Last time: clonal interference

"Staircase" model

← Genome, $L \gg 1$ →

↳ selected mut'ns, $+ S_b$

total rate $U_b \equiv L\lambda_b N$

$\Rightarrow$

$f(k,t)$

$\bar{k}(t)$ $2s_b$ @ rate $1/\tau$

$(q-1)s_b$

q U_b

benefial mutations, k
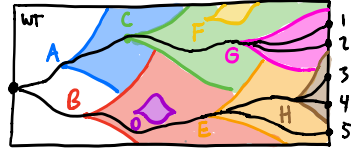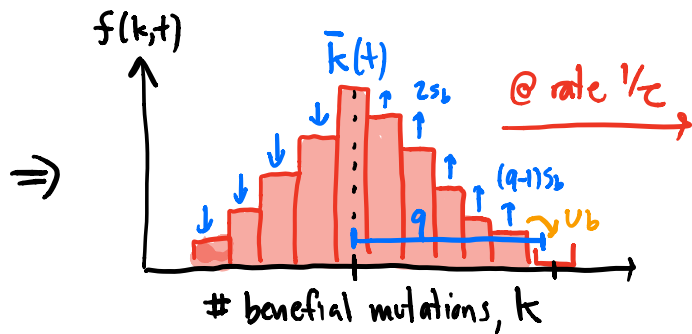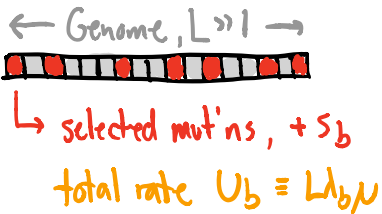
key behavior occurs @ "nose":

freq. $f(k)$

$(q-1)s_b$

qs_b

feeding class   nose

$+ \tau$

$\Rightarrow$

qs_b

feeding class   nose
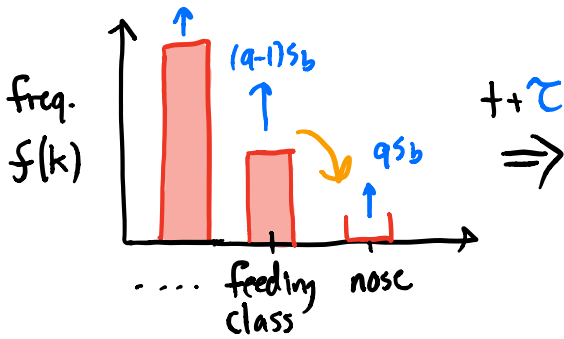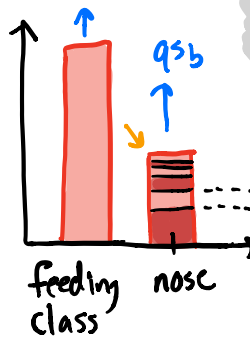
Multiple mutations contribute to nose!

$$f_k(t) \sim \frac{1}{Nqs} e^{qs(t - \tilde{\tau}_k)}$$

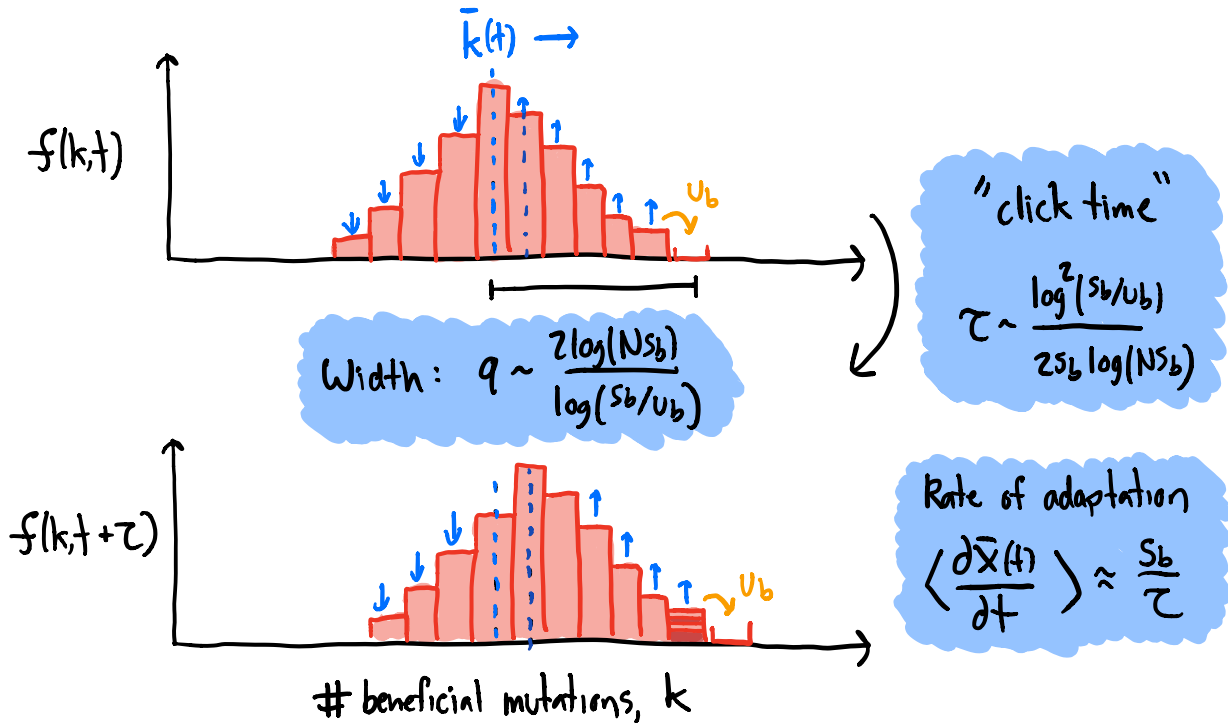$$\tau_k \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b qk}{U_b}\right)$$

Total contribution:

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

⇒ Complete picture of dynamics of **fitness** **dist'n** :

$f(k,t)$

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

U_b

Width: $q \sim \dfrac{2\log(Ns_b)}{\log(s_b/U_b)}$

"click time"

$$\tau \sim \frac{\log^2(s_b/U_b)}{2s_b \log(Ns_b)}$$

$f(k, t+\tau)$

U_b

Rate of adaptation

$$\left\langle \frac{\partial \bar{x}(t)}{\partial t} \right\rangle \approx \frac{s_b}{\tau}$$
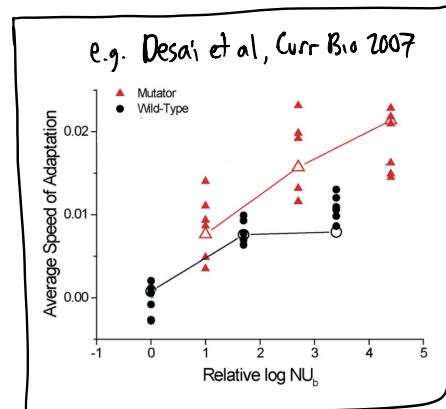
\# beneficial mutations, k

⇒ early tests for clonal interference in lab evolution experiments :
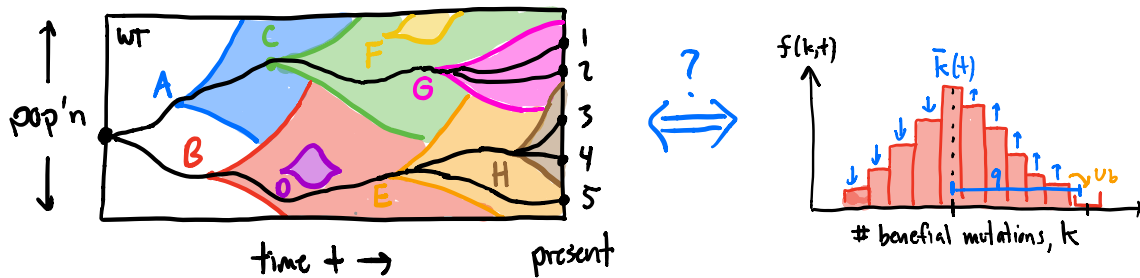
Successive mutations:

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

clonal interference:

$$\left\langle \frac{d\bar{x}}{dt} \right\rangle \sim s_b^2 \cdot \frac{\log(Ns_b)}{\log^2(s_b/U_b)}$$

e.g. Desai et al, Curr Bio 2007

Mutator
Wild-Type

0.02

0.01

0.00

Average Speed of Adaptation

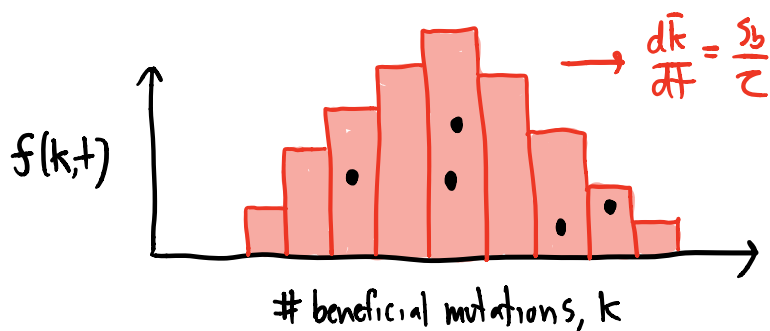-1   0   1   2   3   4   5

Relative log NU_b

Today: 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)

$$\frac{d\bar{k}}{dt} = \frac{s_b}{\tau}$$

Step 2: where was everyone <u>one</u> <u>click</u> ago?

$f(k,t)$ → forward time

one click ago

reverse time

$f(k,t-\tau)$

\# beneficial mutations, k
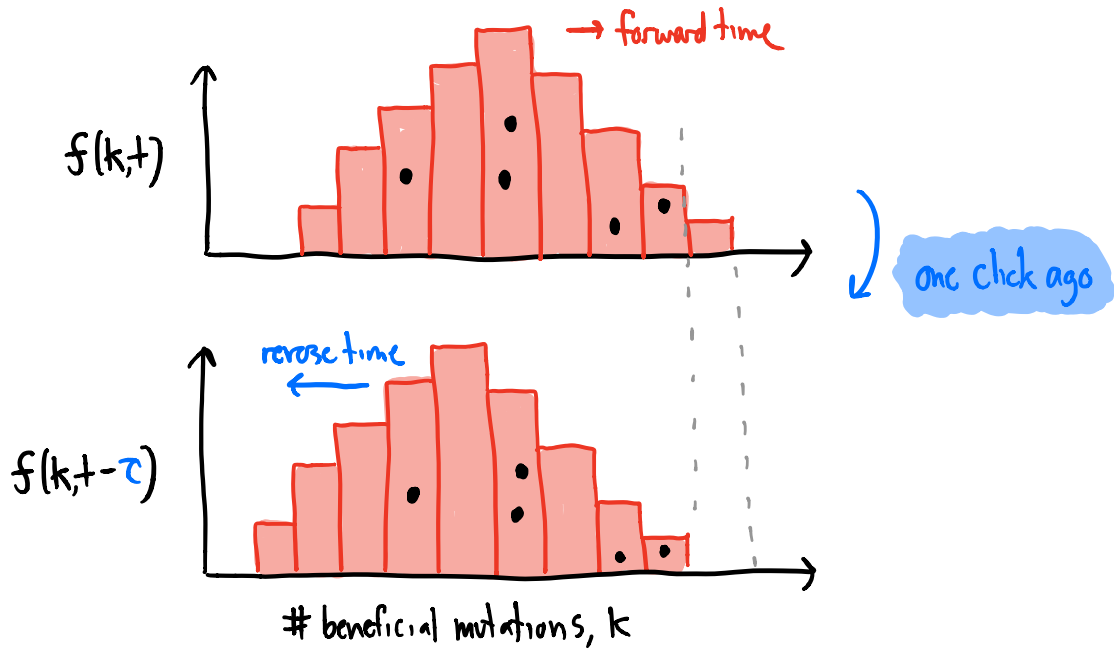
① can only coalesce if in same fitness class

② <u>But</u> little chance of coalescing in "bulk" of dist'n

(since $\tau \ll N f_{q-1}(\tau), N f_{q-2}(\tau)$, etc.)

$f(k,t)$
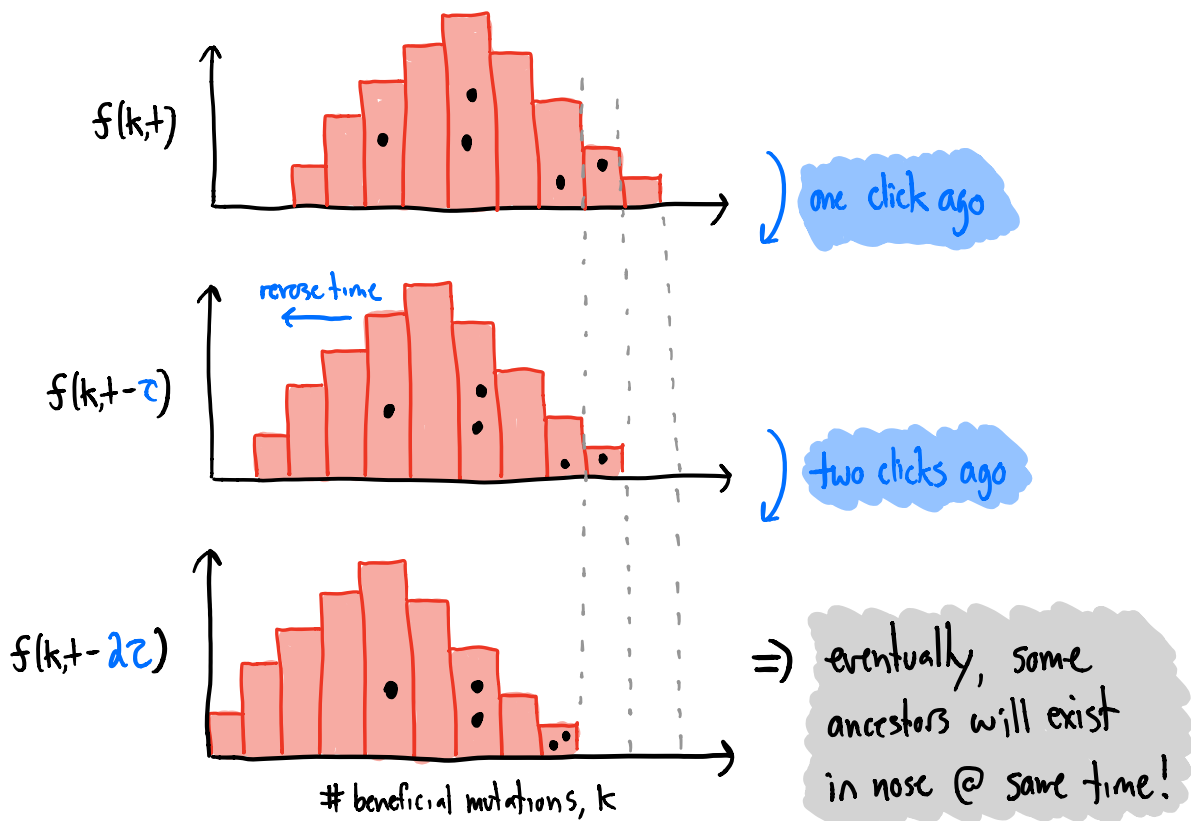

) one click ago

reverse time ←

$f(k,t-\tau)$

) two clicks ago

$f(k,t-2\tau)$

beneficial mutations, k

⟹ eventually, some ancestors will exist in nose @ same time!

Two possible scenarios:

① Individuals are from separate lineages in the nose

$f_{nose}(t)$
$f_k(t)$

feeding class   nose

⟱
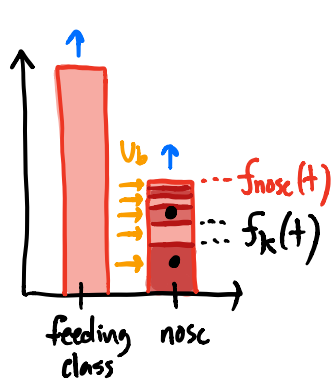
separate ancestors in feeding class

(distinct mut'n events)

② Individuals from <u>same</u> lineage in nose

$\Downarrow$
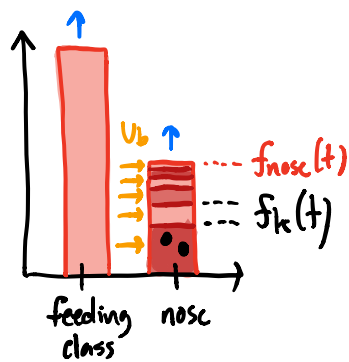
Common ancestor
in feeding class

(coalescence w/in τ gens)

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

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

$\Uparrow$

only depends on establishment times τ_k !

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

$$\tau_k \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b\, qk}{U_b}\right) \; ; \qquad \tau \sim \frac{1}{(q-1)s_b} \log\left(\frac{s_b}{U_b}\right)$$

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

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

$\Rightarrow$ missing key part of puzzle : fluctuations

$\Rightarrow$ coalescence <u>rare</u> for **typical** lineage sizes,
<u>but</u> small chance of having anomalously early mutant
where coalescence is <u>much more likely</u>!

e.g. if <u>first</u> successful mutation occurs when $\tau_1 \lesssim \tau$:

$$f_{rest} \sim \frac{1}{Nqs_b} e^{qs_b(t-\tau)}$$

$$f_1(t) \sim \frac{1}{Nqs_b} e^{qs_b(t-\tau_1)}$$

feeding   nosc
class
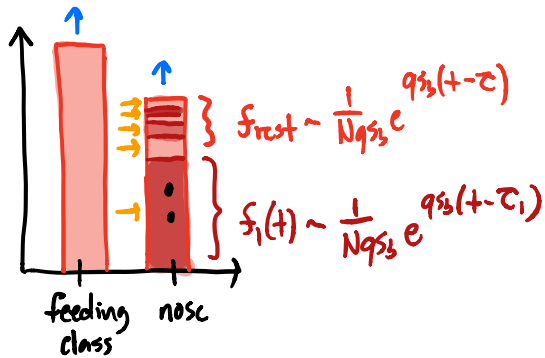
$$\frac{f_1(t)}{f_{nosc}(t)} \gtrsim O(1)$$

$$P_c(2) \gtrsim O(1)!$$

$\Rightarrow$ not a <u>huge</u> shift: typically, $\tau_1 - \tau \sim \frac{\log(q)}{qs_b} \ll \frac{1}{s_b} \ll \tau$

$\Rightarrow$ occurs w/ total probability:

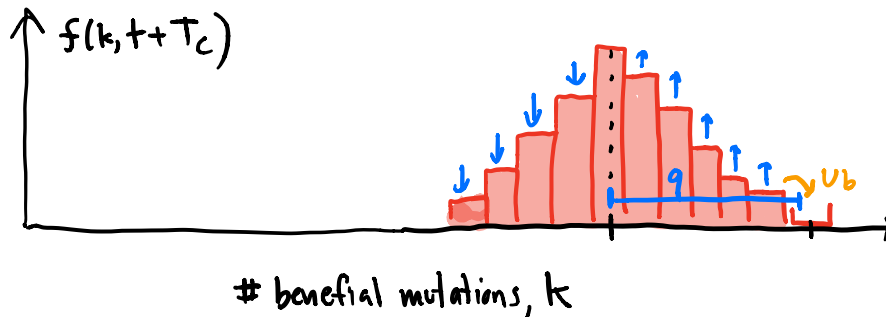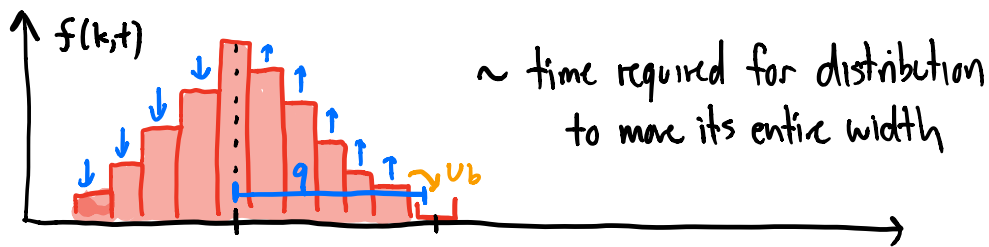$$P_{jackpot} \sim \int_0^\tau d\tau_1 \, NU_b \, f_{q-1}(t) \cdot qs_b$$

$$\sim \int_0^\tau d\tau_1 \, NU_b \cdot \frac{e^{(q-1)s_b t}}{Nqs_b} \cdot qs_b \sim \frac{U_b}{(q-1)s_b} e^{(q-1)s_b \tau}$$

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

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

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

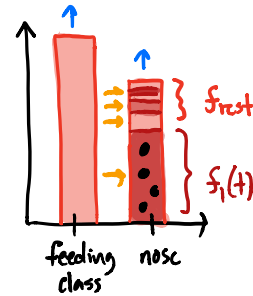$\Rightarrow$ coalescent timescale $T_c \equiv q\tau \sim \frac{1}{s_b} \log\left(\frac{s_b}{U_b} \right)$

$f(k,t)$

$\sim$ time required for distribution to move its entire width

q

U_b

$f(k, t+T_c)$

q

U_b

benefial mutations, k

$\Rightarrow$ fluctuations were crucial for determining T_c !

$\Rightarrow$ coalescence is "bursty":

  e.g. in larger sample size n:

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

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

$$\Rightarrow P_{jackpot}(n) = \int_0^{\tau \sim \log(n)/qs_b} d\tau_1 \; N U_b f_{q-1}(t) \cdot q s_b \sim \frac{1}{qn}$$

  $\Rightarrow$ i.e. multiple mergers likely!

Another interesting feature of genealogies + travelling wave:

⇒ consider same
example:

beneficial mutations
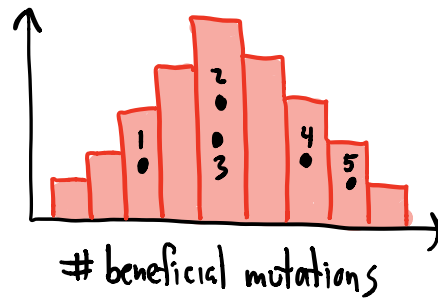
⇒ which individual's <u>descendents</u> are more likely
to take over pop'n in <u>future</u>?

⇒ e.g. 5 → 4 → 2,3 → 1

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

-3τ

-2τ

-τ

present

relative fitness, $k - \bar{k}(t)$

backward in time

MRCA

1 2 3 4 5

3τ
2τ
τ
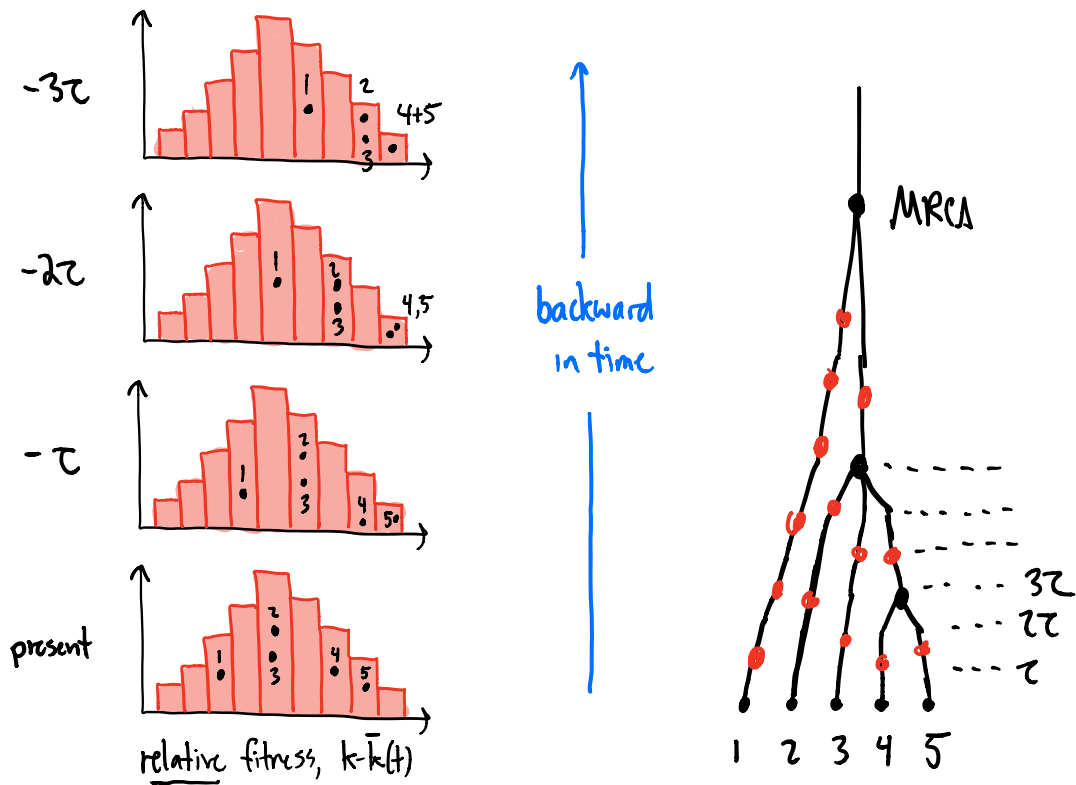
⟹ time (+burstiness) of coalescence in past
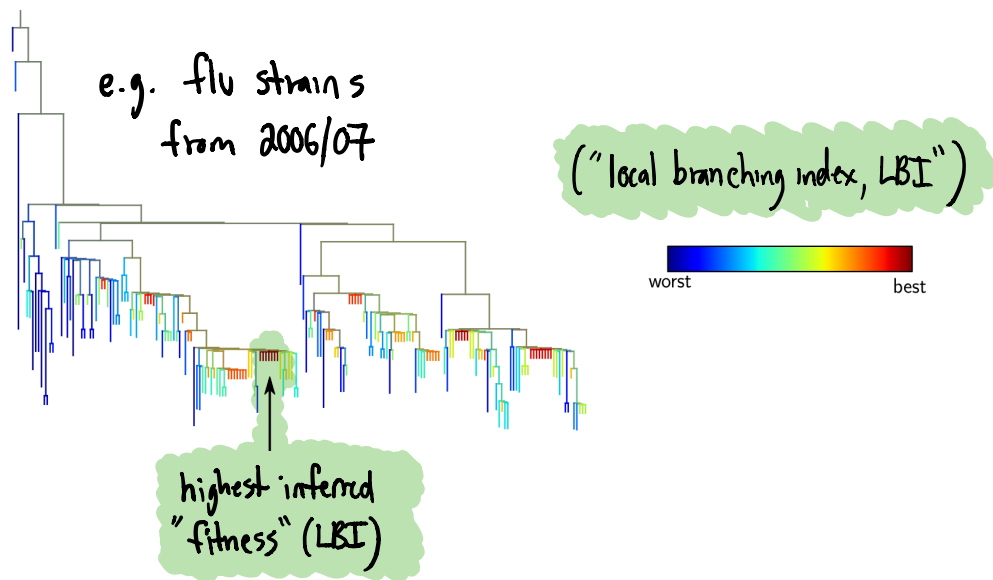
⟹ info about fitness in present

⟹ forecasts about who takes over in future!

Predicting evolution from the shape of genealogical trees

Richard A Neher[1]*, Colin A Russell[2], Boris I Shraiman[3]*

[1]Evolutionary Dynamics and Biophysics, Max Planck Institute for Developmental Biology, Tübingen, Germany; [2]Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom; [3]Kavli 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 HW1)

e.g. flu strains from 2006/07

("local branching index, LBI")

worst   best

highest inferred "fitness" (LBI)

e.g. flu strains from 2006-2007

+ 2007-2008
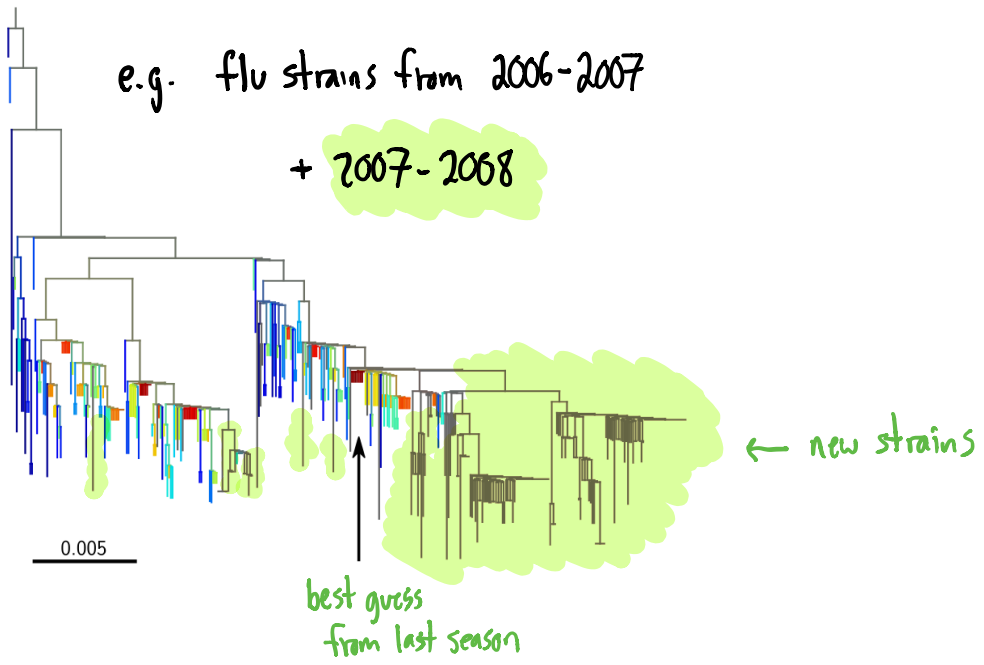
← new strains

0.005

best guess
from last season

What about recombination?

← Genome, $L \gg 1$ →

↳ recombination rate r

↳ selected mut'ns, $+s_b$

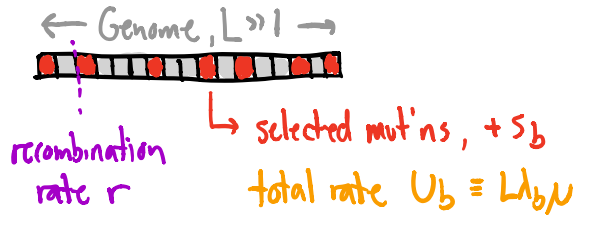total rate $U_b \equiv L\lambda_b N$

$r = 0$

$\bar{k}(t)$

q

$\sim U_b$

beneficial mutations, k

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

⊗ ... ⊗

WT

A

B

C

time →

$r=0$

$\bar{k}(t)$

benefical mutations, k

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

$\leftarrow \ell^* \rightarrow$

In between?

Blocks are ~independent

... but multiple mut'ns / block!

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

WT

A   B   C

time $\rightarrow$
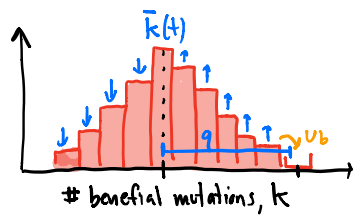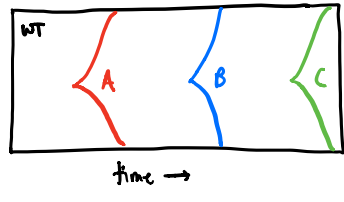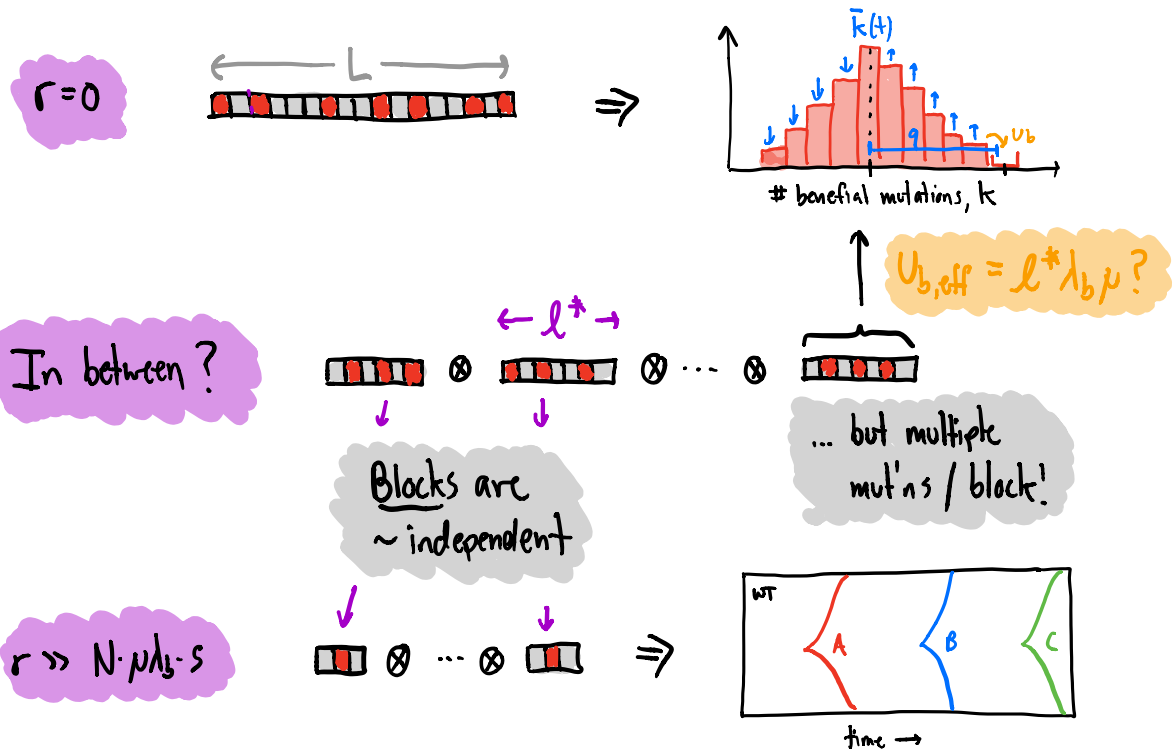
If true, need:

① w/in blocks, recombination should be rare! $(r \hat{=} 0)$

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

② between blocks, recombination should be frequent!

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

$\Rightarrow$ can we (almost) satisfy __both__ w/ $r \ell^* \cdot T_c \sim \mathcal{O}(1)$?

Linkage block ansatz

$\ell^* \sim 1/rT_c$

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

$\rightarrow$

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

benefial mutations, k
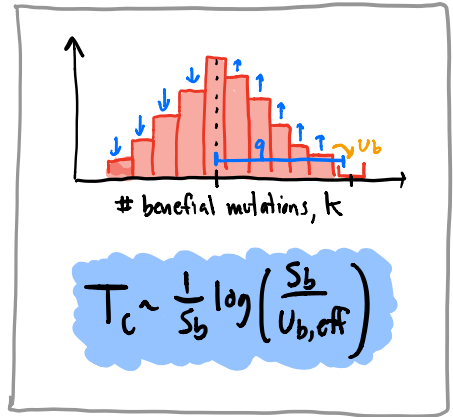
$\Rightarrow$ Self consistency: $\quad T_c \sim \frac{1}{S_b} \log\left(\frac{S_b}{\mu\lambda_b} \cdot rT_c\right)$

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

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

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

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

<u>Supplement</u>: formal analysis of coalescence in the nose

Recall: main result for stochastic size of nose:

$$f_{nose}(t) \equiv \frac{\nu}{2Nqs_b} e^{qs_b t} \quad \Rightarrow \quad H_\nu(z) \approx e^{-\frac{U_b}{s_b} z^{1-\frac{1}{q}}}$$

(supplement of lecture 19)

Let's fine-grain this further:

$$\Rightarrow \quad \text{let} \quad f_\ell(t) \equiv \text{freq of lineage in nose founded by beneficial mutation @ site } \ell$$

$$\Rightarrow \quad \text{then} \quad H_{\nu_\ell}(z) \approx e^{-\frac{\nu}{s_b} z^{1-\frac{1}{q}}}$$

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

$$\Rightarrow \quad P_c(z) = \left\langle \sum_{\ell=1}^{L_{Ab}} \left(\frac{f_\ell(t)}{\sum_{\ell'} f_{\ell'}(t)} \right)^z \right\rangle = \left\langle \sum_{\ell=1}^{L_{Ab}} \left(\frac{v_\ell}{\sum_{\ell'} v_{\ell'}} \right)^z \right\rangle$$

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

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

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

$$= \sum_\ell \int_0^\infty dz \cdot z \cdot \frac{\partial^2 H_{v_\ell}(z)}{\partial z^2} \cdot \prod_{\ell' \neq \ell} H_{v_{\ell'}}(z)$$

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

$$\Rightarrow \quad \frac{\partial H_{v_\ell}}{\partial z} = -\frac{\mu_\ell}{S_b} \left(1 - \frac{1}{q}\right) z^{-\frac{1}{q}} H_{v_\ell}(z)$$

$$\Rightarrow \quad \frac{\partial^2 H_{\nu_\ell}}{\partial z^2} = \frac{1}{q} \frac{N_\ell}{S_b} \left(1 - \frac{1}{q}\right) z^{-1-\frac{1}{q}} H_{\nu_\ell}(z) + \mathcal{O}(N_\ell^2)$$

so that:

$$P_c(2) = \sum_\ell \int_0^\infty dz \cdot z \cdot \frac{\partial^2 H_{\nu_\ell}(z)}{\partial z^2} \cdot \prod_{\ell' \neq \ell} H_{\nu_{\ell'}}(z)$$

$$= \frac{1}{q} \sum_\ell \frac{N_\ell}{U_b} \int_0^\infty dz \, \frac{U_b}{S_b} \left(1 - \frac{1}{q}\right) z^{-\frac{1}{q}} H_\nu(z) \overset{\prod_\ell H_{\nu_\ell}(z)}{}$$

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

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

Can do same thing for larger samples:

$$P_c(n) = \left\langle \sum_\ell \left(\frac{\nu_\ell}{\sum_{\ell'} \nu_{\ell'}} \right)^n \right\rangle = \sum_\ell \int_0^\infty dz \cdot \frac{(-1)^n z^n}{\Gamma(n)} \frac{\partial^n H_{\nu_\ell}(z)}{\partial z^n} \prod_{\ell' \neq \ell} H_{\nu_{\ell'}}(z)$$

$$\Rightarrow \quad \frac{\partial^n H_{\nu_\ell}(z)}{\partial z^n} = \frac{N_\ell}{S_b} \frac{(-1)^n}{q} \left(1 - \frac{1}{q}\right) \prod_{k=1}^{n-2} \left(k + \frac{1}{q}\right) z^{-n+1-\frac{1}{q}} H_{\nu_\ell}(z) + \mathcal{O}(N_\ell^2)$$

$$\approx \frac{N_\ell}{S_b} \frac{(-1)^n (n-2)!}{q} z^{-n+1-\frac{1}{q}} H_{\nu_\ell}(z)$$

$$\Rightarrow \quad P_c(n) = \frac{1}{q} \frac{\Gamma(n-1)}{\Gamma(n)} \cdot \sum_\ell \frac{N_\ell}{U_b} \cdot \int_0^\infty dz \cdot \frac{-\partial H_\nu(z)}{\partial z}$$

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