

Chapter 10

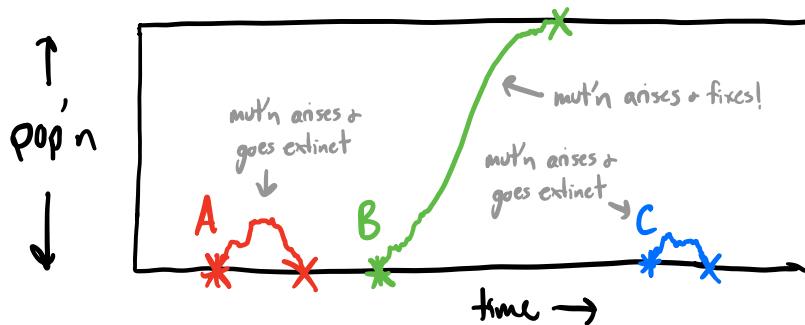
Successive mutations regime

Successive mutations regime

(i.e., mutation is small correction)

$$\frac{d\vec{g}(t)}{dt} = \sim(x - \bar{x}) + \sim L \mu \rightarrow e + \sim e + \sim \frac{\pi}{JN}$$

\Rightarrow i.e. new mutations fix or go extinct before next one occurs...



\Rightarrow @ any given time, only 2 genotypes present:

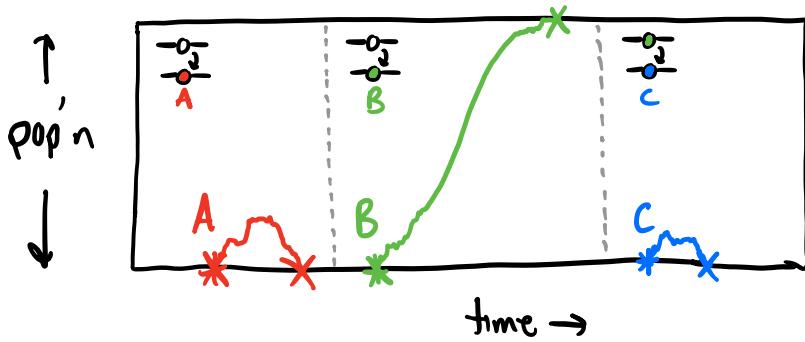
"current wildtype" $\vec{g}_0 = (1, 0, 1, 1, 0, 0, 0)$

\downarrow mutation @ site ℓ

"single mutant" $\vec{g}_m = (1, 0, 1, 1, 0, 1, 0)$

\Rightarrow what can recombination do? Nothing! (on average...)

\Rightarrow then each mut'n looks like $L=1 \begin{pmatrix} -0 \\ -0 \\ +s_{eff} \end{pmatrix}$ w/ $s_{eff} \equiv X(\vec{g}_m) - X(\vec{g}_0)$

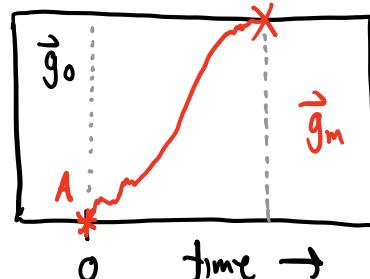


\Rightarrow in this case, know exactly what happens:

(i) w/ prob $P_{fix} = \frac{2s}{1-e^{-2Ns}}$,

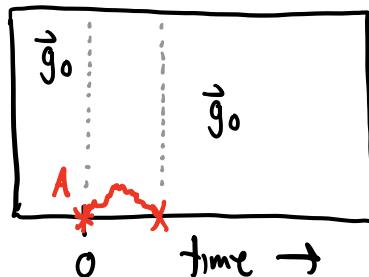
\Rightarrow mutation fixes ("sweeps")

$\Rightarrow \vec{g}_0 \rightarrow \vec{g}_m$; repeat!

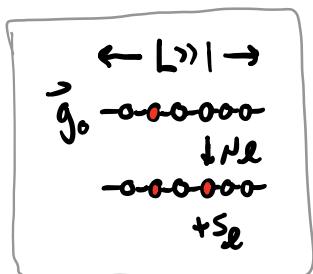


(ii) otherwise, mutation goes extinct

$\Rightarrow \vec{g}_0$ stays put.

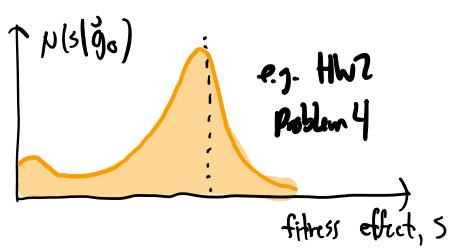


\Rightarrow New: when $L > 1$, multiple different mutations are possible...



\Rightarrow useful to define fitness effects

$$S_l \equiv X(\vec{g}_0 + \underset{\text{at site } l}{\text{mut}}) - X(\vec{g}_0)$$



along w/ distribution of fitness effects ("DFE")

$$N(s|\vec{g}_0) = \sum_{\ell=1}^L N_\ell \delta(s - s_\ell)$$

↑ ↓

technically depends on \vec{g}_0 prob. of drawing a mutation w/ effect s .

\Rightarrow mutations w/ fitness effect s are produced as

Poisson process w/ rate $N\mu(s)$.

\Rightarrow if each successful w/ prob $p_{\text{fix}}(s)$,

then successful mutations also Poisson Process

w/ total rate $R \equiv \int_0^\infty N\mu(s) \cdot p_{\text{fix}}(s) ds = \sum_\ell N\mu_\ell p_{\text{fix}}(s_\ell)$

\Rightarrow ① time until next successful mutation is born is :

$$T_{\text{est}} \sim \text{Exponential}\left(\frac{1}{R}\right)$$

$\left[\begin{array}{l} \text{similar to } N\mu \rightarrow 0 \\ \text{case in lecture 10} \end{array} \right]$

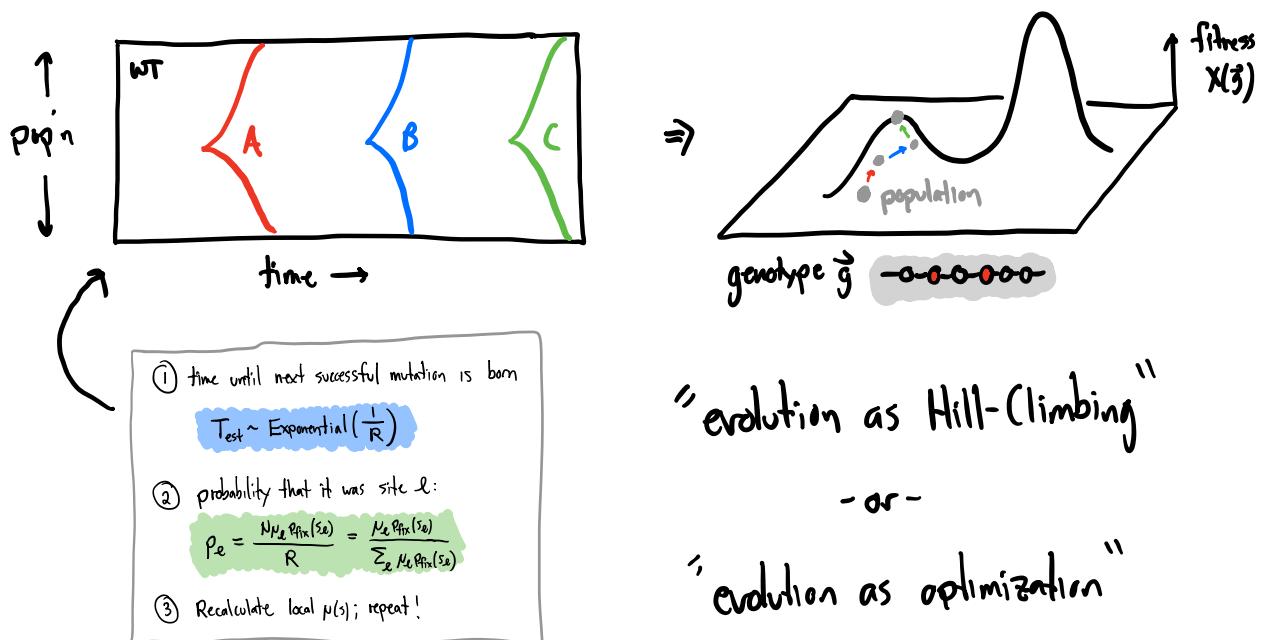
\Rightarrow ② probability that it was site l :

$$P_e = \frac{N\mu_e p_{fix}(s_e)}{R} = \frac{\mu_e p_{fix}(s_e)}{\sum_e N\mu_e p_{fix}(s_e)}$$

\Rightarrow ③ $\vec{g}_0 \rightarrow \vec{g}_m \Rightarrow$ recalculate $\mu(s|\vec{g}_0) \Rightarrow$ repeat from ①

When approx is valid: will check carefully below...

\Rightarrow simple algorithm for modeling evolution (not just pop gen)



"evolution as Hill-Climbing"

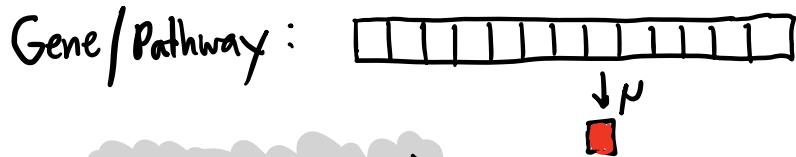
- or -

"evolution as optimization"

Note: even in these simplified settings,

\Rightarrow fundamental limits to optimization picture...

Example: maintaining a useful function

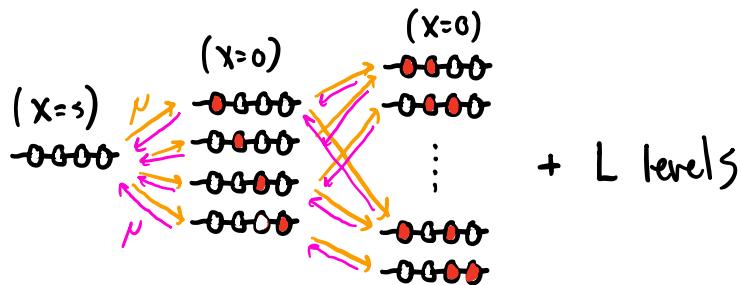


(e.g. for eating some
low-level nutrient)

L ways to "break"
(each \sim rate ν)

$$X(\vec{0}) = S; X(\text{else}) = 0$$

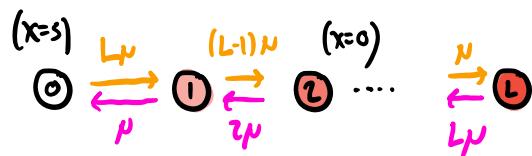
Genotype network:



Key simplification:

"equivalence
class"

$$f_k \equiv \sum_{|\mathbf{j}|=k} f(\mathbf{j})$$



↳ can track: $P_k(t) \equiv \Pr [f_k(t) = 1] = \langle f_k(t) \rangle$

How can $P_k(t)$ change?

$$\underline{k=0}: \quad d_t P_0 = \underbrace{N \nu p_{fix}(s) P_1}_{\text{incoming}} - \underbrace{N L N \cdot p_{fix}(-s) \cdot P_0}_{\text{outgoing}} \xrightarrow{t \rightarrow 0} 0$$

$$\Rightarrow \frac{P_0}{P_1} = \frac{N \nu p_{fix}(s)}{N L N \cdot p_{fix}(-s)} = \frac{1}{L} e^{2Ns}$$

$$\underline{k=1}: \quad d_t P_1 = N L N p_{fix}(-s) P_0 - N \nu p_{fix} P_1 \quad (\text{from 0 class}) \\ + N \cdot 2 \nu \cdot \left(\frac{1}{N}\right) P_2 - N(L-1) N \left(\frac{1}{N}\right) P_1 \xrightarrow{t \rightarrow \infty} 0$$

$$\Rightarrow P_2 = \frac{(L-1)}{2} P_1$$

$$\underline{k=2}: \quad \Rightarrow P_3 = \frac{L-2}{3} \cdot P_2 = \frac{(L-1)(L-2)}{3 \cdot 2} \cdot P_1$$

$$\Rightarrow P_K = \frac{1}{L} \frac{L!}{k!(L-k)!} P_1$$

$$\Rightarrow 1 - P_0 = \sum_{k=1}^L P_k = \frac{1}{L} (2^L - 1) P_1$$

$$\Rightarrow P_1 = (1-P_0) \frac{L}{(2^L - 1)}$$

$$\Rightarrow \frac{P_0}{1-P_0} = \exp \left[\underbrace{2Ns}_{\text{"drift barrier"}} - \underbrace{\log(2^L - 1)}_{\text{"entropy of genotype space"}} \right] = \frac{\Pr(\text{pop'n has function})}{\Pr(\text{broken})}$$

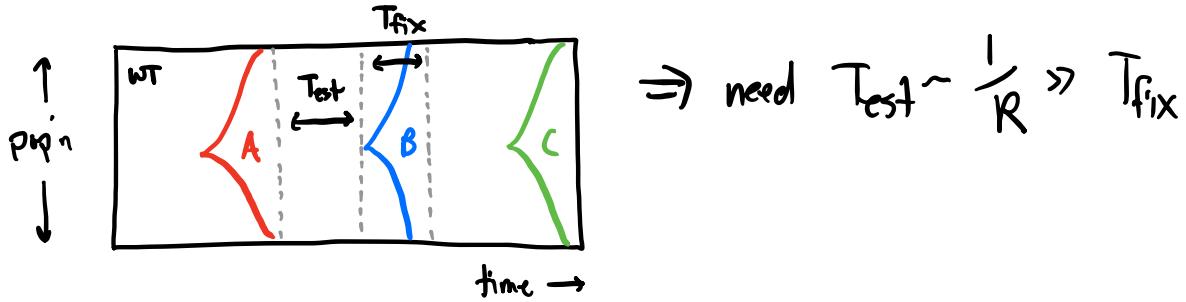
\Rightarrow plug in some #s: if function maintained $\geq 90\%$ of time ...

$$\Rightarrow \frac{0.9}{0.1} \approx e^2 \leq \exp \left[2Ns - \overbrace{\log(2^L - 1)}^{\alpha L} \right]$$

$$S \geq \frac{1}{N} \quad \text{"drift barrier"}$$

compare to deterministic case: $\langle f_0 \rangle = 1 - \frac{L\mu}{S}$

When is successive mutations regime a good approx?



E.g. Neutral mutations ($\mu(s) = U_n \delta(s)$)

$$P_{\text{fix}}(0) = \frac{1}{N} \Rightarrow R = NU_n \left(\frac{1}{N}\right) = U_n ; \quad T_{\text{fix}} \sim O(N)$$

$$\Rightarrow \text{need } \frac{1}{U_n} \gg N \Rightarrow \boxed{NU_n \ll 1} \quad \begin{matrix} \text{"weak mutation"} \\ \text{"weak selection"} \end{matrix}$$

E.g. Strongly beneficial mutations ($\mu(s) = U_b \delta(s-s_b) ; \quad Ns_b \gg 1$)

$$\Rightarrow P_{\text{fix}}(s) \approx 2s \Rightarrow R = 2NU_b s ; \quad T_{\text{fix}} = \frac{2}{s} \log(Ns)$$

$$\Rightarrow \text{need } \frac{1}{2NU_b s} \gg \frac{2}{s} \log(Ns) \Rightarrow \boxed{NU_b \ll \frac{1}{\log(Ns_b)}} \ll 1$$

\Rightarrow What does this look like
for some "real" parameter values?

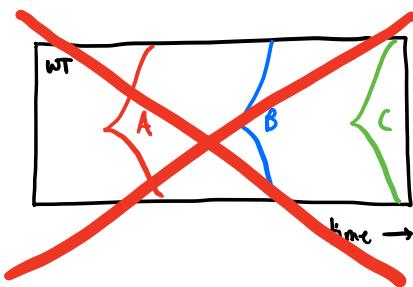
\Rightarrow e.g. HW2 problem 4

$$U_b = 5 \times 10^{-6}, \quad s_b = 0.02$$

just for L.O.F. muts.

$$\Rightarrow \text{e.g. if } N \sim 10^5 \Rightarrow 2Nu_b \log(Ns) = 3$$

\Rightarrow successive mutations picture
does not apply!



\Rightarrow what do things look like instead?