Chapter 10

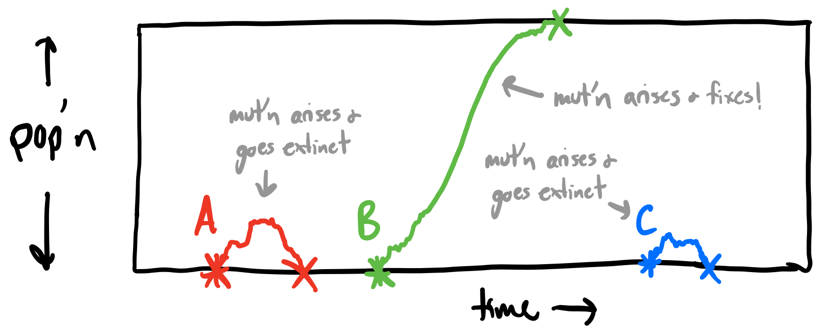
Successive mutations regime

Successive mutations regime

(i.e., mutation is small correction)

$$\frac{\partial f(\vec g)}{\partial t} = \sim(x-\bar{x}) + \cancel{\sim L\mu} \to \epsilon + \sim \rho + \sim \frac{Z}{\sqrt{N}}$$

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



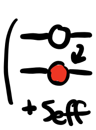
$\Rightarrow$ @ any given time, only 2 genotypes present:

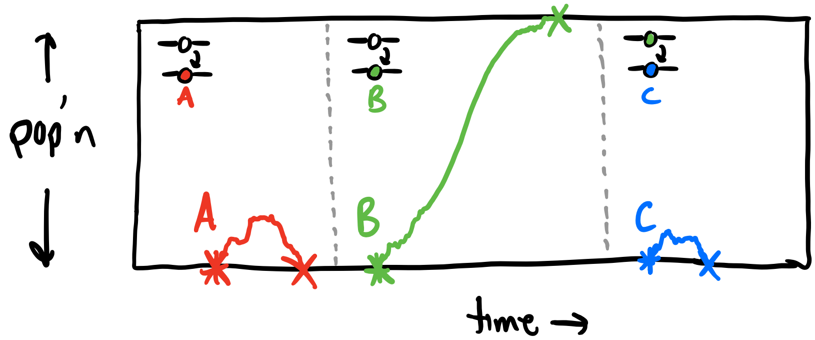
"current wildtype" $\vec{g}_0 = (1,0,1,1,0,0,0)$

$\downarrow$ 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 $L=1$ () w/ $S_{eff} \equiv X(\vec{g}_m) - X(\vec{g}_0)$

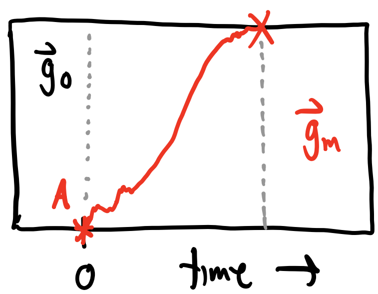


$\Rightarrow$ in this case, know exactly what happens:

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

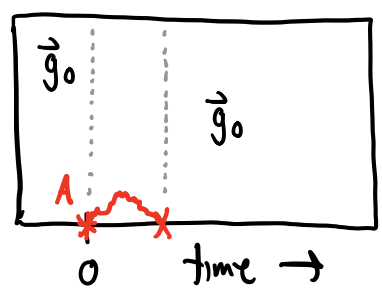
$\Rightarrow$ mutation fixes ("sweeps")

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

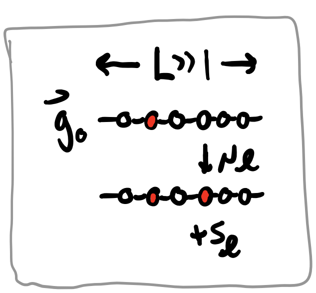


(ii) otherwise, mutation goes extinct

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

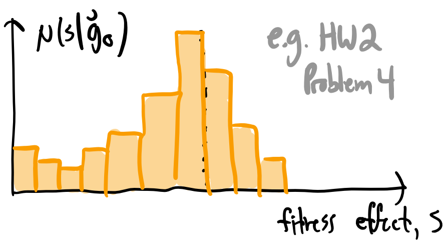


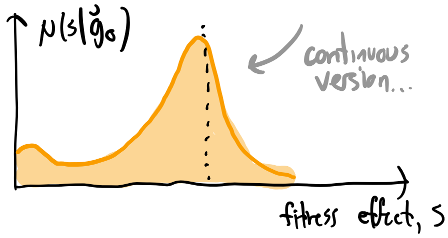
$\Rightarrow$ New: when $L \gg 1$, multiple different mutations are possible...



$\Rightarrow$ useful to define fitness effects

$$S_\ell \equiv X(\vec{g}_0 + \text{mut'n @ site } \ell) - X(\vec{g}_0)$$





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

$$\mu(s|\vec{g}_0) \equiv \underbrace{\sum_{\ell=1}^{L} \mu_\ell \, \delta(s - s_\ell)}_{\text{prob. of drawing a mut'n w/ effect } s}$$

technically depends on $\vec{g}_0$

$\Rightarrow$ mutations w/ fitness effect s are produced as Poisson process w/ rate $N\mu(s)$.

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

then <u>successful</u> <u>mutations</u> also <u>Poisson</u> <u>Process</u>

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

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

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

[similar to $N\mu \to 0$ case in Chapter 7]

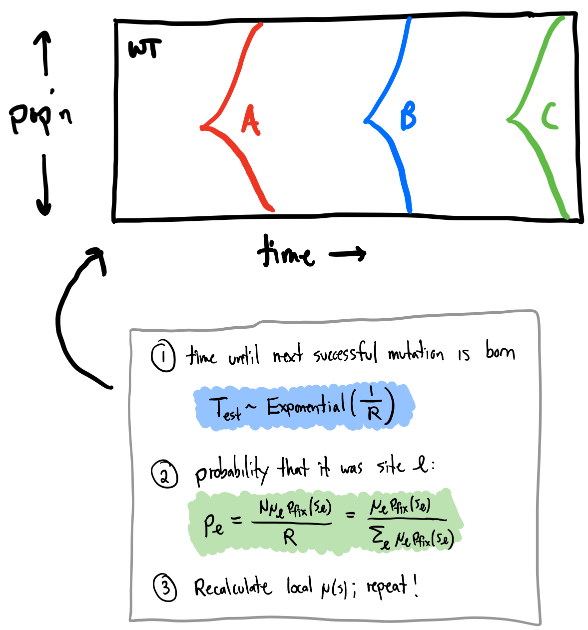
② probability that it was site ℓ:

$$P_\ell = \frac{N\mu_\ell \, P_{fix}(s_\ell)}{R} = \frac{\mu_\ell \, P_{fix}(s_\ell)}{\sum_\ell \mu_\ell \, P_{fix}(s_\ell)}$$

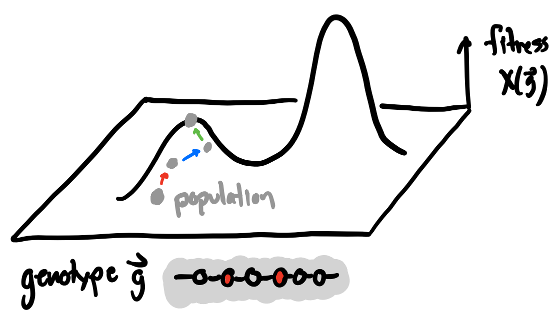
③ $\vec{g}_0 \to \vec{g}_m \Rightarrow$ recalculate $\mu(s|\vec{g}_m) \Rightarrow$ repeat from ①

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

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



$\Rightarrow$



"evolution as hill-climbing"

-or-

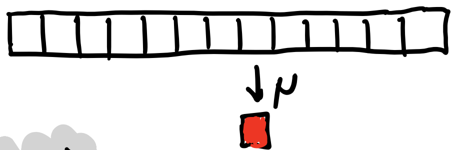
"evolution as optimization"

Note: even in these simplified settings,

$\Rightarrow$ fundamental limits to optimization picture...

Example: maintaining a useful function

Gene/Pathway:

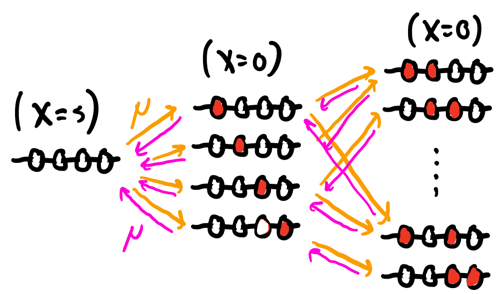


(e.g. for eating some low-level nutrient)

L ways to "break" (each w/ rate μ)

$X(\vec{0}) \equiv s; \quad X(\text{else}) \equiv 0$

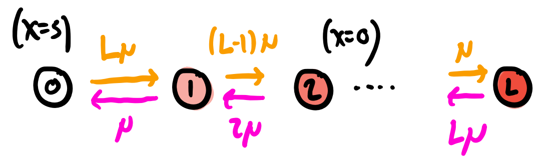
Genotype network:



+ L levels

Key simplification:

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



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

How can $P_k(t)$ change?

$$\underline{k=0}: \quad \partial_t P_0 = \overbrace{N\mu\, p_{fix}(s)\, P_1}^{\text{incoming}} - \overbrace{NL\mu \cdot p_{fix}(-s) \cdot P_0}^{\text{outgoing}} \xrightarrow{t\to\infty} 0$$

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

$$\underline{k=1}: \quad \partial_t P_1 = NL\mu\, p_{fix}(-s) P_0 - N\mu\, p_{fix}(s) P_1 \quad \text{(from 0 class)}$$

$$+ N \cdot 2\mu \cdot \underbrace{\cancel{p_{fix}(0)}}_{\to \frac{1}{N}} P_2 - N(L-1)\mu \cdot \underbrace{\cancel{p_{fix}(0)}}_{\to 1/N} \cdot P_1 \xrightarrow{t\to\infty} 0$$

$$\Rightarrow p_2 = \frac{(L-1)}{2} p_1$$

$$\underline{k=2}: \quad \Rightarrow p_3 = \frac{L-2}{3} \cdot p_2 = \frac{(L-1)(L-2)}{3 \cdot 2} \cdot p_1$$

$$\Rightarrow p_k = \frac{1}{L} \frac{L!}{k!(L-k)!} p_1$$

$$\Rightarrow 1 - p_0 = \sum_{k=1}^{L} p_k = \frac{1}{L}(2^L - 1) p_1$$

combine w/ $k=0$ equation...

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

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

$$\Rightarrow \quad \frac{0.9}{0.1} \approx e^2 \lesssim \exp\Big[2Ns - \overbrace{\log(2^L-1)}^{\propto L}\Big]$$

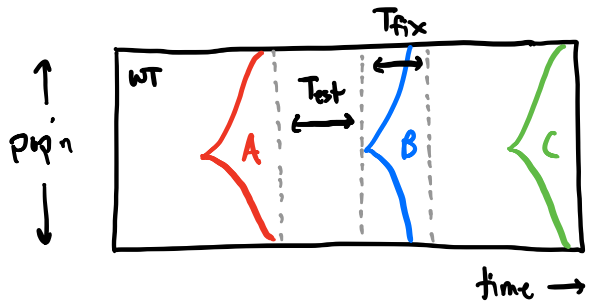
$s \gtrsim \frac{1}{N}$ "drift barrier"

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

$\Rightarrow$ <u>Upshot</u>: evolution is <u>bad</u> @ optimizing shallow fitness gradients...

(<u>hypothesized</u>* to play a role in mut'n rate evolution, protein stability, catalytic efficiency, ...)

When is successive mutations regime a good approx?



$\Rightarrow$ need $T_{est} \sim \frac{1}{R} \gg T_{fix}$

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

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

$\Rightarrow$ need $\frac{1}{U_n} \gg N \Rightarrow$ $NU_n \ll 1$ "weak mutation, weak selection"

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

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

$\Rightarrow$ Need $\frac{1}{2NU_b s} \gg \frac{2}{s} \log(Ns)$ or

$NU_b \ll 1/\log(Ns_b)$ "strong selection, weak mutation" ("SSWM")

⇒ what does this look like for some "real" parameter values?

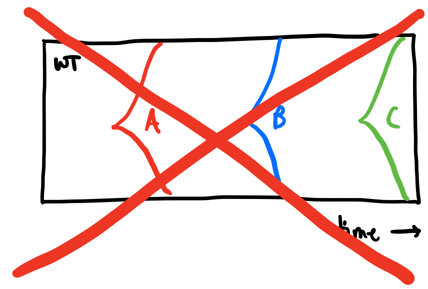
⇒ e.g. HW2 problem 4

$U_b \sim 5\times10^{-6}$, $s_b \sim 0.02$

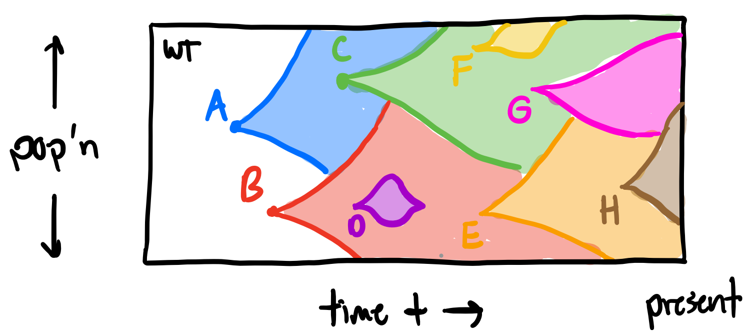
just for L.O.F. muts.

⇒ e.g. if $N \sim 10^5$ ⇒ $4NU_b \log(Ns) \approx 13$

⇒ successive mutations picture does not apply!



⇒ what do things look like instead?



"clonal interference" (will revisit later...)