Genetic variation and the coalescent

In the forward time video on heterozygosity, we showed that the expected heterozygosity in a random sample from an equilibrium population is

$$\mathbb{E}[\pi] = \frac{4N\mu}{4N\mu + 1}$$

N = population size

μ = mutation rate

Now, we'll derive this using coalescent theory

$$\mathbb{E}[\pi] = P\left(\begin{array}{l} \text{at least one of two lineages} \\ \text{mutate before they coalesce} \end{array}\right)$$
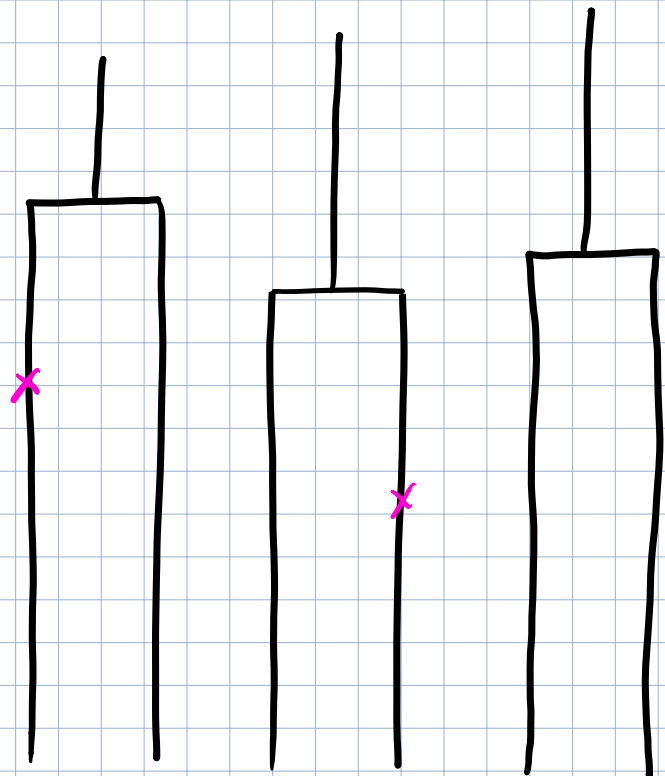
$$= \frac{\text{per generation mutation rate}}{\text{per generation mutation rate + per generation coalescence rate}}$$

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

mutation can occur on either lineage

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

$$\approx 4N\mu \quad \text{if } 4N\mu \ll 1$$

these trees produce a heterozygous site, because they have mutations on they

this one does not

Genetic variation and the coalescent

In the forward time video on heterozygosity, we showed that the expected heterozygosity in a random sample from an equilibrium population is

$$\mathbb{E}[\pi] = \frac{4N\mu}{4N\mu + 1}$$

Now, we'll derive this using coalescent theory

$$\mathbb{E}[\pi] = P\left(\begin{array}{l}\text{at least one of two lineages} \\ \text{mutate before they coalesce}\end{array}\right)$$
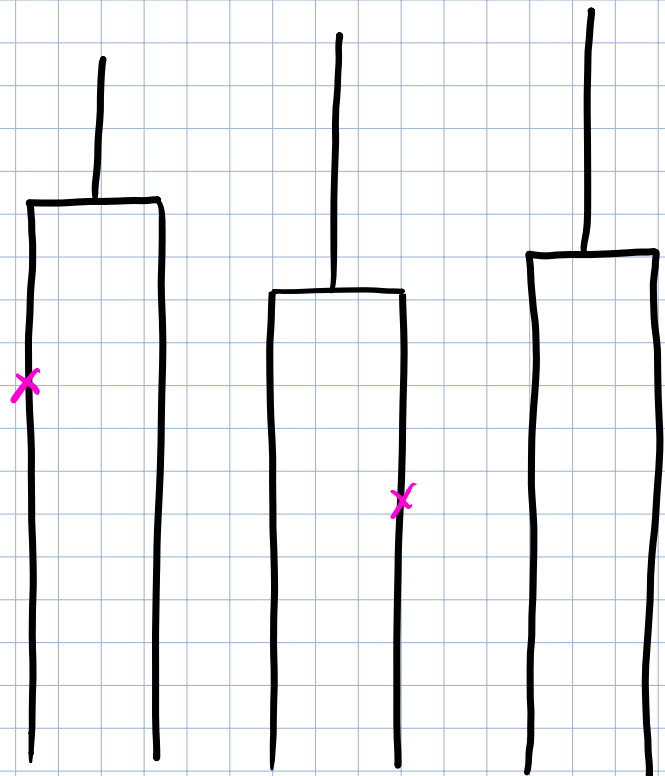
$$= \frac{\text{per generation mutation rate}}{\text{per generation mutation rate} + \text{per generation coalescence rate}}$$

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

mutation can occur on either lineage

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

$$\approx 4N\mu \quad \text{if} \quad 4N\mu \ll 1$$
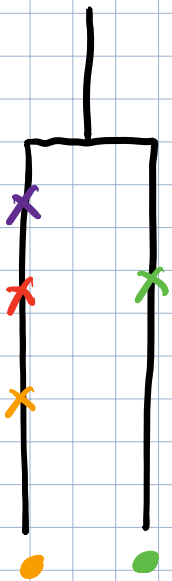

these trees produce a heterozygous site, because they have mutations on they

this one does not

Genetic variation and the coalescent

Infinite alleles model

- Consider only one "site" or "locus" at a time.
- Every time a mutation occurs, it creates a new allele at that site
- μ gives the mutation rate for the locus

Genetic variation and the coalescent

Infinite alleles model

- Consider only one "site" or "locus" at a time.
- Every time a mutation occurs, it creates a new allele at that site
- μ gives the mutation rate for the site

$$\mathbb{E}[\pi] = \frac{4N_a}{4N_a + 1}$$
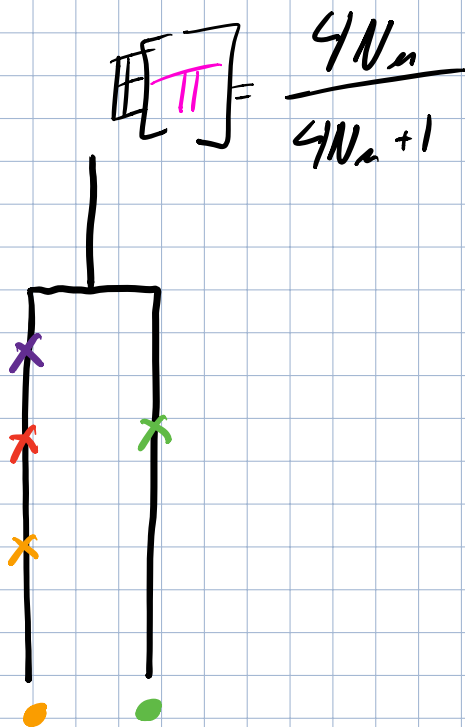
$$\mathbb{E}[T_2] = 2N$$
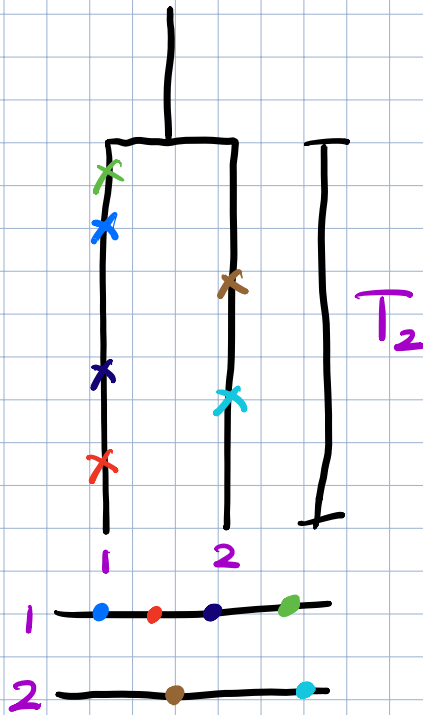
Infinite sites model

- Consider a sequence
- probability of mutation at any given site is infinitesimally small, but there are an infinite number of them
- μ gives the total mutation rate for the sequence

$$\pi = 2T_2 \mu$$

$$\mathbb{E}[\pi] = 2\mathbb{E}[T_2]\mu$$
$$= 4N\mu$$

T_2

1   2

1

2
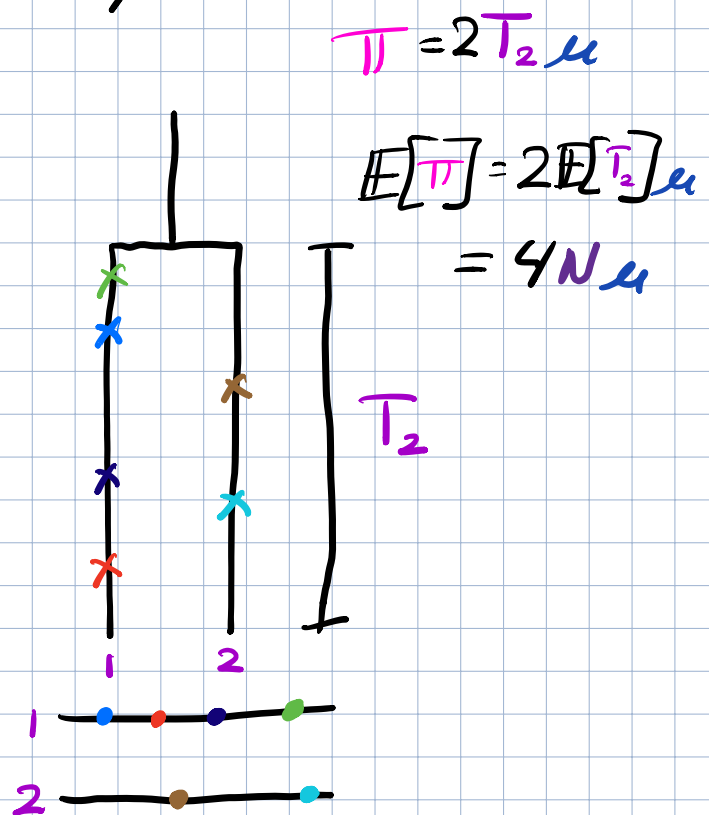
Effective population size

$$\mathbb{E}[T_2] = 2N$$

$$\mathbb{E}[\pi] = 2\mathbb{E}[T_2]\mu = 4N\mu$$

$$N_e = \frac{\hat{\pi}}{4\mu}$$

$$\mathbb{E}[T_2] = \frac{\hat{\pi}}{2\mu}$$

T_2

1   2

1

2