A simple model of evolution

we now have enough background to start thinking about evolutionary dynamics.

$\Rightarrow$ this is traditionally done by starting w/ an abstract mathematical model. (e.g. in pop gen, largely invented before data; 1920's)

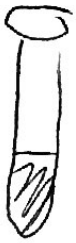
$\Rightarrow$ we're going to take a different approach and try to base our model on experiments we can do in the laboratory.

(this will pay off later, since it will allow us to use operational definitions for quantities that are sometimes difficult to interpret ("fitness", "genetic drift"))

(+ will keep us grounded in some concrete data)

$\Rightarrow$ we need a "population" of organisms that are fast growing & don't take up much space.

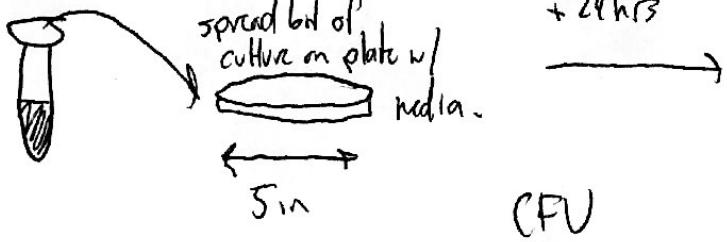
$\Rightarrow$ model microorganisms like E. coli

Can prepare some growth "media" [carbon source, e.g. sugar; salts (N, P, S, etc.); some vitamins]  ~1ml + inoculate w/ lab strain of E. coli (N_0) + 24 hrs →  $N_f \sim 10^7$-10^9 cells/ml (depends on media)

$$\#\text{ of generations} = \log_2\left(\frac{N_f}{N_0}\right)$$

How do we measure N_0, N_f? (in principle hard because need to count tiny things)

(1) Old fashioned way: dilute and grow on plates (Petri dish)



(~100 colonies/plate)

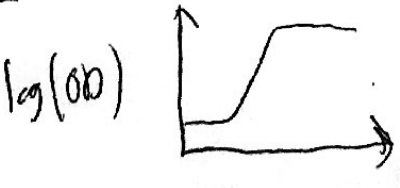
macroscopic "colonies", each inoculated by single cell, can be counted.

$$\#\text{ colonies per plate} \sim \text{Poisson}\left(\underbrace{N_f \times \frac{V_{spread}}{V_{tot}} \times \text{plating efficiency } (P)}_{\text{CFU}}\right)$$

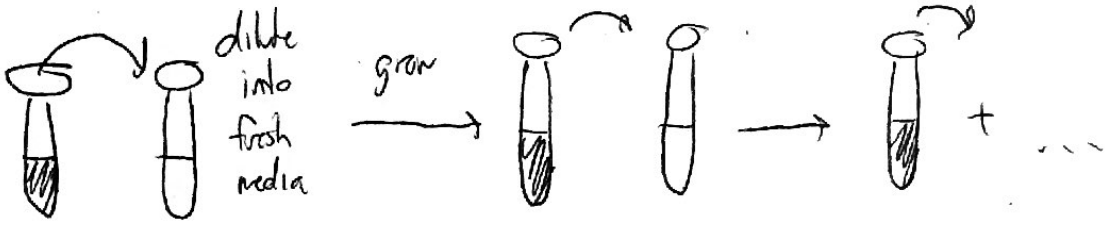
$\frac{V_{spread}}{V_{tot}}$: dilution factor.

→ need conversion factor.

(2) more modern: optical density (measure w/ Lasers)



Basic idea of experimental evolution: we can simulate continually growing pop'n by repeating this process over & over ("serial dilution")



for simplicity, we'll imagine following scenario:

(1) start w N_0 cells grow for fixed time Δt.

$$N(t) = N_0 e^{rt} \rightarrow N_f = N_0 e^{r\Delta t}$$ ($r = \log(2)$ if time measured in gens)

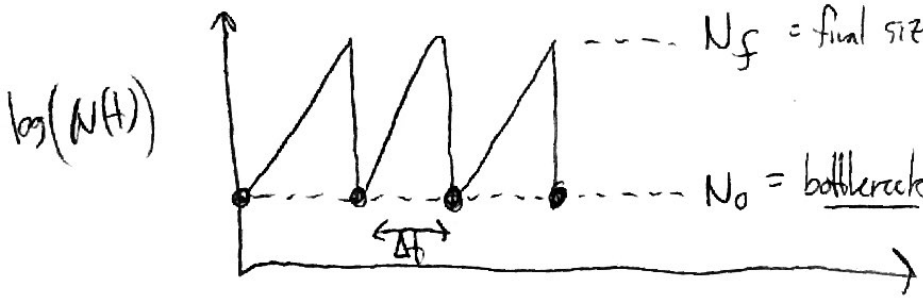
[technically, assumes that Δt is shorter than time where cells start depleting media. we can always set things up such that this will be true — (though in practice, we often don't)]

(2) measure density @ $t = \Delta t$ and choose dilution factor such that we expect $\approx \bar{N}_0$ cells in fresh tube.

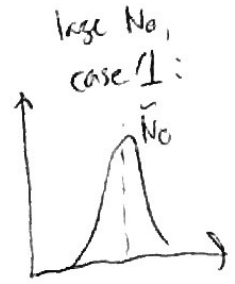
$N_0(k+1)$ = Poisson($\bar{N}_0$) = # of cells in fresh tube at beginning of day.

(3) Repeat over and over.

Visualize pop size over time:



large N_0, case 1:



($\bar{N}_0 \pm \sqrt{\bar{N}_0}$)

$$\#\text{gens per day} = \Delta t = \log_2\left(\frac{N_f}{N_0}\right)$$ ← dilution factor.

e.g. 100-fold dilution = 6.6 gens/day ⇒ 2 weeks ≈ 100 gens
1000-fold dilution = 10 gens/day (2000 yrs in humans)

e.g. 100-fold dilution w/ $N_0 \approx 10^6$ is a reasonable # to have in mind. $10^6 \to 10^8 \to 10^6 \to 10^8$

⇒ this is (purposefully) pretty boring, - just population dynamics.

evolutionary dynamics is about how things change w/in pops.

How do we set this up? let's imagine mixing 2 strains together - in 50-50 ratio.

strain 1 = original strain (WT)
strain 2 = some gene deleted (can't grow on fancy sugar X - not in growth media)
(or e.g. resistance to ABX Y - not in growth media)

Now 2 #s to keep track of $N_1(t), N_2(t)$

(or $N_{tot} = N_1 + N_2$ and frequency $f = \frac{N_2}{N_1+N_2}$)

How do they change over time?

⇒ e.g. suppose that deleting sugar X enzyme frees up some resources (e.g. for ribosomes) & let's strain grow slightly faster in growth media.

e.g. 

$$N_1(t) = N_1(0)\, e^{rt}$$
$$N_2(t) = N_2(0)\, e^{(r+s)t}$$

↳ some empirical param $s > 0$.

ABX example, maybe involves cost w/o ABX, so $s < 0$

if frequency @ beginning of day is $f(0)$, then frequency @ end of day is

$$f(\Delta t) = \frac{N_2(\Delta t)}{N_2(\Delta t) + N_1(\Delta t)} = \frac{N_0 f e^{(r+s)\Delta t}}{N_0 f e^{(r+s)\Delta t} + N_0(1-f)e^{r\Delta t}} = \frac{f e^{s\Delta t}}{f e^{s\Delta t} + (1-f)}$$

# of cells of each type in new flask is then:

$$N_2(k+1) \sim \text{Poisson}\left(N_0 \cdot \frac{f(k)\, e^{s\Delta t}}{1 + f(k)(e^{s\Delta t} - 1)}\right)$$

$$N_1(k+1) \sim \text{Poisson}\left(N_0 \cdot \frac{1 - f(k)}{1 + f(k)(e^{s\Delta t} - 1)}\right)$$

$$f(k+1) = \frac{N_2(k+1)}{N_2(k+1) + N_1(k+1)}$$

"Markov process"

this defines a stochastic process for generating a sequence of frequencies $f_0, f_1, f_2, \ldots$

Simplest case: $s=0$ (no growth rate diffs, or "neutrality")

$$N_2(k+1) \sim \text{Poisson}\left(N_0 f(k)\right)$$
$$N_1(k+1) \sim \text{Poisson}\left(N_0 (1-f(k))\right)$$

$$f(k+1) = \frac{N_2(k+1)}{N_2(k+1)+N_1(k+1)}$$

← still tricky because param of Poisson is random # that depends on frqs @ earlier times

But can derive some properties: ↗ a little tricky to show, but true in this case.

e.g. conditional mean: $E[f(k+1)|f(k)] = f(k)$

$\Rightarrow$ unconditional mean: $E[f(k+1)] = E[f(k)] = E[f(k-1)] \cdots = f(0)$

constant in time!

But in practice, will be fluctuations around this avg value:

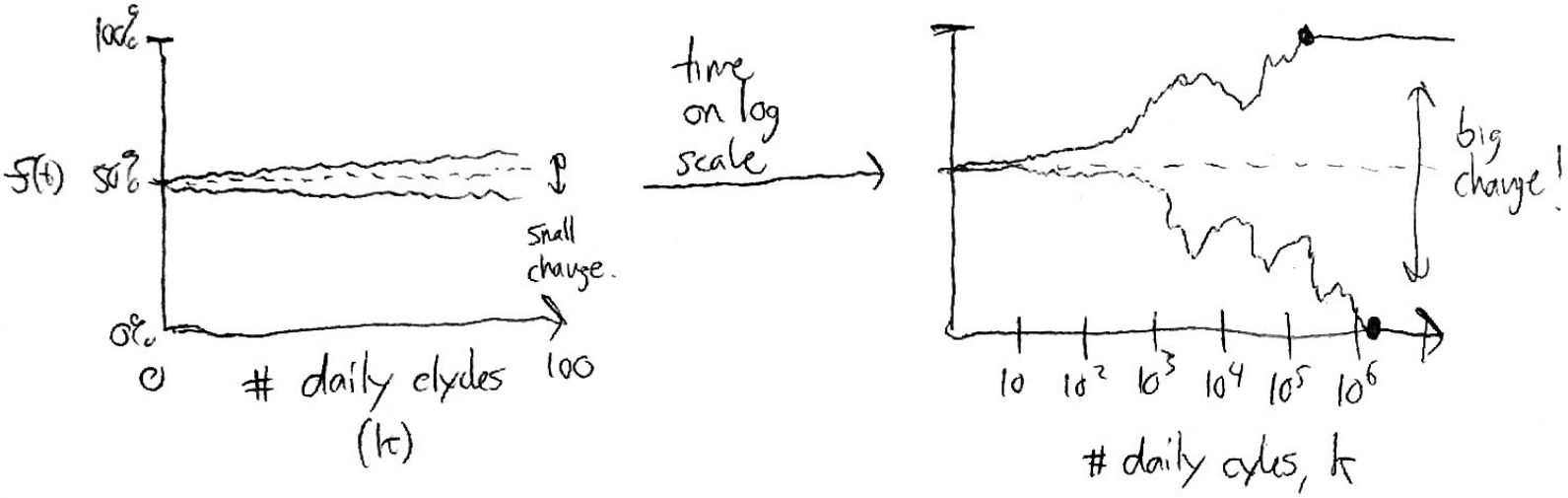
$$f(k+1) \approx \frac{N_0 f \pm \sqrt{N_0 f}}{\left(N_0 f \pm \sqrt{N_0 f}\right) + \left(N_0(1-f) \pm \sqrt{N_0(1-f)}\right)} \approx f(k) \pm O\left(\frac{1}{\sqrt{N_b}}\right)$$

$\Rightarrow$ this is known as <u>genetic drift</u>. in this case, arises purely due to finite sampling @ dilution step.

if N_0 large, drift pretty small ($N_0 \sim 10^5 \Rightarrow \frac{1}{\sqrt{N_0}} \sim 0.3\%$)

$\Rightarrow$ but it is <u>relentless</u>. @ long times,

can see this in computer simulations of our model, e.g. $N_0 = 10^5$, $f(0) = 50\%$, 2 independent replicates



in 2nd case, also notice that something singular happens:

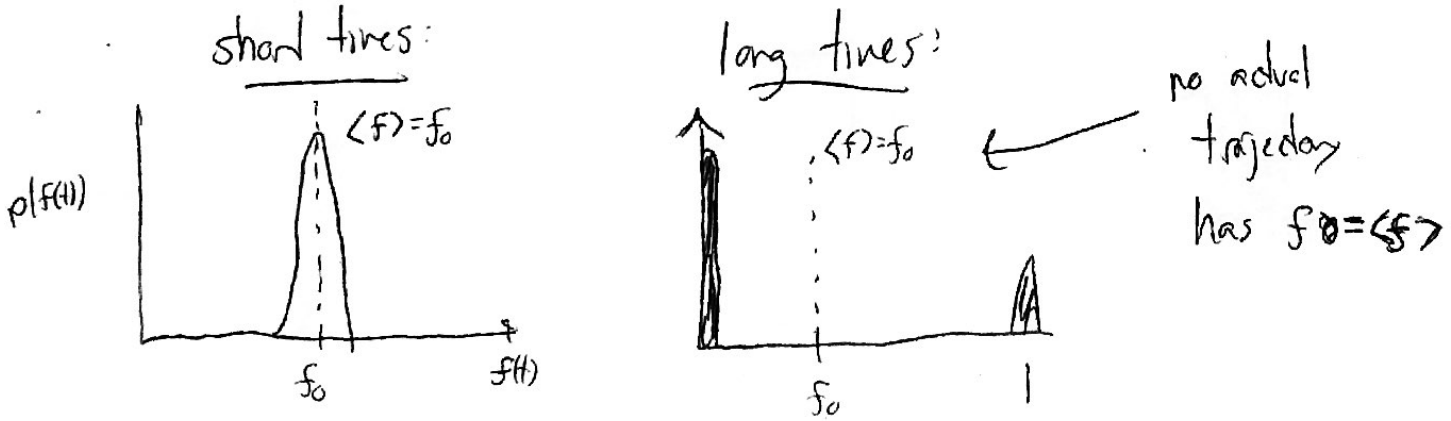
① if $f=0$ @ one time, then $N_2 \sim \text{Poisson}(0) = 0$ @ all later times.

"extinction"

② likewise, if $f=1$ @ one time, $N_1 = \text{Poisson}(0) = 0$ @ all later times

"fixation"

so great illustration of "case 1" & "case 2" distn's from before:



Instead, avg is mixture of 2 outcomes:

$$\langle f \rangle = 0 \times Pr(f=0) + 1 \times Pr(f=1) = f_0$$

(from neutrality assumption)

$\Rightarrow$ can solve for $\boxed{Pr(f=1) = f_0}$

(can also derive from symmetry argument - exchangeability b/w individuals)

the timescale it takes for this is quite long.

↳ will show later that for short times: $f_0(k) \approx f(0) \pm \sqrt{\frac{k}{N_e}}$

"random walk"

$\Rightarrow$ so need $k \sim N_e$ until even start to think about fixation.

(not usually an issue in experiments

e.g. 10^5 days $\approx$ 300 yrs)

$\Rightarrow$ to first approx, drift not relevant for mutations @ high frequency. $\Rightarrow$ instead all about selection

now consider $s>0$, and $N_e = \infty$. (we'll relax this assumption later)

$$f(1) = \frac{f(0)e^{s\Delta t}}{f(0)e^{s\Delta t} + (1-f(0))}; \quad f(2) = \frac{f(1)e^{s\Delta t}}{f(1)e^{s\Delta t} + (1-f(1))} = \frac{f(0)e^{2s\Delta t}}{f(0)e^{2s\Delta t} + (1-f(0))}$$

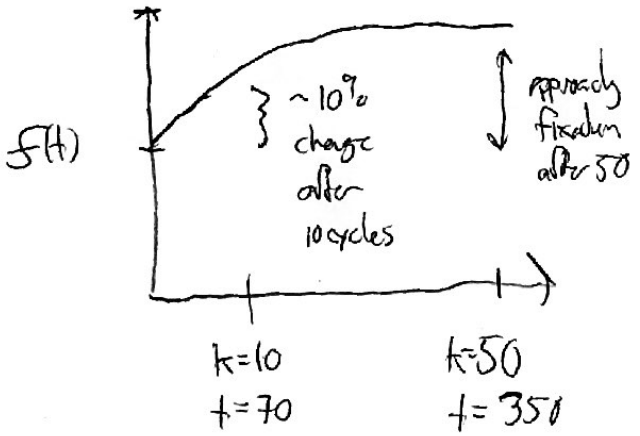
$$1-f(1) = \frac{1-f(0)}{f(0)e^{s\Delta t} + (1-f(0))} \Rightarrow \boxed{f(k) = \frac{f(0)e^{sk\Delta t}}{f(0)e^{sk\Delta t} + (1-f(0))}}$$

e.g. if express time in generations, $t = k\Delta t$

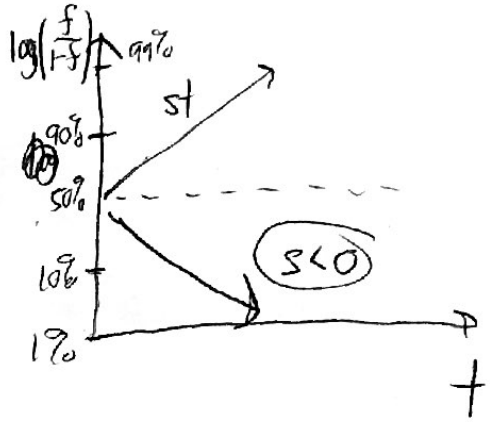
$$\Rightarrow f(t) = \frac{f(0)e^{st}}{f(0)e^{st} + 1 - f(0)} \quad \left[\text{logistic growth, } \partial_t f = f(1-f)\right]$$

now can get big change on lab timescale:

e.g. if $s = 0.01$, $\Delta t = \log_2(100) \approx 7$ (+ $N_e = 10^5$ as before)



On "logit" scale ⟹



can notice change if $st \geq 1$, $t \sim 1/s$ (selection timescale)

↓ big.

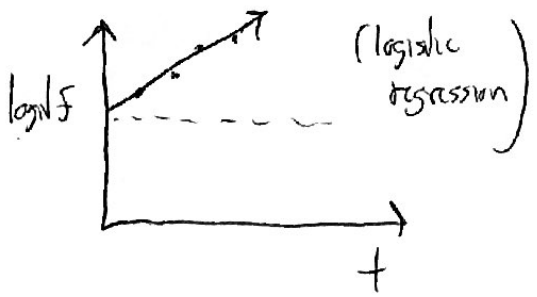
end of lecture 2

⟹ so far, if knew s (e.g. from previous expts on growth rate, $r \to r + s$) can predict $f(t)$

⟹ can turn around and use as definition of s:

$$s = \frac{1}{t}\log\left(\frac{f(t)}{1-f(t)} \cdot \frac{1-f(0)}{f(0)}\right)$$

more than 2 timepoints ⟹



S = "fitness difference" (strictly speaking "competitive fitness")

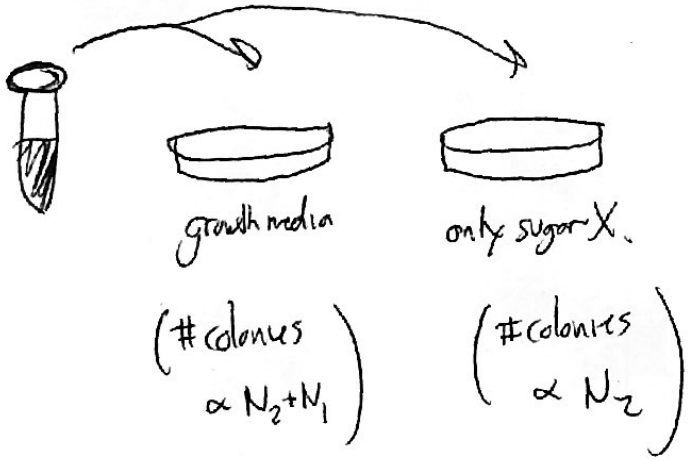
$\Rightarrow$ in this case, we have defined something fuzzy like "fitness" purely operationally based on changes in relative frequency w/in a population.

$\Rightarrow$ in practice, means we can still measure S even when underlying model is different from one we consider here ($r \rightarrow r+s$)

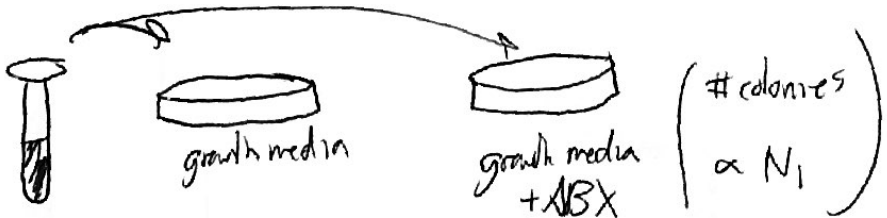
How do we measure $f(t)$? (in principle, hard to distinguish similar looking strains like WT, ΔsugarX)

(1) old fashioned: make them distinguishable & count colonies.

e.g. ΔsugarX $\Rightarrow$

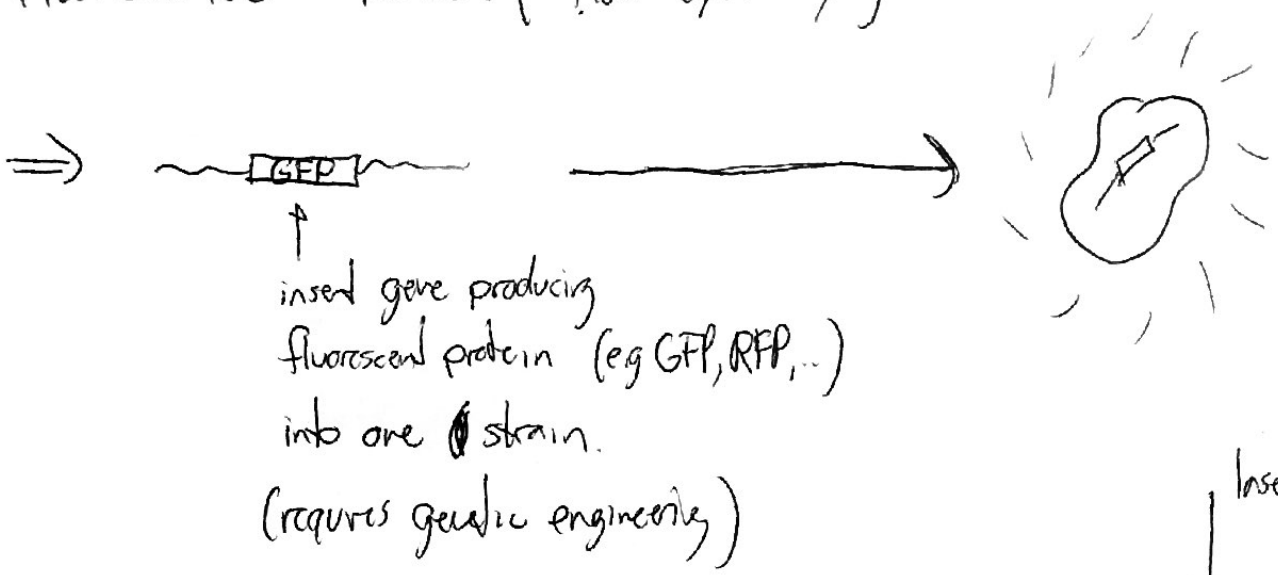


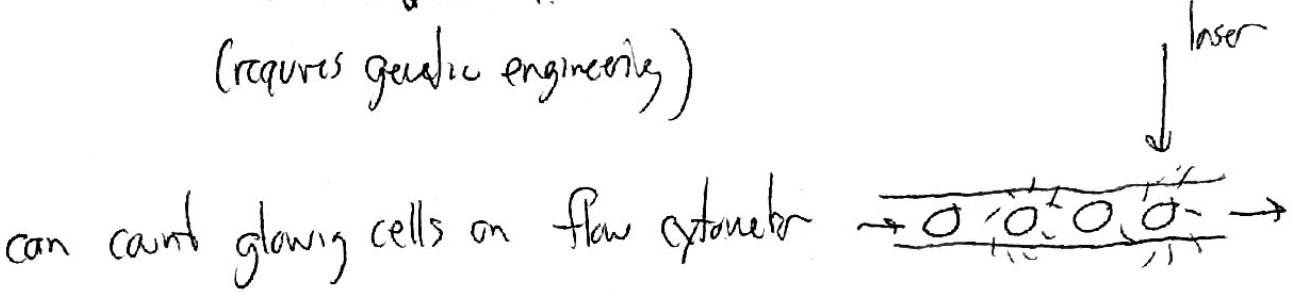
e.g. ABX resistance $\Rightarrow$



more modern:

(2) fluorescence + lasers. (flow cytometry)



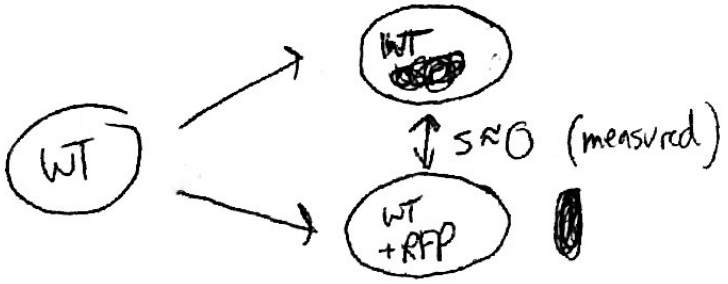


can count glowing cells on flow cytometer

96 well plate/hr (~50,000 cell counts/well)

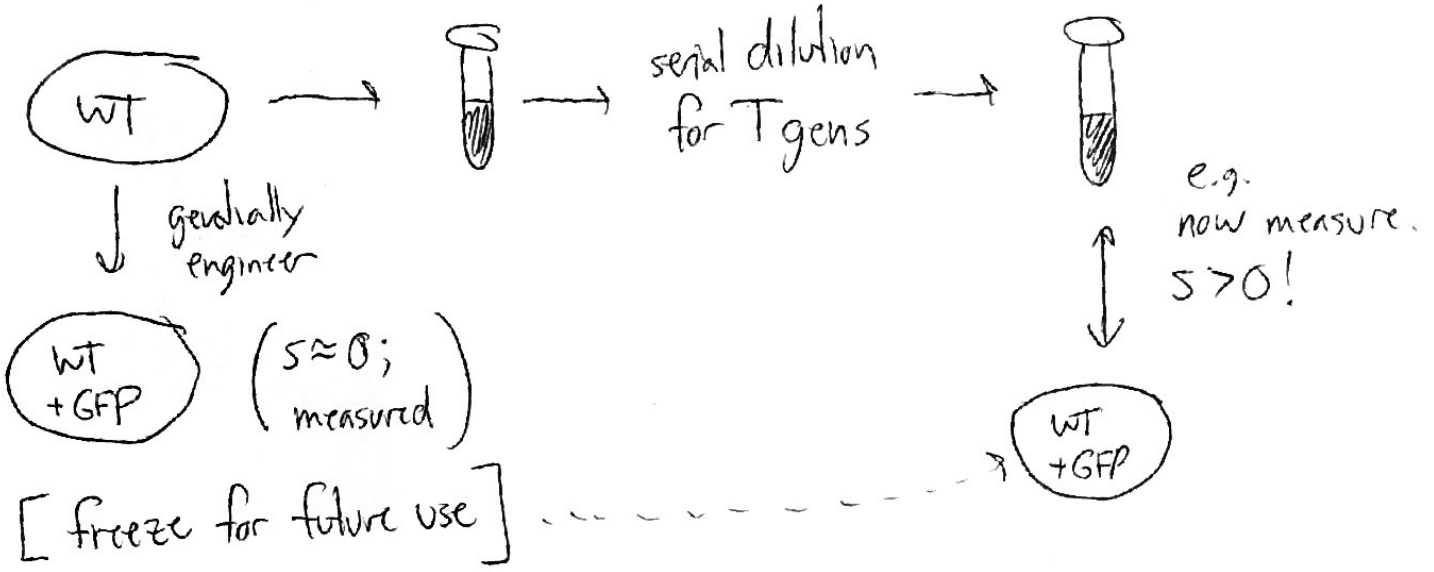
(3) DNA sequencing (will discuss more later)

So now we have way of defining fitness.



So now have way of defining & measuring fitness operationally

⇒ let's consider following experiment:



⇒ must be due to mutations that arose in population during experiment. How to model this process?

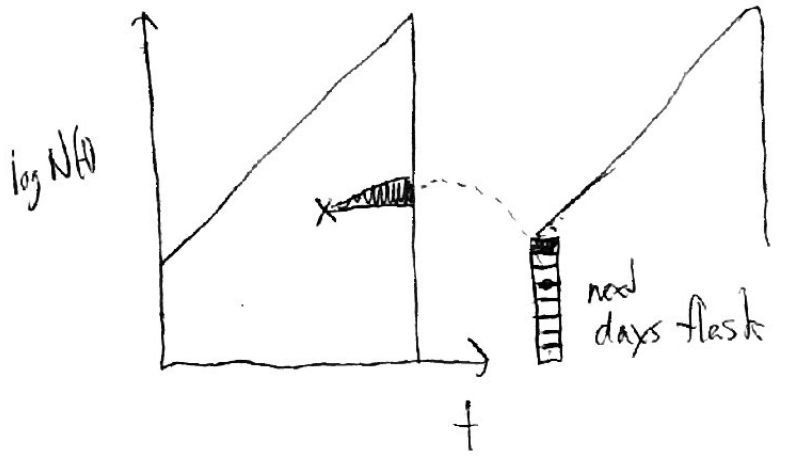
First: suppose there is just a single target for mutations (e.g. WT → ΔsugarX.) that happens w/ probability μ per division. ($\mu \ll 1$)

this is called "single locus" model. (genome w/ a single site ⊸↓⊸)

↳ can learn a lot about evolution from studying this simple case ⇒ will learn how to generalize to bigger genomes later)

Start w/ no mutants in population.

⇒ then at some timepoints during grow-up phase



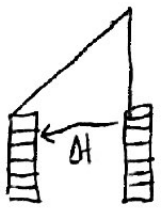
$(f>0) \Rightarrow$ then regular model applies.

distribution of f @ beginning of next day is actually tricky problem (Luria-Delbrück distribution, homework 1)

For simplicity will use following approx: (will show later that it's good one)

(1) mutation doesn't exert fitness benefit until next day's passage (not such a bad assumption biologically... e.g. ΔsugarX, need few gens to dilute out old protein)

(2) every cell at beginning of today's flask traces back to cell alive @ beginning of previous days flask.



by definition, Δt generations between them (Δt divisions)

so probability that single cell acquires mutation is

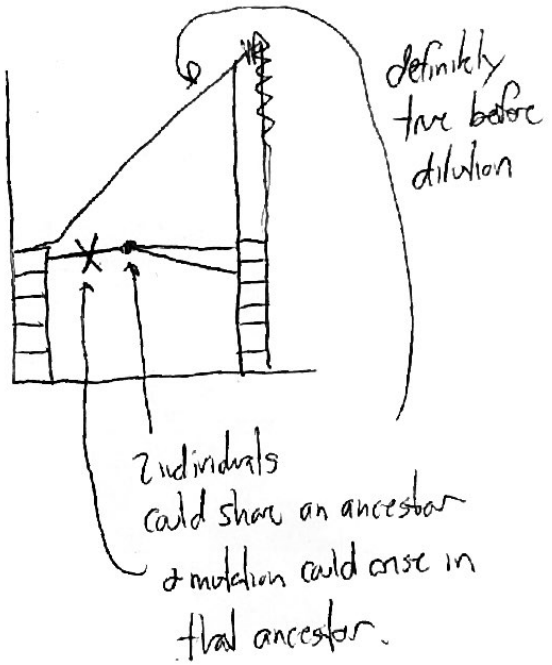
$$P_{mut} = \nu \Delta t$$

$\Rightarrow$ approx is that [redacted] cells acquire mutations $\approx$ independently

$$N_2 \sim \text{Poisson}(N_0 P_{mut})$$
$$N_1 \sim \text{Poisson}(N_0 (1 - P_{mut}))$$

$$\Rightarrow f = \frac{N_2}{N_2 + N_1}$$

why is this not quite right?



will explore consequences of this simple fact in Homework 1

$\Rightarrow$ if $f(k) > 0$ then

$$N_2 \sim \text{Poisson}\left(N_0 \frac{f(k) e^{s\Delta t}}{f(k) e^{s\Delta t} + (1 - f(k))}\right) + \text{Poisson}\left(N_0 P_{mut} \left([redacted] \frac{1 - f(k)}{f(k) e^{s\Delta t} + (1 - f(k))}\right)\right)$$

$$N_1 \sim \text{Poisson}\left(N_0 (1 - P_{mut}) \left(\frac{1 - f(k)}{f(k) e^{s\Delta t} + (1 - f(k))}\right)\right)$$

if you want, could add back-mutation (mut $\to$ wt) at rate ν as well.