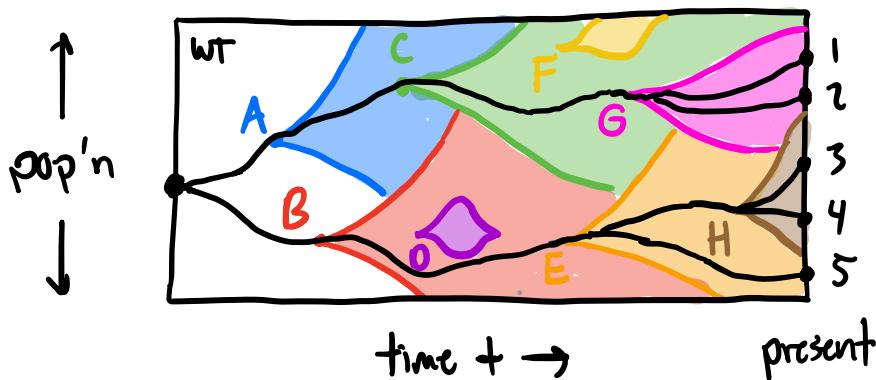


Chapter 15

Linked selection and clonal interference

Linked selection & clonal interference

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



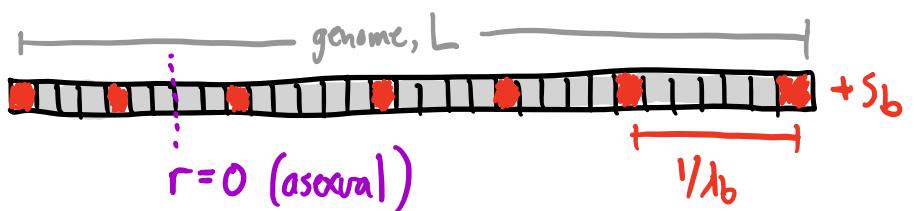
⇒ can't be reduced to $L=1$ or $L=2$ model (collective phase)

⇒ Most progress only recently, w/ big contribution from physicists

[e.g. Tsimring et al PRL '96, Rouzine et al '03, Desai + Fisher '07, ...]

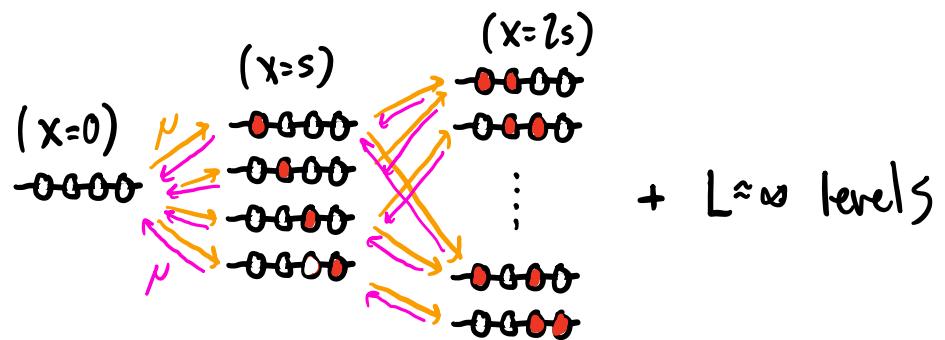
⇒ Analytical progress enabled by starting w/ very simple model:

"Staircase" Model



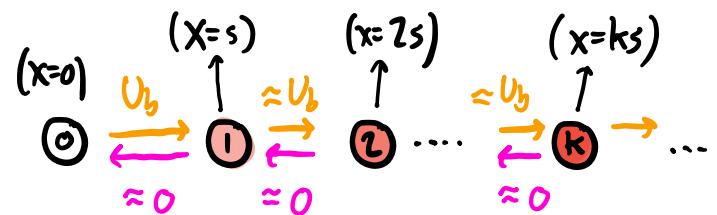
- ① All mutations provide same benefit (s_b)
- ② Occur @ total rate $U_b \equiv L\lambda_b N$
- ③ Never run out (e.g. $L\lambda_b \rightarrow \infty, N \rightarrow 0$)

Genotype network:



Key simplification:

"fitness class" $f(k,t) \equiv \sum_{|\vec{g}|=k} f(\vec{g},t)$



\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

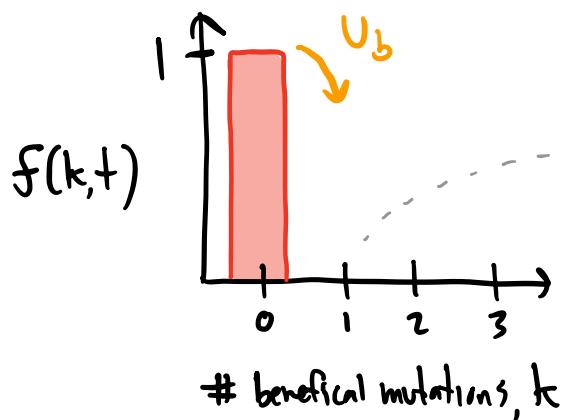
$$+ \sqrt{\frac{f(k)}{N}}\eta(k) - f(k)\sum_{k'} \sqrt{\frac{f(k')}{N}}\eta(k')$$

genetic drift
(stochastic)

\Rightarrow let's consider behavior when $Ns_b \gg NU_b \gg 1$

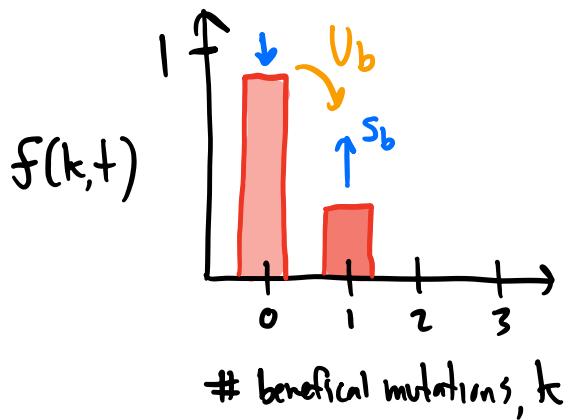
(e.g. yeast barcode experiment in HW 4 Problem #1)

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



$$\frac{df(1)}{dt} \approx Sf(1) + U_b + \sqrt{\frac{f(1)}{N}} \eta_1(t)$$

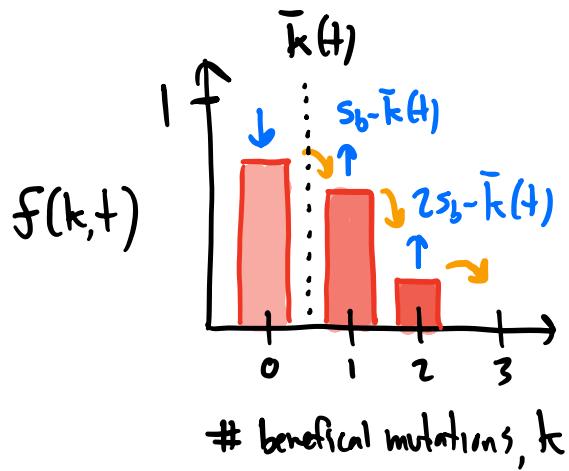
② First-step mutations ($k=1$) establish & grow exponentially



$$\Rightarrow f(1,t) \approx \frac{U_b}{S_b} (e^{S_b t} - 1)$$

(deterministic approx good
@ first, since $NU_b \gg 1$)

③ Double mutants establish before single mutants take over,



\Rightarrow clonal interference!

(since $\int_0^{k(t)} Nf(1,t) \cdot U_b \cdot S_b dt \sim NU_b \gg 1$)

\Rightarrow Is deterministic approx still useful?

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

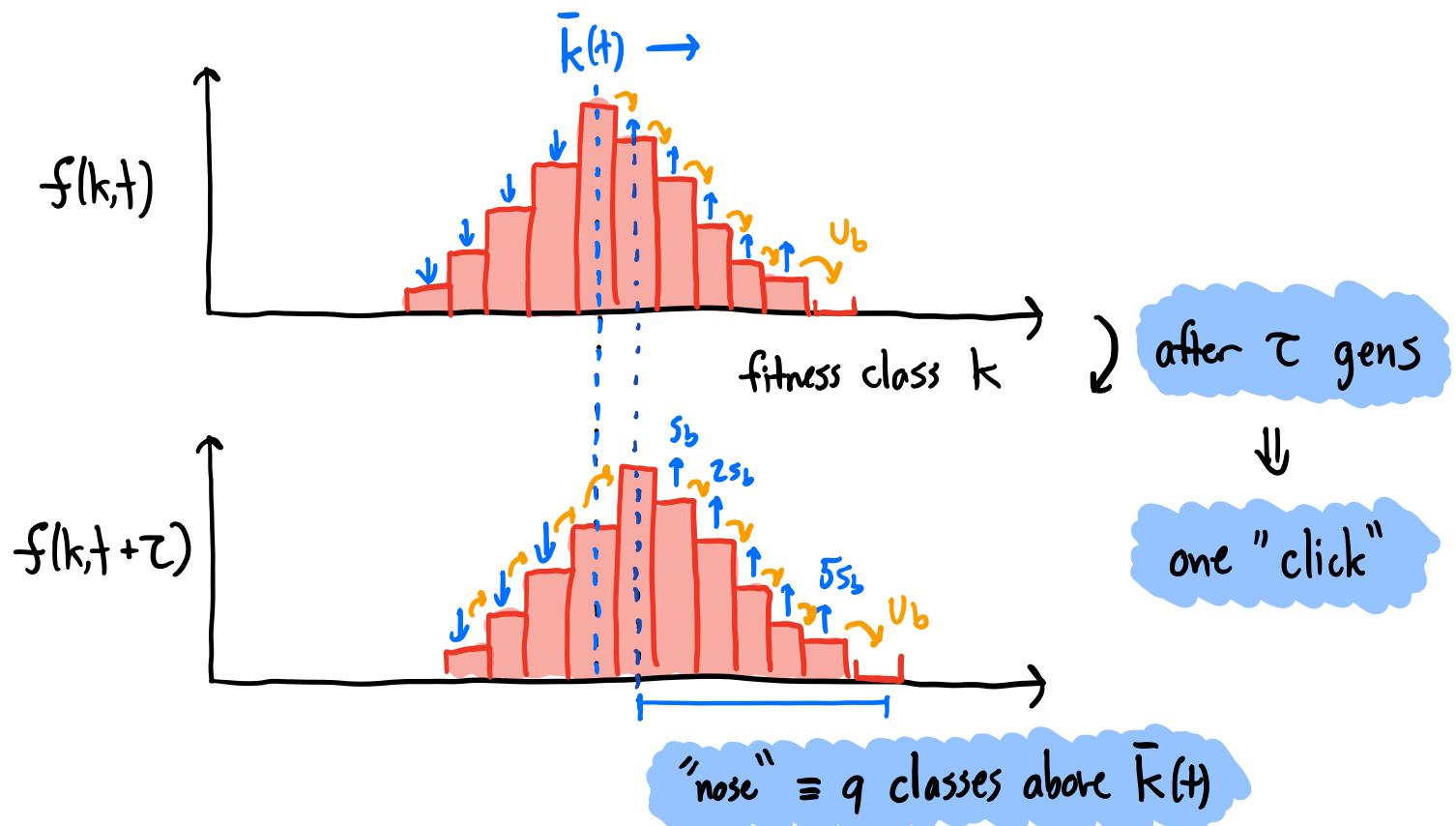
selection (nonlinear) mutation genetic drift $\rightarrow 0$

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

\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 $\tau(N, s_b, U_b) + q(N, U_b, s_b)$?

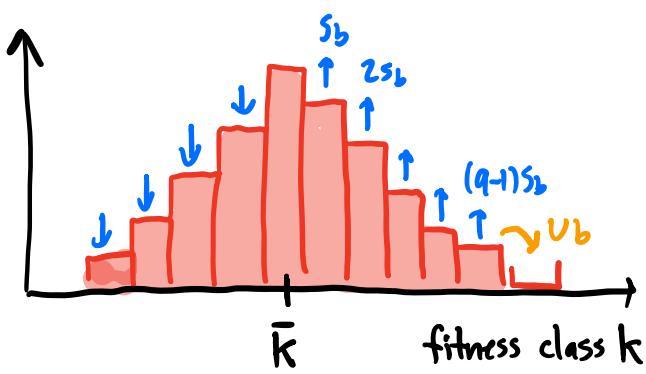
\Rightarrow Today: heuristic analysis [\sim Desai & Fisher 2007]

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"

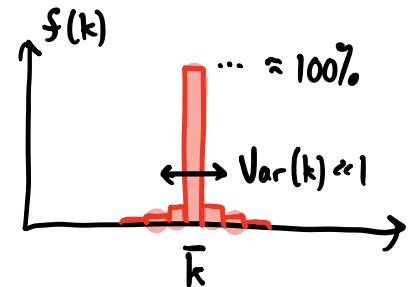
(since $s_b \gg U_b$)



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 $k \approx \bar{k}(t)$

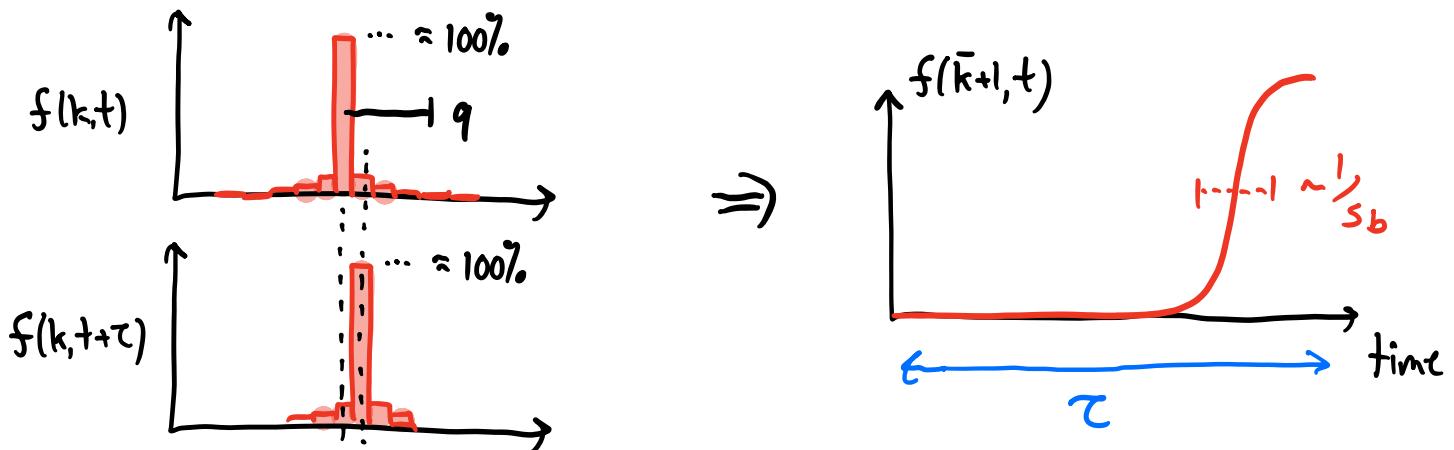


Problem 3 of HW 4:

$$\frac{d\langle \bar{k} \rangle}{dt} = \left\langle \sum_k k \frac{\partial f(k)}{\partial t} \right\rangle = \underbrace{\left\langle \sum_k s_b (k - \bar{k})^2 f(k, t) \right\rangle}_{\equiv 1/\tau}$$

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

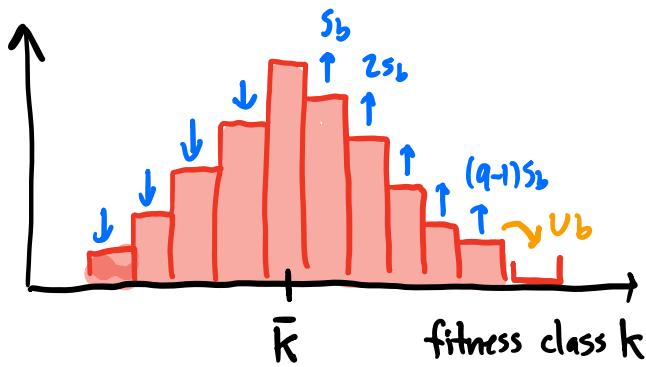
④ Also implies that $\bar{k}(t)$ clicks suddenly:



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

\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:

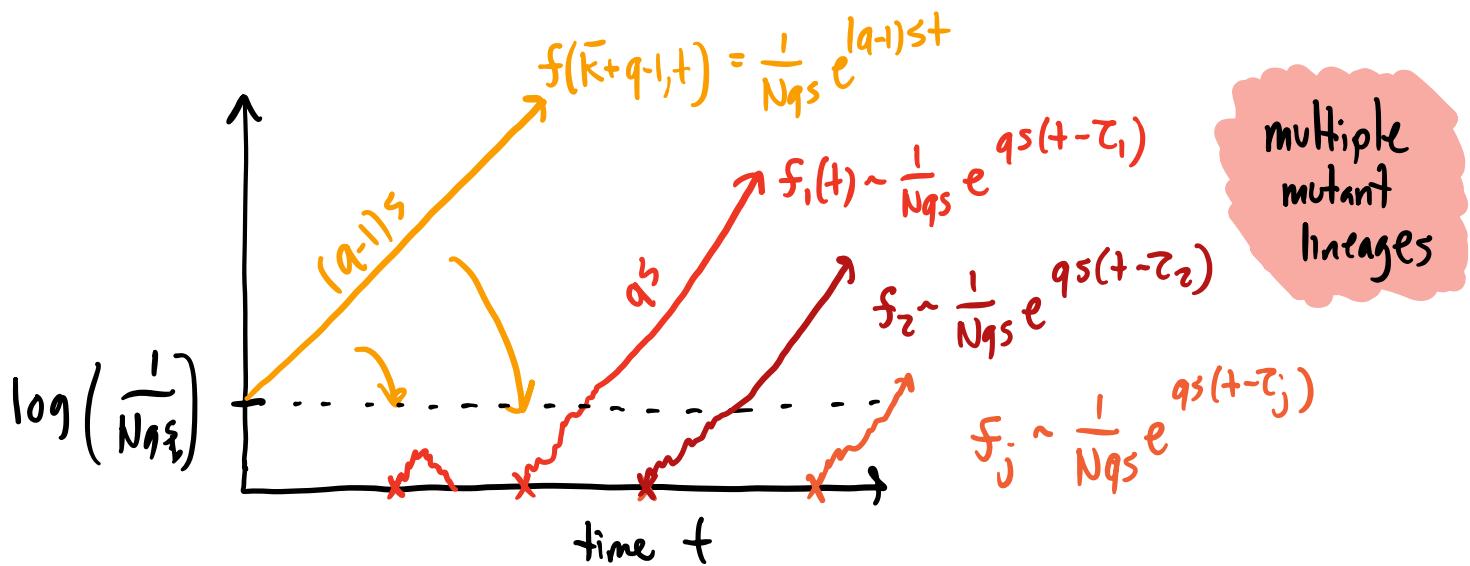


\Rightarrow in one click (τ), must establish new nose



$$f(\bar{k}+q, t) = \frac{1}{Nqs} e^{qs(t-\tau)}$$

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



$$\Rightarrow f(\bar{r} + q, t) = \sum_{j=0}^{J_{\max}} f_j(t) \equiv \frac{1}{Nq^s} e^{qs(\bar{r} - \tau)}$$

↓

establishment time
for whole class.

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

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

Note: extra little bit will be important below!

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

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

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

(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 j^{th} lineage:

$$\Rightarrow f_j(t) = \frac{1}{Nqs} e^{qs(t-\tau_j)} = \frac{e^{qst}}{Nqs} \left(\frac{s_b q j}{U_b} \right)^{-1 - \frac{1}{q-1}}$$

↑ extra bit will be important!

\Rightarrow Size of entire nose class:

$$f(\bar{k}+q, t) = \sum_{j=1}^{J_{\max}} f_j(t) = \frac{1}{Nqs} e^{qst} \left(\frac{s_b}{U_b} \right)^{-\frac{q}{q-1}} \sum_{j=1}^{J_{\max}} \frac{1}{(q \cdot j)^{1 + \frac{1}{q-1}}} \rightarrow 1$$

set equal!

\Rightarrow Time to establish new nose: $\tau = \frac{1}{(q-1)s} \log \left(\frac{s_L}{U_b} \right)$

vs $\tau_j = \frac{1}{(q-1)s_b} \log \left(\frac{s_b}{U_b} \cdot q \cdot j \right)$

Note: $\tau < \tau_j$; b.c.
multiple mutations contribute ρ once

One task remaining... how to determine $q(N, s_b, v_b)$?

\Rightarrow follow new nose over time:

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

(right after est.)

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

$$f(q\tau) \sim \frac{1}{Nqs_b} e^{\frac{-2s + \dots + 2s(2-q) + 2s(1-q)}{Nqs_b}} \sim \frac{1}{Nqs_b} e^{\frac{-q^2 s^2}{2}} \sim O(1)$$

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

$$\frac{q^2 s^2 \tau}{2} \approx \log(Ns_b)$$

+

$$\tau = \frac{1}{qs_b} \log\left(\frac{s_b}{v_b}\right)$$

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

$$\Rightarrow \left\langle \frac{d\bar{x}}{dt} \right\rangle = \frac{s_b}{\tau} = \frac{2s_b^2 \log(Ns_b)}{\log^2(s_b/u_b)}$$

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

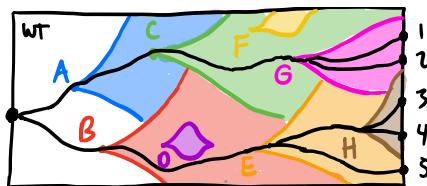
\Rightarrow Self consistency: $s\tau \gg 1 + q \gg 1$

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

Note: used heuristic derivation here...

for formal analysis (using branching processes)
see Appendix A and B below

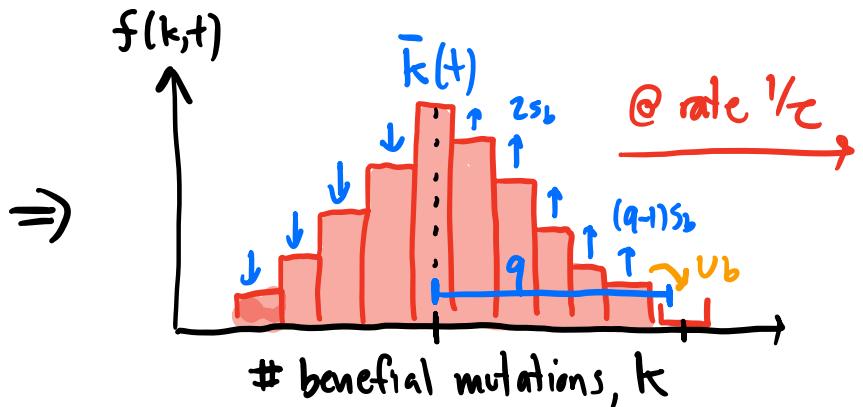
Recap : clonal interference



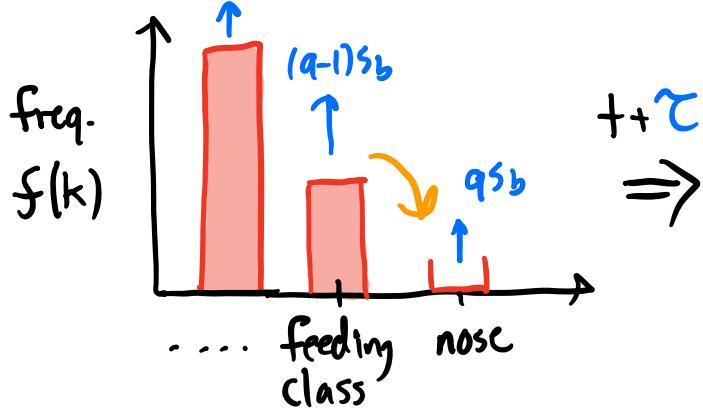
"Staircase" model

← Genome, $L \gg 1$ →

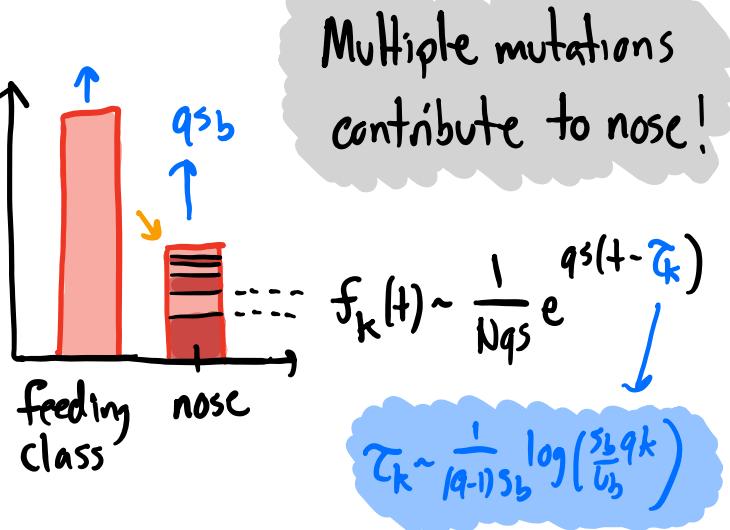
 ↳ selected mut's, + s_b
 total rate $U_b = L\lambda_b N$



key behavior occurs @ "nose":



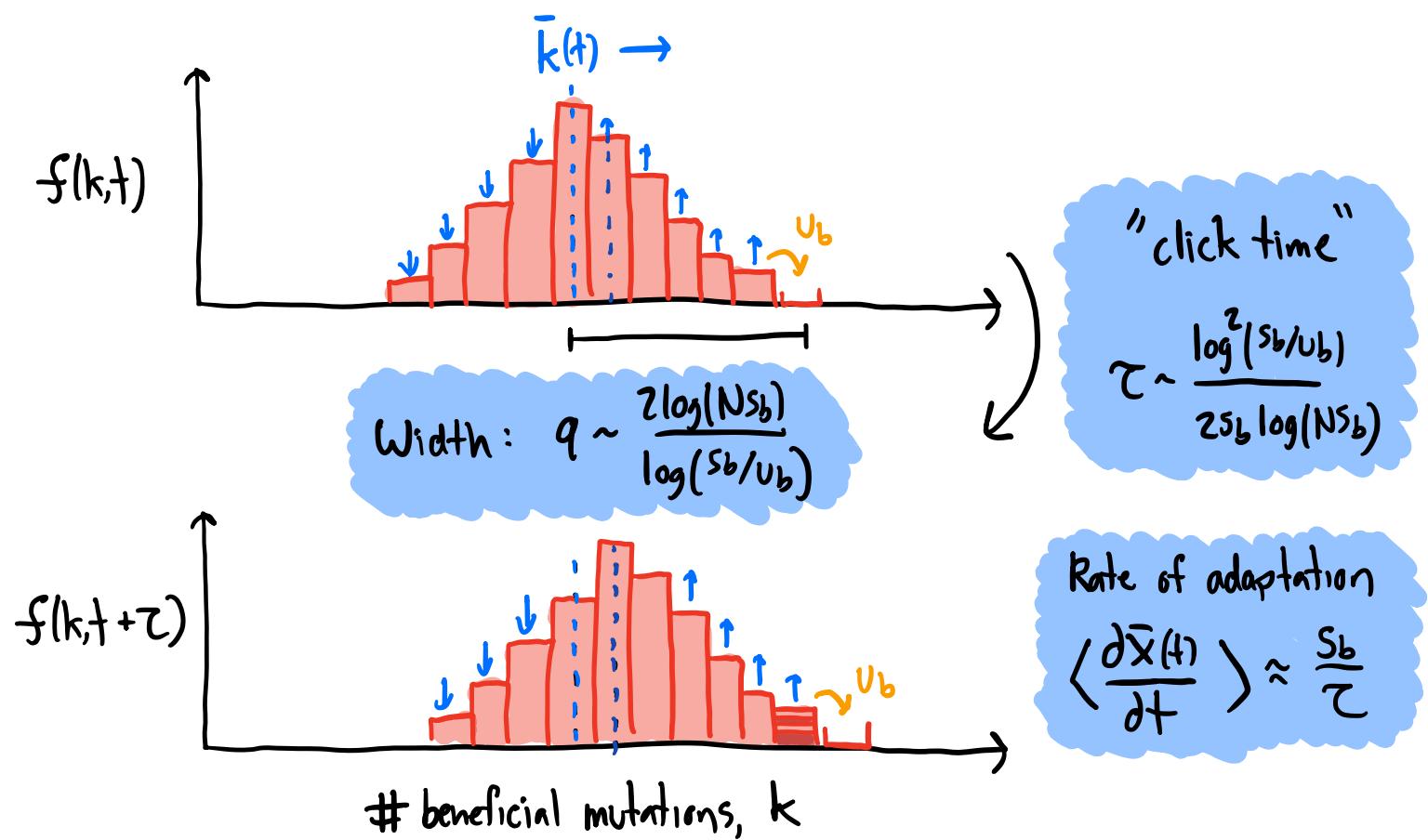
+ τ \Rightarrow



Total contribution:

$$f_{\text{nose}}(t) \equiv \sum_{k=1}^{\infty} f_k(t) \equiv \frac{1}{Nqs} e^{qs(t-\tau)} \Rightarrow \tau \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b}\right)$$

\Rightarrow Complete picture of dynamics of fitness dist'n:



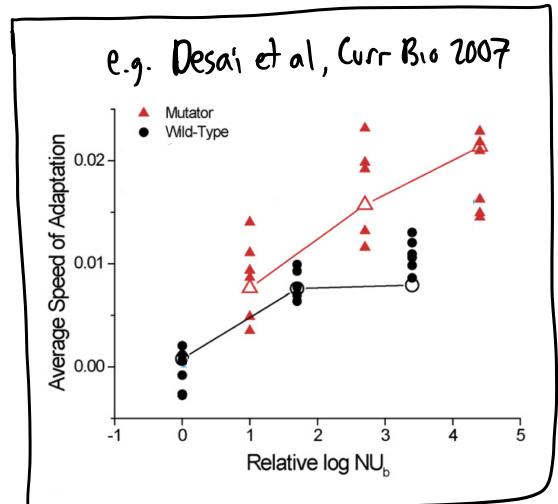
\Rightarrow early tests for clonal interference in lab evolution experiments:

Successive mutations:

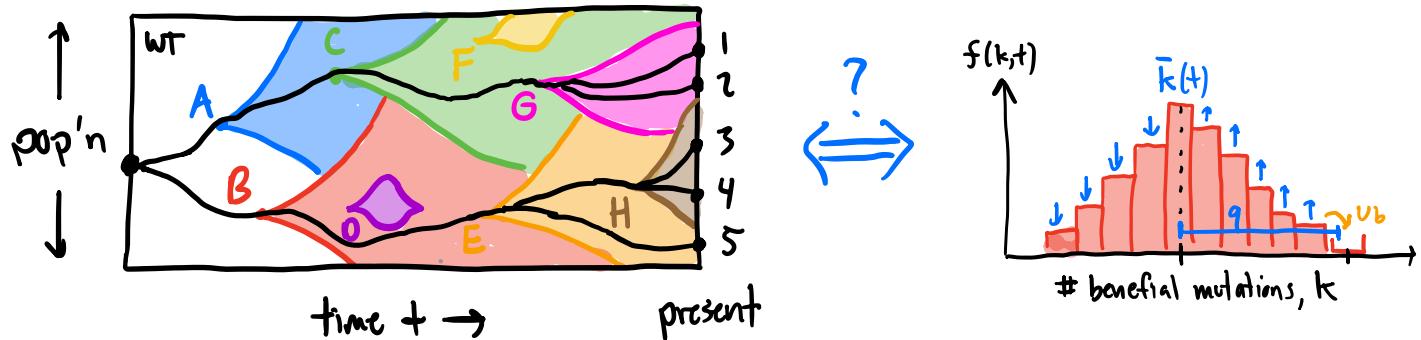
$$\langle \frac{d\bar{x}}{dt} \rangle \sim S_b^2 \cdot N U_b$$

clonal interference:

$$\langle \frac{d\bar{x}}{dt} \rangle \sim S_b^2 \cdot \frac{\log(NS_b)}{\log^2(S_b/U_b)}$$

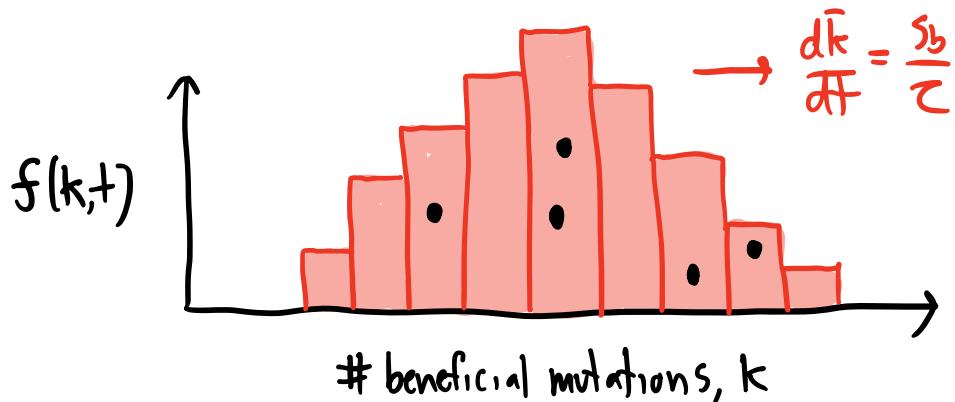


Next: Can we use this picture to understand
genetic diversity backwards in time?

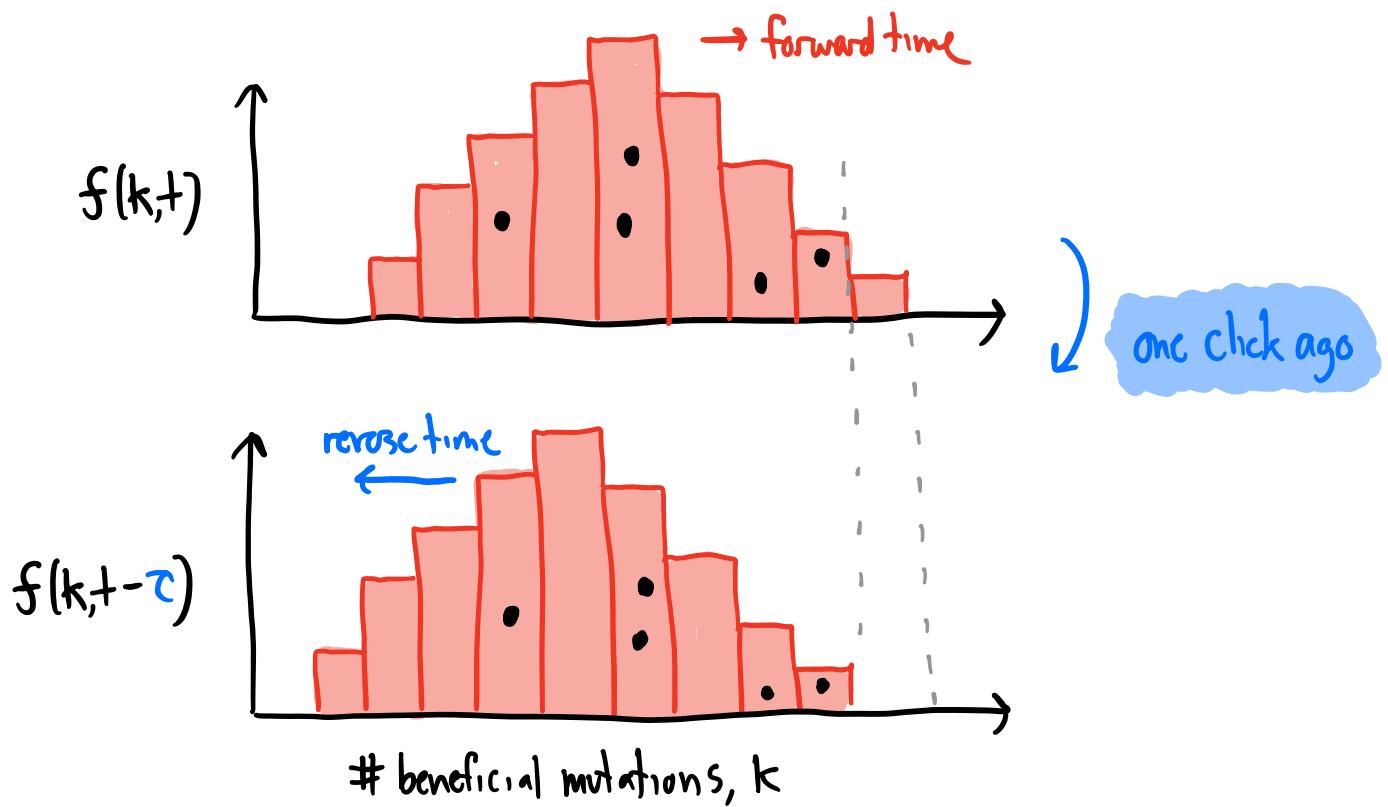


Answer: Yes we can! Let's start w/ some cartoons...

Step 1: draw sample of individuals from pop'n (present day)

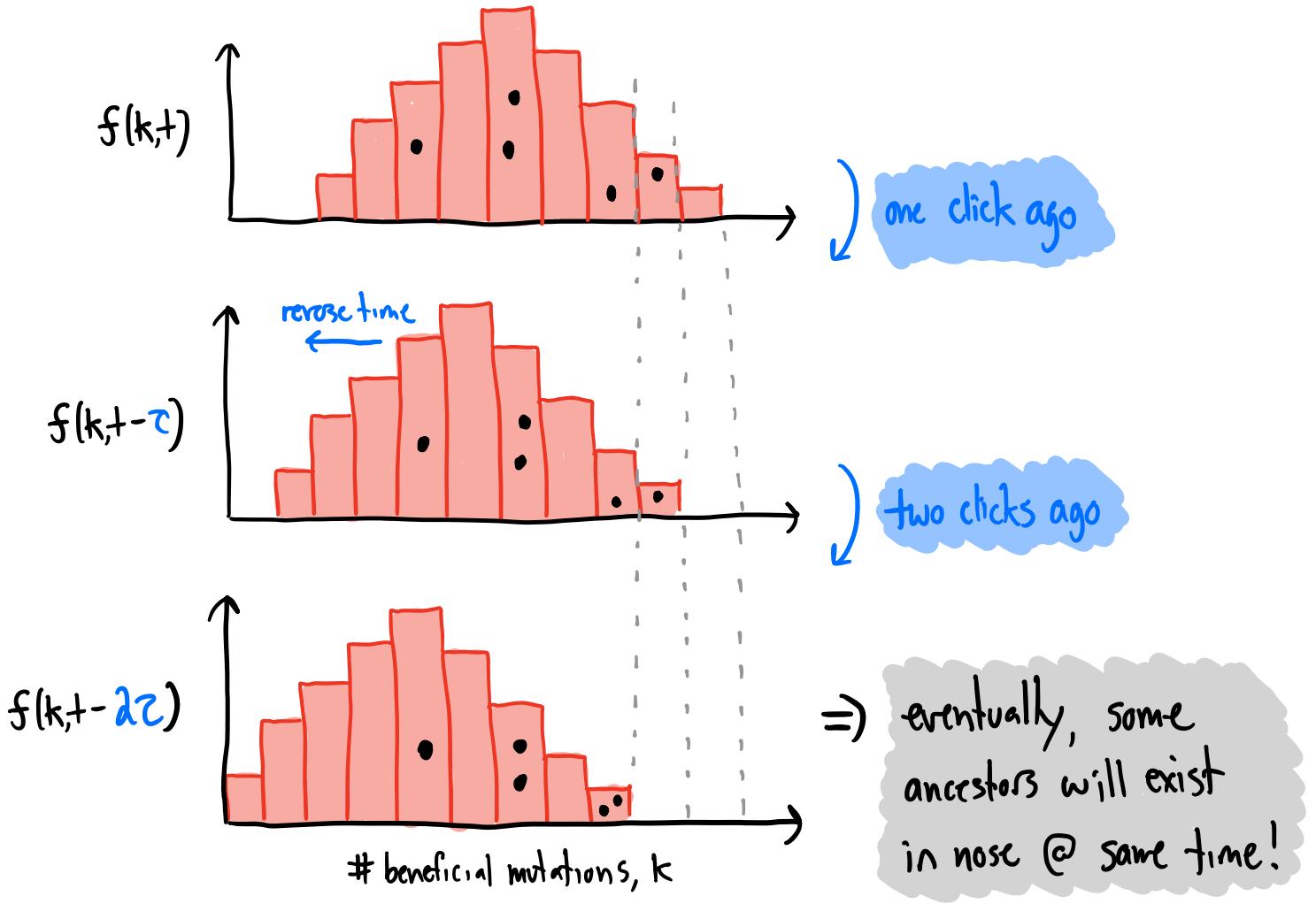


Step 2: where was everyone one click ago?



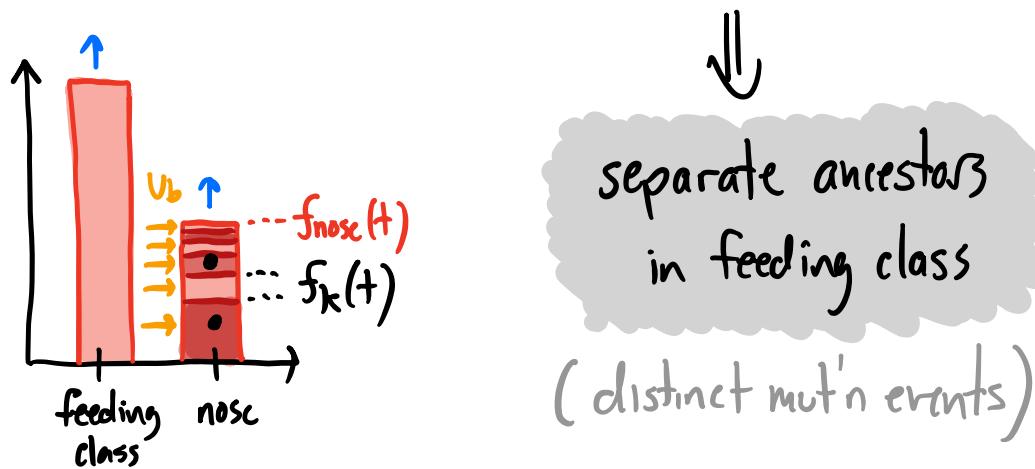
① can only coalesce if in same fitness class

② But little chance of coalescing in "bulk" of dist'n
(since $\tau \ll Nf_{q-1}(\tau), Nf_{q-2}(\tau)$, etc.)

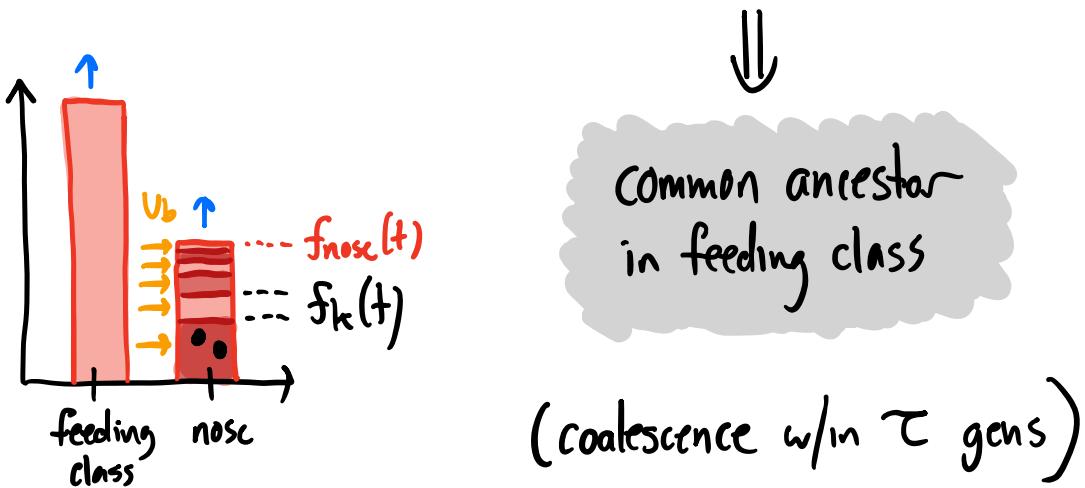


Two possible scenarios:

① Individuals are from separate lineages in the nose



② Individuals from same lineage in nose



$$\Rightarrow \text{Probability: } p_c(\lambda) = \sum_{k=1}^{\infty} \left(\frac{f_k(t)}{f_{\text{nose}}(t)} \right)^2 = \sum_{k=1}^{\infty} \left[\frac{\frac{1}{NqS_b} e^{qS_b(t - \tau_k)}}{\frac{1}{NqS_b} e^{qS_b(t - \tau)}} \right]^2$$

$$= \sum_{k=1}^{\infty} e^{-2qS_b(\tau_k - \tau)}$$



only depends on establishment times τ_k !

\Rightarrow if we plug-in typical values of τ_k & τ from heuristics:

$$\tau_k \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b q k}{U_b}\right); \quad \tau \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b}\right)$$

$$\Rightarrow P_c(z) = \sum_{k=1}^{\infty} e^{-2qs_b(\tau_k - \tau)} = \sum_{k=1}^{\infty} (qk)^{-\frac{2q}{q-1}} \approx \frac{1}{q^2}$$

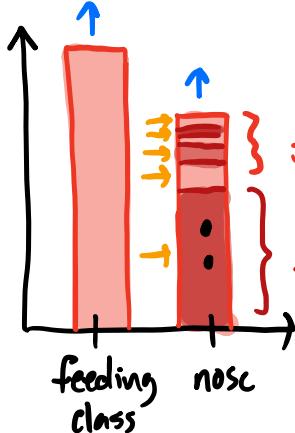
\Rightarrow suggests coalescence after $\sim q^2$ clicks ($T_{MRU} \sim q^2 \tau$)

\Rightarrow missing key part of puzzle: fluctuations

\Rightarrow coalescence rare for typical lineage sizes,

but small chance of having anomalously early mutant
where coalescence is much more likely!

e.g. if first successful mutation occurs when $\tau_1 \leq \tau$:



$$f_{\text{first}} \sim \frac{1}{NqS_b} e^{qS_b(\tau - \tau_1)}$$

$$f_1(\tau) \sim \frac{1}{NqS_b} e^{qS_b(\tau - \tau_1)}$$



$$\frac{f_1(\tau)}{f_{\text{nosc}}(\tau)} \gtrsim O(1)$$



$$P_c(\tau) \gtrsim O(1)!$$

\Rightarrow not a huge shift: typically, $\tau_1 - \tau \sim \frac{\log(q)}{qS_b} \ll \frac{1}{S_b} \ll \tau$

\Rightarrow occurs w/ total probability:

$$P_{\text{jackpot}} \sim \int_0^{\tau} d\tau_1 N U_b f_{q-1}(\tau) \cdot q S_b$$

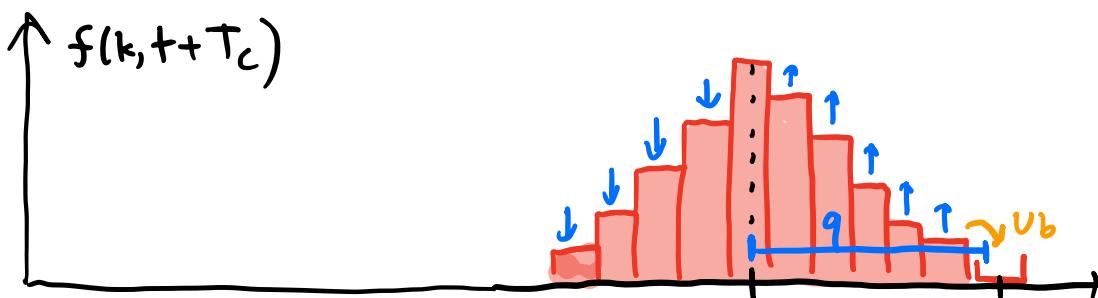
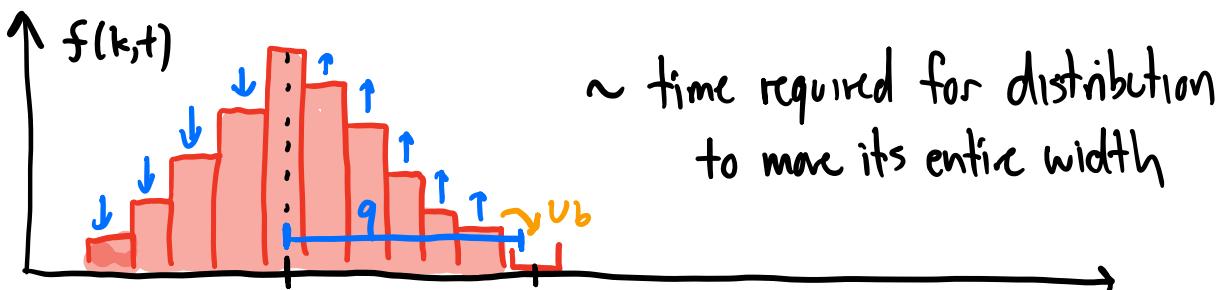
$$\sim \int_0^{\tau} d\tau_1 N U_b \cdot \frac{e^{(q-1)S_b \tau}}{NqS_b} \cdot \cancel{qS_b} \sim \frac{U_b}{(q-1)S_b} e^{(q-1)S_b \tau}$$

$$\sim \frac{1}{q-1} \approx \frac{1}{q}$$

$$\Rightarrow P_{\text{jackpot}} \sim \frac{1}{q} \quad (\gg \frac{1}{q^2})$$

\Rightarrow typical coalescence after $\frac{1}{P_{\text{jackpot}}} \sim q$ clicks

\Rightarrow coalescent timescale $T_c \equiv q\tau - \frac{1}{S_b} \log\left(\frac{S_b}{U_b}\right)$



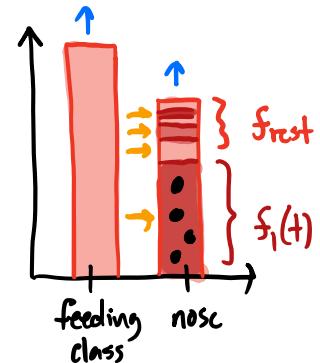
beneficial mutations, k

\Rightarrow fluctuations were crucial for determining T_c !

\Rightarrow coalescence is "bursty":

e.g. in larger sample size n :

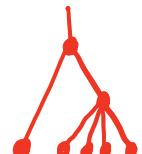
$$P_c(n \rightarrow 1) \sim \left(\frac{f_1(t)}{f_1(t) + f_{rest}} \right)^n$$



$$\Rightarrow O(1) \text{ if } f_{rest}(t) / f_1(t) \lesssim 1/n$$

$$\Rightarrow P_{\text{jackpot}}(n) = \int_0^{\tau - \log(n)/q s_b} d\tau, N V_b f_{q-1}(t) \cdot q s_b \sim 1/q n$$

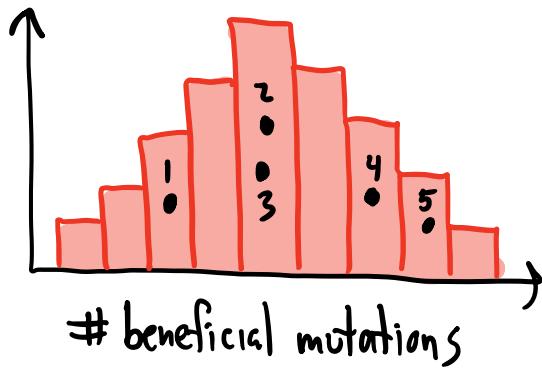
\Rightarrow i.e. multiple mergers likely!



* For "formal" treatment, see Appendix C ...

Another interesting feature of genealogies + travelling wave:

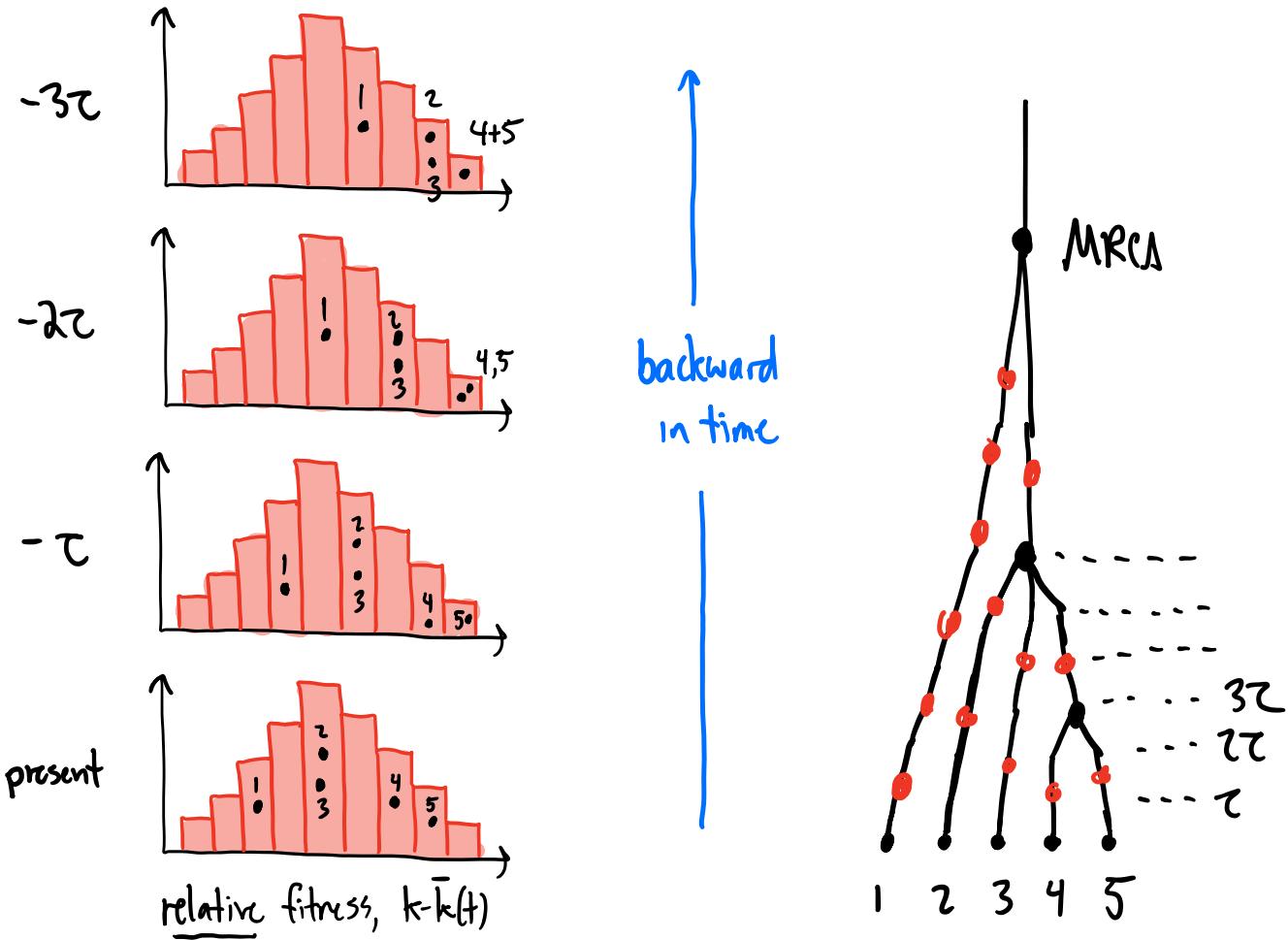
⇒ consider same example:



⇒ which individual's descendants are more likely to take over pop'n in future?

⇒ e.g. $5 \rightarrow 4 \rightarrow 2, 3 \rightarrow 1$

⇒ now let's try to "simulate" genealogy...



\Rightarrow time (+burstiness) of coalescence in past

\Rightarrow info about fitness in present

\Rightarrow forecasts about who takes over in future!

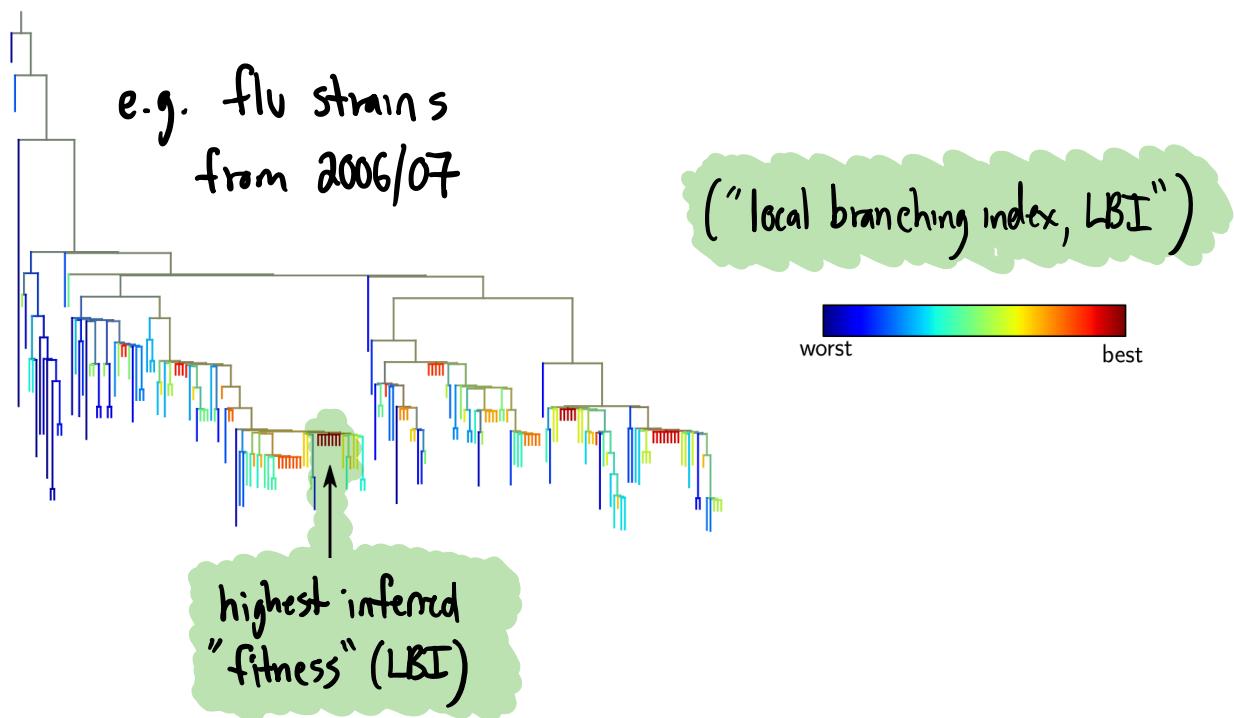


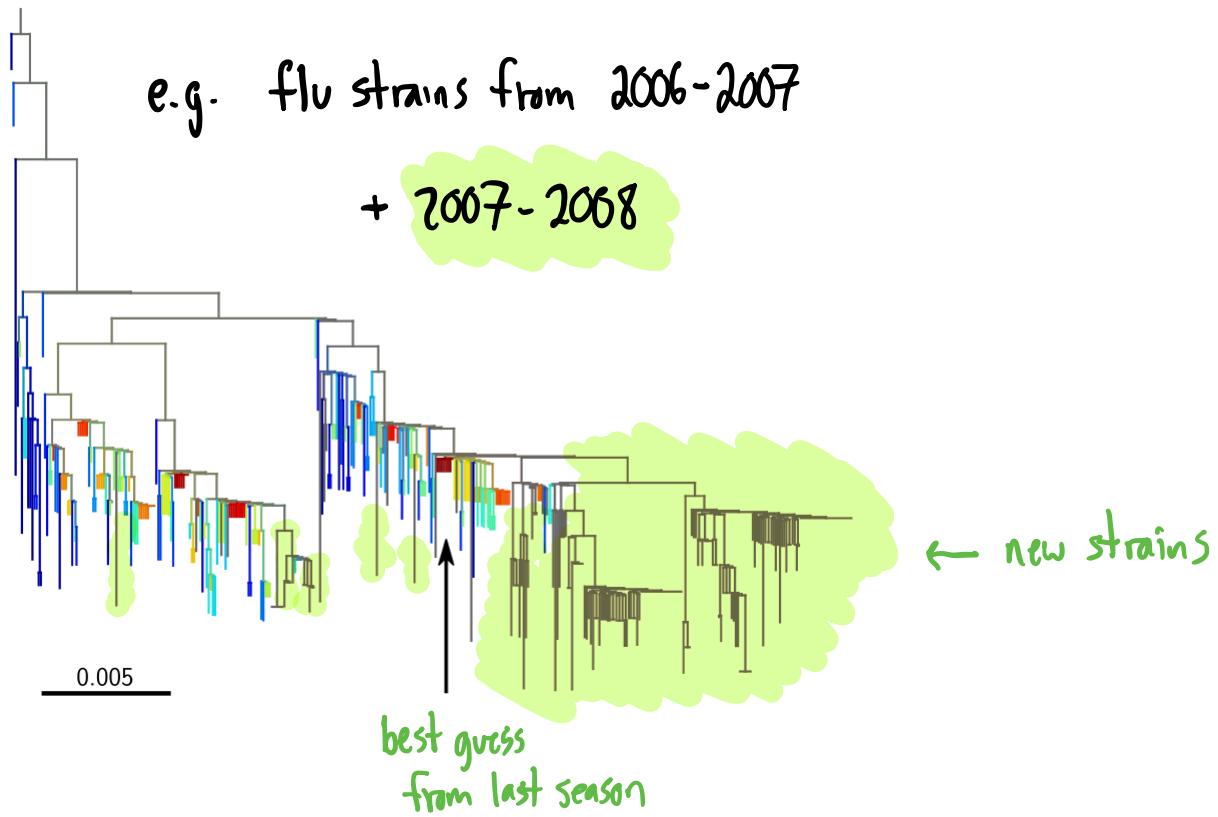
Predicting evolution from the shape of genealogical trees

Richard A Neher^{1*}, Colin A Russell², Boris I Shraiman^{3*}

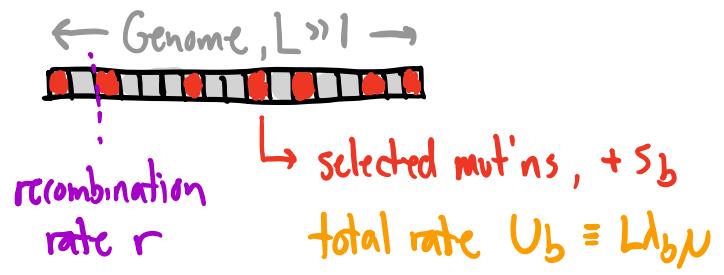
¹Evolutionary Dynamics and Biophysics, Max Planck Institute for Developmental Biology, Tübingen, Germany; ²Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom; ³Kavli Institute for Theoretical Physics, University of California, Santa Barbara, Santa Barbara, United States

⇒ implemented this idea for HA gene in influenza
(data from Problem #1 in HW 1)





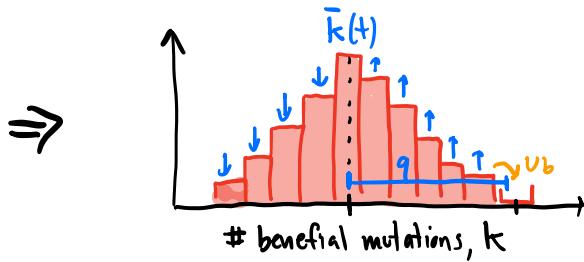
What about recombination?



$$r=0$$



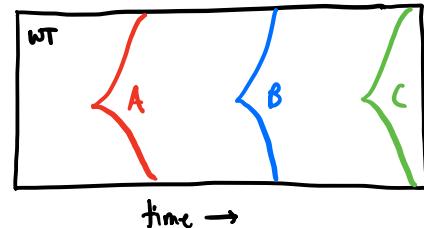
⇒

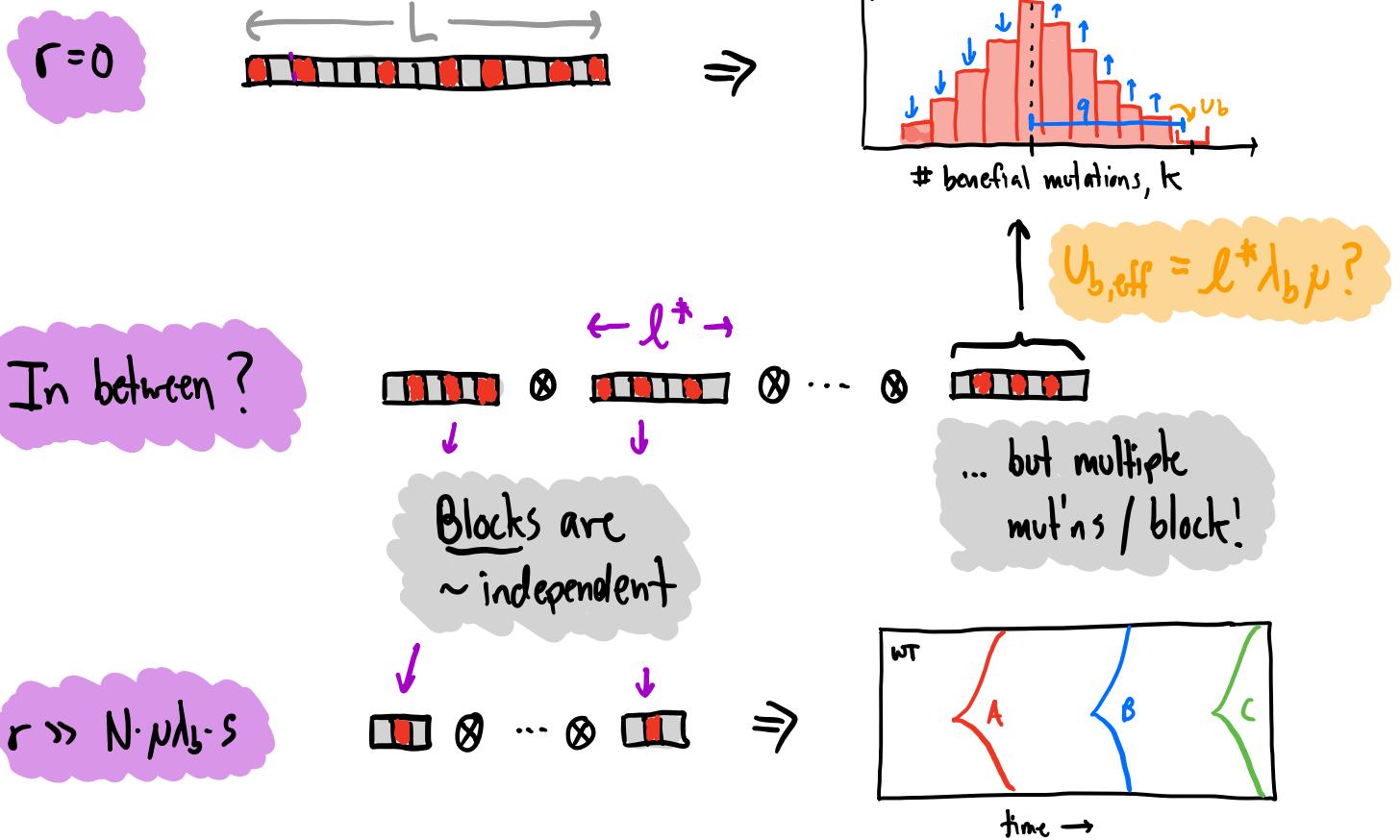


$$r \gg N \cdot \mu \lambda_b \cdot S$$

$$\boxed{\text{red}} \otimes \dots \otimes \boxed{\text{red}}$$

⇒





If true, need:

① w/in blocks, recombination should be rare! ($r \approx 0$)

$$\Rightarrow r l^* \cdot T_c(N, s_b, U_{\text{eff}}(l^*)) \ll 1$$

② between blocks, recombination should be frequent!

$$\Rightarrow r l^* \cdot T_c \gg 1$$

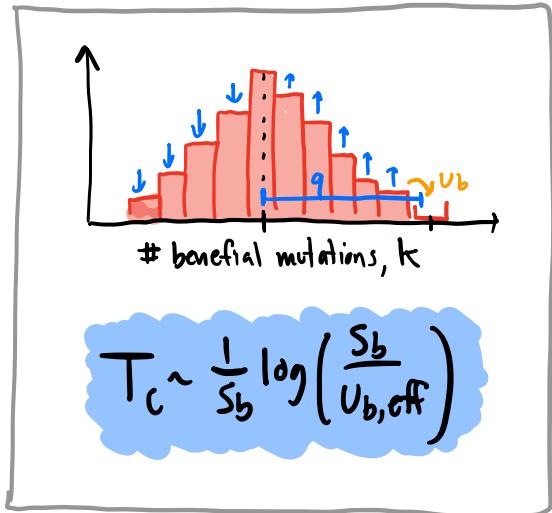
\Rightarrow can we (almost) satisfy both w/ $r l^* \cdot T_c \sim O(1)$?

Linkage block ansatz



$$\ell^* \sim 1/r T_c$$

$$U_{b,\text{eff}} = \ell^* \lambda_b \mu$$



$$T_c \sim \frac{1}{S_b} \log \left(\frac{S_b}{U_{b,\text{eff}}} \right)$$

$$\Rightarrow \text{Self consistency: } T_c \sim \frac{1}{S_b} \log \left(\frac{S_b}{\mu \lambda_b} \cdot r T_c \right)$$

$$\Rightarrow \text{solution: } T_c \sim \frac{1}{S_b} \log \left(\frac{r}{\mu \lambda_b} \right)$$

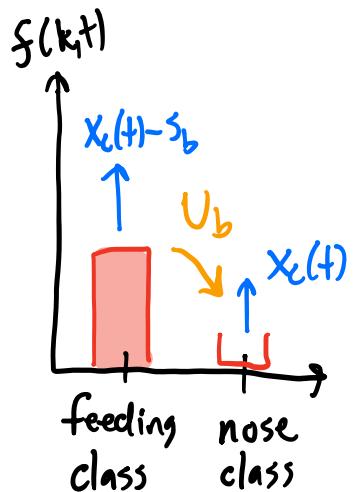
$$\Rightarrow \ell^* \sim \frac{S_b}{r} \log^{-1} \left(\frac{r}{\mu \lambda_b} \right)$$

\Rightarrow self consistent if $N U_{b,\text{eff}} \log(N S_b) \gg 1$ & $U_{\text{eff},b} \ll S_b$

$$\Rightarrow N \cdot N \lambda_b \cdot S_b \gg r \gg \mu \lambda_b$$

Appendix A: 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: $U_b f_{q-1}(t) = \frac{U_b}{2Ns_b} e^{\int_0^t (X_c(\tau) - s_b) d\tau}$

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

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

\Rightarrow will be harder because time-varying fitness

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

\Rightarrow From our discussion in class, can take

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

where t_c is the time that $\bar{F}(t)$ 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:

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

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

\Rightarrow can solve w/ method of characteristics:

define: $\Psi(t_R) = \log [H_f(z(t_R), t - t_R)]$

w/ $\Psi(t) = 0$, $z(0) = z$,

$$\Psi(t) = \log H_f(z, t)$$

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

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

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

$$\Rightarrow \psi(t_R) = \psi(0) + \int_0^{t_R} z(t') v_b f_{q-1}(t-t') dt'$$

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

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:

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

(can plug in & check...)

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

$$= \exp \left[- \int_0^+ \frac{z \cdot U_b f_{q-1}(u) e^{\int_u^+ x_c(u') du'}}{1 + \frac{z}{2N} \int_u^+ e^{\int_{u'}^+ x_c(u'') du''} du'} du' \right]$$

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

$$\Rightarrow H_v(z, t) \equiv \langle e^{-z \cdot v(t)} \rangle \equiv H_f \left(2NqS_b e^{-\int_0^+ x_c(t') dt'}, z, t \right)$$

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

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

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

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

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

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

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

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

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

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 $t = -\infty$
 \nrightarrow 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 $\xi = (2z)^{1/q} e^{-S_b t}$,

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

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

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

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

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

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

\Rightarrow typical value of establishment time:

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

\Rightarrow consistent w/ results from
simpler heuristic argument!

Appendix B: How many lineages contribute to new nose?

Recall in heuristic argument, we had:

$$f_{\text{nose}}(t) \equiv \sum_{k=1}^{K_{\max}} f_k(t) = \frac{1}{Nq_s} e^{qS_b(t-\tau)} \cdot \sum_{k=1}^{K_{\max}} \frac{1}{q} k^{1+\frac{1}{q}}$$

& argued that sum over k converged to ≈ 1 .

\Rightarrow Let's look @ this more carefully...

\Rightarrow if $K_{\max} \gg 1$ (will revisit below)

$$\Rightarrow \sum_{k=1}^{K_{\max}} \frac{1}{q} \frac{1}{k^{1+\frac{1}{q}}} \approx \int_1^{K_{\max}} \frac{dk}{q k^{1+\frac{1}{q}}} = 1 - e^{-\frac{1}{q} \log K_{\max}}$$

Thus, sum converges to 1 provided that

$\log K_{\max}$ is large compared to q

\Rightarrow how does this translate to establishment times τ_k ?

recall that $\tau_k - \tau_1 \sim \frac{1}{qS_b} \log(k)$, so condition becomes:

$$\Rightarrow \tau_{k_{\max}} - \tau_1 \sim \frac{1}{qS_b} \log(K_{\max}) \gg \frac{1}{S_b}$$

Thus, mutations that establish $\gg \frac{1}{S_b}$ after τ_1
have negligible contribution to $f_{\text{base}}(t)$, τ , etc

\Rightarrow since $\frac{1}{S_b} \ll \tau$, this happens long before next click.

\Rightarrow can take $K_{\max} \approx \infty$ w/o losing any accuracy

i.e., can pretend that infinite # of muts
contribute to establishment of new nose.

Appendix C: formal analysis of coalescence in the nose

Recall: main result for stochastic size of nose:

$$f_{\text{nose}}(t) \equiv \frac{v}{2Ns_b} e^{qs_b t} \Rightarrow H_v(z) \approx e^{-\frac{v}{s_b} z^{1-\frac{1}{q}}}$$

(supplement of lecture 19)

Let's fine-grain this further:

\Rightarrow let $f_e(t) \equiv$ freq of lineage in nose founded by beneficial mutation @ site e

$$\Rightarrow \text{then } H_{v_e}(z) \approx e^{-\frac{N}{s_b} z^{1-\frac{1}{q}}}$$

\Rightarrow Probability that 2 individuals coalesce = probability that they came from same lineage:

$$\Rightarrow P_c(z) = \left\langle \sum_{l=1}^{L_b} \left(\frac{f_l(t)}{\sum_{e'} f_{e'}(t)} \right)^2 \right\rangle = \left\langle \sum_{l=1}^{L_b} \left(\frac{v_l}{\sum_{e'} v_{e'}} \right)^2 \right\rangle$$

\Rightarrow Trick: using $\int_0^\infty \frac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-\lambda z} dz = 1$, can write as

$$P_c(z) = \left\langle \sum_e \left(\frac{v_e}{\sum_{e'} v_{e'}} \right)^2 \right\rangle = \left\langle \sum_e v_e^2 \int_0^\infty dz z^{-\sum_{e'} v_{e'}} e^{-z} \right\rangle$$

$$= \sum_e \left\langle \int_0^\infty dz \cdot z \cdot \left(v_e^2 e^{-z v_e} \right) \cdot \prod_{e' \neq e} e^{-v_{e'} z} \right\rangle$$

$$= \sum_e \int_0^\infty dz \cdot z \cdot \frac{d^2 H_{v_e}(z)}{dz^2} \cdot \prod_{e' \neq e} H_{v_{e'}}(z)$$

\Rightarrow using results above for $H_{v_e}(z)$ and $H_v(z)$,

$$\Rightarrow \frac{d H_{v_e}}{dz} = -\frac{v_e}{S_b} (1 - \frac{1}{q}) z^{-\frac{1}{q}} H_{v_e}(z)$$

$$\Rightarrow \frac{\partial^2 H_{V_L}}{\partial z^2} = \frac{1}{q} \frac{N_L}{S_b} \left(1 - \frac{1}{q}\right) z^{-1-\frac{1}{q}} H_{V_L}(z) + O(N_L^2)$$

so that:

$$P_c(z) = \sum_l \int_0^\infty dz \cdot z \cdot \frac{\partial^2 H_{V_L}(z)}{\partial z^2} \cdot \prod_{l' \neq l} H_{V_{L'}}(z)$$

$$= \frac{1}{q} \sum_l \frac{N_L}{S_b} \int_0^\infty dz \frac{N_L}{S_b} \left(1 - \frac{1}{q}\right) z^{-\frac{1}{q}} H_V(z) = \prod_l H_{V_L}(z)$$

$$= \frac{1}{q} \int_0^\infty - \frac{\partial H_V(z)}{\partial z} = \frac{1}{q} \left[H_V(0) - H_V(\infty) \right]$$

$$= \frac{1}{q} \quad \text{as desired}$$

Can do same thing for larger samples:

$$P_c(n) = \left\langle \sum_l \left(\frac{v_L}{\sum v_L} \right)^n \right\rangle = \sum_l \int_0^\infty dz \cdot \frac{(-1)^n}{\Gamma(n)} \frac{d^n H_{V_L}(z)}{dz^n} \prod_{l' \neq l} H_{V_{L'}}(z)$$

$$\Rightarrow \frac{d^n H_{V_e}(z)}{dz^n} = \frac{N_e}{S_b} \frac{(-1)^n}{q} \left(1 - \frac{1}{q}\right) \prod_{k=1}^{n-2} \left(k + \frac{1}{q}\right) z^{-n+1-\frac{1}{q}} H_{V_e}(z) + O(N_e^2)$$

$$\approx \frac{N_e}{S_b} \frac{(-1)^n (n-2)!}{q} z^{-n+1-\frac{1}{q}} H_{V_e}(z)$$

$$\Rightarrow p_c(n) = \frac{1}{q} \frac{\Gamma(n-1)}{\Gamma(n)} \cdot \sum_e \frac{N_e}{U_b} \cdot \int_0^\infty dz \cdot -\frac{dH_V(z)}{dz}$$

$$\approx \frac{1}{q(n-1)}$$