Selective sweeps in structured populations

Cornelia Pokalyuk
joint work with Andreas Greven, Peter Pfaffelhuber and Anton Wakolbinger

June 14, 2012

We are interested in populations living on one island or being distributed equally on two islands with symmetric migration.

Assume a beneficial allele is introduced once to the population and eventually goes fixed under strong constant selection and the population is evolving according to the diffusion setting, i.e.

- Selection strength α
- Migration rate μ
- frequency path $(Y_1(t), Y_2(t))_{t\geq 0}$ of the beneficial allele conditioned on fixation solves the SDE

$$dY_1 = (\alpha Y_1(1-Y_1)\coth(\alpha(Y_1+Y_2)) + \mu(Y_2-Y_1))dt + \sqrt{Y_1(1-Y_1)}dW_1$$

$$dY_2 = (\alpha Y_2(1-Y_2)\coth(\alpha(Y_1+Y_2)) + \mu(Y_1-Y_2))dt + \sqrt{Y_2(1-Y_2)}dW_2,$$

started with $Y_1(0) = Y_2(0) = 0$ for two independent standard Brownian motions W_1, W_2.

Define the fixation time

$$T_{\text{fix}} := \inf\{t | Y_1(t) = Y_2(t) = 1\}.$$

Analogously define the fixation time for a panmictic population.

Define the fixation time

$$T_{\text{fix}} := \inf\{t | Y_1(t) = Y_2(t) = 1\}.$$

Analogously define the fixation time for a panmictic population.

Theorem:

One island

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha\to\infty} 2 \text{ in probability, i.e. for large } \alpha$$

$$T_{\text{fix}} \approx 2\frac{\log(\alpha)}{\alpha}.$$

Two islands

- If $\mu \in \Omega(\alpha)$,

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha\to\infty} 2 \text{ in probability.}$$

- If $\mu \in \Theta(\alpha^p)$ for $p \in [0,1)$,

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha\to\infty} (3-p) \text{ in probability, i.e. for large } \alpha$$

$$T_{\text{fix}} \approx (1-p+2)\frac{\log(\alpha)}{\alpha} = (3-p)\frac{\log(\alpha)}{\alpha}.$$

- If $\mu = c/\log(\alpha)$, for some $c \in \mathbb{R}_+$

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha\to\infty} 3+X \text{ in distribution, where } X \sim \text{Exp}(2c).$$

  That is for large α

$$T_{\text{fix}} \approx (1+X+2)\frac{\log(\alpha)}{\alpha} = (3+X)\frac{\log(\alpha)}{\alpha}.$$

Heuristic explanation: One island

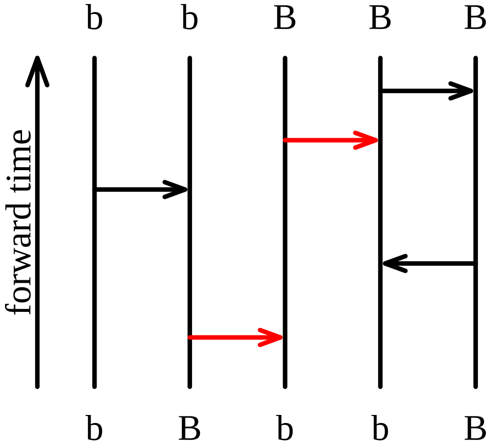
Main tool:

Ancestral selection graph (Krone and Neuhauser, 1997)

- analog to Kingman coalescent in setting with selection:
- gives potential genealogies of a sample of a large population evolving under constant selection.

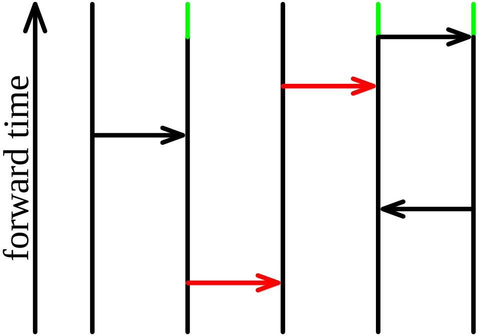
Motivation ASG

Moran model with selection



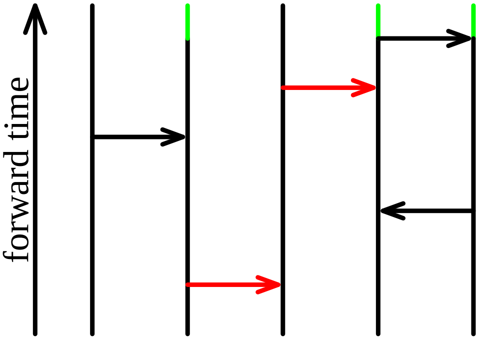
- Alleles are b and B
- Each pair resamples at rate 1
- Each line creates red arrows at rate α
- Black arrows can be used by any allele
- Only B alleles can use red arrows

Motivation ASG



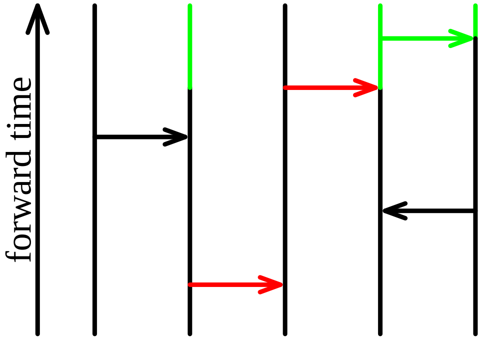
- Consider a sample of the population
- Forget types of individuals

Motivation ASG



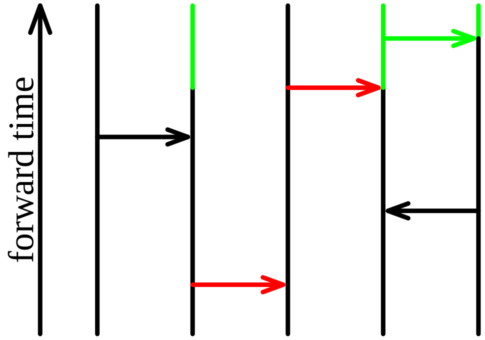
- If two lines meet a common black arrow, they coalesce

Motivation ASG



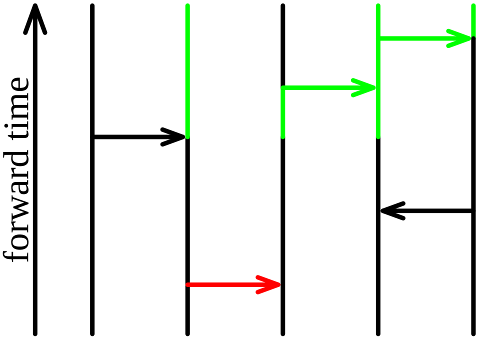
- If two lines meet a common black arrow, they coalesce

Motivation ASG



- If two lines meet a common black arrow, they coalesce
- If a line meets a red arrow, the line splits

Motivation ASG



- If two lines meet a common black arrow, they coalesce
- If a line meets a red arrow, the line splits

Definition: Ancestral selection graph (ASG)

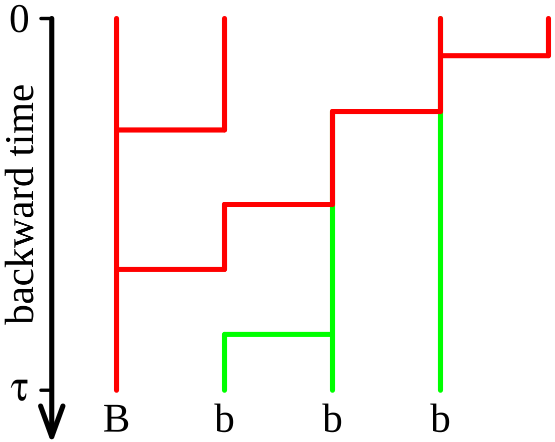
- each pair of lines coalesces at rate 1
- each line splits at rate α

Fixation in the ASG:

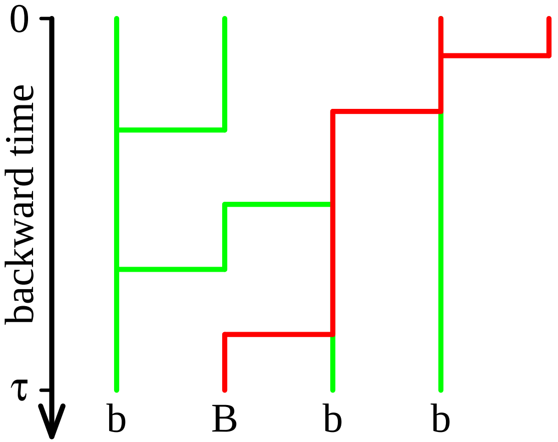
Mark one random individual i_0 with good type B, others with wild type b at time τ in the past.

The beneficial allele is fixed at time 0 (the present), if there exist directed paths between the individual (τ, i_0) and all individuals $(0, k)$.

Fixed



Not fixed



Definition: Ancestral selection graph (ASG)

- each pair of lines coalesces at rate 1
- each line splits at rate α

Fixation in the ASG:

Mark one random individual i_0 with good type B, others with wild type b at time τ in the past.
The beneficial allele is fixed at time 0 (the present), if there exist a directed paths between the individual (τ, i_0) and all individuals $(0, k)$.
Write $(0, k) \leftrightarrow (\tau, i_0)$.

Proposition:
"Fixation time in ASG is distributed as fixation time in diffusion:"

$$\mathbf{P}(T_{\text{fix}} \leq \tau) = \mathbf{P}\Big(\text{for all } k = 1, 2, \ldots \text{ it holds } (0, k) \leftrightarrow (\tau, 1) \text{ in ASG}\Big)$$

Important property: Reversibility of ASG

- Line counting process of ASG has reversible equilibrium.
  In equilibrium have $\approx n = 2\alpha$ lines
  (splitting rate $n \cdot \alpha = 2\alpha \cdot \alpha \approx \frac{2\alpha(2\alpha-1)}{2} = \binom{2\alpha}{2} = \binom{n}{2}$ coalescing rate)
- Reversibility can be conferred to ASG $\rightarrow$ in equilibrium the reversed ASG is also an ASG

Sketch of the proof

Step 1: Start ASG in equilibrium instead of with ∞-ly many individuals. Have at most 2α ancestors within time of order $1/\alpha$

Sketch of the proof

Step 1: Start ASG in equilibrium instead of with ∞-ly many individuals. Have at most 2α ancestors within time of order $1/\alpha$

Step 2: "Forward" and "backward" graphs spanned by single individuals are also ASGs

Sketch of the proof

Step 1: Start ASG in equilibrium instead of with ∞-ly many individuals. Have at most 2α ancestors within time of order $1/\alpha$

Step 2: "Forward" and "backward" graphs spanned by single individuals are also ASGs

Step 3: Exponential growth at rate α: forward ASG of (τ, i_0) reaches $\varepsilon\alpha$ lines at time $t = \log(\varepsilon\alpha)/\alpha$ for any $\varepsilon > 0$

Sketch of the proof

- Step 1: Start ASG in equilibrium instead of with ∞-ly many individuals. Have at most 2α ancestors within time of order $1/\alpha$
- Step 2: "Forward" and "backward" graphs spanned by single individuals are also ASGs
- Step 3: Exponential growth at rate α: forward ASG of (τ, i_0) reaches $\varepsilon\alpha$ lines at time $t = \log(\varepsilon\alpha)/\alpha$ for any $\varepsilon > 0$
- Step 4: Approximately deterministic growth according to differential equation $dQ = Q(1 - Q/2)dt$ from $\varepsilon\alpha$ lines to $2\alpha - \varepsilon\alpha$ lines within time frame $\mathcal{O}(1/\alpha)$.

Sketch of the proof

- Step 1: Start ASG in equilibrium instead of with ∞-ly many individuals. Have at most 2α ancestors within time of order $1/\alpha$
- Step 2: "Forward" and "backward" graphs spanned by single individuals are also ASGs
- Step 3: Exponential growth at rate α: forward ASG of (τ, i_0) reaches $\varepsilon\alpha$ lines at time $t = \log(\varepsilon\alpha)/\alpha$ for any $\varepsilon > 0$
- Step 4: Approximately deterministic growth according to differential equation $dQ = Q(1 - Q/2)dt$ from $\varepsilon\alpha$ lines to $2\alpha - \varepsilon\alpha$ lines within time frame $\mathcal{O}(1/\alpha)$.
- Step 5: Backward ASG of typical individual $(0, k)$ reaches $\varepsilon\alpha$ lines after time $t = \log(\varepsilon\alpha)/\alpha$
  Size of underlying ASG $\approx 2\alpha \Rightarrow$ Forward ASG of (τ, i_0) meets with backward ASG of typical individual $(0, k)$.

Two islands

Definition: Ancestral selection graph (ASG)

- each pair of lines coalesces within islands at rate 1
- each line splits at rate α within its island
- each line migrates to the neighboring island at rate μ

Note:

- ASG on two islands has also a reversible equilibrium
- to compute fixation time can again start ASG in equilibrium

Migration rate $\mu \gtrsim \alpha$

First migrant to island 2: immediate migrant, i.e. within time of order $\mathcal{O}(1/\alpha)$.

Show: Other migrants do not speed up the sweep on island 2 and do not slow down the sweep on island 1.

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha \to \infty} 2 \text{ in probability.}$$

Migration rate $\mu \approx \alpha^p$ for $p \in [0, 1)$

First migrant to island 2:
Size of (τ, i_0)-ASG $\approx e^{\alpha \cdot \log(\alpha)(1-p)/\alpha} = \alpha^{1-p}$ at time $\log(\alpha)(1-p)/\alpha$
$\Rightarrow$ first migrant approximately at time $\log(\alpha)(1-p)/\alpha$

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha \to \infty} 1 - p + 2 = 3 - p \text{ in probability.}$$

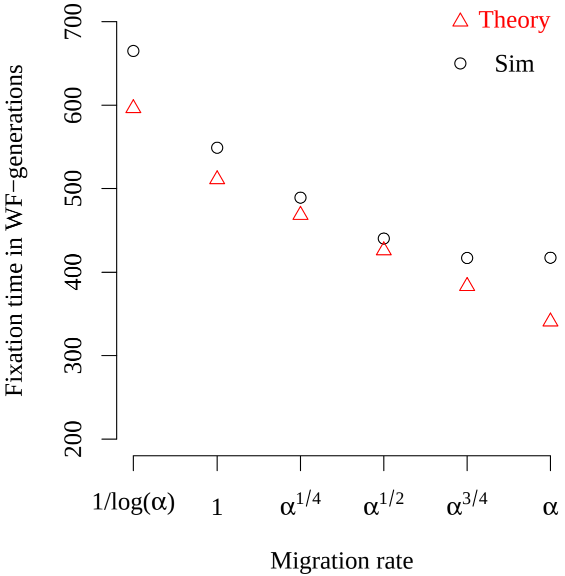
Migration rate $\mu = c/\log(\alpha)$

First migrant to island 2:

- no migrant until time $\log(\alpha)/\alpha$
- after time $\log(\alpha)/\alpha$: Size of (τ, i_0)-ASG $\approx 2\alpha$

$\Rightarrow$ At ("constant") rate $2\alpha \cdot c/\log(\alpha)$ individuals migrate to island 2
$\Rightarrow$ waiting time scaled by $\log(\alpha)/\alpha$ is exponentially distributed with rate $2c$.

$$\frac{\alpha}{\log(\alpha)} T_{\text{fix}} \xrightarrow{\alpha \to \infty} 1 + X + 2 \text{ in distribution, where } X \sim \text{Exp}(2c).$$



$N = 10^5$,
$s = 0.05$, i.e.
$\alpha = N \cdot s = 5000$
simulations based on 1000 draws