Adaptation in structured populations

Peter Pfaffelhuber

University of Freiburg

joint work with Andreas Greven, Cornelia Pokalyuk and Anton Wakolbinger

LMU, July 2014

Population genetic models

Populations of constant size have been modelled by

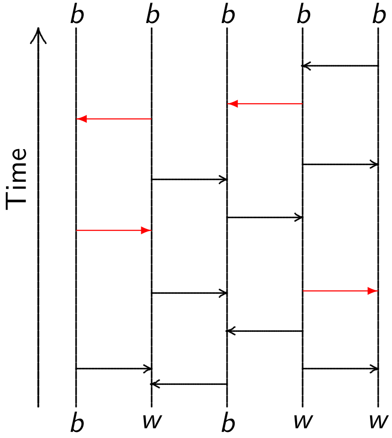
- **Markov Chains** (Wright-Fisher-model, Moran model)
- **Diffusion approximations** (Fisher-Wright diffusion)

$$dX = \alpha X(1-X)dt + \sqrt{X(1-X)}dW$$

or measure-valued diffusions (Fleming-Viot superprocess)

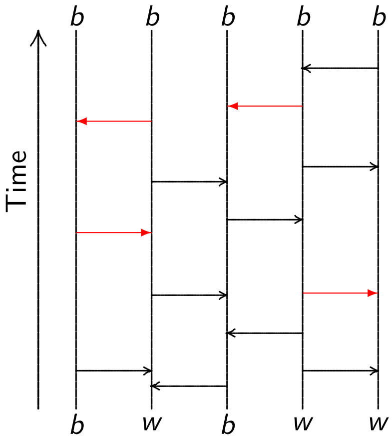
- **Keywords:** Superprocess, duality of Markov processes, random tree

The Moran model with selection



- Alleles are b and w ($b$eneficial, $w$ild-type)
- Each pair resamples with rate 1
- Each line creates red arrows with rate α

The Moran model with selection



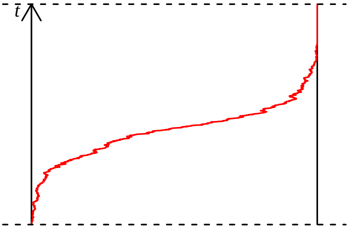
- Black arrows can be used by any allele
- Only b alleles can use red arrows
- The state at all times can be read from this graphical representation

The Wright-Fisher diffusion

- **Frequency path** of beneficial allele is

$$dX = \alpha X(1-X)dt + \sqrt{X(1-X)}dW, \qquad X_0 = x$$

N	population size
s	selective advantage
α	$:= sN$
dt	$\equiv Ndt$ generations

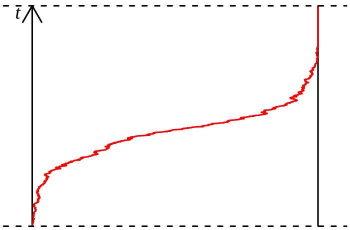


- **Conditioned on fixation**

$$dX = \alpha X \coth(\alpha X)(1-X)dt + \sqrt{X(1-X)}dW, \qquad X_0 \downarrow 0$$

Selective sweeps: Classical case

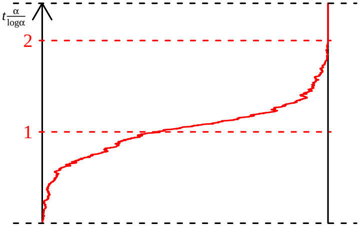
- A highly beneficial mutant eventually fixes in a panmictic population
- Fixation time $T := \inf\{t \geq 0 : X(t) = 1\}$



- Classical result: $\frac{\alpha}{\log \alpha} T \xrightarrow{\alpha \to \infty} 2.$

Selective sweeps: Classical case

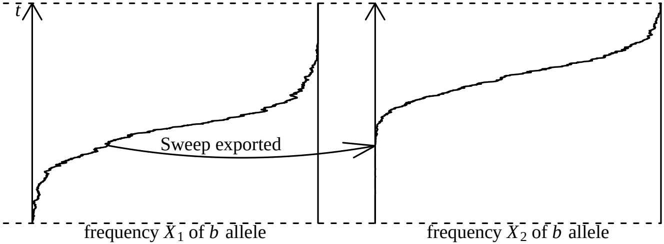
- A highly beneficial mutant eventually fixes in a panmictic population
- Fixation time $T := \inf\{t \geq 0 : X(t) = 1\}$



- Classical result: $\frac{\alpha}{\log \alpha} T \xrightarrow{\alpha\to\infty} 2.$

Selective sweeps: Case considered in this talk

- A highly beneficial mutant eventually fixes in a structured population
- Fixation time $T := \inf\{t \geq 0 : X_i(t) = 1, i = 1, ..., \nu\}$



- What is the distribution of T?

Selective sweeps: Case considered in this talk

- Conditioned on eventual fixation, on 2 islands

$$dX_1 = \big(\mu(X_2 - X_1) + \alpha X_1(1 - X_1)\coth\left(\alpha(X_1 + X_2)\right)\big)dt + \sqrt{X_1(1 - X_1)}dW_1$$
$$dX_2 = \big(\mu(X_1 - X_2) + \alpha X_2(1 - X_2)\coth\left(\alpha(X_1 + X_2)\right)\big)dt + \sqrt{X_2(1 - X_2)}dW_2$$

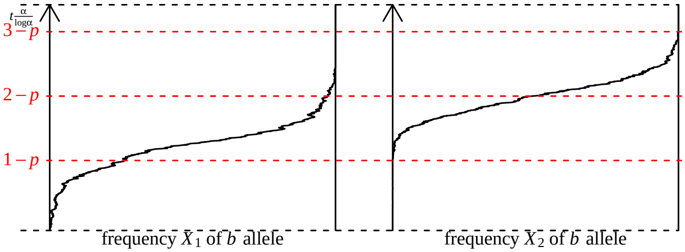
- Theorem: Let $T := \inf\{t \geq 0 : X_1(t)_t = X_2(t) = 1\}$.

$$\frac{\alpha}{\log \alpha} T \xRightarrow{\alpha \to \infty} \begin{cases} 3 - p, & \text{if } \mu \approx c\alpha^p \text{ for } p \in [0, 1], \\ 3 + X, & \text{if } \mu \approx \frac{c}{\log \alpha}, \end{cases}$$

where $X \sim \exp(2c)$.

Selective sweeps: Case considered in this talk

- Fixation time $T := \inf\{t \geq 0 : X_i(t) = 1, i = 1, ..., \nu\}$



$$\frac{\alpha}{\log \alpha} T \xrightarrow{\alpha\to\infty} 3 - p$$

Selective sweeps on ν islands

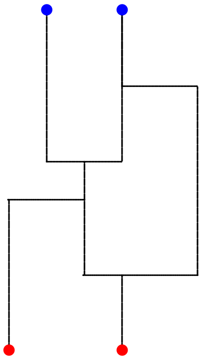
- Theorem: Condition on eventual fixation, on ν islands.
  Let $T := \inf\{t \geq 0 : X_i(t) = 1, i = 1, ..., \nu\}$. Then,

$$\frac{\alpha}{\log \alpha} T \xrightarrow[\alpha\to\infty]{} \begin{cases} 2 + (1-p)S_Y, & \text{if } \mu \approx c\alpha^p \text{ for } p \in [0,1], \\ 1 + S_Z, & \text{if } \mu \approx \frac{c}{\log \alpha}. \end{cases}$$

Take home

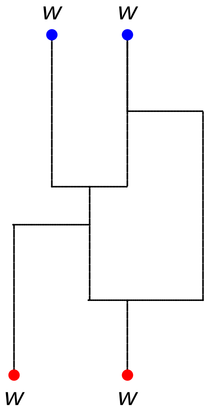
Even if the population is only weakly structured (i.e. μ is large), fixation can take a lot longer than under panmixia.

The Ancestral Selection Graph



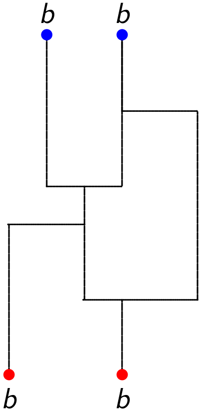
- Two lines coalesce at rate 1
- At rate α each line is hit by a red arrow; thus it produces a new line in the ancestral graph

The Ancestral Selection Graph



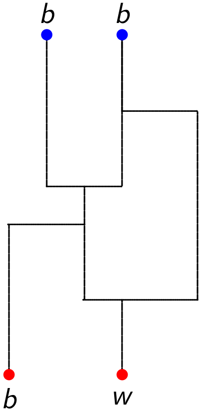
- Two lines coalesce at rate 1
- At rate α each line is hit by a red arrow; thus it produces a new line in the ancestry graph

The Ancestral Selection Graph



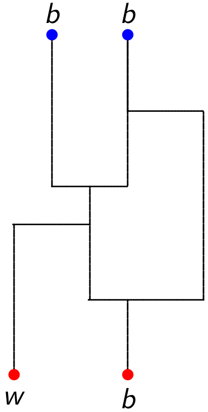
- Two lines coalesce at rate 1
- At rate α each line is hit by a red arrow; thus it produces a new line in the ancestry graph

The Ancestral Selection Graph



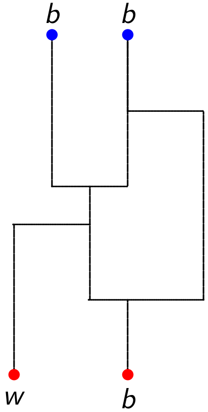
- Two lines coalesce at rate 1
- At rate α each line is hit by a red arrow; thus it produces a new line in the ancestry graph

The Ancestral Selection Graph



- Two lines coalesce at rate 1
- At rate α each line is hit by a red arrow; thus it produces a new line in the ancestry graph

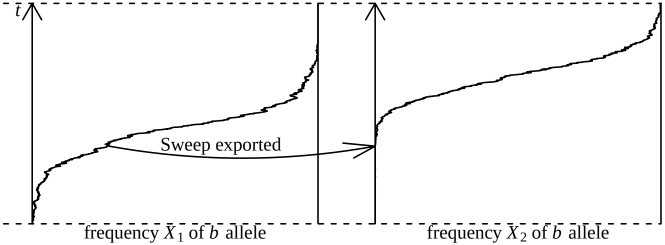
The Ancestral Selection Graph



- An individual • has b iff there is an ancestor • which has b
- Reason: only b-alleles can use every branch

Fixation under strong selection in the ASG

- Consider 2 demes.
  Assume that deme 1 carries the immortal line at time 0.



Fixation under strong selection in the ASG

- **Lemma 1:** Start ASG with ∞ many lines in all demes. Wait until time t and mark a random line from deme 1 with b. Let

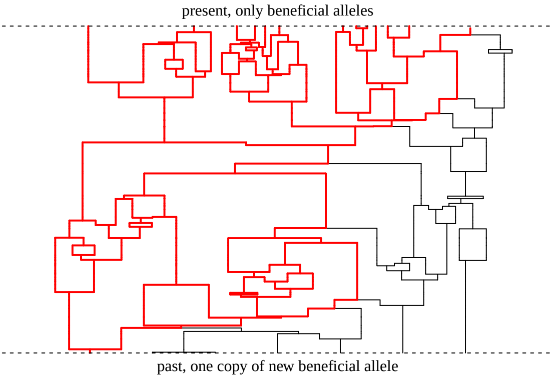
$$C_t := \{\text{marked line potential ancestor of all starting lines}\}.$$

Then,

$$\lim_{\varepsilon \to 0} \mathbf{P}_\varepsilon(T < t | \text{fixation}) = \mathbf{P}(C_t).$$

- Reason: Conditioning on eventual fixation means that at least one line is b at time 0.

Fixation under strong selection in the ASG



Fixation under strong selection in the ASG

- $\mathcal{K}$: line counting process of the ASG has the dynamics

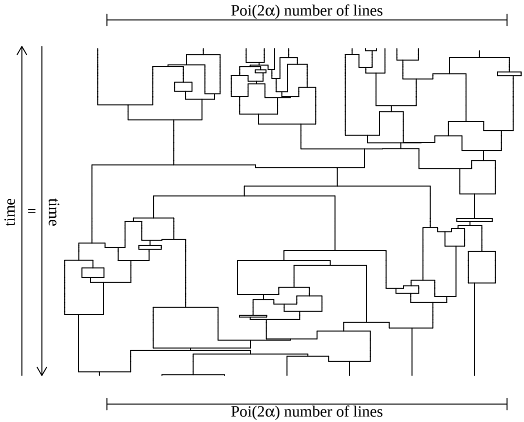
$$K_t \to \begin{cases} K_t + 1 & \text{at rate } \alpha K_t, \\ K_t - 1 & \text{at rate } \binom{K_t}{2} \end{cases}$$

  in each deme, and migration events.

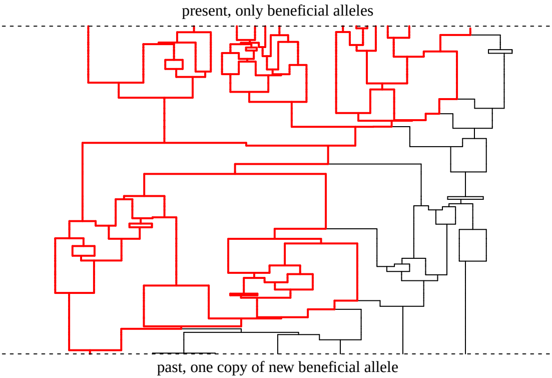
- **Lemma 2:** $K_\infty \sim \pi_{2\alpha}$, a vector of Poi$(2\alpha)$ distributions, conditioned on being positive.
  - $\pi_{2\alpha}$ is reversible for $\mathcal{K}$.
  - The full graph is reversible in its equilibrium.

- **Corollary:** Up to a small error, Lemma 1 still holds if ASG is started in equilibrium.

Fixation under strong selection in the ASG



Fixation under strong selection in the ASG



Some observations on a single deme

- Use reversibility of equilibrium in order to construct ASG from past to present.
- Start ASG in the past with equilibrium number of lines, one *b* allele in deme 1
- *b*-allele increases on deme 1 approximately according to a Yule process with rate α, up to having $\alpha^{1-\varepsilon}$ copies of *b*.
- If *w*-allele has $\sim \alpha^{1-\varepsilon}$ copies, *w*-allele decreases on deme 1 approximately according to a subcritical branching process.
- Fixation occurs iff all lines carry allele *b*.

Some observations for $\mu = c\alpha^p$

- Number of lines with allele b is about $e^{\alpha t}$ for $t < \frac{\log \alpha}{\alpha}$.
- Q: At what time $\tilde{T}$ does the first migrant (occurring at rate α^p) move from deme 1 to 2 in the ASG?

Some observations for $\mu = c\alpha^p$

- Number of lines with allele b is about $e^{\alpha t}$ for $t < \frac{\log \alpha}{\alpha}$.
- Q: At what time $\tilde{T}$ does the first migrant (occurring at rate α^p) move from deme 1 to 2 in the ASG?
- A: Not before $t' = (1 - p - \varepsilon)\frac{\log \alpha}{\alpha}$:
  Expected number of migration marks on Yule process before t' is

$$\mu \sum_{i=1}^{\alpha^{1-p-\varepsilon}} i \frac{1}{\alpha i} = \alpha^p \frac{\alpha^{1-p-\varepsilon}}{\alpha} = \alpha^{-\varepsilon} \xrightarrow{\alpha \to \infty} 0.$$

Some observations for $\mu = c\alpha^p$

- Number of lines with allele b is about $e^{\alpha t}$ for $t < \frac{\log \alpha}{\alpha}$.
- Q: At what time $\tilde{T}$ does the first migrant (occurring at rate α^p) move from deme 1 to 2 in the ASG?
- A: Not after time $t'' = (1 - p + \varepsilon)\frac{\log \alpha}{\alpha}$:
  Expected number of migration marks on rate-α Yule process up to time t'' is

$$\mu \sum_{i=1}^{\alpha^{1-p+\varepsilon}} i \frac{1}{\alpha i} = \cdots = \alpha^{\varepsilon} \xrightarrow{\alpha \to \infty} \infty.$$

Some observations for $\mu = c\alpha^p$

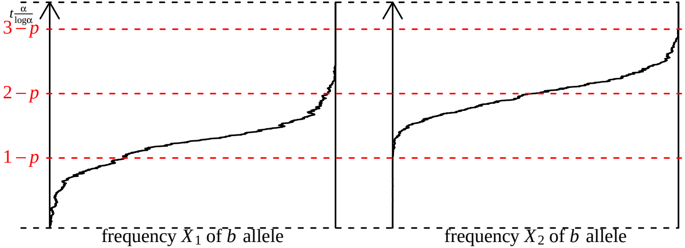
- Number of lines with allele b is about $e^{\alpha t}$ for $t < \frac{\log \alpha}{\alpha}$.
- Q: At what time $\tilde{T}$ does the first migrant (occurring at rate α^p) move from deme 1 to 2 in the ASG?
- A: Combining the arguments,

$$\frac{\alpha}{\log \alpha} \tilde{T} \xRightarrow{\alpha \to \infty} 1 - p.$$

Some observations for $\mu = c\alpha^p$

- b allele increases in frequency up to time $\tilde{T} \approx (1-p)\frac{\log \alpha}{\alpha}$.
- First migrants infects deme 2 at time $\tilde{T}$
- Fixation on island 2 after duration $2\frac{\log \alpha}{\alpha}$.

$$\implies \qquad \frac{\alpha}{\log \alpha} T \xrightarrow{\alpha\to\infty} 3 - p.$$



Summary and outlook

- If the migration rate is $\sim \alpha^p$, the fixation time of a beneficial mutant is $\frac{\log \alpha}{\alpha}(2 + (1-p)S_Y)$, where S_Y is maximal graph distance from the founder island.

- Ancestral selection graph can be extended to include recombination in order to compute other statistics.