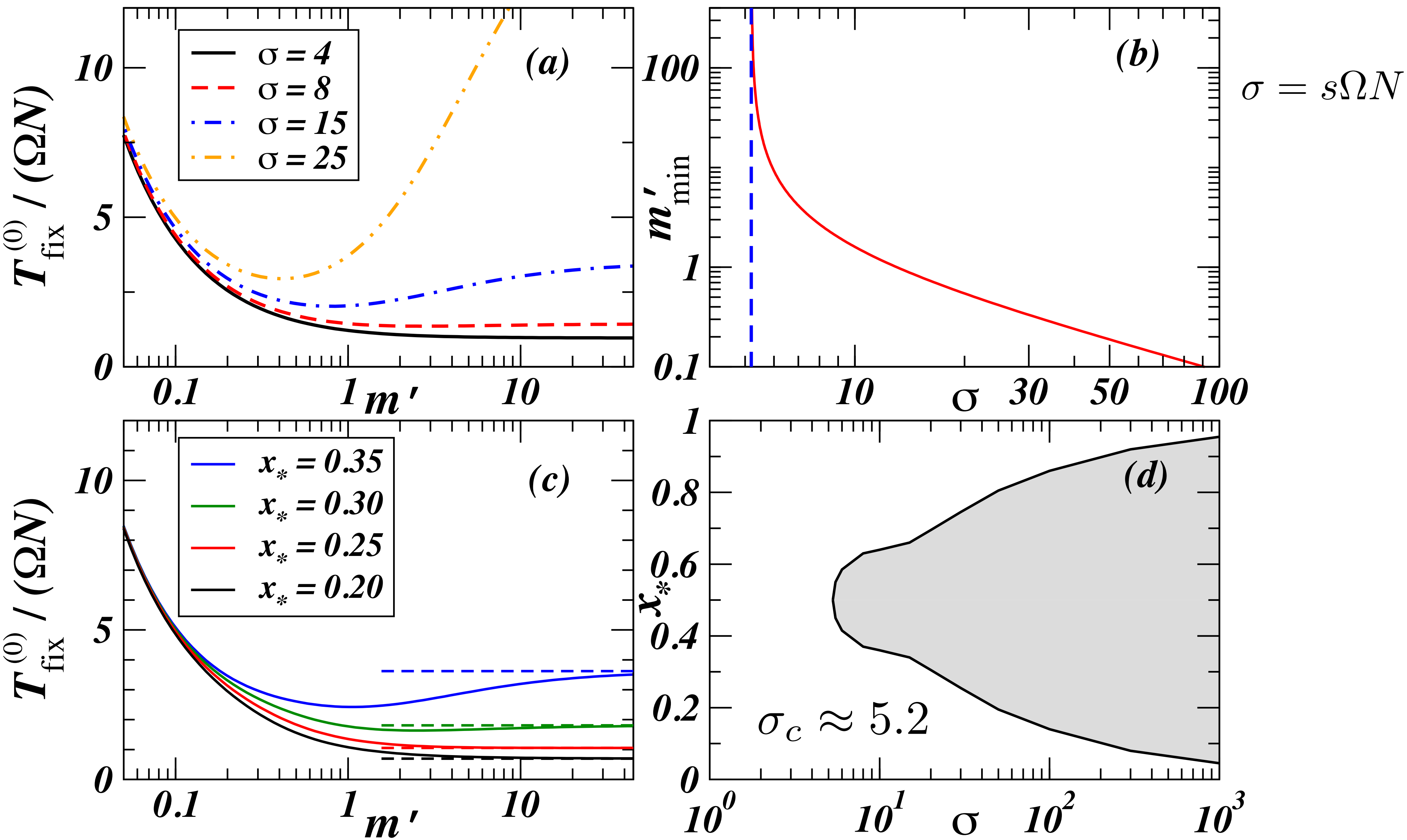
with asymptotic behavior

$$\frac{T_{\mathrm{fix}}^{(0)}}{\Omega N} \approx \begin{cases} \text{const.} & \text{for } m\Omega \gg 1 \\ \log 2/2m\Omega & \text{for } m\Omega \ll 1 \end{cases}$$



Mean Fixation Time
$T_{fix}/(\Omega N)$
m'
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
0^{th}-order
1^{st}-order
WF model
bounds

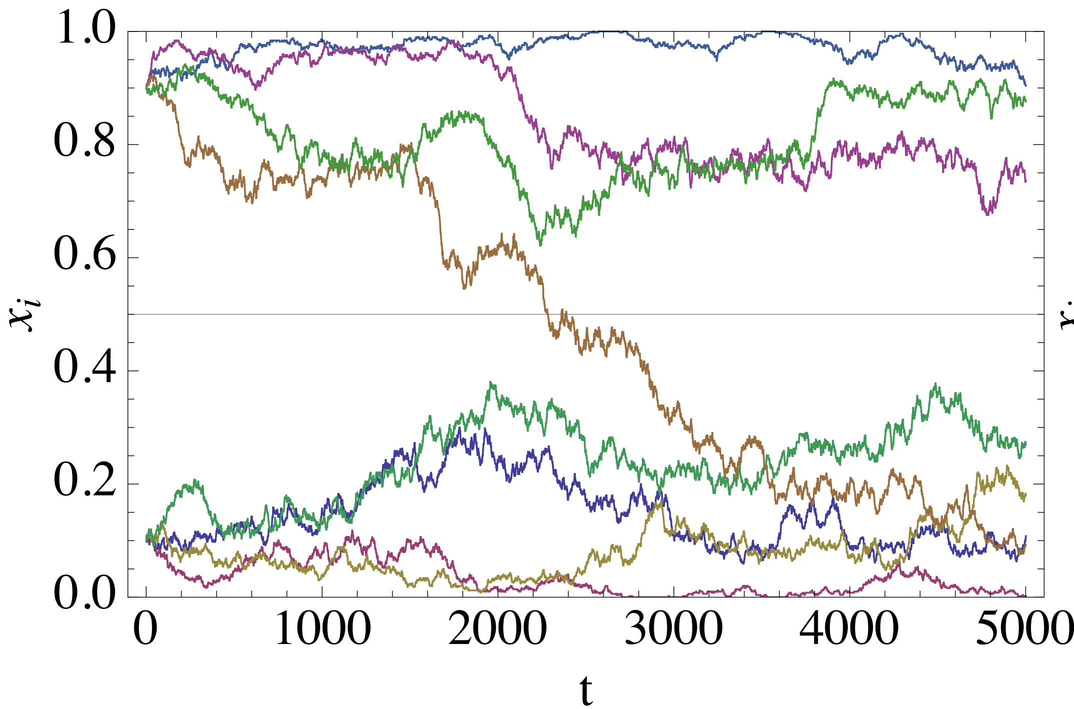
Mean Fixation Time



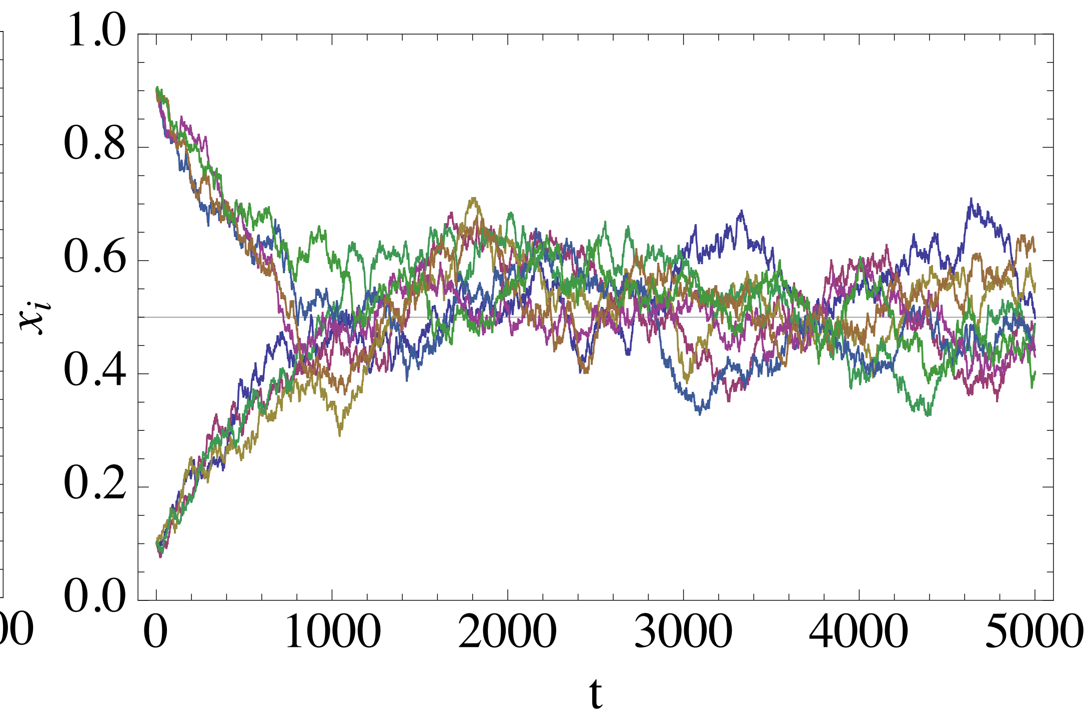
Biodiversity

1) global heterozygosity $H = 2\bar{x}(1-\bar{x})$ $H \in [0, 1/2]$

2) local heterozygosity $h = (2/N)\sum_i x_i(1-x_i)$ $h \in [0, H]$



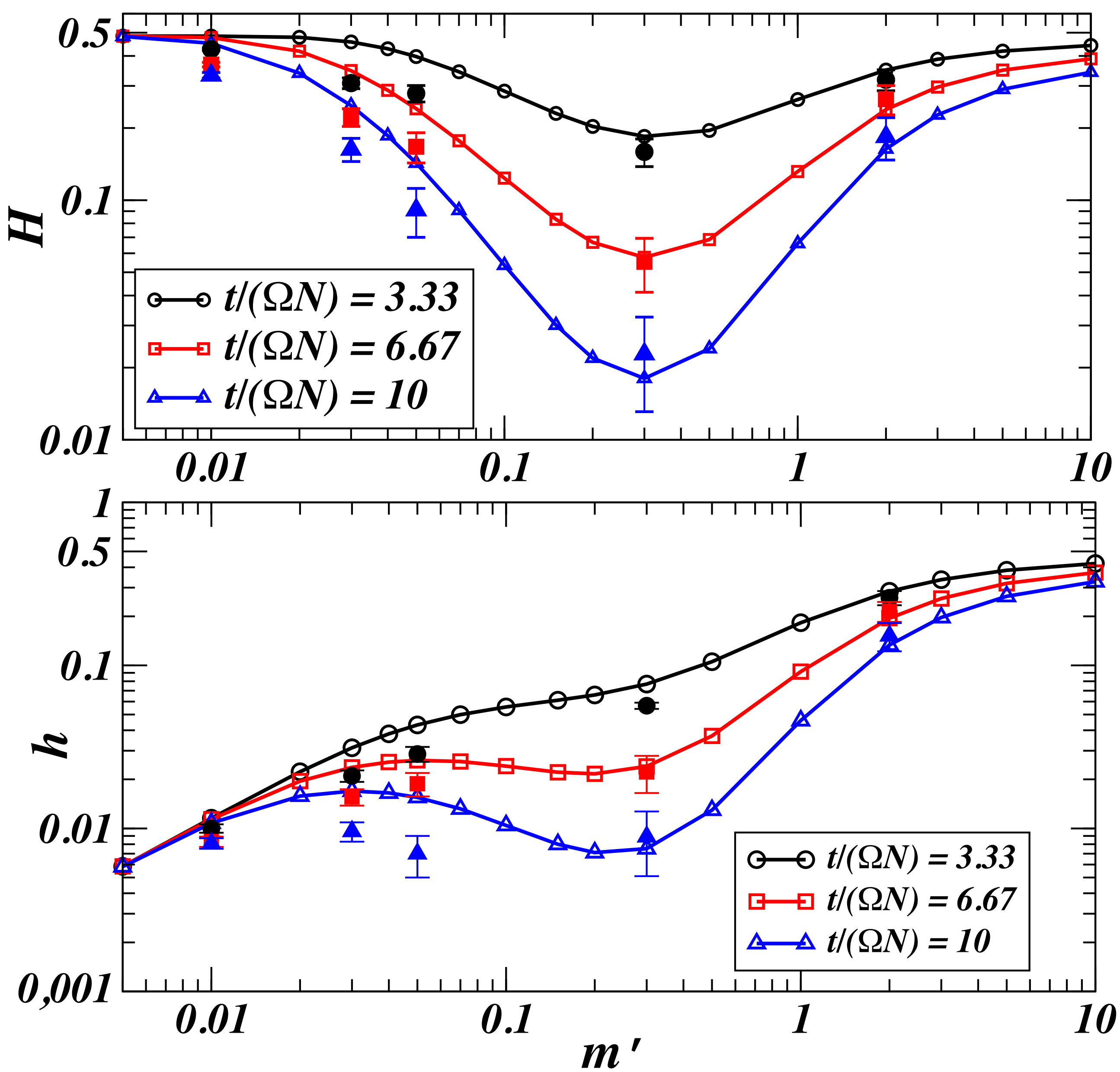
$H \simeq \frac{1}{2},\ h \simeq 0$



$H \simeq h \simeq \frac{1}{2}$

Biodiversity

$$H(t) \simeq H_{\mathrm{met}} P_{\mathrm{surv}}(t)$$



$\Omega = 100, N = 30, s' = 1, x_* = 0.5$

Approx. does not work well anymore for $m' \lesssim 1/\sigma$

Consistent with results for models with many alleles and mutation

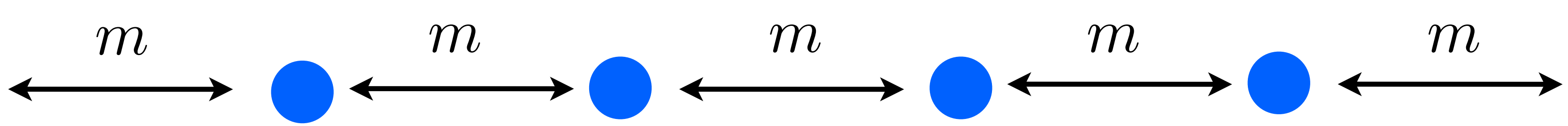
(Schierup 1998, Schierup et al. 2000, Muirhead 2001, Nishino and Tajima 2005)

2. Fixation-Coexistence phase transition in 1d populations with balancing selection.

In collaboration with: F. Caccioli, Univ. of Cambridge (UK)
D. Beghè, University of Parma (Italy)

See: L. Dall'Asta, F. Caccioli and D. Beghè, *Europhys. Lett.* **101**, 18003 (2013)

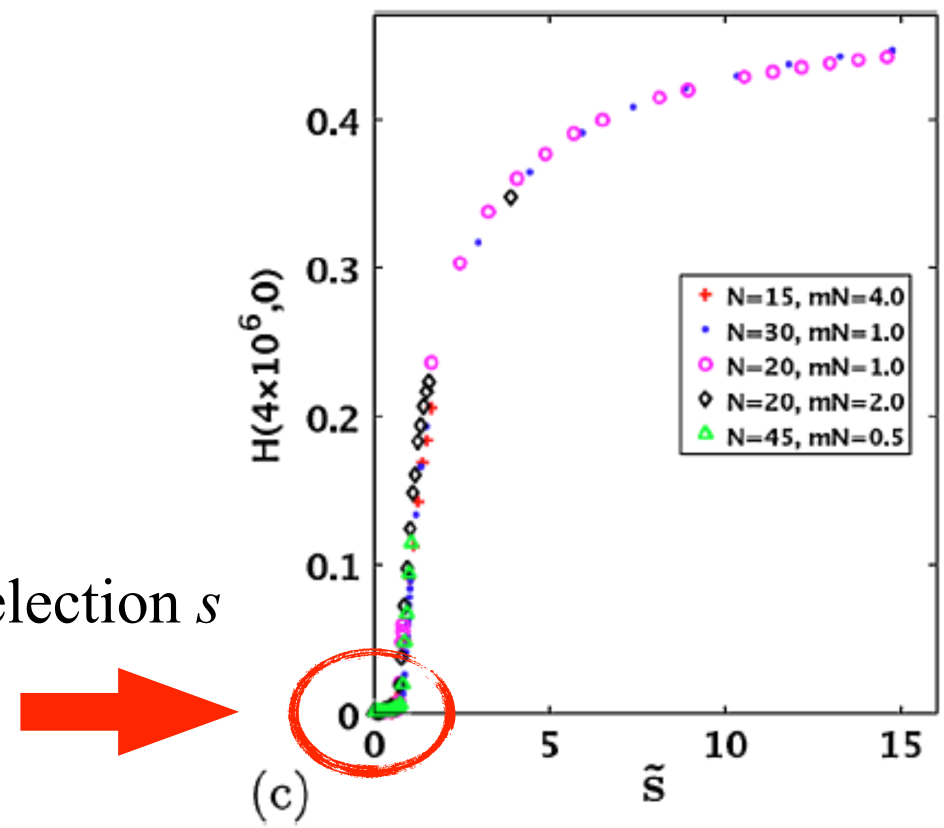
Consider an infinite one-dimensional stepping-stone model:



$$x_i = s x_i(1 - x_i)(x_* - x_i) + m(x_{i+1} + x_{i-1} - 2x_i) + \sqrt{\frac{x_i(1-x_i)}{\Omega}}\,\eta_i$$

but we cannot use previous approximations...

Numerical simulations suggest a **phase-transition** at finite values of balancing selection *s* (Korolev and Nelson, 2011)



A spin model of balancing selection

We consider a one-dimensional voter-like model $S_i \in \{-1,1\}$ with heterozygosity advantage (Sturm and Swart, *Ann. App. Prob.* **18**, 2008)

The model:

- with rate λ_0, each individual is replaced by a new one equal to a randomly selected neighbor (*voter-like process*);

- with rate $\lambda_b/2$, each individual *i* is replaced by the less frequent type in the triplet $\{S_i, S_{i+1}, S_{i+2}\}$(and the same for the symmetric case)

Similar to the celebrated Neuhauser-Pacala model of spatial plant ecology.
(C. Neuhauser and S. Pacala, Ann. Appl. Prob. **9**,1999)

From the master equation for the distribution P(S_1,..,S_N;t), we get a hierarchy of coupled equations for multispin correlation functions.

$$\frac{d\langle S_i\rangle}{dt} = \frac{\lambda_0}{2}\left[\langle S_{i-1}\rangle + \langle S_{i+1}\rangle - 2\langle S_i\rangle\right] + \frac{\lambda_b}{2}\left[\langle S_i S_{i-1} S_{i-2}\rangle + \langle S_i S_{i+1} S_{i+2}\rangle - 2\langle S_i\rangle\right]$$

$$\frac{d\langle S_i S_{i+1}\rangle}{dt} = \frac{\lambda_0}{2}\left[\langle S_i S_{i+2}\rangle + 2 + \langle S_{i-1} S_{i+1}\rangle - 4\langle S_i S_{i+1}\rangle\right] + \frac{\lambda_b}{2}\left[\langle S_{i-2} S_{i-1} S_i S_{i+1}\rangle + \langle S_i S_{i+2}\rangle + \langle S_{i-1} S_{i+1}\rangle + \langle S_i S_{i+1} S_{i+2} S_{i+3}\rangle - 4\langle S_i S_{i+1}\rangle\right]$$

we close the equations using (Kirkwood) factorization approximation:

$$\langle S_i S_{i+1} S_{i+2}\rangle \approx \langle S_i\rangle\langle S_{i+1} S_{i+2}\rangle$$
$$\langle S_i S_j S_{i\pm1} S_{i\pm2}\rangle \approx \langle S_i S_j\rangle\langle S_{i\pm1} S_{i\pm2}\rangle$$

We get equations for two point correlations $c_k(t) = \langle S_i S_{i+1+k} \rangle$

$$\dot{c}_0(t) = \lambda_0(c_1 + 1 - 2c_0) + \lambda_b(c_0^2 + c_1 - 2c_0)$$
$$\dot{c}_1(t) = \lambda_0(c_2 + c_0 - 2c_1) + \lambda_b(c_1 c_0 + c_0 - 2c_1)$$
$$\dot{c}_k(t) = \lambda_0(c_{k+1} + c_{k-1} - 2c_k) + 2\lambda_b(c_0 - 1)c_k,$$

The approximation is correct only if we define a separation of scales between the local dynamics and the large-scale behavior.

At large times:

1) we assume $c_1 \simeq c_0$,

2) we solve the equation for c_0

3) use the solution to solve the equation for c_k

We find: $c_0^{\infty} = 1, \quad \text{for} \quad \lambda_b < \lambda_0$

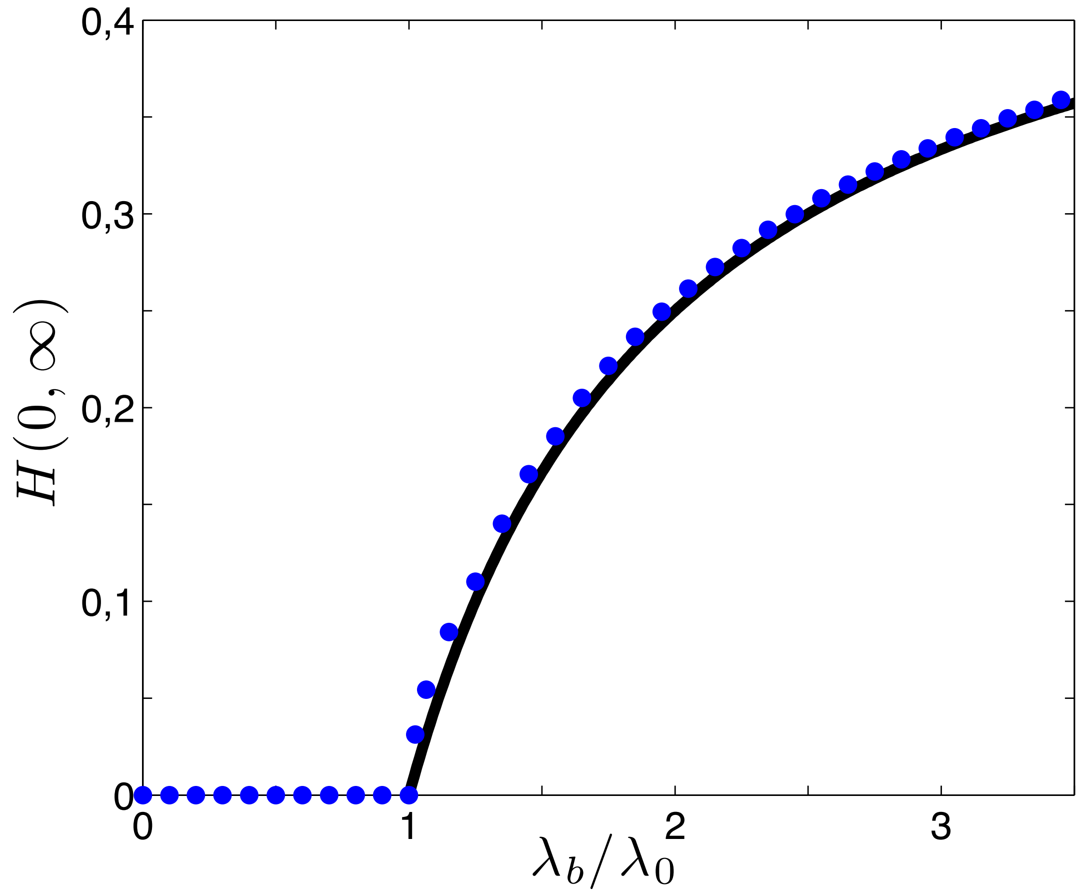
$$c_0^{\infty} = \frac{\lambda_0}{\lambda_b}, \quad \text{for} \quad \lambda_b \geq \lambda_0$$

Local Heterozygosity

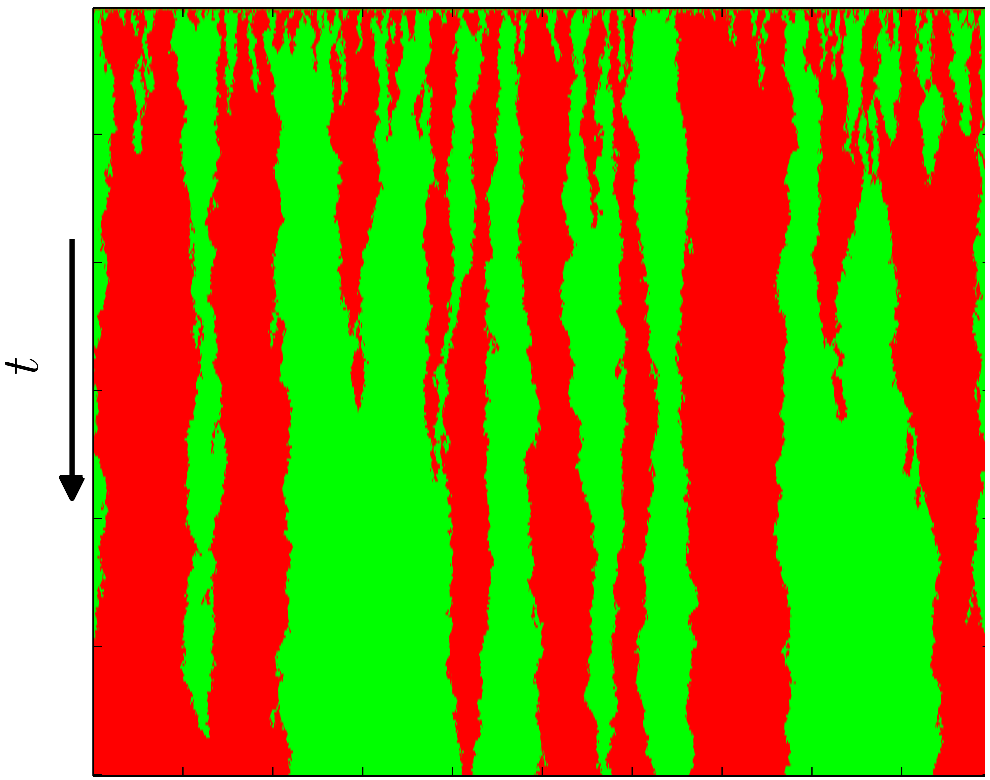
$$H(0,\infty) = (1 - c_0^{\infty})/2$$

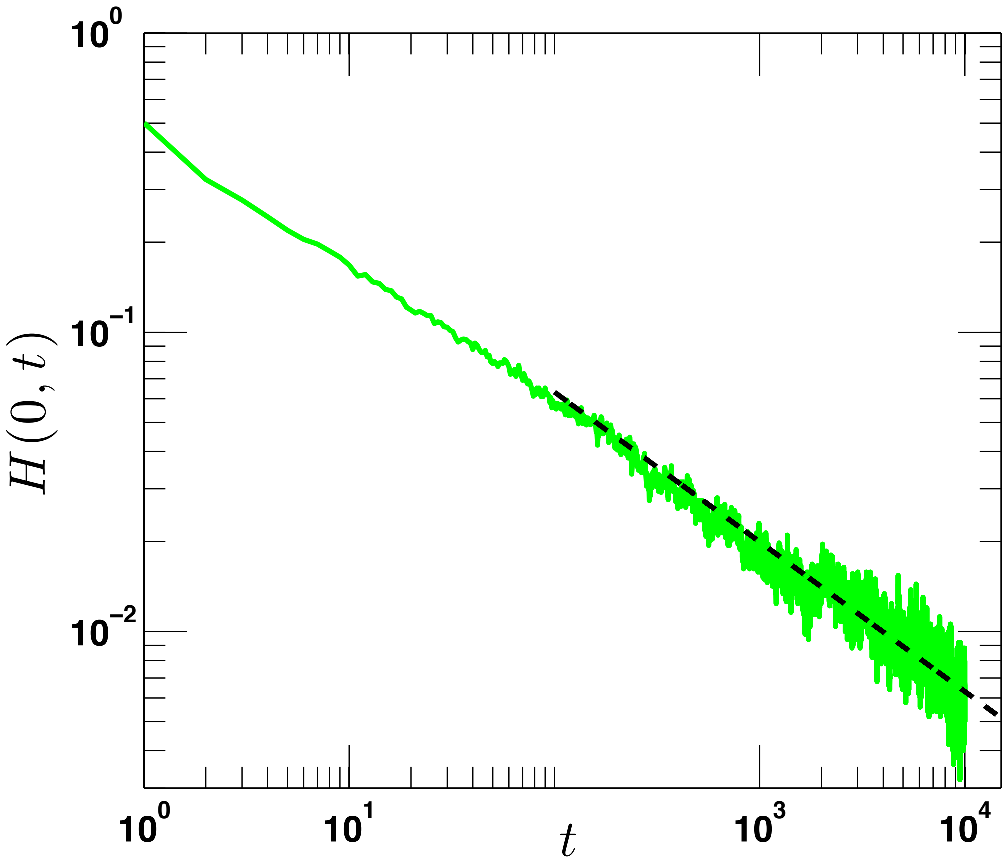
Fixation-Coexistence Phase Transition

(infinite size limit)



Dynamics in the Fixation Phase: $H(0,t) \propto (4\lambda_0 t)^{-1/2} \quad \text{for } t \to \infty$





Genetic demixing through algebraic coarsening (as for neutral populations)

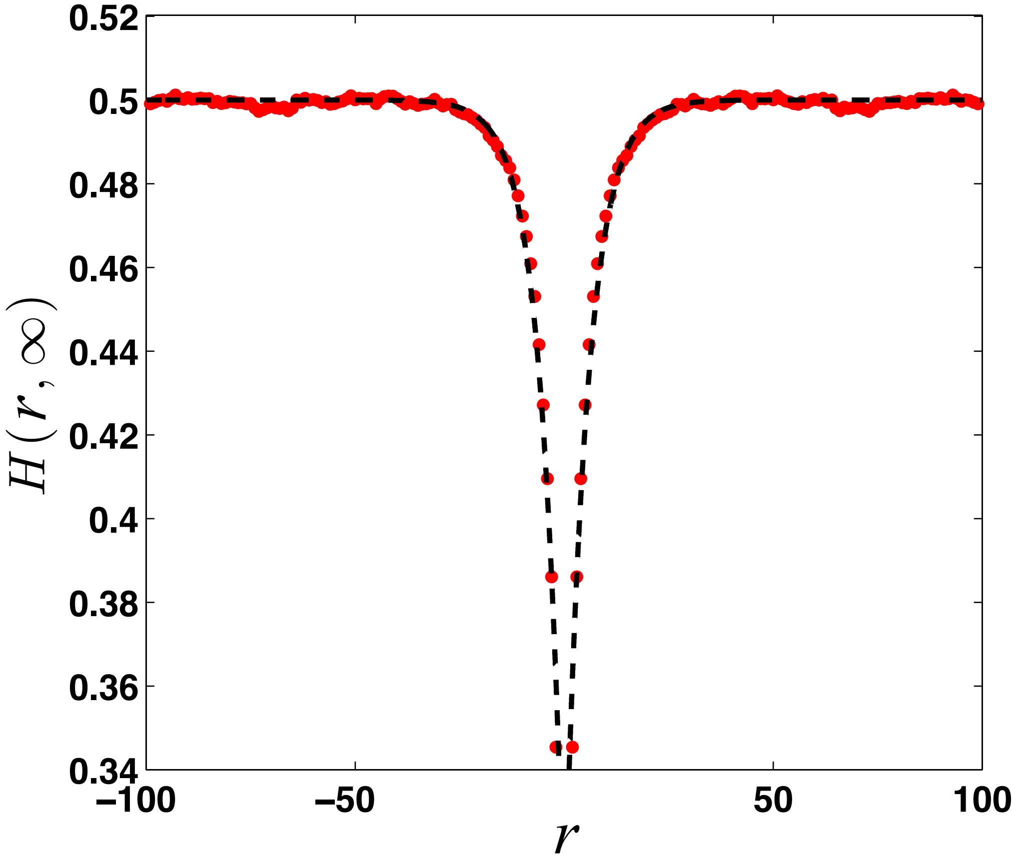
Dynamics in the Coexistence Phase:

Exponential relaxation to the stationary spatial profile of heterozygosity

$$c(r,t) \approx c_{st}(r) + \frac{\lambda_0}{2\lambda_b} e^{-2(\lambda_b - \lambda_0)t} \left[-\frac{\Phi(r)}{\sqrt{t}} + O(t^{-3/2}) \right],$$

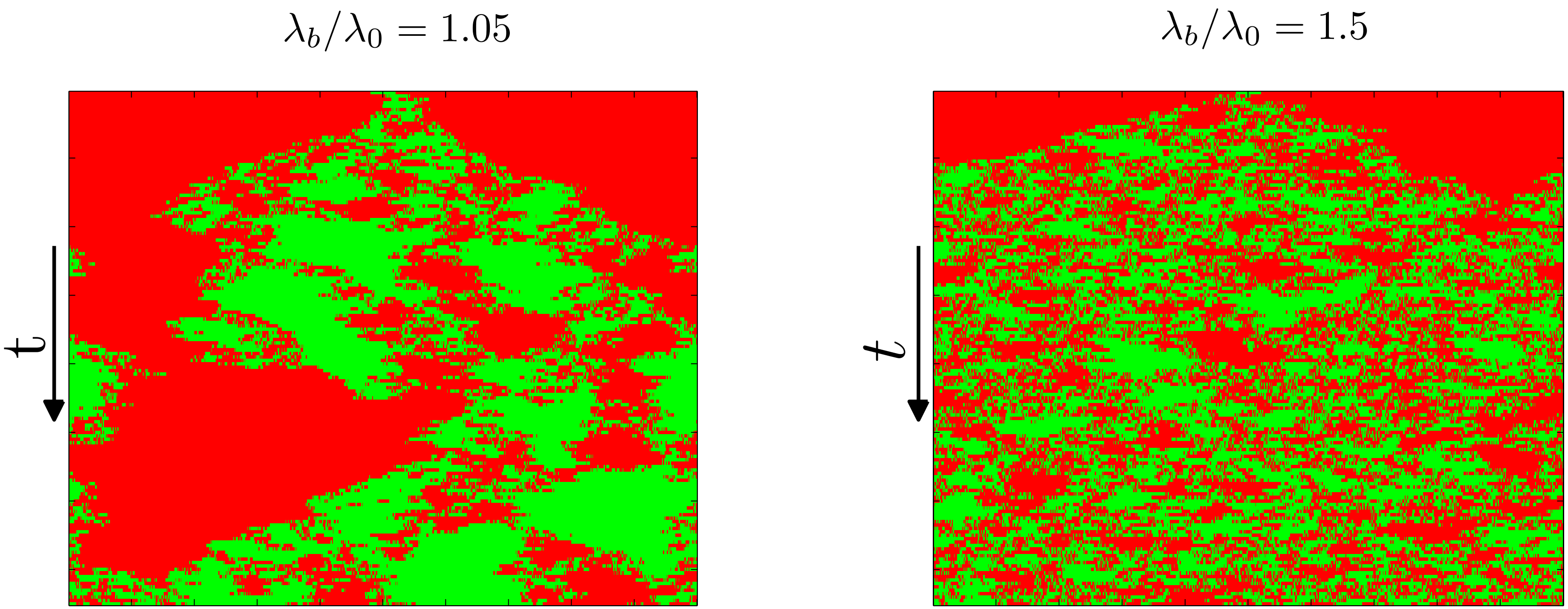
$$c_{st}(r) = (\lambda_0/\lambda_b) e^{-|r|/\xi}$$

$$\text{with } \xi^2 \propto 1/(\lambda_b - \lambda_0)$$



Dynamics in the Coexistence Phase:

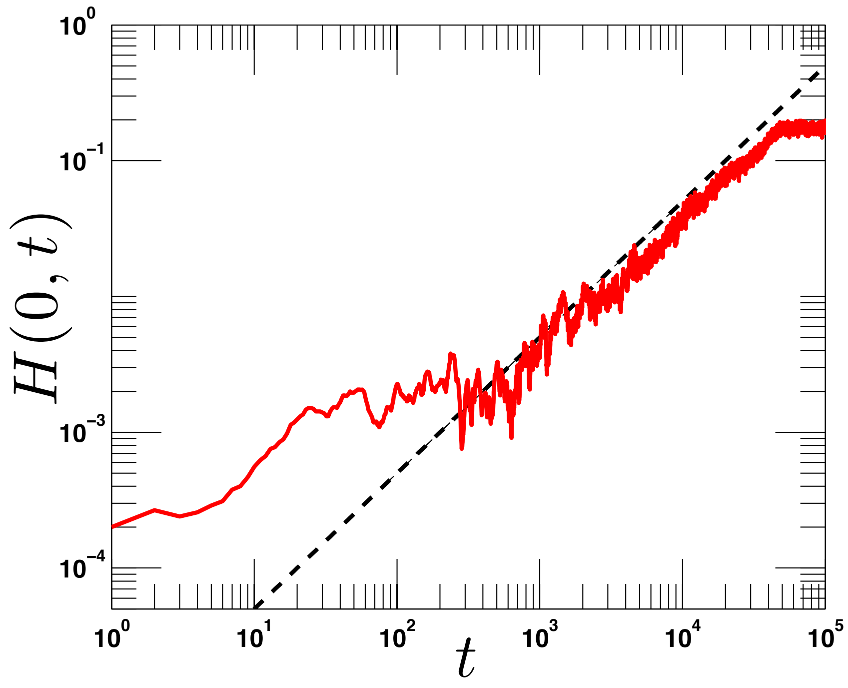
Balancing selection could favor the propagation of polymorphism in a completely homogenous environment (Fisher-like waves)



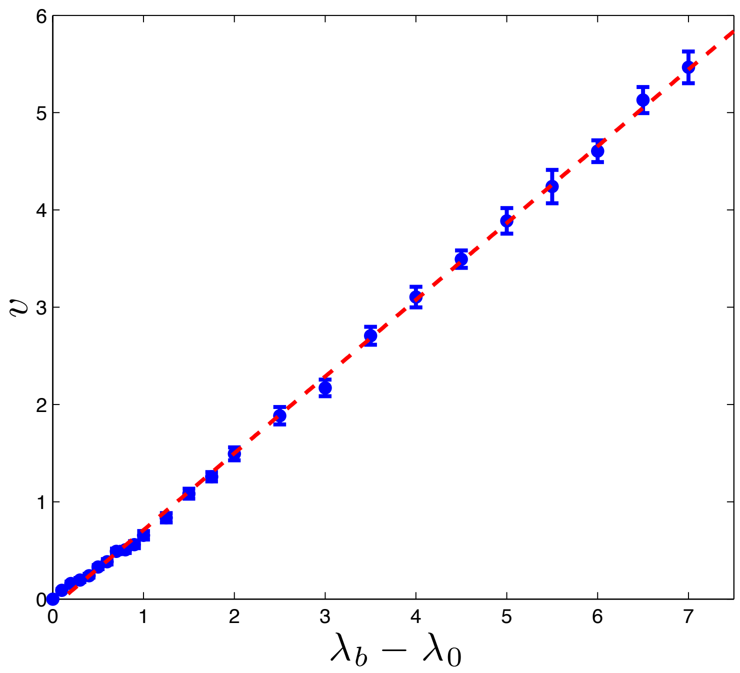
Dynamics in the Coexistence Phase:

Balancing selection could favor the propagation of polymorphism in a completely homogenous environment (Fisher-like waves)

ballistic propagation



velocity is linear with $\lambda_b - \lambda_0$



Conclusions

We studied the effects of balancing selection in subdivided populations

1) MFT in unstructured populations is a **non-monotonic** function of the migration rate

2) **coexistence-fixation phase transition** in 1d populations

Both phenomena should be very general (other dynamics with an internal attractive equilibrium with two symmetric absorbing states)

Future works:

- Populations on networked structures.
- Range expansion.

Microscopic Model Simulations

fitness $\quad w_A = 1 + \tilde{s}, \quad w_B = 1$

$$p_{\mathrm{r}}(x) = \frac{w_A \Omega_A}{w_A \Omega_A + w_B \Omega_B} = \frac{(1+\tilde{s})x}{1+\tilde{s}x}$$

$$p_{\mathrm{m}}(x_i, \bar{x}) = m\bar{x} + (1-m)x_i$$

$$(x_i, \bar{x}) \xrightarrow{\text{migr}} x_i' \xrightarrow{\text{repr}} x_i'',$$

$$\mu(x_i) \equiv \langle \Delta x_i \rangle = \tilde{s}x_i(1-x_i) + m(\bar{x} - x_i) + O(\tilde{s}^2, \tilde{s}m)$$

$$v(x_i) \equiv \langle \Delta x_i^2 \rangle = [x_i(1-x_i) + O(m, \tilde{s})]/\Omega + O(\tilde{s}^2, m^2, m\tilde{s})$$