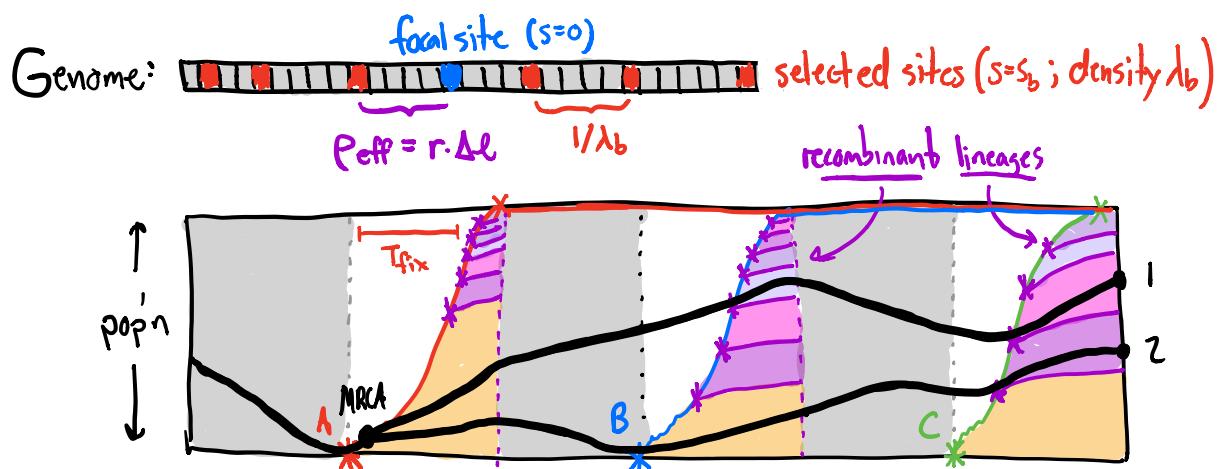


- Announcements:
- ① No weekly feedback next week (thanks for effort!)
 - ② No office hrs this week (always happy to chat!)

Last time: Linked selection via "classic selective sweeps"



Coalescence Prob Per Sweep:

$$P_c(n, \Delta l) = \exp[-n \cdot \Delta l \cdot \frac{f}{s_b} \cdot \log(N s_b)]$$

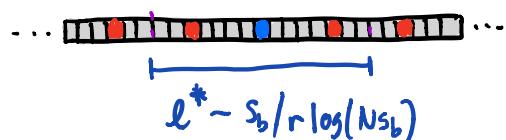
Total coalescence Rate from sweeps

$$R = \int_{-\infty}^{\infty} P_c(2, 1 \Delta l) \cdot 2 N \mu d_b s_b \cdot d \Delta l$$

When $N \gg \frac{1}{R} \gg T_{fix}$:

$$\langle T_{MRCA} \rangle \approx \frac{1}{R} = \frac{r \log(N s_b)}{2 N \mu d_b s_b^2}$$

$$\approx \int_{-l^*/2}^{l^*/2} O(1) \cdot 2 N \mu d_b s_b \cdot d \Delta l$$



Today: ① wrap up this case ② clonal interference

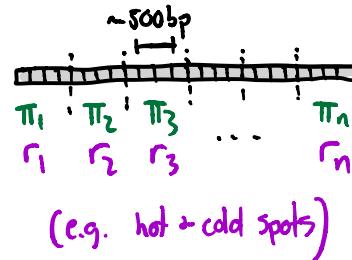
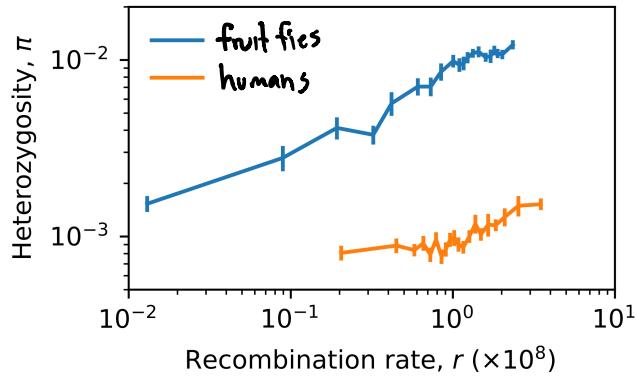
\Rightarrow Key prediction: genetic diversity (e.g. π) @ neutral (e.g. syn) sites

depends on local recombination rate r !

(since controls linkage to selected sites)

$$\langle \pi \rangle \approx \frac{r \log(Ns_b)}{s_b \cdot Ns_b \cdot \lambda_b}$$

\Rightarrow can test using natural variation in r along genome:

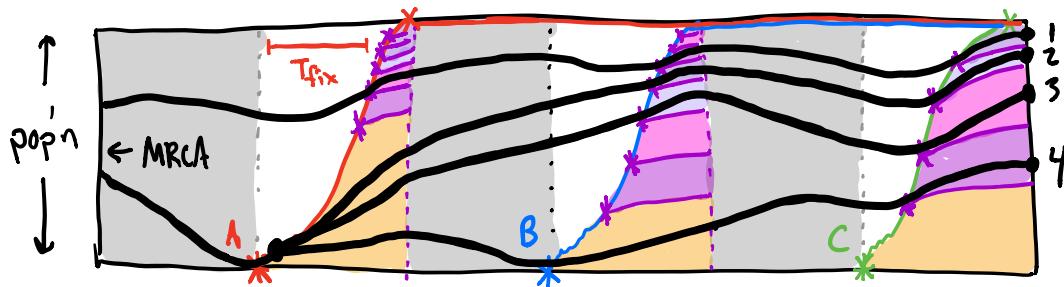


\Rightarrow sometimes spun as "linked selection \Rightarrow local reductions in N_e "

$$\text{e.g. } \pi_e \approx 2N_e(\ell) \pi$$

\Rightarrow WRONG!

\Rightarrow can see by examining larger sample sizes:



Recall: Coalescence Prob Per Sweep: $p_c(n, \Delta t) = \exp[-n \cdot \Delta t \cdot \frac{r}{s_b} \cdot \log(N s_b)]$

\Rightarrow Total rate of sweeps w/ n lineages coalescing:

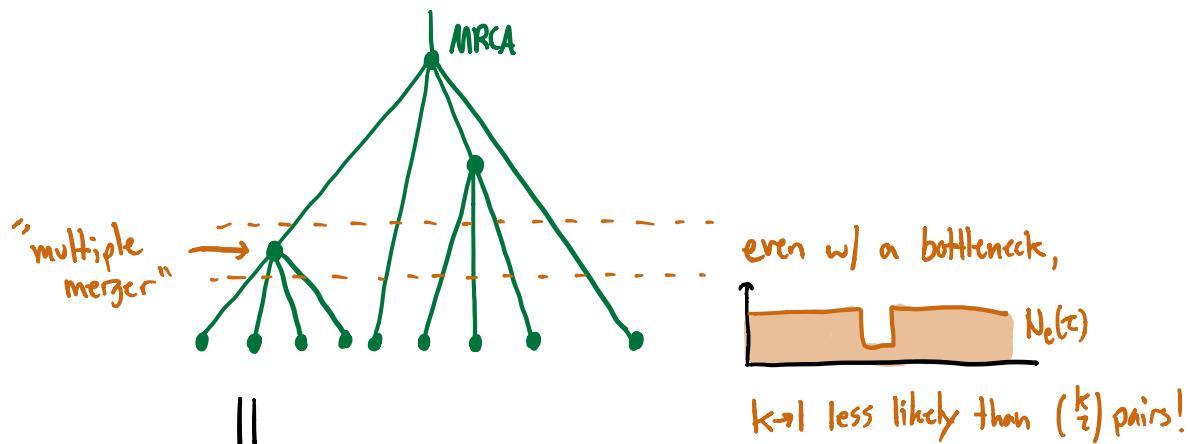
$$R(n \rightarrow 1) = \int_0^{\infty} e^{-n \frac{\Delta t r}{s_b} \log(N s_b)} \cdot 4N \nu d_b s_b d \Delta t = \frac{4N n \lambda_b s_b}{\frac{n r}{s_b} \log(N s_b)}$$

$$\Rightarrow R(n \rightarrow 1) = \frac{2}{n} R \quad \Rightarrow \text{Decays very slowly w/ n!}$$

[compare to $N \cdot \left(\frac{1}{N}\right)^n$ for neutral (Kingman) coalescent]

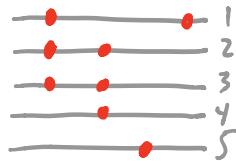
Upshot: if 2 lineages coalesce in a given timestep,
⇒ likely multiple lineages coalesce into same block!

⇒ can produce genealogies like:

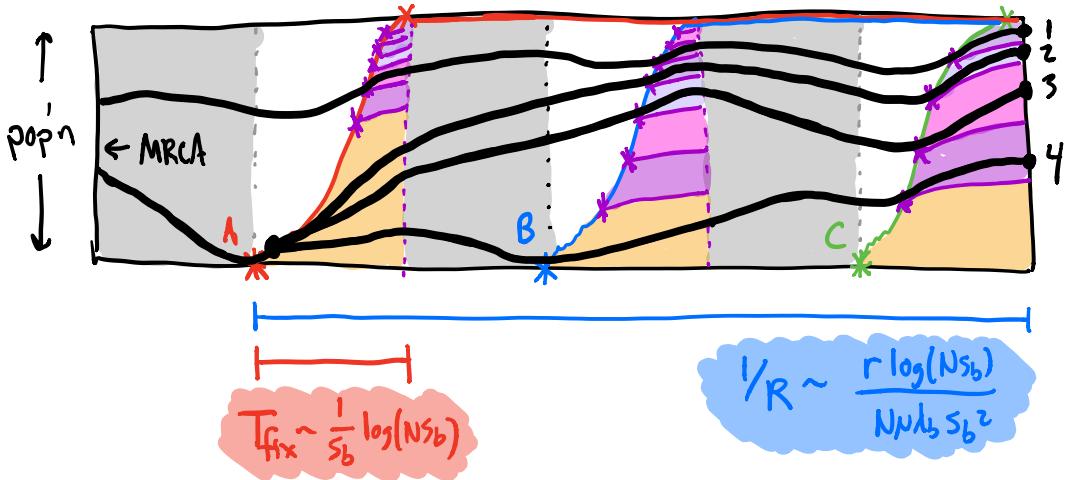


↳ Important open question:

How can we detect these
effects from mutation data?



\Rightarrow when is this successive mutations-like picture a good approx?



\Rightarrow check self consistency:

Each coalescence-causing sweep should fix before next one occurs

$$\Rightarrow R T_{fix} \ll 1 \Rightarrow \frac{N \lambda_b s_b^2}{r \log(N_{s_b})} \cdot \frac{1}{s_b} \cdot \log(N_{s_b}) = \frac{N}{r} \cdot \lambda_b \cdot N s_b \ll 1$$

Alternative interpretation: multiple sweeps cannot occur w/in l^*



$$l^* \sim D_s \log(N_s)$$

of each other in a single fixation time:

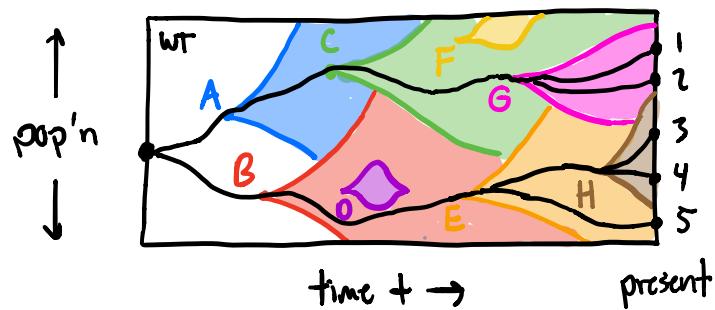
$$N \cdot \mu \lambda_b l^* \cdot s_b \cdot T_{fix} \ll 1$$

$$\Leftrightarrow \frac{N}{r} \cdot \lambda_b \cdot N s_b \ll 1$$

\Rightarrow if $\frac{\mu}{r} \sim O(1) \Rightarrow$ need $\lambda_b \ll \gamma_{NS_b} \ll 1$

\Rightarrow will always break down in sufficiently large pop'n's!

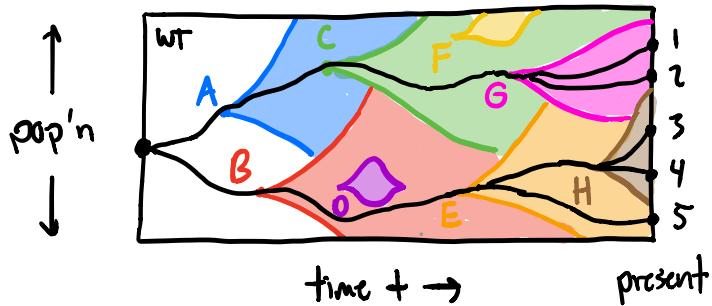
\Rightarrow Back to clonal interference regime!



\Rightarrow Finally time to consider in detail...

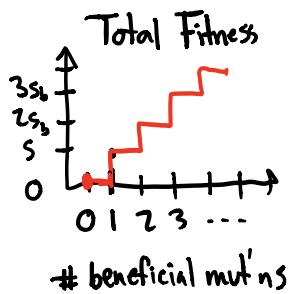
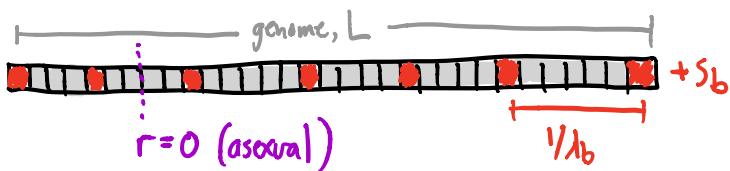
Linked selection & clonal interference

(a.k.a. "Hill-Robertson Interference")



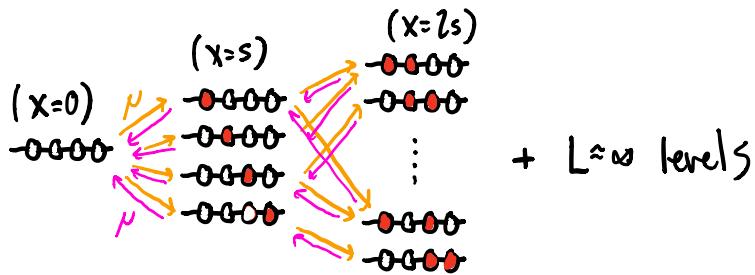
- ⇒ can't be reduced to $L=1$ or $L=2$ model (collective phase)
- ⇒ Most progress only recently, w/ big contribution from physicists
[e.g. Tsimring et al PRL '96, Rouzine et al '03, Desai + Fisher '07, ...]
- ⇒ Analytical progress enabled by starting w/ very simple model:

"Staircase" Model



- ① All mutations provide same benefit (s_b)
- ② Occur @ total rate $U_b \equiv L\lambda_b N$
- ③ Never run out (e.g. $L\lambda_b \rightarrow \infty$, $N \rightarrow 0$)

Genotype network:

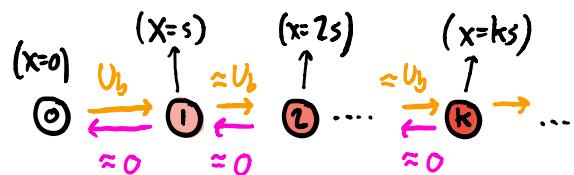


+ $L \approx \infty$ levels

Key simplification:

"fitness class"

$$f_k \equiv \sum_{|\vec{g}|=k} f(\vec{g})$$



\Rightarrow coarse-grained SDE (1+1 dimensional vs $2^L + 1$ dim.)

$$\frac{df(k)}{dt} = S_b(k - \bar{k}(t))f(k) + U_b[f(k-1) - f(k)]$$

selection (nonlinear) mutation

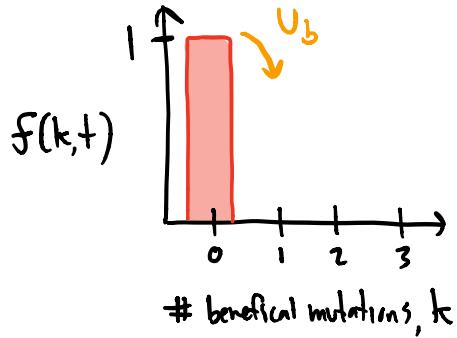
$$+ \sqrt{\frac{f(k)}{N}} \eta(k) - f(k) \sum_{k'} \sqrt{\frac{f(k')}{N}} \eta(k')$$

genetic drift
(stochastic)

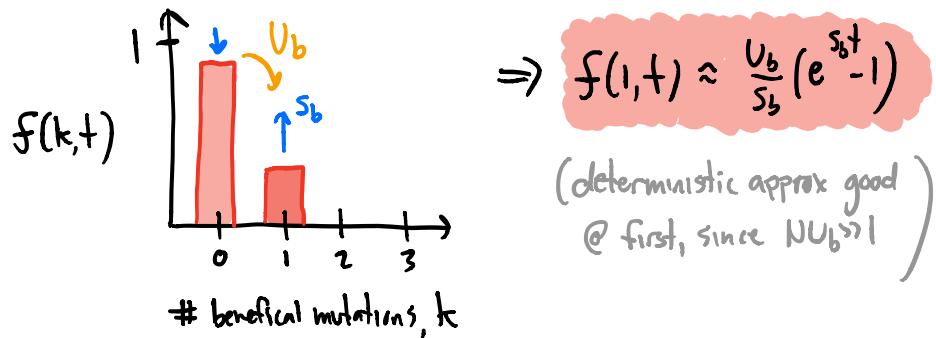
\Rightarrow let's consider behavior when $N S_b \gg N U_b \gg 1$

(e.g. yeast barcode experiment in HW 4 Problem #1)

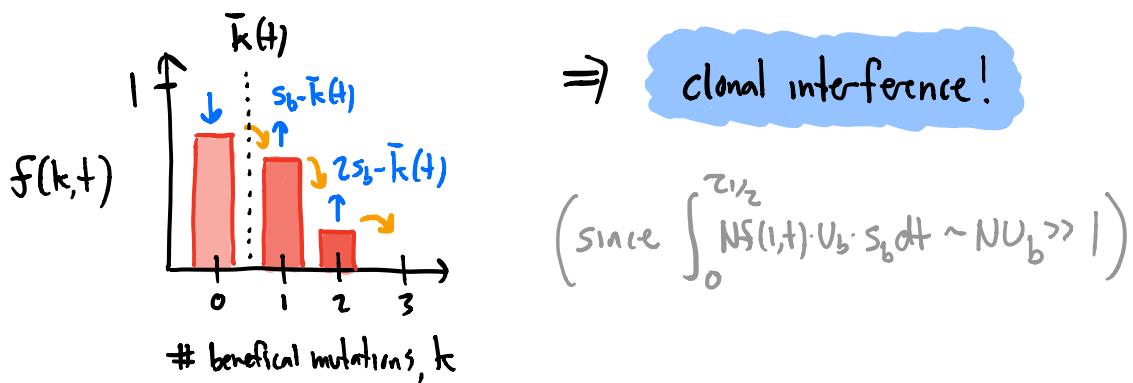
① Start w/ wildtype population @ $t=0$



② First-step mutations ($k=1$) establish & grow exponentially



③ Double mutants establish before single mutants take over,



\Rightarrow Is deterministic approx still useful?

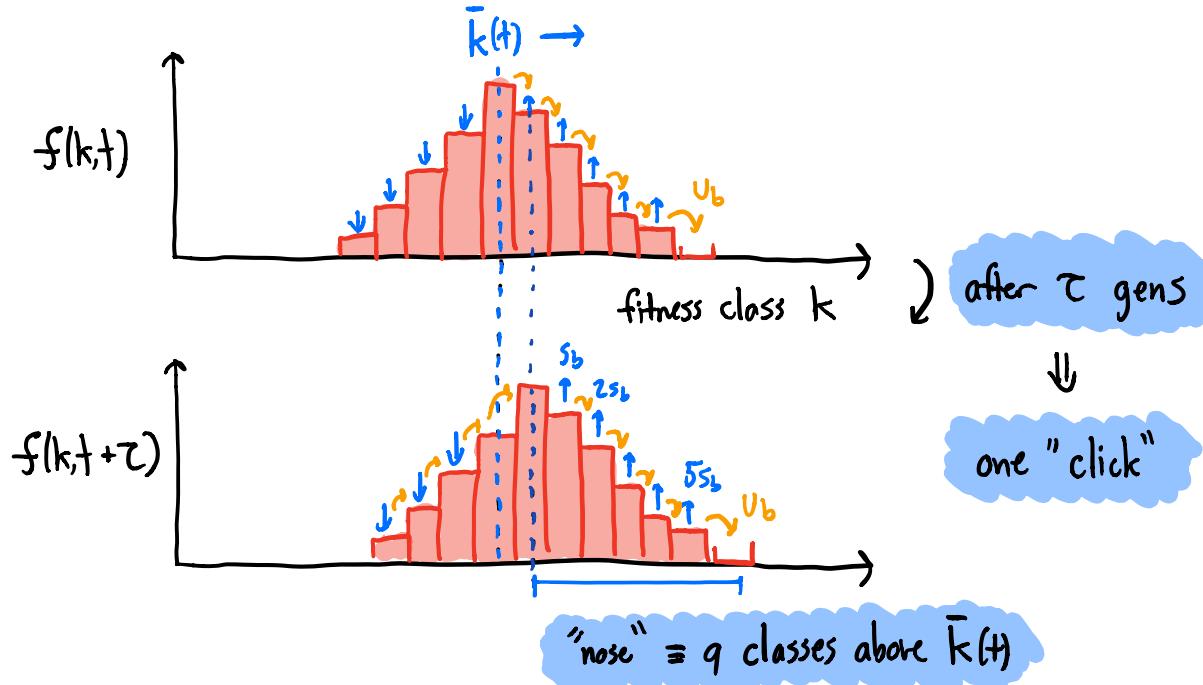
$$\frac{df(k)}{dt} = s_b(k - \bar{k}(t))f(k) + u_b[f(k-1) - f(k)] + \sqrt{\frac{f(k)}{N}}\eta(k) - f(k)\sum_k \sqrt{\frac{f(k')}{N}}\eta(k') \xrightarrow{0}$$

selection (nonlinear) mutation genetic drift

\Rightarrow can show: $f_{det}(k,t) = \frac{1}{k!} \left[\frac{u_b}{s_b} (e^{s_b t} - 1) \right]^k \cdot e^{-\frac{u_b}{s_b} (e^{s_b t} - 1)}$

\Rightarrow Not self-consistent! \Rightarrow Predicts $s_b \bar{k}(t) \approx u_b e^{s_b t}$
 (eventually all $f(k,t) \ll 1/N$!)

\Rightarrow Instead, if we simulate model, observe "travelling wave":



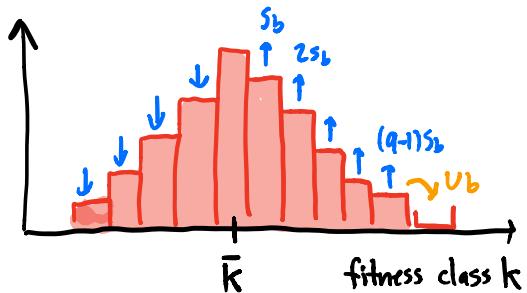
\Rightarrow What determines $\tau(N, s_b, u_b) + q(N, u_b, s_b)$?

\Rightarrow Today: heuristic analysis [\sim Desai + Fisher 2007]

applies when: $Ns_b \gg Nu_b \gg 1$ + $s_b^2 \gg 1$ & $q \gg 1$

Leads to simplifications:

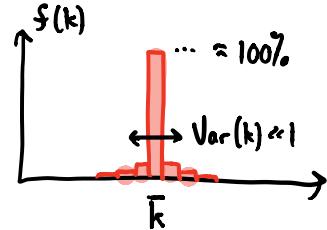
- ① mutations only important for establishing new "nose"
(since $s_b \gg u_b$)



② Genetic drift only important for establishing new noise

(since $\tau \gg 1/s_b$, individual mutations establish before next click.)

③ most of pop'n is near $k \approx \bar{k}(t)$

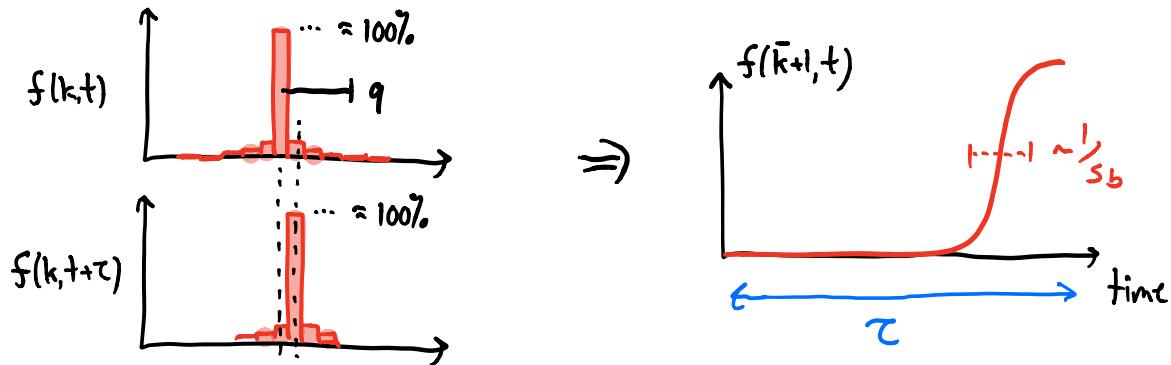


Problem 3 of HW 4:

$$\frac{d\langle \bar{k} \rangle}{dt} = \left\langle \sum_k k \frac{df(k)}{dt} \right\rangle = \underbrace{\left\langle \sum_k s_b (k - \bar{k})^2 f(k, t) \right\rangle}_{s_b \text{Var}(k)}$$

$$\Rightarrow \text{Var}(k) = \frac{1}{s_b \tau} \ll 1 \quad (\text{by assumption})$$

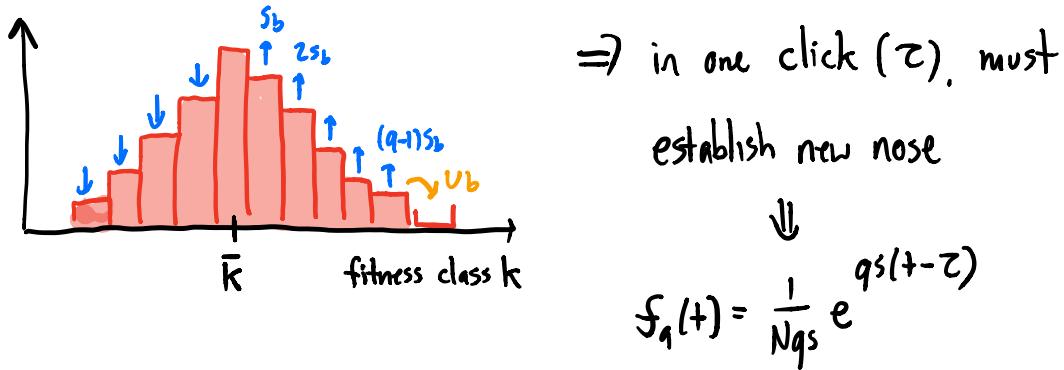
④ Also implies that $\bar{k}(t)$ clicks suddenly:



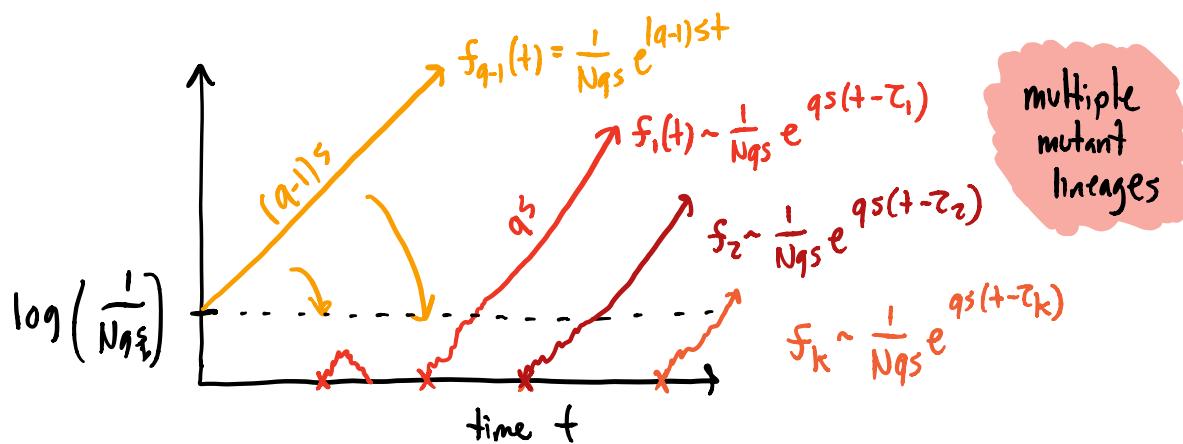
\Rightarrow i.e. for most $t \in [0, \tau]$ $\Rightarrow \bar{k}(t) = \bar{k}(0)$

\Rightarrow everyone grows as $f(k, t) \sim f(k, 0) e^{(k - \bar{k}(0))st}$

\Rightarrow Now we have all ingredients to understand wave:



$\Rightarrow \tau \approx$ establishment time of nose class!



$$\Rightarrow f_q(t) = \sum_{k=0}^{K_{\max}} f_k(t) \equiv \frac{1}{Nq^s} e^{qs(t-\tau)}$$

↓

Establishment time
for whole class.

$\Rightarrow k^{\text{th}}$ successful mutant establishes when:

$$\int_0^{\tau_k} N U_b \cdot f_{q-1}(t) \cdot q s_b dt \sim O(k)$$

Note: extra little bit will be important below!

$$\int_0^{\tau_k} N U_b \cdot \frac{1}{Nq^s} e^{(q-1)s_b t} \cdot q s_b dt = \frac{U_b}{q s_b} e^{(q-1)s_b \tau_k} \sim O(k)$$

$$\Rightarrow \tau_k = \frac{1}{(q-1)s_b} \log \left(\frac{s_b}{U_b} \cdot q \cdot k \right)$$

$$\Rightarrow \text{Note: } \tau_k = \underbrace{\frac{1}{(q-1)s_b} \log \left(\frac{s_b}{U_b} \cdot q \right)}_{\tau_1} + \underbrace{\frac{1}{(q-1)s_b} \log(k)}_{\tau_k - \tau_1}$$

(most time spent waiting for first mut'n)

\Rightarrow many mutations establish in quick succession ($\delta t \sim \frac{1}{qs_b} \ll \tau$)

\Rightarrow Typical size of k^{th} lineage:

$$\Rightarrow f_k(t) = \frac{1}{Nqs} e^{qs(t-\tau_k)} = \frac{e^{qst}}{Nqs} \left(\frac{s_b q k}{U_b} \right)^{-1 - \frac{1}{q}}$$

important!

\Rightarrow Size of entire nose class:

$$\Rightarrow f_q(t) = \sum_{k=1}^{K_{\text{nose}}} f_k(t) = \frac{1}{Nqs} e^{qst} \left(\frac{s_b}{U_b} \right)^{-1 - \frac{1}{q}} \sum_{k=1}^{K_{\text{nose}}} \frac{1}{q} \frac{1}{k^{1 + \frac{1}{q}}}$$

set equal!

$$\Rightarrow \tau = \frac{\left(1 + \frac{1}{q}\right)}{qs_b} \log \left(\frac{s_b}{U_b} \right) \approx \frac{1}{qs_b} \log \left(\frac{s_b}{U_b} \right)$$

vs $\tau_k = \frac{1}{qs_b} \log \left(\frac{s_b}{U_b} \cdot q \cdot k \right)$

$\left[\begin{array}{l} \tau < \tau_k \text{ because} \\ \text{multiple mutations} \\ \text{contribute } \neq \text{once} \end{array} \right]$

\Rightarrow how to determine q ?

\Rightarrow follow new noise over time:

$$f(\bar{k}+q, \tau) \approx \frac{1}{Nq\varsigma_b} e^{\frac{2S(1-q)}{Nq\varsigma_b}} \rightarrow \frac{1}{Nq\varsigma_b} e^{\frac{2S(1-q)}{Nq\varsigma_b}} \rightarrow \dots$$

(right after est.)

\Rightarrow After q clicks, old noise is new mean! (majority of pop'n)

$$f(q\tau) \sim \frac{1}{Nq\varsigma_b} e^{-\frac{q^2 S^2 \tau}{2}} \sim O(1)$$

\Rightarrow system of 2 eqs for $\tau + q$:

$$\frac{q^2 S^2 \tau^2}{2} \approx \log(Nq\varsigma_b)$$

$$\tau = \frac{1}{q\varsigma_b} \log\left(\frac{S^2}{\varsigma_b^2}\right)$$

$$\Rightarrow \text{solution: } q = \frac{2 \log(Ns_b)}{\log\left(\frac{s_b}{v_b}\right)} ; \quad \tau = \frac{1}{2s_b} \frac{\log^2\left(\frac{s_b}{v_b}\right)}{\log(Ns_b)}$$

$$\Rightarrow \left\langle \frac{\partial \bar{x}}{\partial t} \right\rangle = \frac{s_b}{\tau} = \frac{2s_b^2 \log(Ns_b)}{\log^2(s_b/v_b)}$$

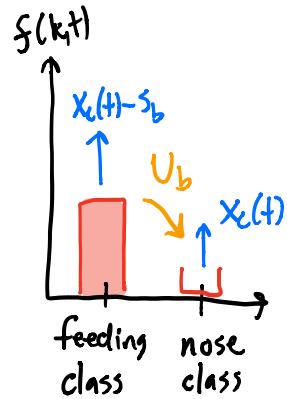
(compare to $\sim Nv_b s_b^2$ in *successive mutations regime*)

\Rightarrow Self consistency: $s\tau \gg 1 + q \gg 1$

$$\Rightarrow \log\left(\frac{s_b}{v_b}\right) \ll \log(Ns_b) \ll \log^2\left(\frac{s_b}{v_b}\right)$$

Supplement: Formal analysis of the nose class

⇒ we can understand the establishment of the nose class more formally using the branching process framework that we studied in the 1st half of the course



⇒ Under our assumptions, nose can be described by LBP model:

$$\frac{df_q}{dt} = X_c(t)f_q + U_b f_{q-1}(t) + \sqrt{\frac{f_q}{N}} \eta(t)$$

w/ $f_q(0)=0$ & time-varying:

selection: $X_c(t) = [q - \bar{k}(t)]s_b$

+
mutation: $U_b f_{q-1}(t) = \frac{U_b}{2Nqs_b} e^{\int_0^t (X_c(t') - s_b) dt'}$

\Rightarrow In their analysis, Desai & Fisher (2007) assumed that $\bar{k}(t) \approx 0$ throughout the establishment period, so that $X_c(t) \equiv q s_b$ + $f_{q-1}(t) = \frac{1}{Nq} e^{(q-1)s_b t}$

\Rightarrow Let's see how far we can get by relaxing this approx & explicitly modeling the "click" of $\bar{k}(t)$...

\Rightarrow will be harder because time-varying fitness

$$X_c(t) = q s_b - s_b \bar{k}(t)$$

\Rightarrow From our discussion in class, can take

$$\bar{k}(t) \equiv \frac{e^{s_b(t-t_c)}}{1 + e^{s_b(t-t_c)}}$$

where t_c is the time that $\bar{k}(t)$ clicks.

(later we will imagine that t_c is close to T ...)

\Rightarrow From SDE, the generating function $H_f(z, t) = \langle e^{-z \cdot f_1(t)} \rangle$ satisfies the PDE:

$$\frac{\partial H_f}{\partial t} = \left[X_t(t) z - \frac{z^2}{2N} \right] \frac{\partial H_f}{\partial z} - z v_b f_{q-1}(t) H_f$$

w/ initial condition $H_f(z, 0) = 1$

\Rightarrow can solve w/ method of characteristics:

$$\text{define: } \psi(t_R) = \log [H_f(z(t_R), t - t_R)]$$

$$\text{w/ } \psi(t) = 0, \quad z(0) = z,$$

$$\psi(t) \equiv \log H_f(z, t)$$

$$\Rightarrow \psi \text{ satisfies: } \frac{d\psi}{dt_R} = - \frac{\partial H_f}{\partial t} + \frac{\partial H_f}{\partial z} \left(\frac{\partial z}{\partial t_R} \right)$$

$$\Rightarrow \frac{dy}{dt_k} = \left\{ \frac{dz}{dt_k} - \left[x_c(t-t_k)z - \frac{z^2}{2N} \right] \right\} \frac{d \log H_f}{dz} + z(t_k) v_b f_{q-1}(t-t_k)$$

$$\Rightarrow \text{if } \frac{dz}{dt_k} = x_c(t-t_k)z - \frac{z^2}{2N} \quad \text{and} \quad z(0) = z$$

$$\Rightarrow y(t_k) = y(0) + \int_0^{t_k} z(t_k') v_b f_{q-1}(t-t_k') dt_k'$$

$$\Rightarrow \log H_f(z,t) = - \int_0^t z(\tau) v_b f_{q-1}(t-\tau) d\tau$$

where

$$\frac{dz}{dt} = x_c(t-t)z - \frac{z^2}{2N}, \quad z(0) = z$$

\Rightarrow solution for characteristic curve is given by:

$$z(t) = \frac{ze^{\int_0^t x_c(t-t') dt'}}{1 + \frac{z}{2N} \int_0^t e^{\int_0^{t'} x_c(t-t'') dt''} dt'}$$

(can plug in & check...)

$$\text{so } H_f(z, t) = \exp \left[- \int_0^+ \frac{z \cdot u_b f_{q-1}(t-u) e^{\int_0^{t-u} x_c(u') du'}}{1 + \frac{z}{2N} \int_u^+ e^{\int_{u'}^{t-u} x_c(u'') du''}} du \right]$$

$$= \exp \left[- \int_0^+ \frac{z \cdot u_b f_{q-1}(u) e^{\int_u^+ x_c(u') du'}}{1 + \frac{z}{2N} \int_u^+ e^{\int_{u'}^{t-u} x_c(u'') du''}} du \right]$$

$$\Rightarrow \text{again, helpful to define } v(t) \text{ s.t. } f_q(t) \equiv \frac{v(t)}{2Nq s_b} e^{\int_0^t x_c(t') dt'}$$

$$\Rightarrow H_v(z, t) \equiv \langle e^{-z \cdot v(t)} \rangle = H_f(2Nq s_b e^{-\int_0^t x_c(t') dt'}, z, t)$$

$$\Rightarrow H_v(z, t) = \exp \left[- \int_0^+ \frac{z \cdot u_b f_{q-1}(u) 2Nq s_b e^{-\int_0^u x_c(u') du'}}{1 + q s_b z \cdot \int_u^+ e^{-\int_{u'}^{t-u} x_c(u'') du''}} du \right]$$

\Rightarrow Similar to single-locus case, we expect $v(t)$ to approach constant value v @ long times

$$\Rightarrow H_v(z) \equiv \lim_{t \rightarrow \infty} H_v(z, t)$$

$$\Rightarrow \log H_v(z) = - \int_0^\infty \frac{z \cdot U_b e^{-s_b t}}{1 + z \cdot q s_b \int_t^\infty e^{s_b t'} x(t') dt'} dt$$

\Rightarrow Now we have to plug in our expression for $x_c(t)$:

$$x_c(t) = q s_b - \frac{s_b e^{s_b(t-t_c)}}{1 + e^{s_b(t-t_c)}}$$

$$\Rightarrow \int_0^t x_c(t') dt' = q s_b t - \log \left[\frac{1 + e^{s_b(t-t_c)}}{1 + e^{-s_b t_c}} \right]$$

$$\Rightarrow e^{-\int_0^t x_c(t') dt'} = e^{-q s_b t} \left[\frac{1 + e^{s_b(t-t_c)}}{1 - e^{-s_b t_c}} \right]$$

$$\Rightarrow q s_b \int_t^\infty e^{-\int_0^t x_c(t') dt'} e^{-s_b t''} dt'' = \frac{e^{-q s_b t}}{1 + e^{-s_b t_c}} + \left(\frac{q}{q-1} \right) \frac{e^{-q s_b t} e^{s_b(t-t_c)}}{1 + e^{-s_b t_c}}$$

and hence :

$$\log H_V(z) = - \int_0^\infty \frac{z \cdot U_b e^{-S_b t} dt}{1 + z \cdot e^{-qS_b t} \left[1 + e^{S_b(t-t_c)} \left(\frac{q}{q-1} \right) \right]}$$

(where we have assumed that the click time t_c is $\gg \frac{1}{S_b}$)

\Rightarrow for large q & relevant values of z , this integral will be dominated by times w/in $O(\frac{1}{S_b})$ of τ .

\Rightarrow can extend lower limit of integral to $+ = -\infty$
 w/o much error...

\Rightarrow if t_c is also w/in $O(\frac{1}{S_b})$ of τ , we can expand $e^{S_b(t-t_c)}$ term in denominator, so that

$$\log H_V(z) \approx - \int_{-\infty}^{\infty} \frac{z \cdot U_b e^{-S_b t} dt}{1 + 2 \cdot z \cdot e^{-qS_b t}}$$

changing variables to $\xi = (2z)^{\frac{1}{q}} e^{-\frac{s_b t}{2}}$, $1 + \Theta(\frac{1}{q})$

$$\log H_v(z) = \exp \left[-\frac{U_b}{S_b} \cdot z^{1-\frac{1}{q}} \cdot \left(2^{\frac{1}{q}} \int_0^\infty \frac{\xi d\xi}{1+\xi^q} \right) \right]$$

$$\Rightarrow H_v(z) = e^{-\frac{U_b}{S_b} z^{1-\frac{1}{q}}}$$

\Rightarrow typical value of v occurs when $H_v(z = \frac{1}{v^*}) = e^{-1}$

$$\Rightarrow v^* = \left(\frac{S_b}{U_b} \right)^{\frac{q}{q-1}}$$

$$\Rightarrow \text{substituting into } f_q(t) = \frac{v}{2NqS} e^{qst} \equiv \frac{e^{qs(t-\tau)}}{2NqS}$$

$$\Rightarrow \text{typical value of } f_q^*(t) = \frac{e^{s_b t}}{NqS} \left(\frac{S_b}{U_b} \right)^{\frac{q}{q-1}}$$

\Rightarrow typical value of establishment time:

$$\tau^* = \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{v_b}\right)$$

\Rightarrow consistent w/ results from
simpler heuristic argument!