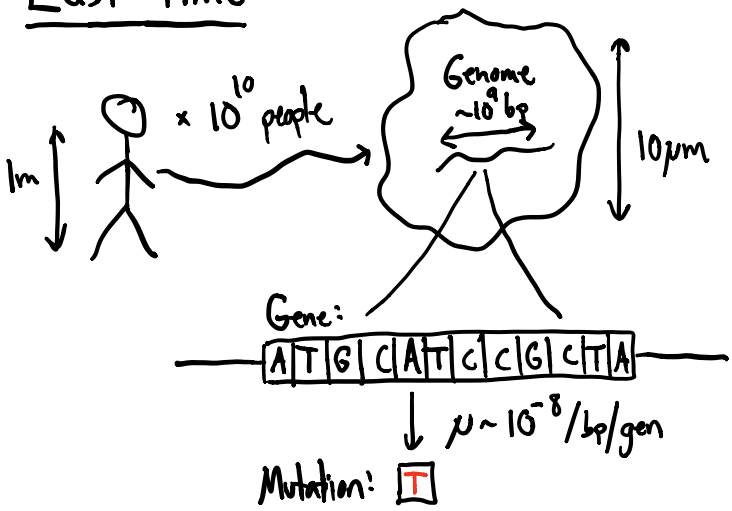
Last time:



"Fermi problem" (mutation supply)

(# individuals in population) × (Pr[mutation] per site per generation) = (# new mutations produced in pop'n per site per generation)

E.g. Humans: $N \sim 10^{10}$ × $\mu \sim 10^{-8}$ = ~ 100 /bp/gen

Empirical observation:

Avg # differences between my genome and yours is $\sim 10^{-3}$ /bp

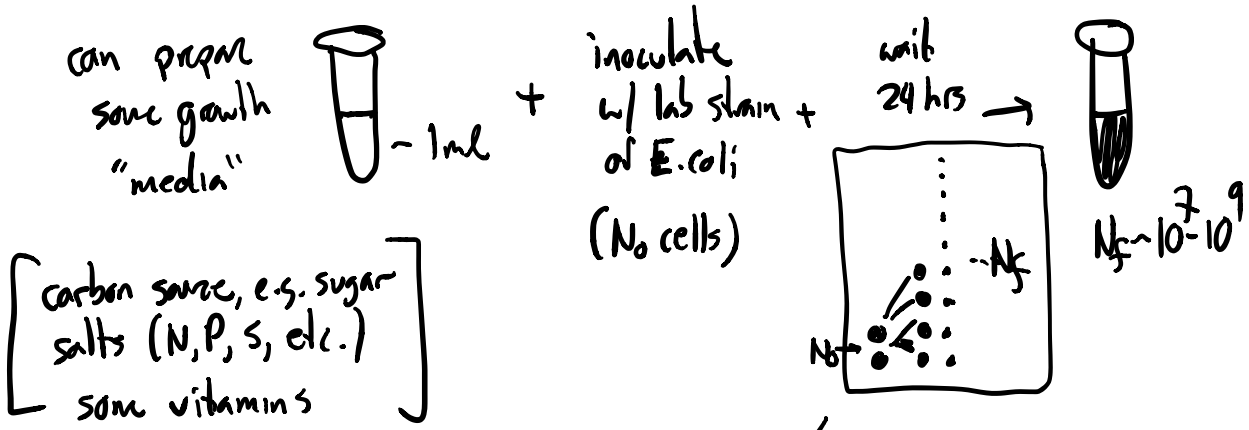
How do we connect these 2 observations?

↓

Evolutionary dynamics!

Today: A Simple Model of Evolution

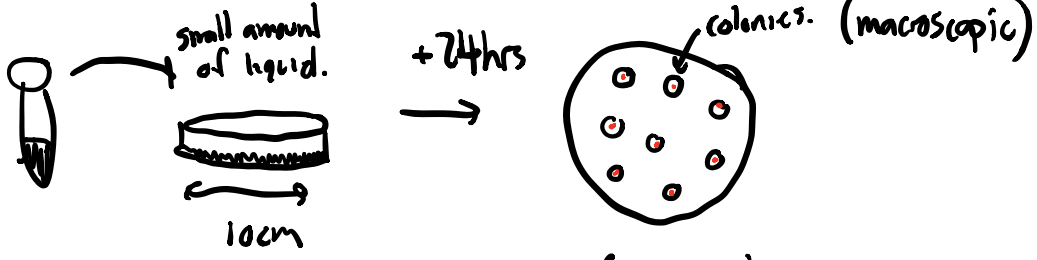
⇒ need a "population" ⇒ model microorganisms (E. coli)



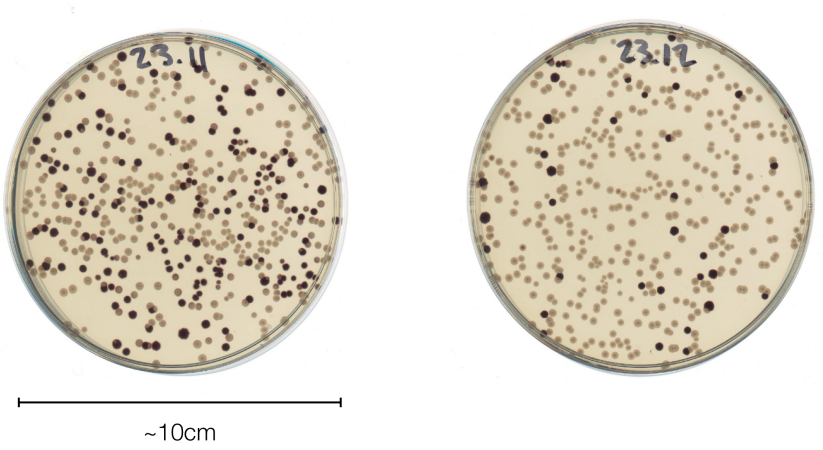
$$\#\ \text{generations} = \log_2\left(\frac{N_f}{N_0}\right)$$

How can we measure N_0, N_f?

① Old fashioned way: diluting & growing on plates (Petri dish)



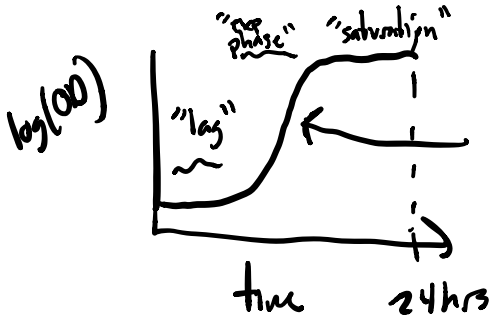
Example: *E. coli* colonies on plates



colonies on plate $\sim$ Poisson $\left(N_f \times \frac{V_{spread}}{V_{tot}} \times \text{plating efficiency, } \rho \right)$

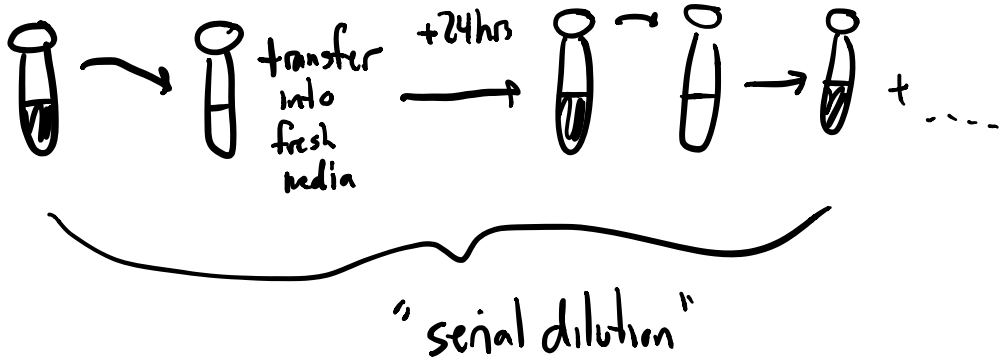
- # colonies on plate: ↓ measure
- V_{spread}/V_{tot}: $\sim$ ↓ measure
- plating efficiency, ρ: ↓ scaling factor.
- N_f: ↑ can inferring N_f

(2) More modern: measure "optical density" ("OD") (measure w/ "Lasers")



growth curve. ⇒ e.g. KC Huang's group.

Basic idea of experimental evolution:



For simplicity, imagine the following scenario:

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

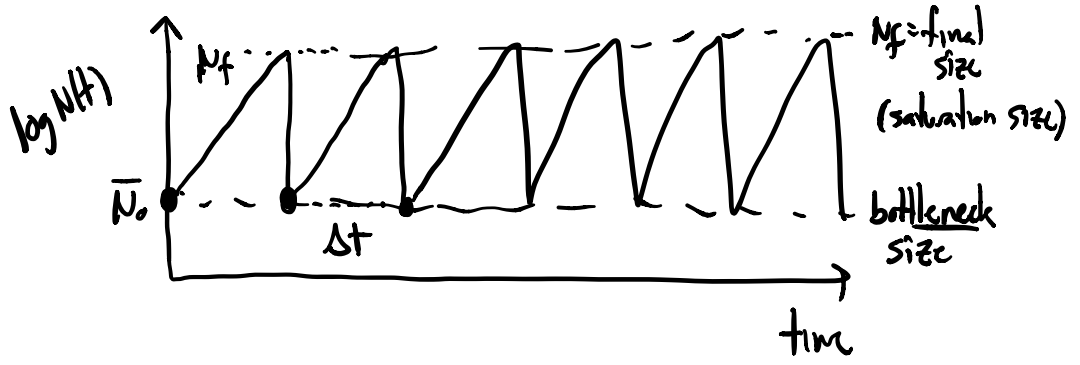
$$N(t) = N_0 e^{rt} \longrightarrow N_f = N_0 e^{r\Delta t}$$

"growth rate" $(r = \log(2)$ if Δt in generations$)$

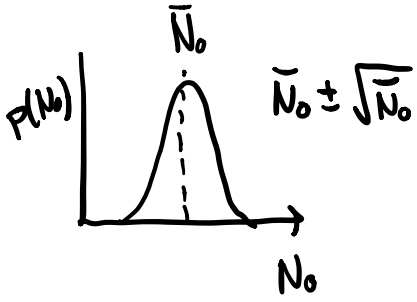
(2) measure density @ time $\Delta t \Rightarrow$ choose dilution factor such that expect $\simeq \bar{N}_0$ cells in fresh tube.

$$N_0(k+1) \sim \text{Poisson}(\bar{N}_0) = \#\text{ cells in fresh tube on day } k+1$$

(3) Repeat. (over & over)



$$\#\text{ gens} \sim \Delta t = \log_2\left(\frac{N_f}{N_0}\right)$$ → "dilution factor"

randomness: 

"case 1" dist'n.

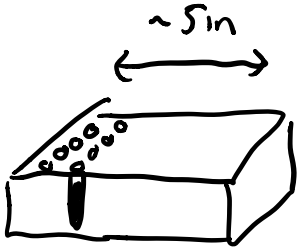
e.g. 100-fold dilution ⇒ 6.6 gens/day ⇒ 2 weeks to get 100 gens.

1000-fold ⇒ 10 gens/day

↳ $N_0 \approx 10^6$ is reasonable #

→ $N_f \approx 10^8$ cells

⇒ not just test tube ⇒ "96 well plates"



How do we think about evolution in this scenario?

let's imagine mixing 2 E. coli strains together in 50-50 ratio

strain 1: normal lab strain (WT)

strain 2: = some gene deleted ("Δsugar X") (can't grow on some fancy sugar X → not in growth media).

(e.g. resistance to ABX Y not in growth media)

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

[or $N_{tot}(t) = N_1(t) + N_2(t)$ ⇒ look @ frequency $f = \frac{N_2}{N_1 + N_2}$]

⇒ e.g. $N_1(t) = N_1(0)\, e^{rt}$

$N_2(t) = N_2(0)\, e^{(r+s)t}$ ⟵ some empirical parameter $s > 0$

⇒ if freq @ beginning of day is $f(0)$

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

$\Rightarrow$ # of cells of each type @ beginning of <u>next</u> day:

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

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

$$f(k+1) \sim \frac{N_2(k+1)}{N_1(k+1)+N_2(k+1)} \Rightarrow f_0, f_1, f_2, \ldots, f_k$$

"Markov process"

"simple model of evolution"

$\Rightarrow$ simplest case: $s=0$ (no growth rate diffs.)
"neutrality"

$$N_2(k+1) \sim \text{Poisson}(N_0 f(k))$$

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

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

$\Rightarrow$ can derive some properties:

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

$\Rightarrow$ unconditional mean:

$$E[f(k+1)] = \sum_{f(k)} E[f(k+1)|f(k)]\,p(f(k)) = E[f(k)]$$
$$= E[f(k-1)]$$
$$= f_0$$

constant in time!

$\Rightarrow$ in practice: fluctuations around avg value

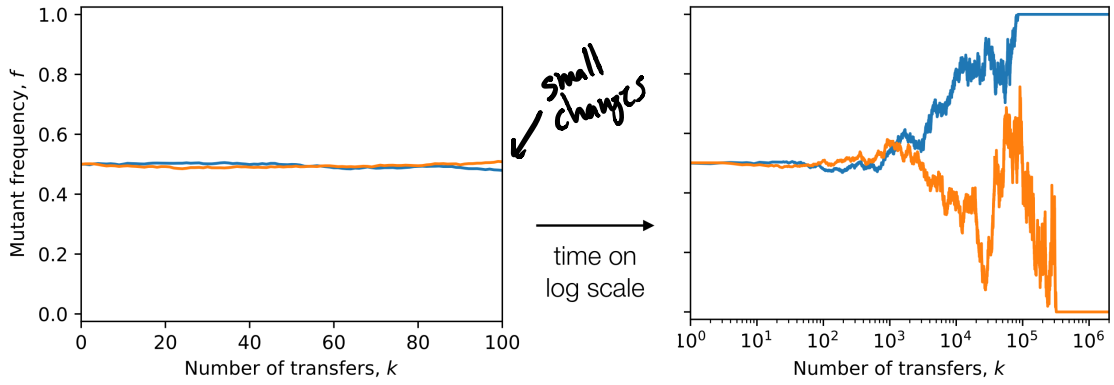
$$f(k+1) = \frac{N_0 f \pm \sqrt{N_0 f}}{\left(N_0 f \pm \sqrt{N_0 f}\right) + \left(N_0(1-f) \pm \sqrt{N_0(1-f)}\right)} \overset{\text{Taylor expand}}{\approx} f(k) \pm \underbrace{\sqrt{\frac{C}{N_0}}}_{\text{"genetic drift"}}$$

$\Rightarrow$ if N_0 is large $\Rightarrow$ drift is pretty small!

$\left(N_0 \sim 10^5 \text{ cells}, \frac{1}{\sqrt{N_0}} \sim 0.3\%\right)$

$\Rightarrow$ but it is relentless (compounds over time)

Computer simulations of model with $s = 0$, $N_0 = 10^5$, $f(0) = 50\%$



⇒ in 2nd case, something "singular" happens:

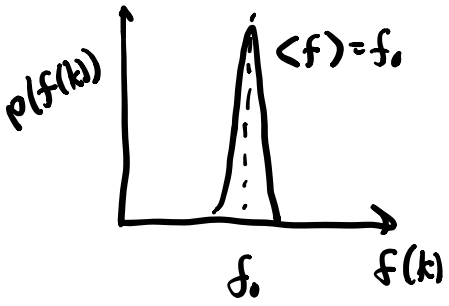
① if $f=0$ @ one time ⇒ $f=0$ @ all later times t.

② if $f=1$ @ " ⇒ $f=1$ @ all "

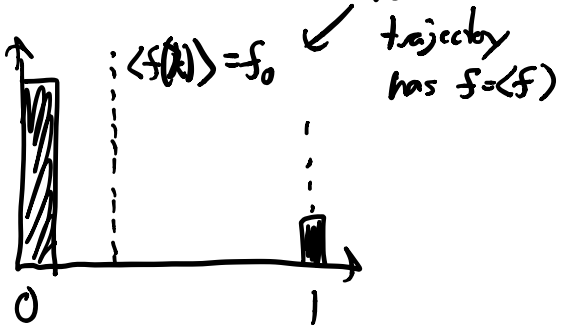
↳ "fixation"

↳ "extinction"

Short times:



Long times:



instead:

$$\langle f \rangle = 0 \times \underbrace{Pr(f=0)}_{1-Pr(f=1)} + 1 \times Pr(f=1) = f_0$$

↗ from neutrality

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

$\Rightarrow$ timescale for this is quite long.

$\Rightarrow$ will show for short times $f(k) \approx f_0 \pm \sqrt{\frac{kC}{N_0}}$

"random walk"

$\Rightarrow$ need $k \sim N_0$ until we can start to think about fixation.

$\Rightarrow$ if $N_0 \sim 10^5 \Rightarrow 10^5$ days $\sim$ 300 yrs.

$\Rightarrow$ genetic drift is weak $\Rightarrow$ all about <u>selection</u>

for $f_0 = 50\%$ on lab timescales

Now consider $s>0$, for simplicity $N_0=\infty$ (no drift for now)

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

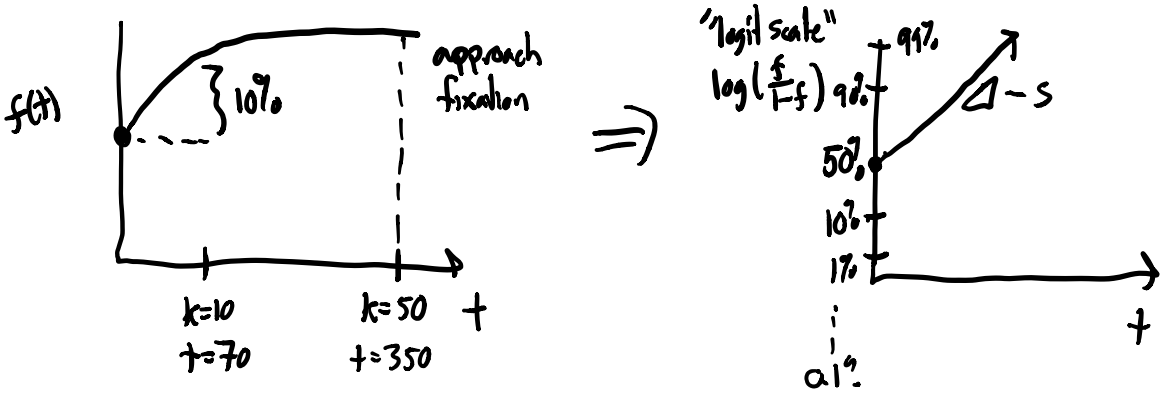
$$\Rightarrow \boxed{f(k)=\frac{f(0)e^{sk\Delta t}}{f(0)e^{sk\Delta t}+(1-f)}}$$

$\Rightarrow$ if time t in generations, $t=k\Delta t$

$$\boxed{f(t)=\frac{f(0)e^{st}}{f(0)e^{st}+1-f}} \Leftrightarrow \boxed{\text{logistic growth} \quad \partial_t f = sf(1-f)}$$

now can get by change:

e.g. if $s=0.01$, $\Delta t = \log_2(100)=7$ (+ if $N_0=10^5$ as before)



can notice a big change when $st \gtrsim 1 \Rightarrow t \gtrsim 1/s$

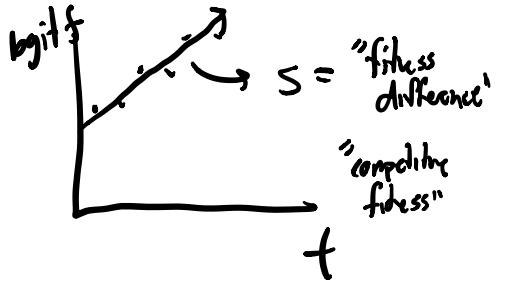
("selection timescale")

$\Rightarrow$ so far: if know $s \Rightarrow$ predict $f(t)$

$\Rightarrow$ can turn around & use as <u>definition</u> of s.

$\Rightarrow$ if measure $f(t) \Rightarrow s = \frac{1}{t} \log\left(\frac{f(t)}{1-f(t)} \cdot \frac{1-f(0)}{f(0)}\right)$ (2 timepoints)

$\Rightarrow$ multiple timepoints



How do we measure $f(t)$?

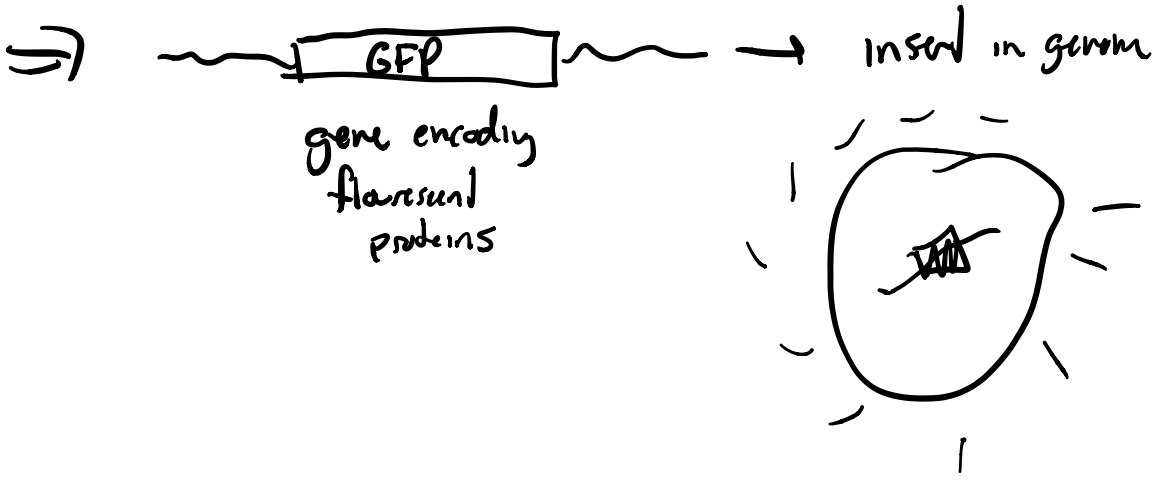
① old fashioned: make them distinguishable & count colonies.

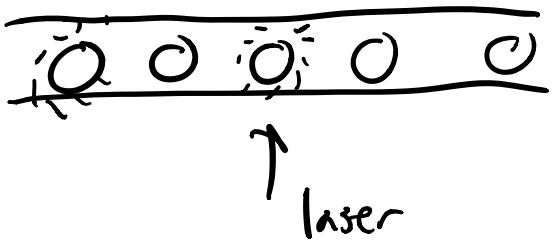
e.g. Δsugar X

growth media. (N_2+N_1)

only sugar X (N_1)

(2) flourescence + lasers (flow cytometry)





⇒ 96 well plate
~ 1 hr
~ 50,000 cell counts/well

(3) DNA sequencing (later)

⇒ consider the following experiment: