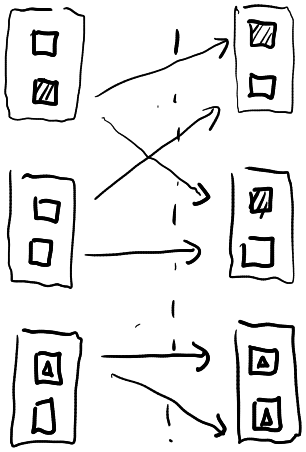
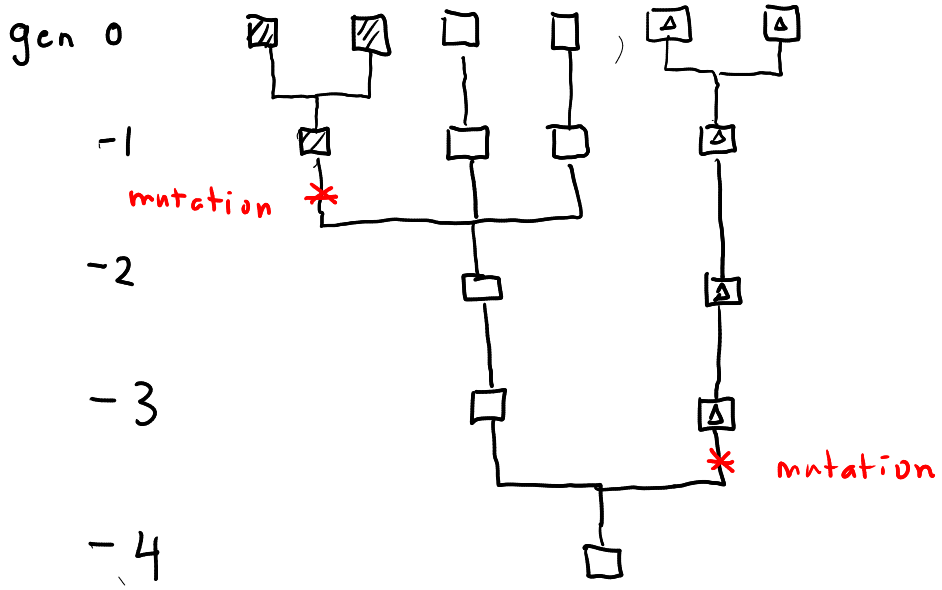
WF model: look at one generation



2N alleles

family tree:



MRCA:

most recent common ancestor for all alleles in curr. gen.

initial focus: 1) neutral mutations (have no effect on fitness)

example: GCG / GCC / ... ⇒ arginine

mutations in third site of the sequence for each amino acid

⇒ same amino acid

⇒ same protein ⇒ little or no biological effect

2) every mutation leads to a new genetic variant (new allele)

$\Rightarrow$ "infinite alleles" model

imagine a locus of $L \gg 1$ sites

4^L possible alleles

$L = 100 \Rightarrow 4^L = 10^{60}$

for a single mutation there are

$3L$ possibilities

chance of undoing a single mutation (going back to original) $= \frac{1}{3L}$

$\Rightarrow \frac{1}{300}$ for $L = 100$

(generally small for large L)

3) probability of mutation in a lineage going back one generation $= \mu \ll 1$

bp = base pair

$\mu = 2.6 \times 10^{-8}$ / bp / gen

Introduce a variable to measure genetic diversity at a locus in current gen:

heterozygosity H = prob. that any two alleles chosen at random from current gen. are diff.

= rephrase: did a mutation occur before coalescence for lineages of these two alleles?

Choose two alleles in curr. generation.

Going back each gen:

$\frac{1}{2N}$ = prob. that lineages coalesced in previous gen.

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

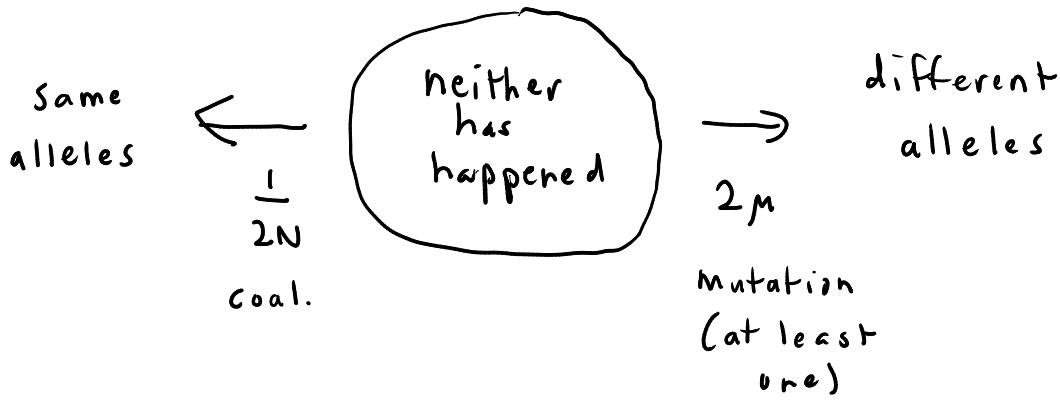
$1-(1-\mu)^2$ = prob. that there is at least one mutation in the two lineages going back one gen.

$(1-\mu)^2$: prob. of no mut. along both lines

$\mu \ll 1 \Rightarrow 1-(1-\mu)^2 \approx 2\mu$

Taylor series

Prob. that one happens before the other as you go back in time in tree:



$\mathcal{H}$ = prob. of mut. before coalescence

$$= \frac{2\mu}{\frac{1}{2N} + 2\mu} = \frac{4N\mu}{1 + 4N\mu}$$