

Initial focus:

- 1) neutral mutations (have no effect on fitness)

example: • many mutations in portions of genome that do not code for proteins [98% of human genome]

- even in sections that code for proteins, there are synonymous mutations: mutation in the third letter of the 3 letter code $\Rightarrow$ same amino acid

GC	G	$\Rightarrow$	arginine
GC	C		"
GC	T		"
GC	A		"

- 2) every mutation leads to a new genetic variant (alleles)
 $\Rightarrow$ "infinite alleles model"

imagine a locus of length L

4^L possible alleles

$L=100$

$\Rightarrow 4^L = 10^{60}$

initial string of L letters

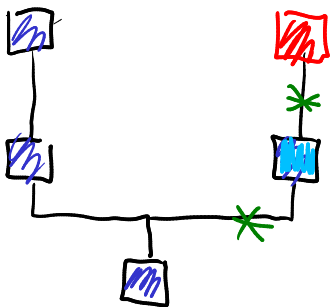
$\Rightarrow 3L$ "neighboring" strings differing by a single letter

$L = 100 \Rightarrow 300$ neighbors

$\frac{1}{3L}$ chance of returning via mutation to starting string
(small for large L)

Measure genetic diversity in current generation:

heterozygosity H = prob. that at a given locus any two alleles chosen at random are diff. in current generation



Rephrase definition of H :
for a pair of alleles in current gen., prob. that a mutation occurred in one or both lineages before coalescence

Going back one generation:

$\frac{1}{2N}$ = prob. that lineages coalesce going back one gen.

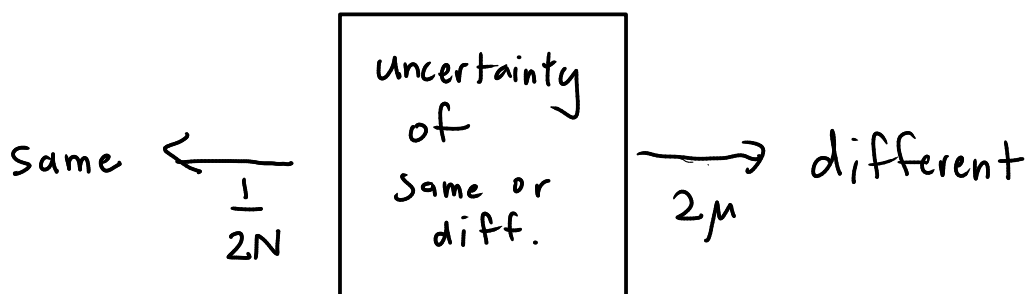
(first allele has some parent;
prob. that second allele has
same parent = $\frac{1}{2N}$)

$\mu \equiv$ prob. of a mutation in prev.
generation along a lineage line

$1 - \underbrace{(1-\mu)^2}_{\substack{\text{prob. of} \\ \text{no mutation} \\ \text{in both lineages}}}$ = prob. of at least one mutation
in the two lineages
going back one gen.

$\mu \ll 1$ typically: $\mu = 2.6 \times 10^{-8} / \text{bp} / \text{gen}$

$1 - (1-\mu)^2 \approx 2\mu$ Taylor expansion in humans



from our Markov state theory:

prob. of exiting thru
"different" door

$$H = \frac{2\mu}{\frac{1}{2N} + 2\mu}$$

$$\Theta \equiv 4N\mu$$

measure H

$\Rightarrow$ extract Θ

measure μ

$$= \frac{4N\mu}{1 + 4N\mu}$$

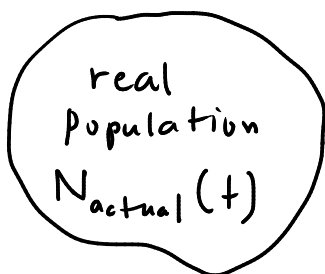
$$= \frac{\Theta}{1 + \Theta}$$

humans; $\Theta \sim 1.1 - 2 \times 10^{-3}$ for $L=1$
 $\mu \sim 2.6 \times 10^{-8}$

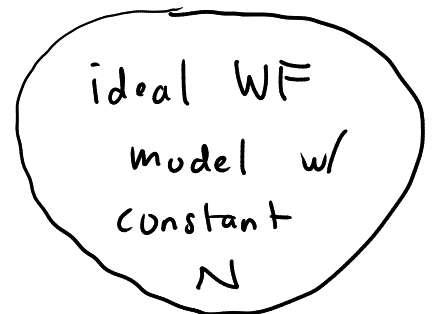
$$\Rightarrow N = 1 - 2 \times 10^4$$

Compare to $N_{\text{actual}} = 8 \times 10^9$

$\Rightarrow N$ in WF model $\equiv$ effective population size



$\Rightarrow$
mathematical
map



large $N \Rightarrow$ large H

small $N \Rightarrow$ small H

$\Rightarrow N$ as a
measure of gen.
diversity

"size of gene pool"

Focus on another quantity:

prob. that two lineages of a pair of alleles have not coalesced going back n generations

$$U_c(n) = \left(1 - \frac{1}{2N}\right)^n \quad \text{in ideal WF}$$

$$\begin{aligned} t = n t_g &\Rightarrow \left(1 - \frac{1}{2N}\right)^{\frac{t}{t_g}} \\ &\downarrow \\ &\text{time per} \\ &\text{gen.} \approx 20 \text{ yrs} \\ &\text{for humans} \end{aligned} \quad \begin{aligned} &\approx \left(e^{-\frac{1}{2N}}\right)^{\frac{t}{t_g}} \\ &= e^{-\frac{t}{2N t_g}} \end{aligned} \quad N \gg 1$$