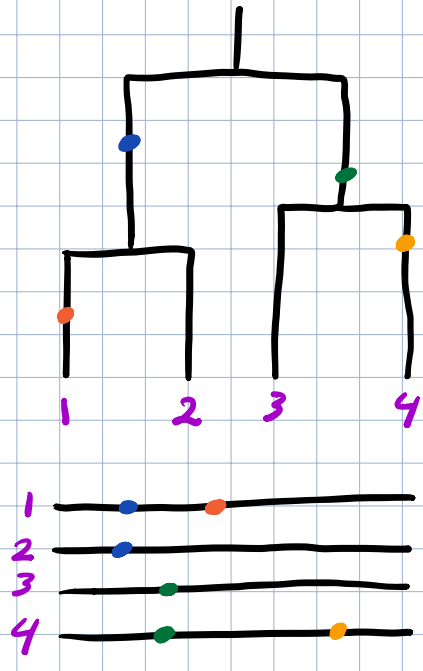
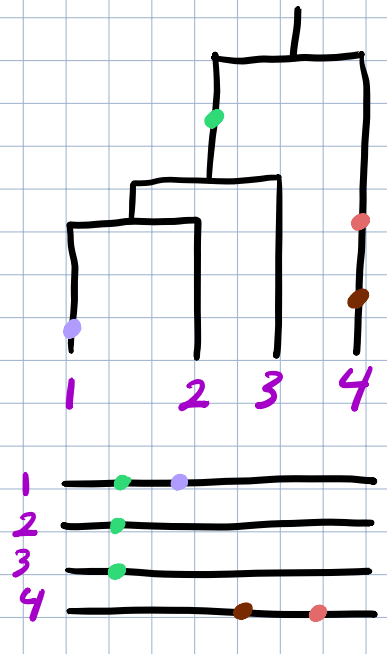
# Genetic variation and the coalescent

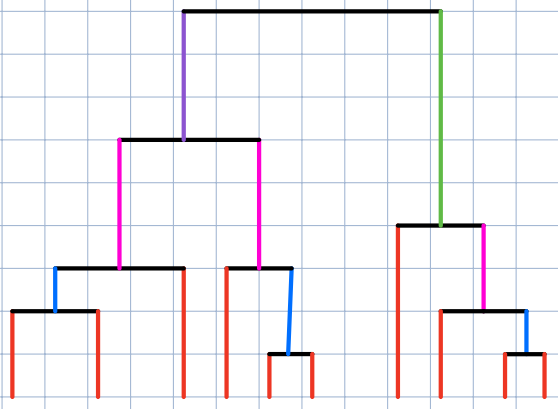
# Genetic variation and the coalescent

- Mutations "fall randomly" on the branches of each coalescent tree
- An individual's sequence in a particular part of the genome depend on the set of mutations that fall on branches of the coalescent tree that are ancestral to it
- the likelihood that a mutation will fall on a particular branch is proportional to its length
- tree shape determines what patterns of variation are possible



# Tree shape determines the distribution of allele frequencies



$\tau_k$ = total length of branches that have $k$ descendants in the sample

$\mu$ = mutation rate

$$\mathbb{E}[\#\text{ of singleton mutations} \mid \text{tree}] = \mu\tau_1$$

$$\mathbb{E}[\#\text{ of doubleton mutations} \mid \text{tree}] = \mu\tau_2$$

$$\mathbb{E}[\#\text{ of tripleton mutations} \mid \text{tree}] = \mu\tau_3$$

$$\mathbb{E}[\#\text{ of quadrupleton mutations} \mid \text{tree}] = \mu\tau_4$$

$$\mathbb{E}[\#\text{ of 5-ton mutations} \mid \text{tree}] = 0$$ ← because $\tau_5 = 0$ in this tree

$$\mathbb{E}[\#\text{ of 6-ton mutations} \mid \text{tree}] = \mu\tau_6$$

# Expected distribution of allele frequencies depends on average tree shape

$$\mathbb{E}[\#\text{ of } k\text{-ton mutations}] = \sum_{\text{trees}} \mathbb{E}[\#\text{ of } k\text{-ton mutations} \mid \text{tree}]\, P(\text{tree})$$

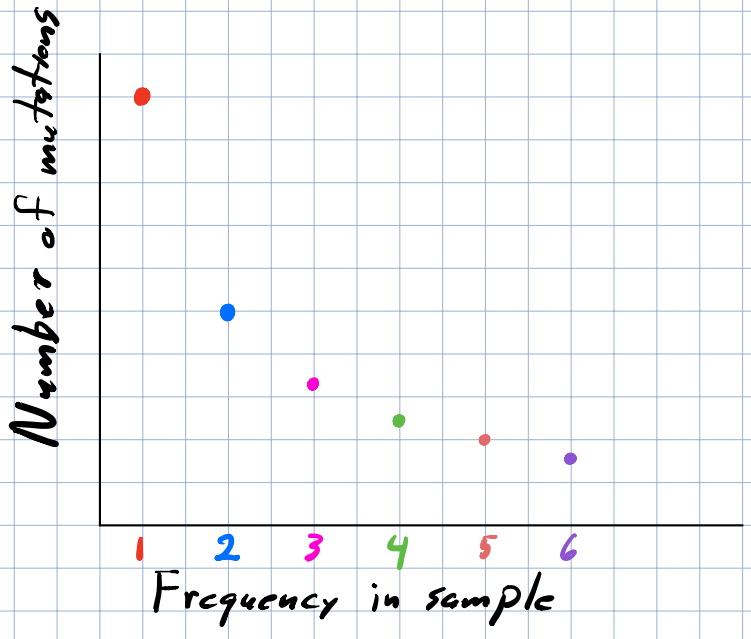
$$= C\mu\, \mathbb{E}[\tau_k]$$

$$\mathbb{E}[\tau_2] = \frac{\mathbb{E}[\tau_1]}{2}$$

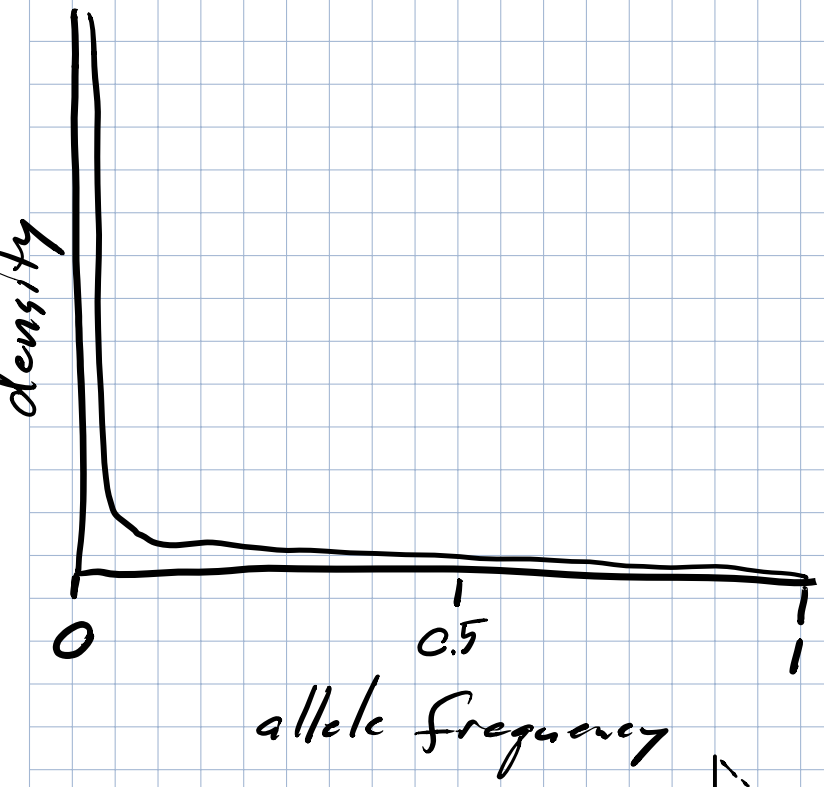
$$\mathbb{E}[\tau_3] = \frac{\mathbb{E}[\tau_1]}{3}$$

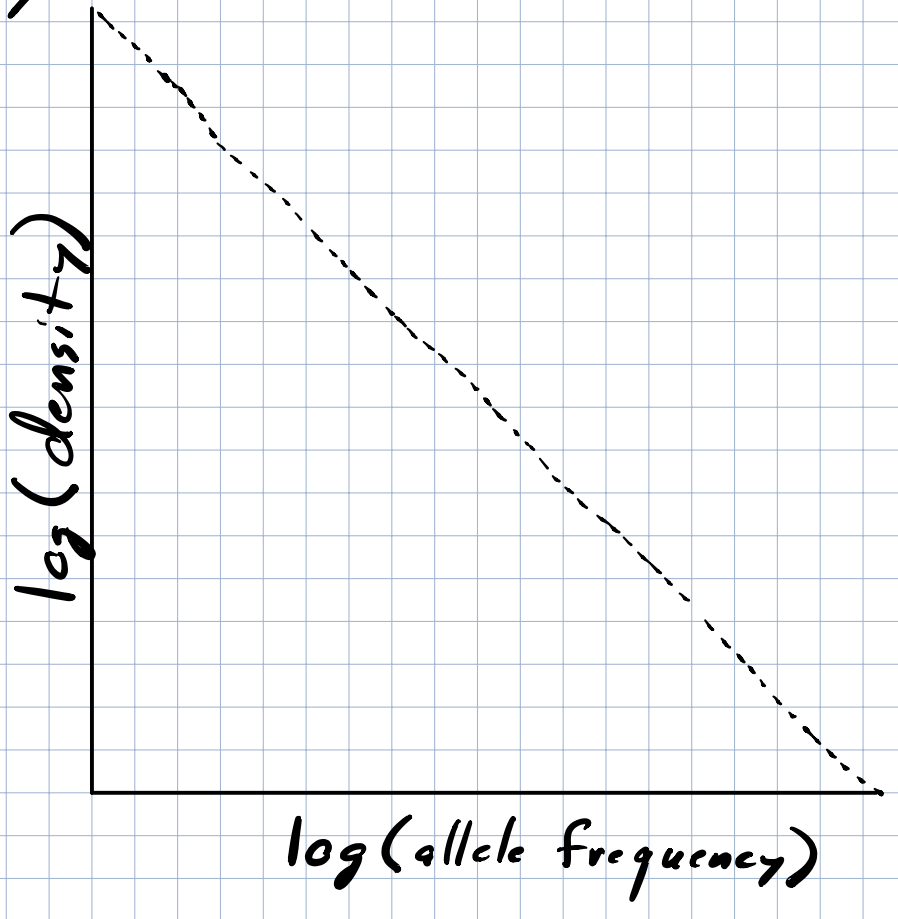
$$\vdots$$

$$\mathbb{E}[\tau_k] = \frac{\mathbb{E}[\tau_1]}{k}$$



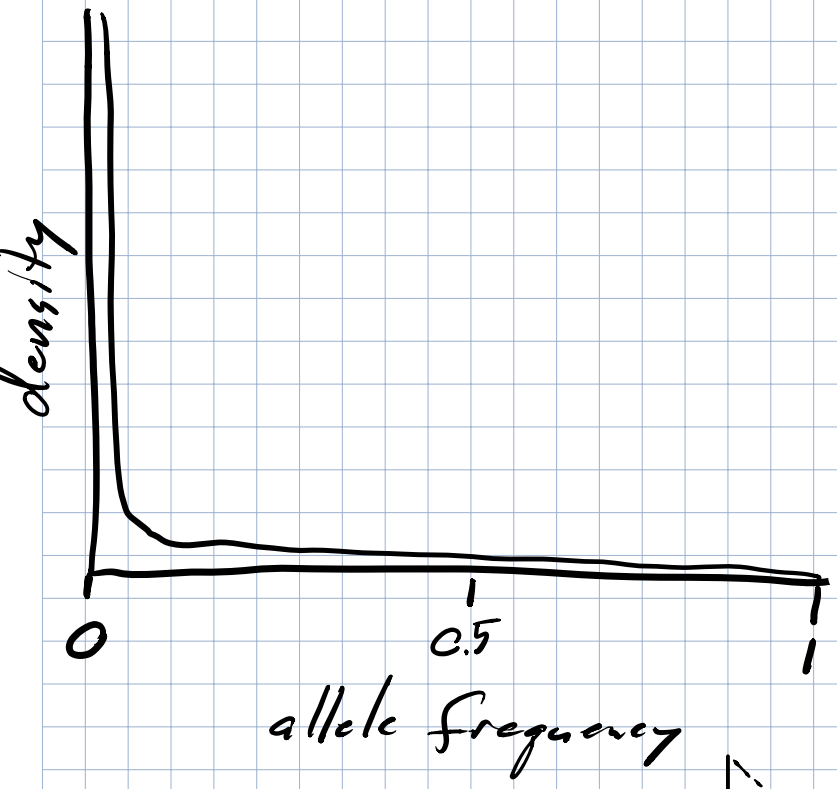

The number of neutral mutations found at frequency $i$ in sample from a population of constant size is proportional to $\frac{1}{i}$

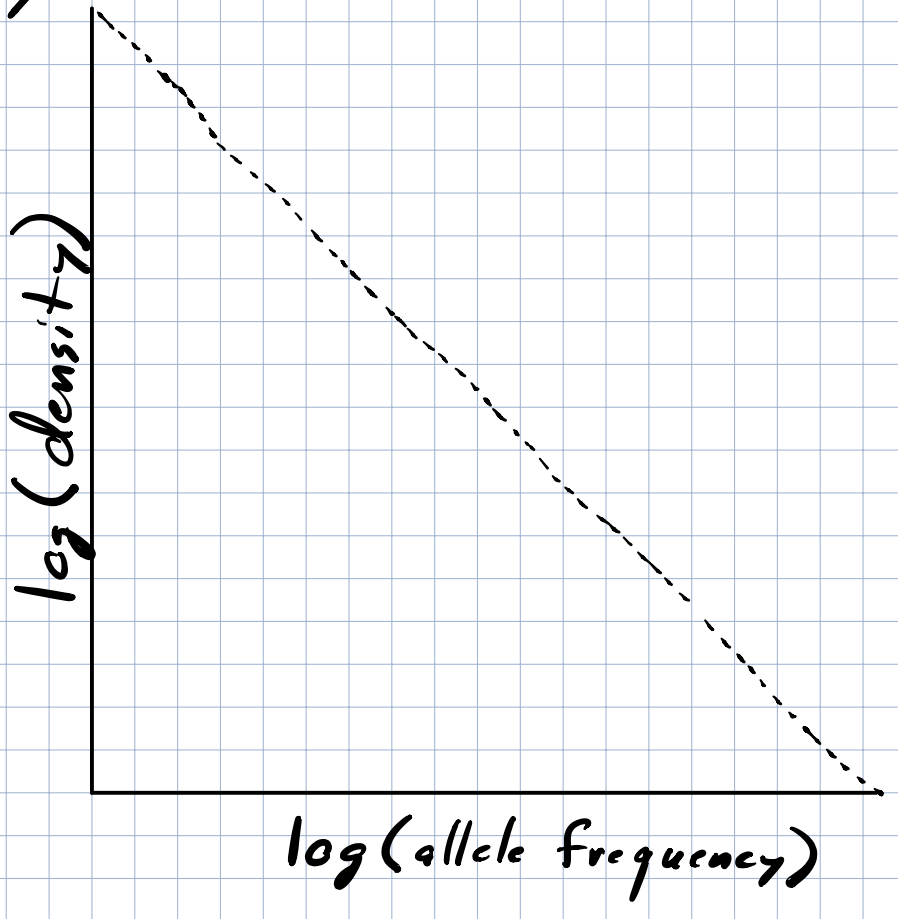

density
0
0.5
1
allele frequency


log (density)
log (allele frequency)

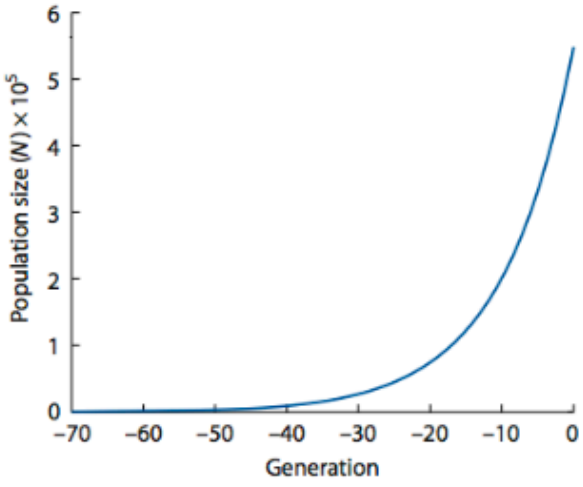
$$\mathbb{E}[\#\text{ of } k\text{-ton mutations}] = \mu \sum_{\text{trees}} \mathbb{E}[\#\text{ of } k\text{-ton mutations} \mid \text{tree}]\, P(\text{tree})$$

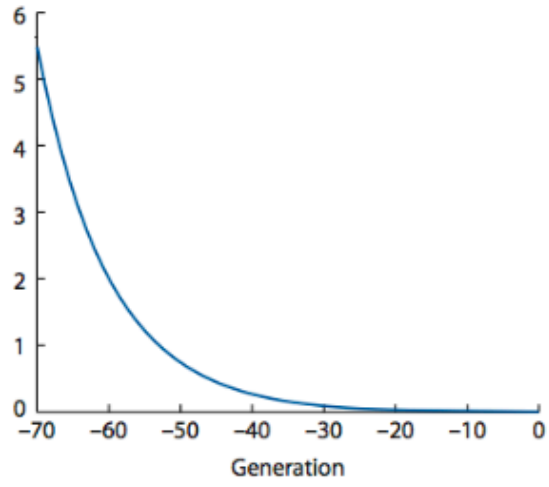
$$= \mu\, \mathbb{E}[\tau_k]$$

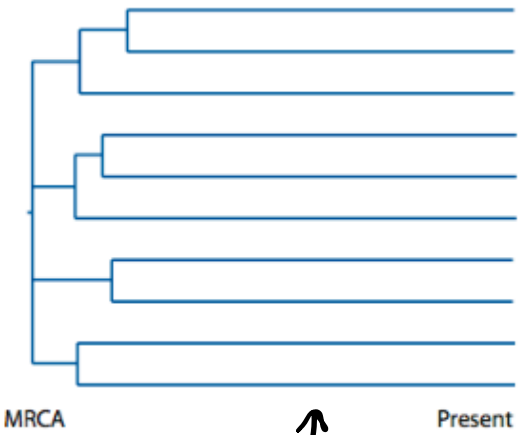


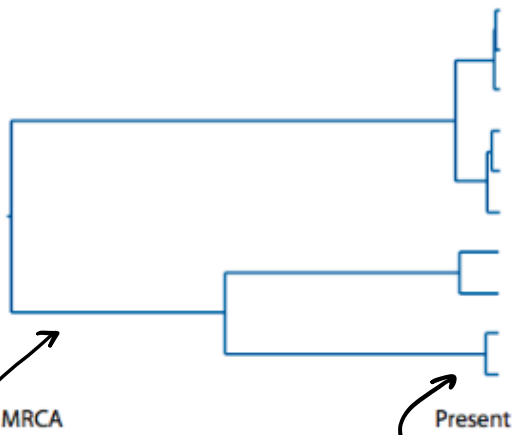



# Population history has a strong influence on tree shape




Recent growth creates gene trees that have relatively long branches closer to the present, and therefore an excess of low frequency alleles.

Recent population decline causes lots of coalescence in the recent past. It is therefore the deeper branches in the tree that are relatively long, so we can expect an excess of intermediate frequency mutations.