

Neutral models backward in time

Neutral models backward in time

Forward time recap

- Track evolution of allele frequency from one generation to the next
- We have to follow the entire population in each generation, even if the sample we are interested in is only a small subset of the population

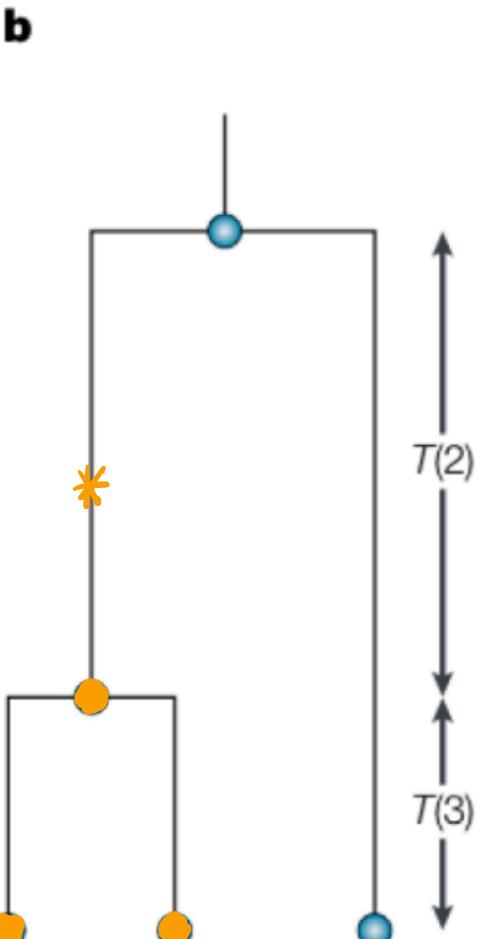
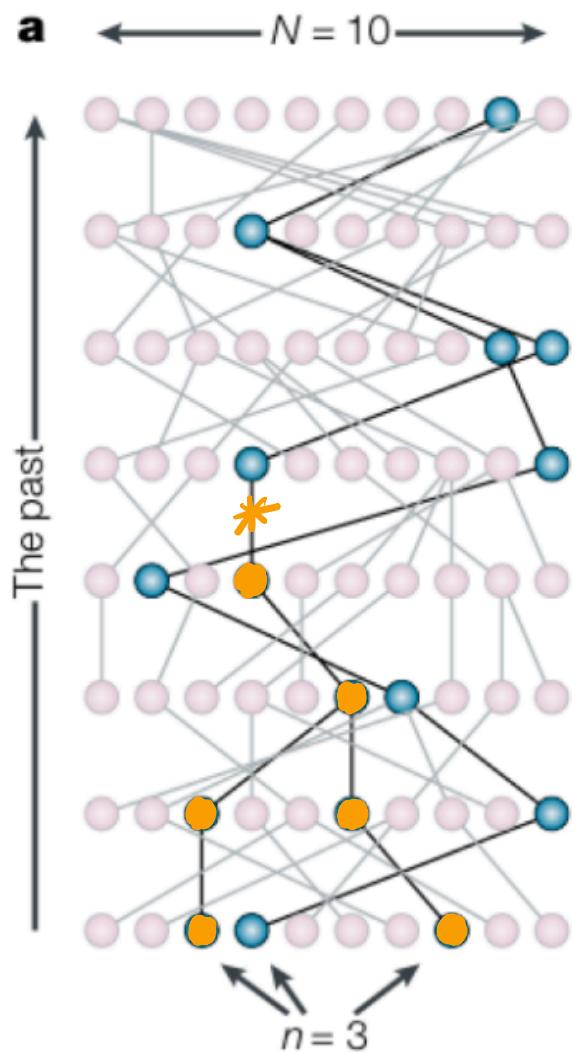
Neutral models backward in time

Forward time recap

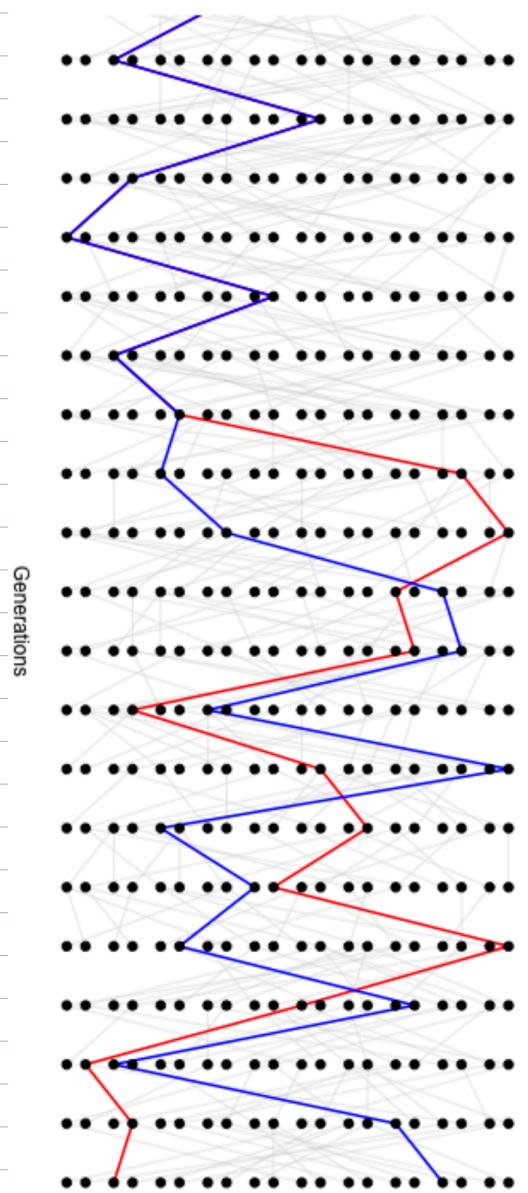
- Track evolution of allele frequency from one generation to the next
- We have to follow the entire population in each generation, even if the sample we are interested in is only a small subset of the population

Backward time preview

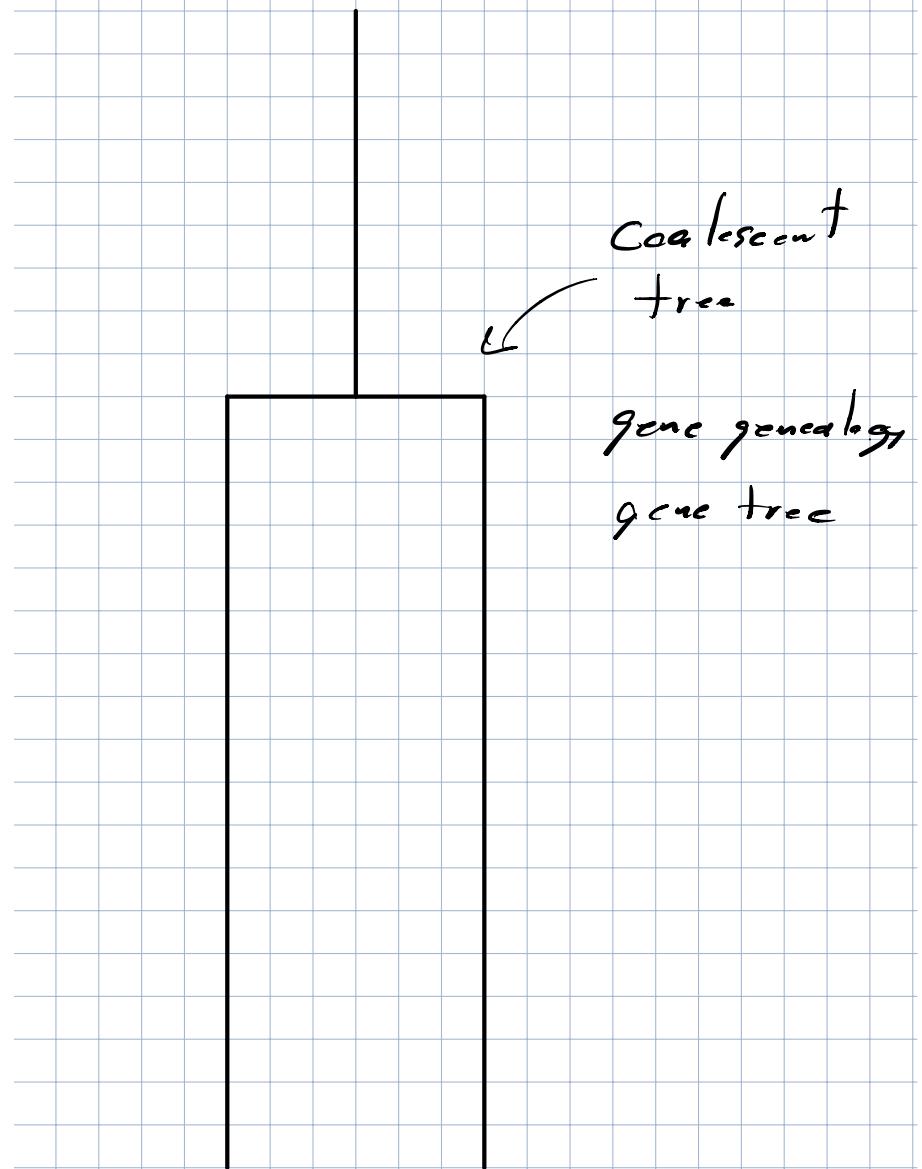
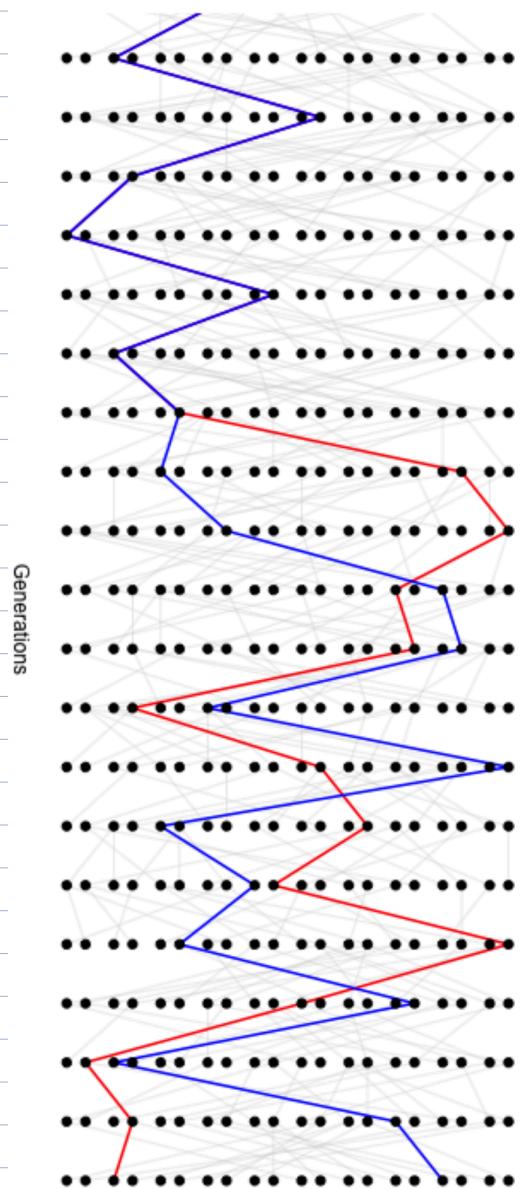
- Start with the sample in the present day and follow the "ancestral lines" of all individuals backward in time to find their ancestors
- Can treat this "ancestral process" as separate from the "mutation process"
- Only have to consider the ancestry of the sample. Allows us to ignore events in the history of the population that don't impact the sample
 - **HUGE simplification!!!**



Coalescent trees are embedded in the population history



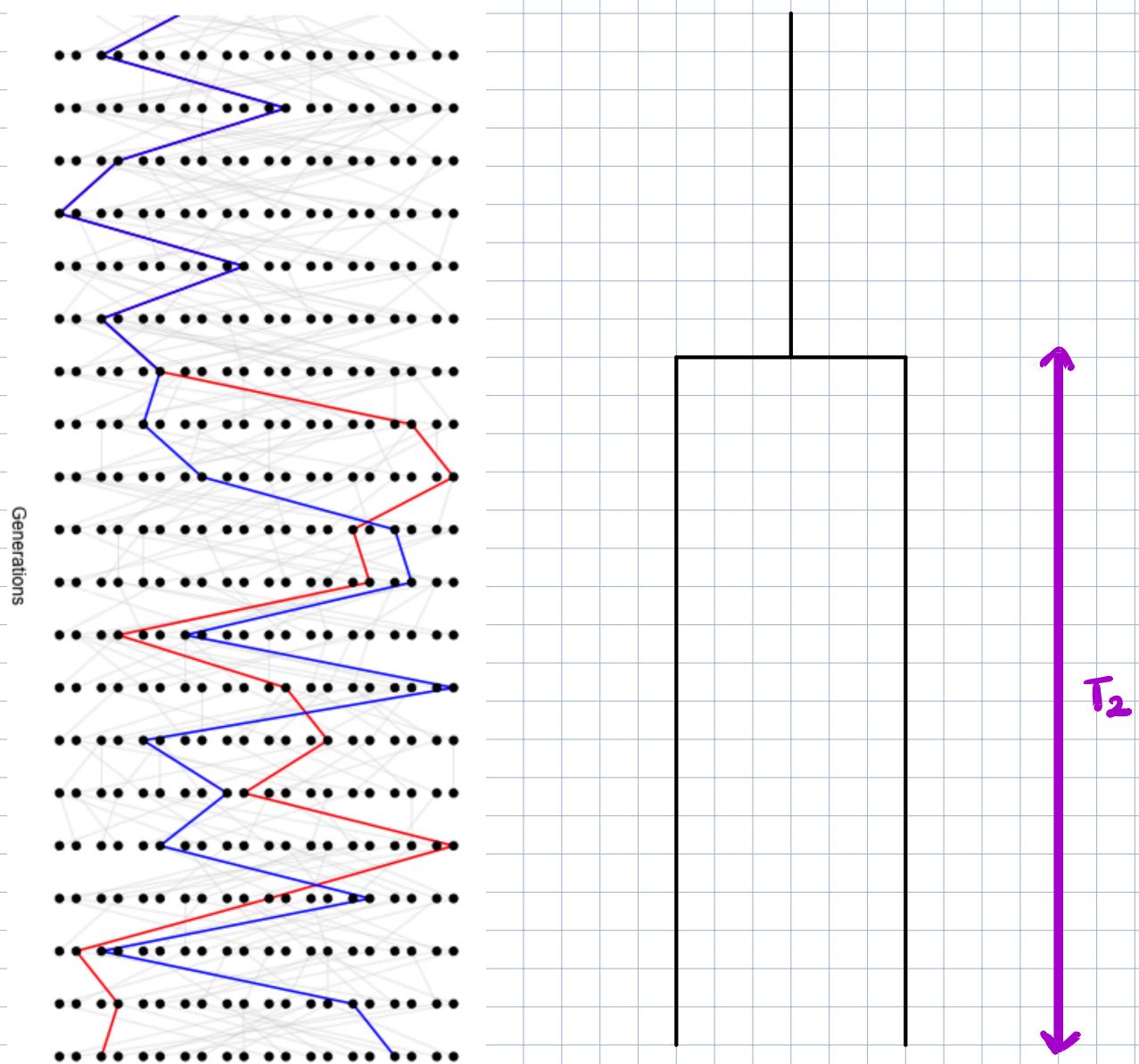
Coalescent trees are embedded in the population history



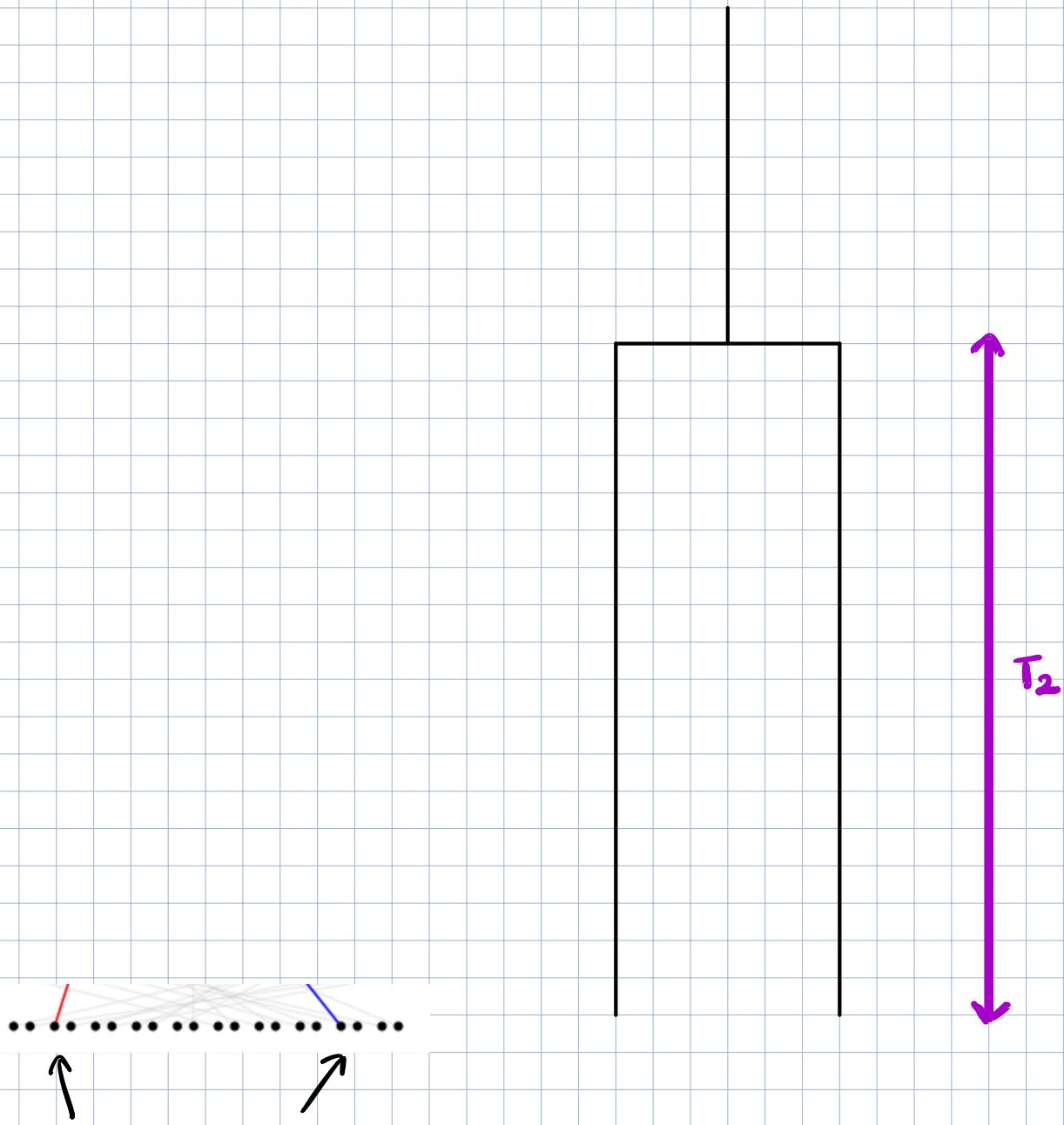
Coalescent
tree

gene genealogy
gene tree

Coalescent trees are embedded in the population history

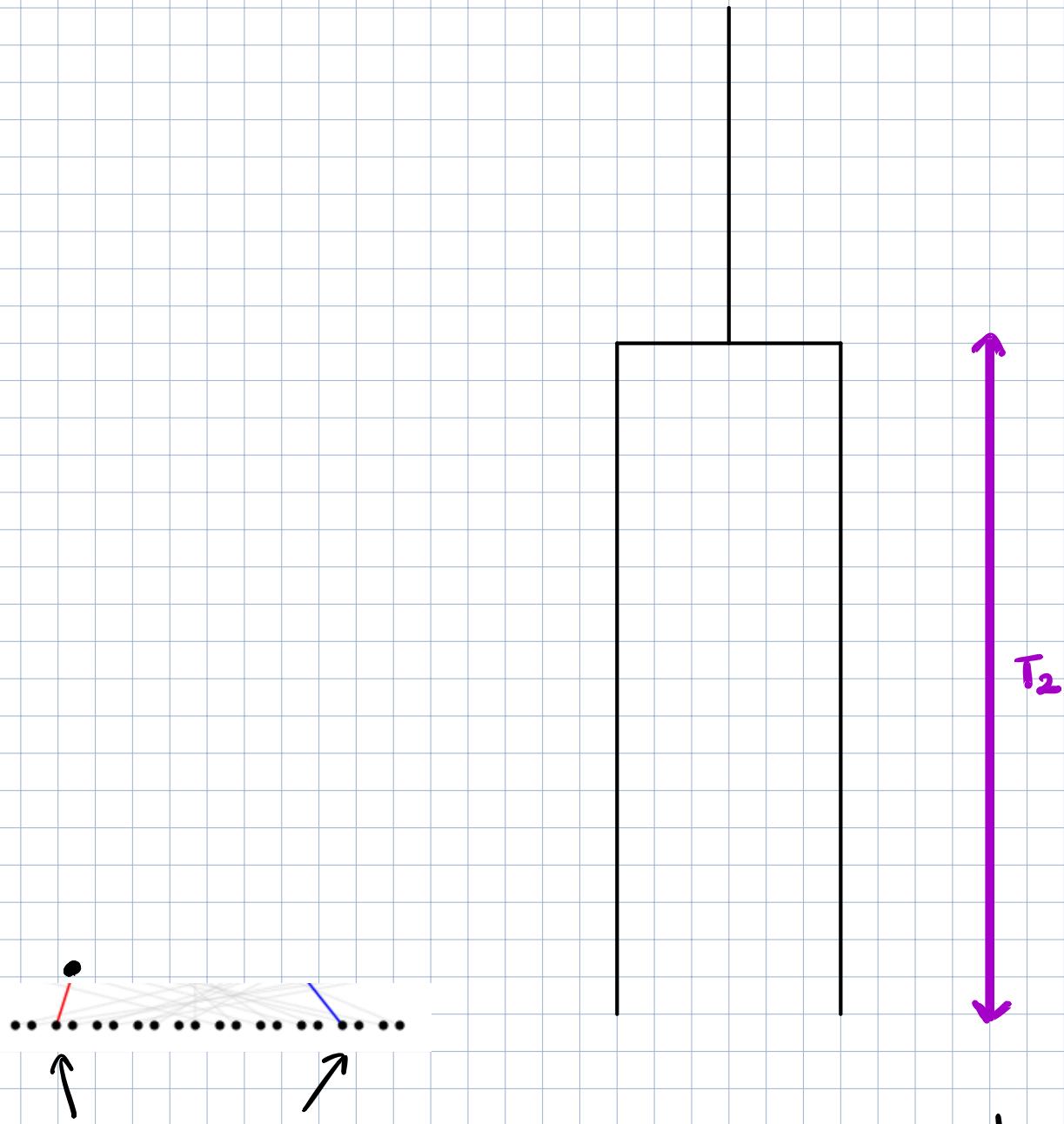


Coclescent theory: First steps



Assuming this is a randomly mating Wright-Fisher population, what is the distribution of T_2 , i.e. the "pairwise coalescent time"?

Coevolution theory: First steps

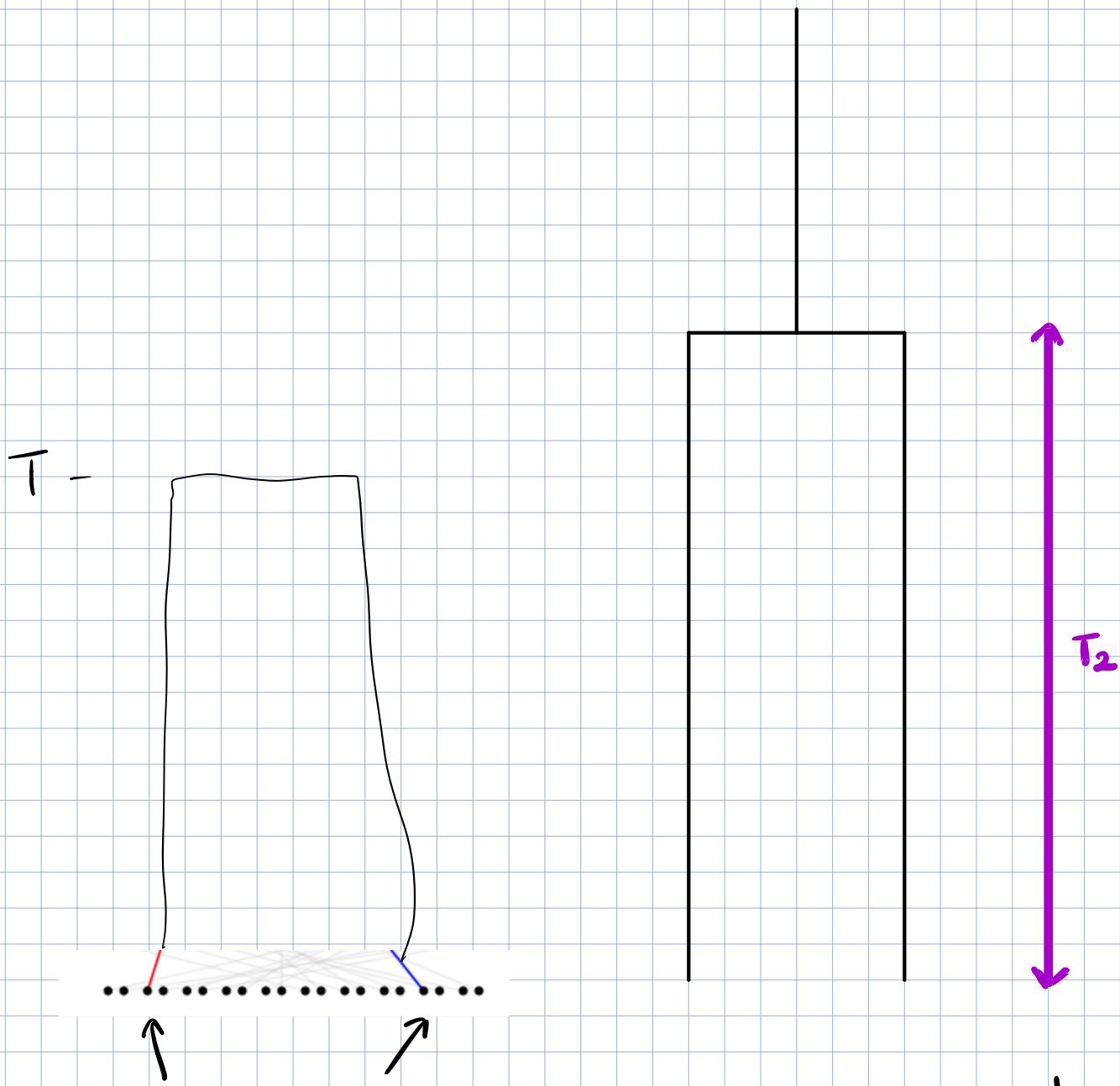


Assuming this is a randomly mating Wright-Fisher population, what is the distribution of

$$P(\text{two genes coalesce in the previous generation}) = \frac{1}{2N}$$

T_2 , i.e. the "pairwise coalescent time"?

Coalescent theory: First steps



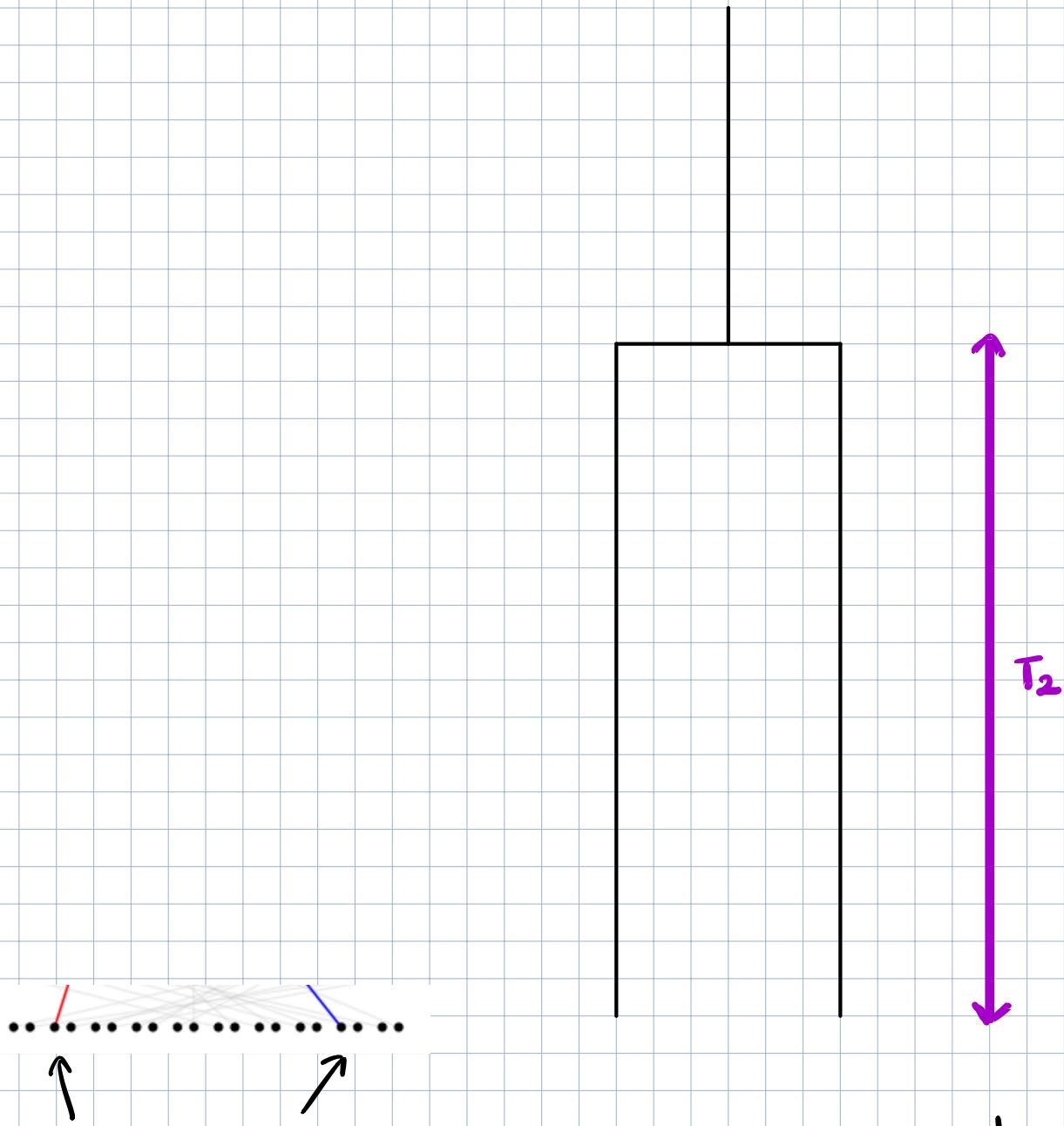
Assuming this is a randomly mating Wright-Fisher population, what is the distribution of T_2 , i.e. the "pairwise coalescent time"?

$$P(\text{two genes coalesce in the previous generation}) = \frac{1}{2N}$$

$$P(\text{two genes coalesce } T \text{ generations ago}) = \left(1 - \frac{1}{2N}\right)^{T-1} \cdot \frac{1}{2N}$$

$$\approx \frac{1}{2N} e^{-\frac{T}{2N}}$$

Coelescent theory: First steps



Assuming this is a randomly mating Wright-Fisher population, what is the distribution of T_2 , i.e. the "pairwise coalescent time"?

$$P(\text{two genes coalesce in the previous generation}) = \frac{1}{2N}$$

$$P(\text{two genes coalesce } T \text{ generations ago}) = \left(1 - \frac{1}{2N}\right)^{T-1} \cdot \frac{1}{2N}$$

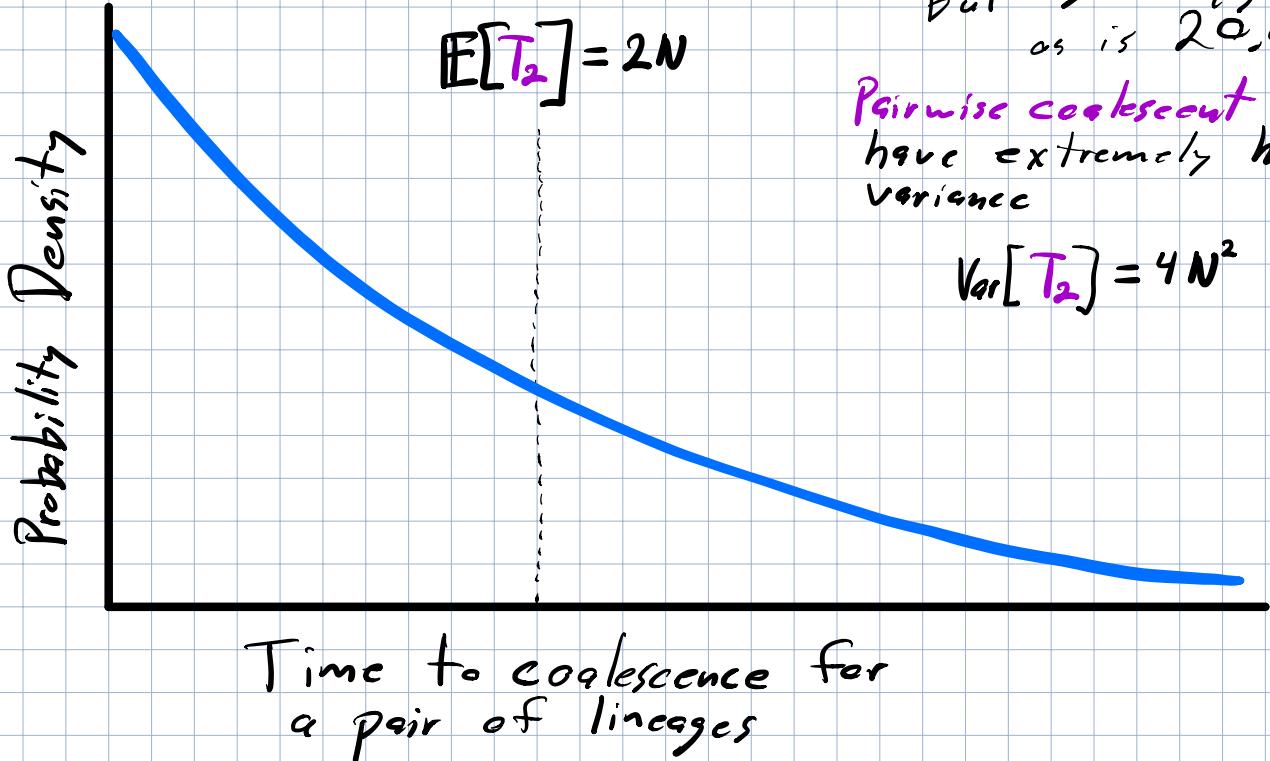
$$\approx \frac{1}{2N} e^{-\frac{T}{2N}}$$

Pairwise coalescent times are exponentially distributed with rate $\frac{1}{2N}$

$$T_2 \sim \exp(\frac{1}{2N})$$

$$E[T_2] = 2N$$

Cohescent theory: First steps



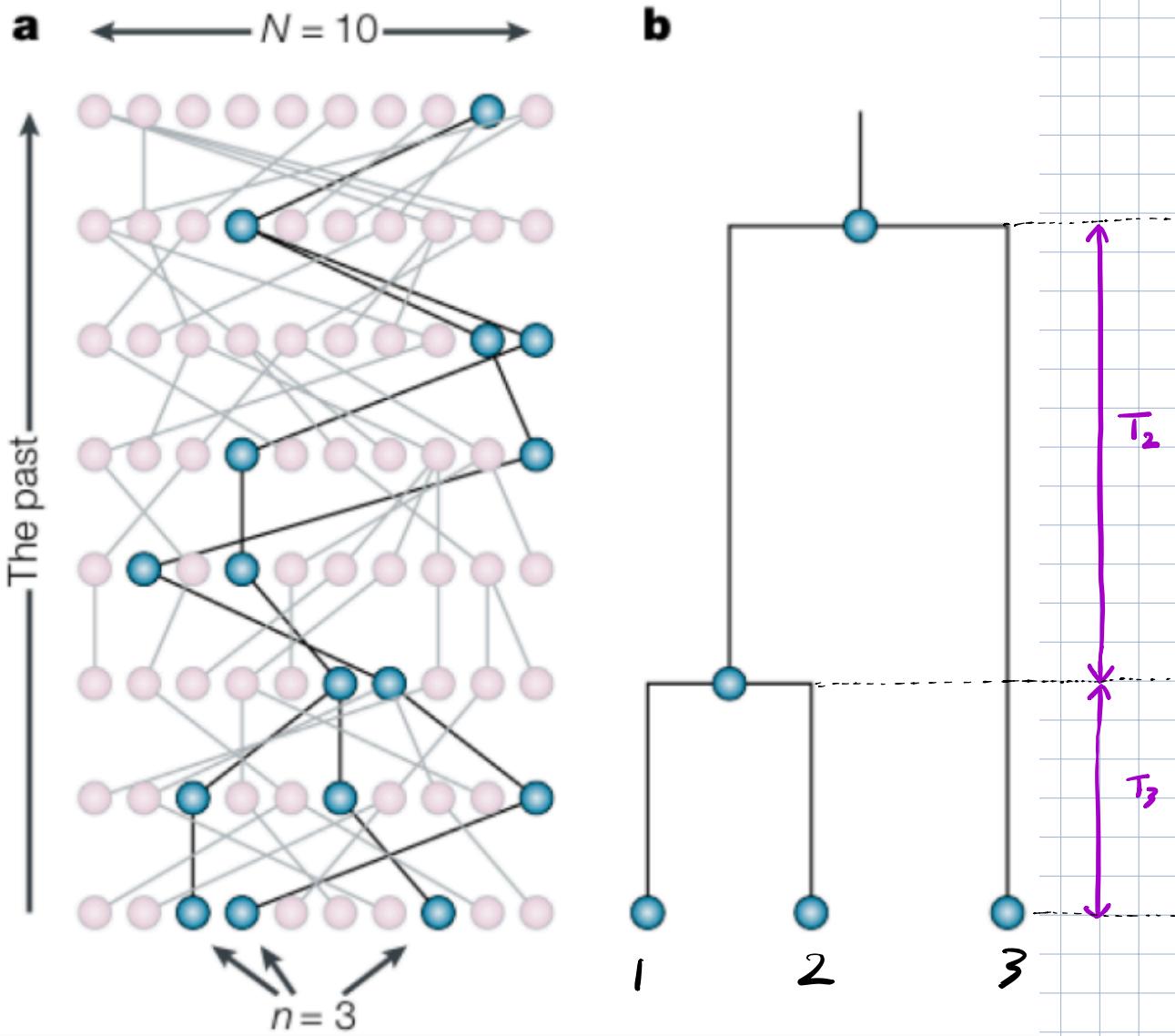
$$; f \\ N = 1000$$

$$E[T_2] = 2000$$

but 200 is plausible
as is 20,000

Pairwise coalescent times
have extremely high
variance

$$\text{Var}[T_2] = 4N^2$$



$$P(\text{any pair coalesces in previous generation} \mid \text{there are three lineages}) = \frac{\binom{3}{2}}{2N}$$

$$\binom{3}{2} = 3$$

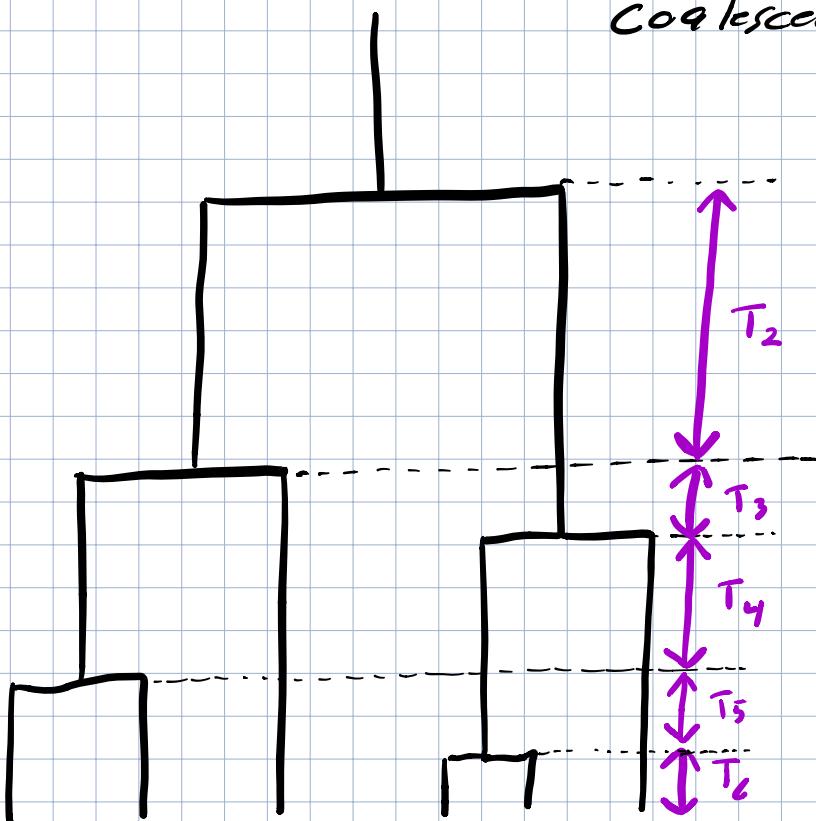
$$P(\text{next coalescence is in } T \text{ generations} \mid \text{there are three lineages}) = \left(1 - \frac{\binom{3}{2}}{2N}\right)^T - \frac{\binom{3}{2}}{2N}$$

$$\approx \frac{\binom{3}{2}}{2N} e^{-T \frac{\binom{3}{2}}{2N}}$$

$$T_3 \sim \exp\left(\frac{\binom{3}{2}}{2N}\right) \implies \mathbb{E}[T_3] = \frac{2N}{\binom{3}{2}} = \frac{2N}{3}$$

The coalescent is a continuous time Markov chain

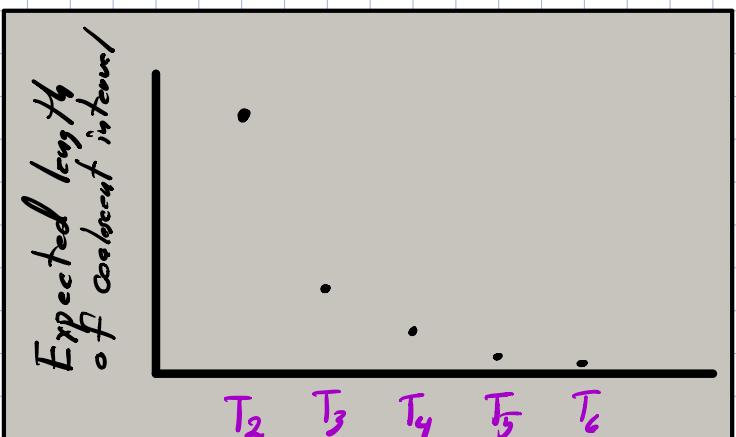
- the time to the next coalescent event is exponentially distributed with a rate that depend on the number of surviving lineages and the population size
- the amount of time during which there are k lineages is independent of how long the other coalescent intervals are



$$T_R \sim \exp\left(\frac{\binom{k}{2}}{2N}\right)$$

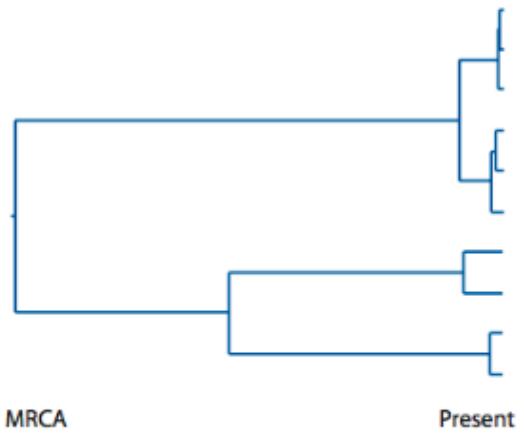
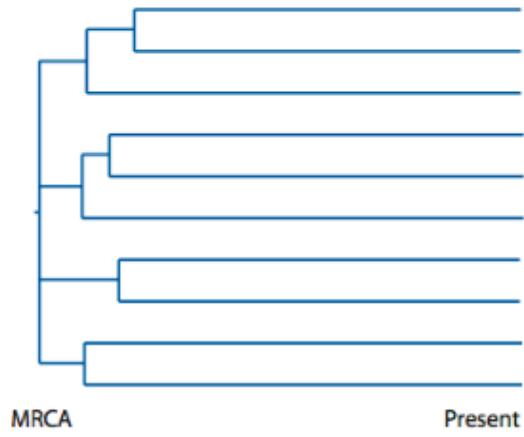
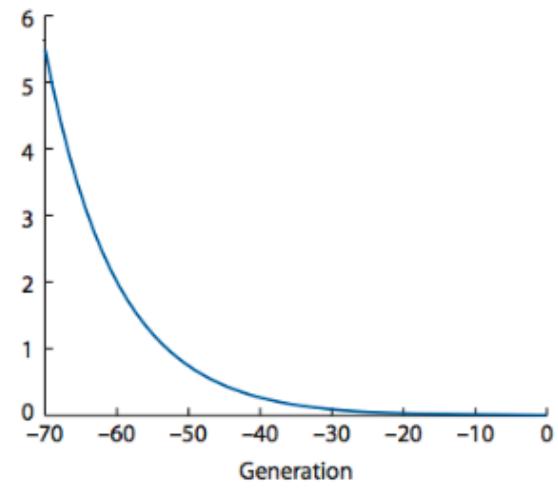
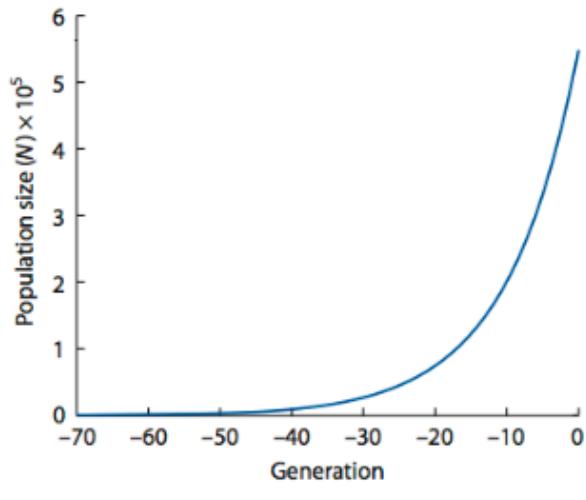
$$\mathbb{E}[T_R] = \frac{2N}{\binom{k}{2}}$$

$$\text{Var}[T_R] = \left(\frac{2N}{\binom{k}{2}}\right)^2$$



Population history has a strong influence on tree shape

- If the population size changes, then the coalescent rates change



Population structure and the "structured coalescent"

- two lineages can only coalesce with one another when they are in the same population, and this creates structure in the coalescent trees
- migration events can move lineages between populations
 - high migration rates: less structured trees
 - low migration rates: more structured trees

