

Announcements: Office hours Today 12:30-2pm

Last time: Linkage equilibrium approx ("independent sites")



$$\frac{ds(\vec{g})}{dt} = \sim(x-\bar{x}) + \sim L\mu + \sim e + \sim \frac{\sum z_e}{\sqrt{L}}$$

Selection on genotypes

$$e \rightarrow \infty \rightarrow \prod_{e=1}^L$$

$$\frac{df_e}{dt} = \sim s_e + \sim \mu_e + \sim \frac{z_e}{\sqrt{L}}$$

Selection on alleles

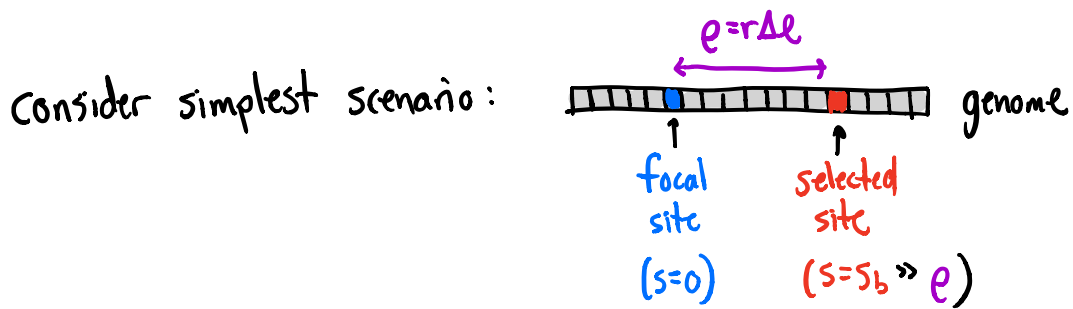
(\sim the "ideal gas" of evolutionary dynamics)

\Rightarrow A victory for reductionism?

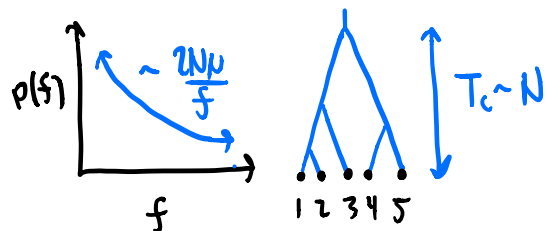
\Rightarrow Requires $\rho_{\text{eff}} = r\Delta e \gg \frac{1}{N}, s \Rightarrow \frac{\Gamma}{\mu} \gg 1, Ns$

\Rightarrow empirically, $\frac{\Gamma}{\mu} \sim \mathcal{O}(1) \Rightarrow$ breaks down for strong beneficial mut'n's!

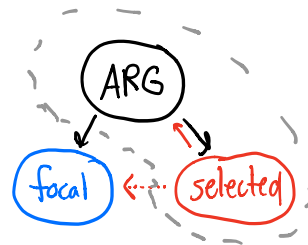
Today: what happens when this condition breaks down?



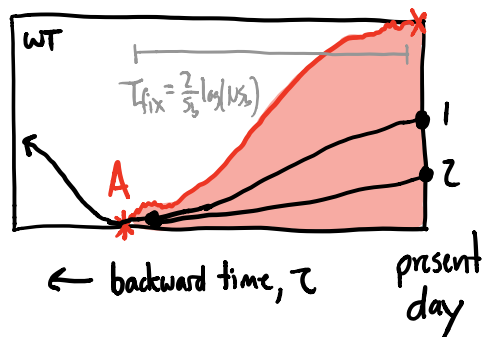
\Rightarrow when $e \rightarrow \infty$ focal site looks like neutral model:



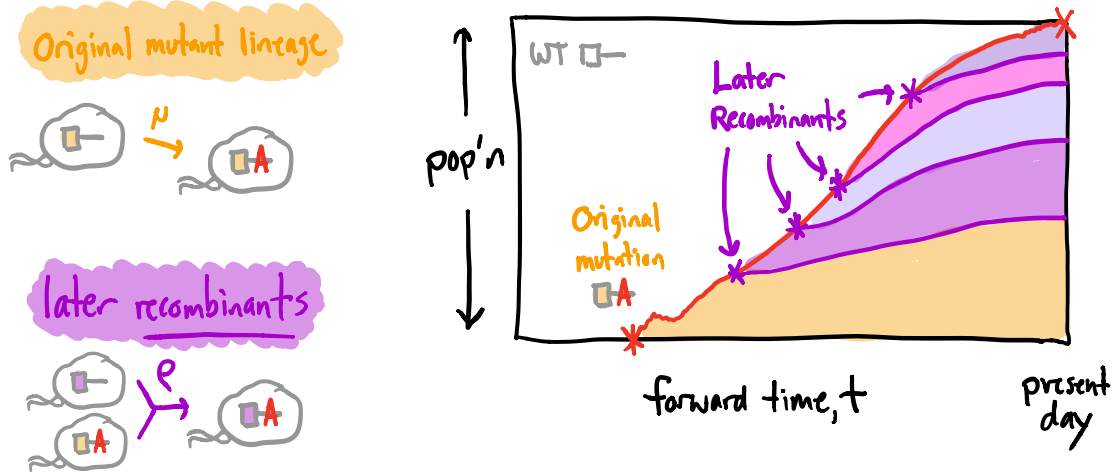
\Rightarrow these patterns can change when $e < \infty$ ("linked selection")



\Rightarrow behavior @ selected site is easy:

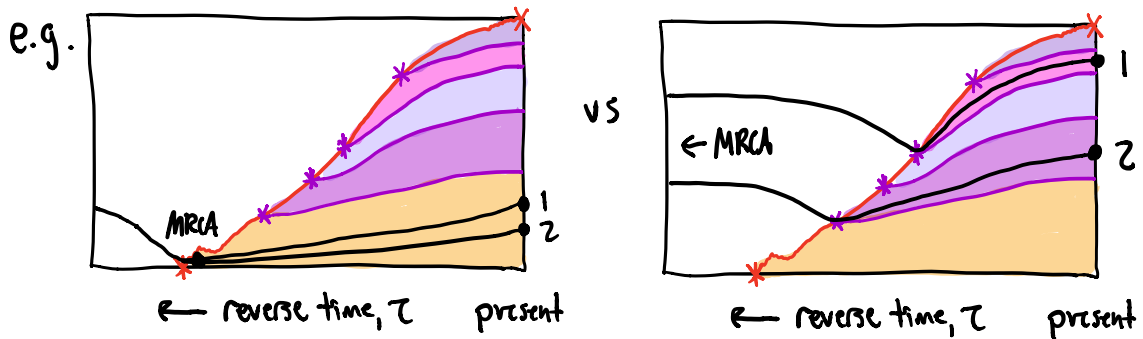


⇒ @ linked neutral site, must now distinguish between:



Why is this important?

⇒ individuals only coalesce during sweep if drawn from same lineage!

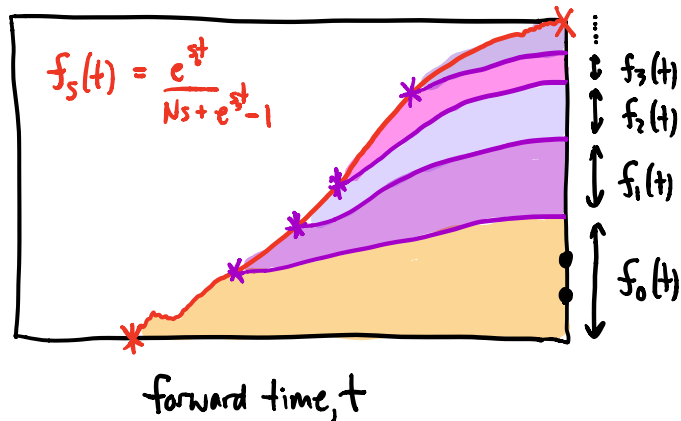


⇒ otherwise, trace back to different pre-sweep ancestors

⇒ neutral coalescence ($T_{MRCA} \sim N \gg T_{fix}$)

⇒ Total probability that 2 individuals coalesce during sweep:

$$P_c \equiv \sum_{k=0}^{K_{\max}} f_k(t)^2$$



where $f_k(t)$ = size of k^{th} recombinant lineage.

⇒ How do we predict $f_k(t)$?

⇒ can learn a lot by focusing on short times

$\frac{1}{s_b} \ll t \ll T_{\text{fix}} = \frac{2}{s_b} \log(N_s)$ when A is still rare.

$$f_s \sim \frac{1}{N_s} e^{s_b t} \ll 1$$

⇒ recombinant lineages are founded @ total rate:

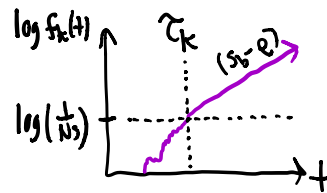
$$\Theta_r(t) \sim N e \underbrace{f_s(t) \cdot (1 - f_s(t))}_{\approx 1} \approx \frac{e}{s_b} e^{s_b t}$$

⇒ each recombinant lineage satisfies:

$$\frac{df_k}{dt} = \underbrace{s_b f_k}_{\text{selection}} - \underbrace{e f_k}_{\text{outflow due to recomb. (w/UT)}} + \underbrace{\sqrt{\frac{f_k}{N}} \eta_k(t)}_{\text{genetic drift.}}$$

⇒ we know how these behave:

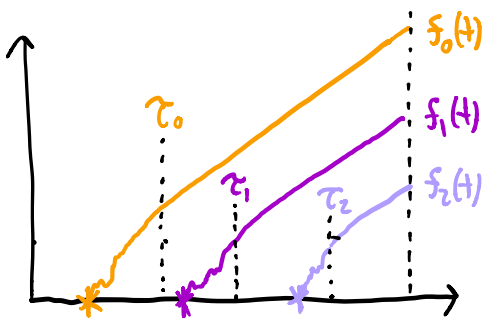
w/ probability $P_{\text{est}} \sim s_b e \sim s_b$



lineage establishes + grows as $f_k(t) \sim \frac{1}{N_s} e^{(s_b - e)(t - \tau_k)}$

where τ_k = establishment time of lineage k.

⇒ key insight: all recombinants grow @ same rate $(s_b - e)$
so relative sizes independent of time!



$$\frac{f_k(t)}{f_0(t)} = e^{-(s_b - e)(\tau_k - \tau_0)}$$

⇒ holds even for $t \geq T_{\text{fix}}$!

⇒ crucial step: How do we find τ_k ?

① By convention, set $\tau_0 \equiv 0$ (i.e. $t =$ time since start of sweep)

② Successful recombinants are produced @ rate

$$\Theta_{r,est}(t) = \Theta_r(t) \times P_{est} \sim \frac{\rho}{s_b} e^{s_b t} \cdot s_b \sim \rho e^{s_b t}$$

total # recombinants produced @ gen t

prob. that each survives drift

⇒ avg # of successful recombinants by time t :

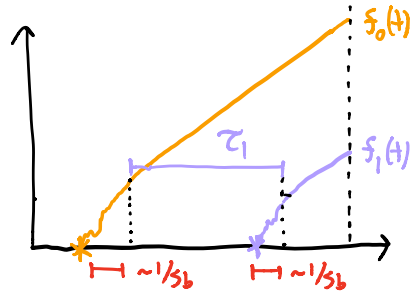
$$\langle k \rangle = \int_0^t \Theta_{r,est}(t') dt'$$

③ Heuristically, time to first successful recomb occurs when:

$$\int_0^{\tau_1} \Theta_{r,est}(t') dt' \sim \mathcal{O}(1)$$

$$\Rightarrow 1 \sim \int_0^{\tau_1} \rho e^{s_b t'} dt' = \frac{\rho}{s_b} (e^{s_b \tau_1} - 1) \Rightarrow \tau_1 \approx \frac{1}{s_b} \log \left(\frac{s_b}{\rho} + 1 \right)$$
$$\approx \frac{1}{s_b} \log \left(\frac{s_b}{\rho} \right) \quad [\text{when } s_b \gg \rho]$$

Note: $\tau_1 \gg \frac{1}{s_b}$ when $s_b \gg e$:



④ Similarly, k^{th} successful recombinant typically occurs when:

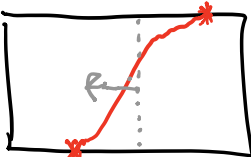
$$\int_0^{\tau_k} \Theta_{r, \text{est}}(t) dt \sim k \quad \Rightarrow \quad \tau_k \approx \frac{1}{s_b} \log\left(\frac{s_b k}{e}\right)$$

$$\Rightarrow \frac{f_k(t)}{f_0(t)} = e^{(s_b - e)(\tau_0 - \tau_k)} = e^{- (s_b - e) \frac{1}{s_b} \cdot \log\left(\frac{s_b k}{e}\right)} = \left(\frac{e}{s_b k}\right)^{1 - e/s_b}$$

\Rightarrow @ end of sweep, size of k^{th} lineage is given by

$$f_k(\infty) = \frac{f_k(t)}{f_0(t) + \sum_{j=1}^{K_{\text{max}}} f_j(t)} = \frac{f_k(t)/f_0(t)}{1 + \sum_{j=1}^{K_{\text{max}}} f_j(t)/f_0(t)}$$

$$\Rightarrow f_k(\infty) = \begin{cases} \left[1 + \sum_{j=1}^{K_{\max}} \left(\frac{e}{s_b j} \right)^{1 - e/s_b} \right]^{-1} & \text{if } k=0 \\ f_0(\infty) \left(\frac{e}{s_b k} \right)^{1 - e/s_b} & \text{if } k \geq 1 \end{cases}$$

\Rightarrow what is K_{\max} ?  $f_s = \frac{e^{s_b t}}{N s_b + e^{s_b t} - 1}$

\Rightarrow total # of successful recombinants:

$$\leq \int_0^{\infty} N e f_s(t) (1 - f_s(t)) \cdot s_b dt \sim N e$$

\Rightarrow decent approx to take $K_{\max} \sim N e$

\Rightarrow Two regimes:

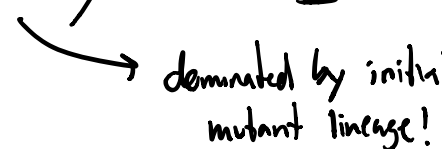
① $N e \ll 1 \Rightarrow K_{\max} \ll 1 \Rightarrow$ typically no recombinants
 \Rightarrow like asexual case

② $N e \gg 1 \Rightarrow$ many recombinants contribute!

$$\begin{aligned}
\Rightarrow \frac{1}{f_0(\infty)} &= 1 + \sum_{j=1}^{K_{\max}} \left(\frac{e}{s_b j}\right)^{1-e/s_b} \approx 1 + \int_1^{Ne} \left(\frac{e}{s_b j}\right)^{1-e/s_b} dj \\
&= 1 + \left(\frac{e}{s_b}\right)^{1-e/s_b} \frac{s_b}{e} (j)^{e/s_b} \Big|_1^{Ne} = 1 + \frac{e^{-e/s_b}}{s_b} \left[Ne^{e/s_b} - 1 \right] \\
&\approx \exp\left[+\frac{e}{s_b} \log\left(Ne \cdot \frac{s_b}{e} \right) \right] = \exp\left[+\frac{e}{s_b} \log(Ns_b) \right]
\end{aligned}$$

Finally, probability that 2 individuals coalesce during sweep:

$$\begin{aligned}
p_c &= \sum_{k=0}^{K_{\max}} f_k(\infty)^2 = f_0(\infty)^2 \left[1 + \sum_{k=1}^{K_{\max}} \left(\frac{e}{s_b k}\right)^{2(1-e/s_b)} \right] \\
&= \exp\left(-\frac{2e}{s_b} \log(Ns_b) \right)
\end{aligned}$$



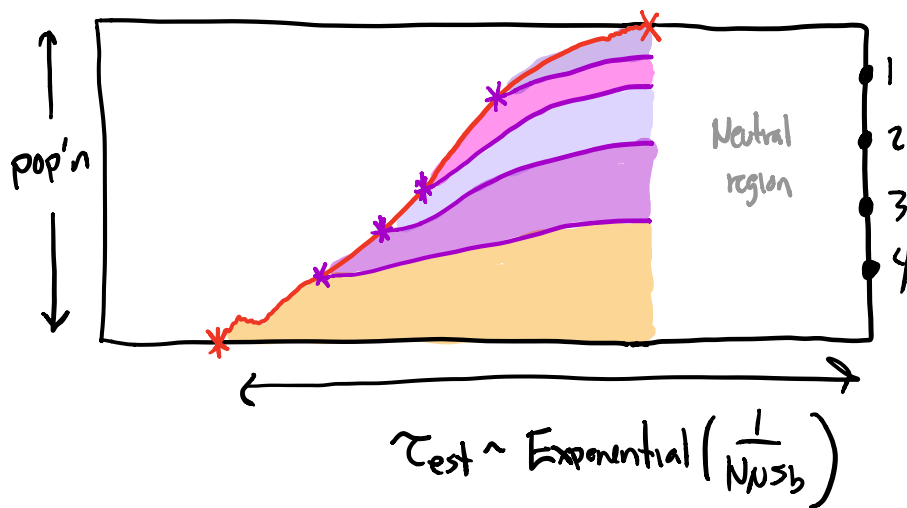
dominated by initial mutant lineage!

$$\begin{aligned}
\Rightarrow \langle T_{MRCA} \rangle &= T_{fix} \cdot p_c + N \cdot (1-p_c) \approx N(1-p_c) \quad \xrightarrow{\text{when } Ne \gg 1} \\
&\approx N \left(1 - e^{-\frac{2e}{s_b} \log(Ns_b)} \right) \approx \begin{cases} N & \text{if } eT_{fix} \gg 1 \\ \frac{2Ne}{s_b} \log(Ns_b) & \text{if } eT_{fix} \ll 1 \end{cases}
\end{aligned}$$

⇒ works for larger sample sizes:

$$p_c(n) = \sum_{k=0}^{K_{\max}} f_k(\infty)^n \approx e^{-\frac{n\theta}{s_b} \log(Ns_b)}$$

⇒ what happens if sweep fixed earlier?



Two regimes:

① $N \ll \tau_{est} \Rightarrow$ neutral coalescence before sweep! → ("after"?)

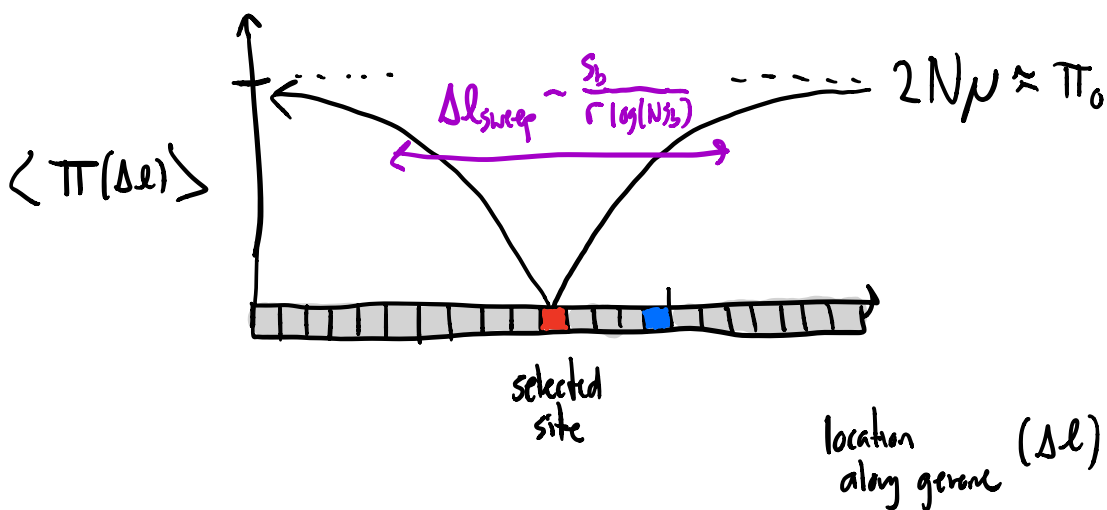
② $N \gg \tau_{est} \Rightarrow$ coalesce like before!

$$\langle T_{MRCA} \rangle \approx N \left(1 - e^{-\frac{2e}{s_b} \log(Ns_b)} \right)$$

\Rightarrow since $e = r\Delta l$:

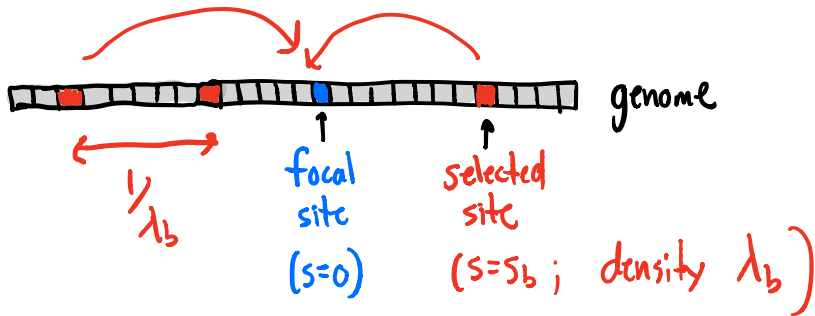
$$\langle T_{MRCA} \rangle = N \left(1 - e^{-\Delta l \cdot \frac{2r}{s_b} \log(Ns_b)} \right)$$

\Rightarrow can visualize as distance from selected site:



\Rightarrow major signal that people try to look for in data!
("selection scans")

Recurrent sweeps: can extend to multiple selected sites as long as they don't interfere...

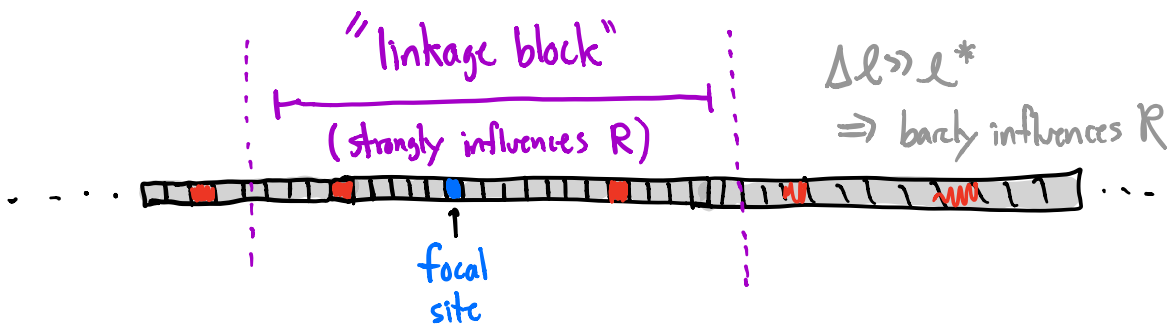


⇒ per generation rate of sweeps that lead to coalescence:

$$R = \int_0^{\infty} e^{-\frac{2r\Delta l}{s_b} \cdot \log(Ns_b)} \cdot 2N\mu\lambda_b \cdot s_b \cdot d\Delta l = \frac{N\mu\lambda_b s_b^2}{r \log(Ns_b)}$$

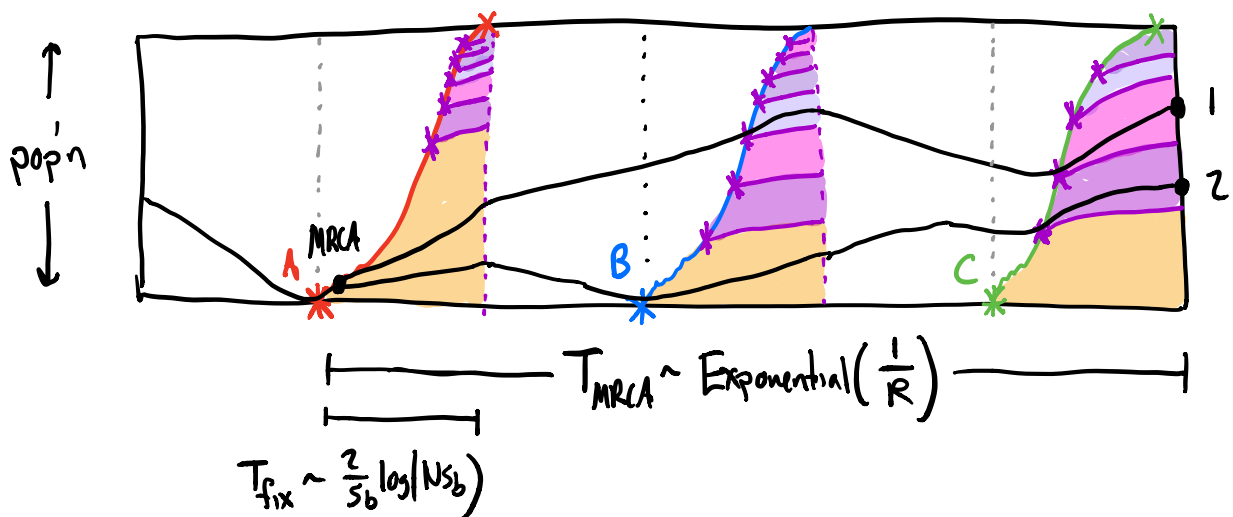
dominated by probability of really close sweep:

$$\left[\Delta l \lesssim l^* \equiv \frac{s_b}{r \log(Ns_b)} \right] \text{ when } p_c(z) \sim \mathcal{O}(1)$$



⇒ if time between sweeps ($\frac{1}{R}$) is $\gg T_{\text{fix}}$ but $\ll N$

$$\Rightarrow \langle T_{\text{MRCA}} \rangle = \frac{1}{R} = \frac{r \log(Ns)}{N\mu\lambda_b s_b^2}$$



⇒ Using definition of e^* can also write as:

$$\langle T_{\text{MRCA}} \rangle = \frac{1}{N \cdot e^* \mu \lambda_b \cdot s_b} = \frac{1}{N U_{b,\text{eff}} s_b}$$

⇒ looks like asexual case w/ $U_{b,\text{eff}} \equiv e^* \mu \lambda_b$

⇒ differences emerge in larger samples...