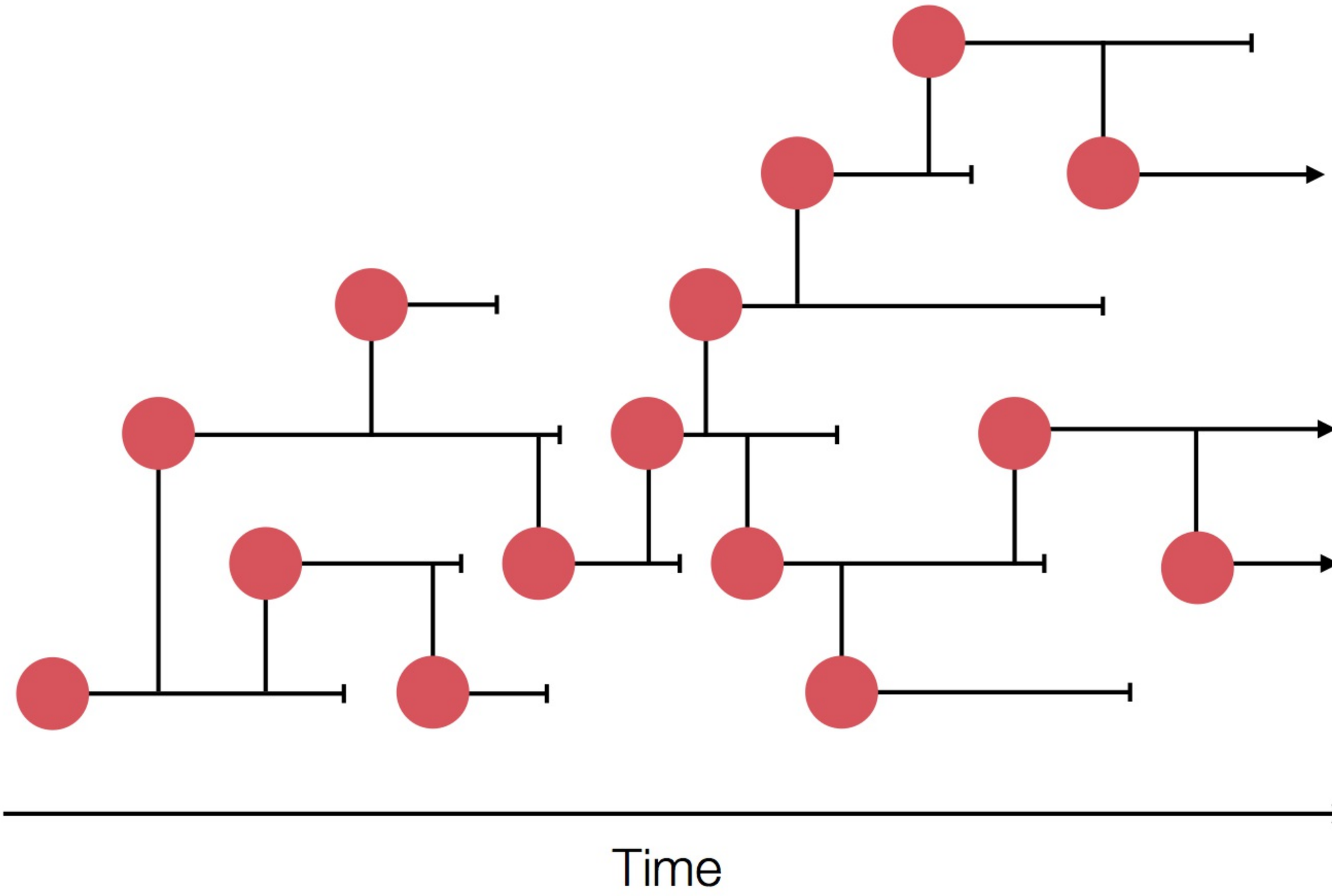
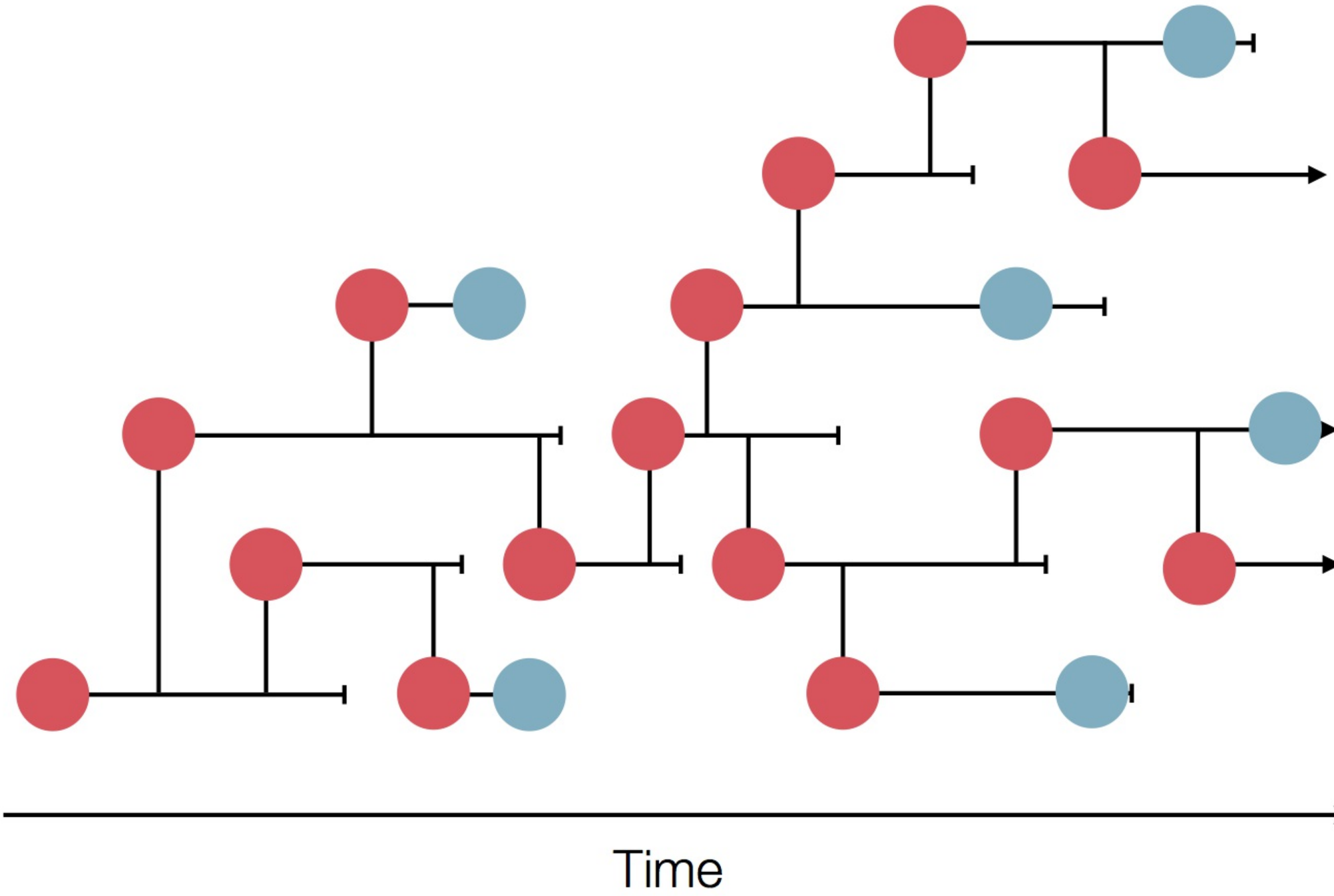
The coalescent

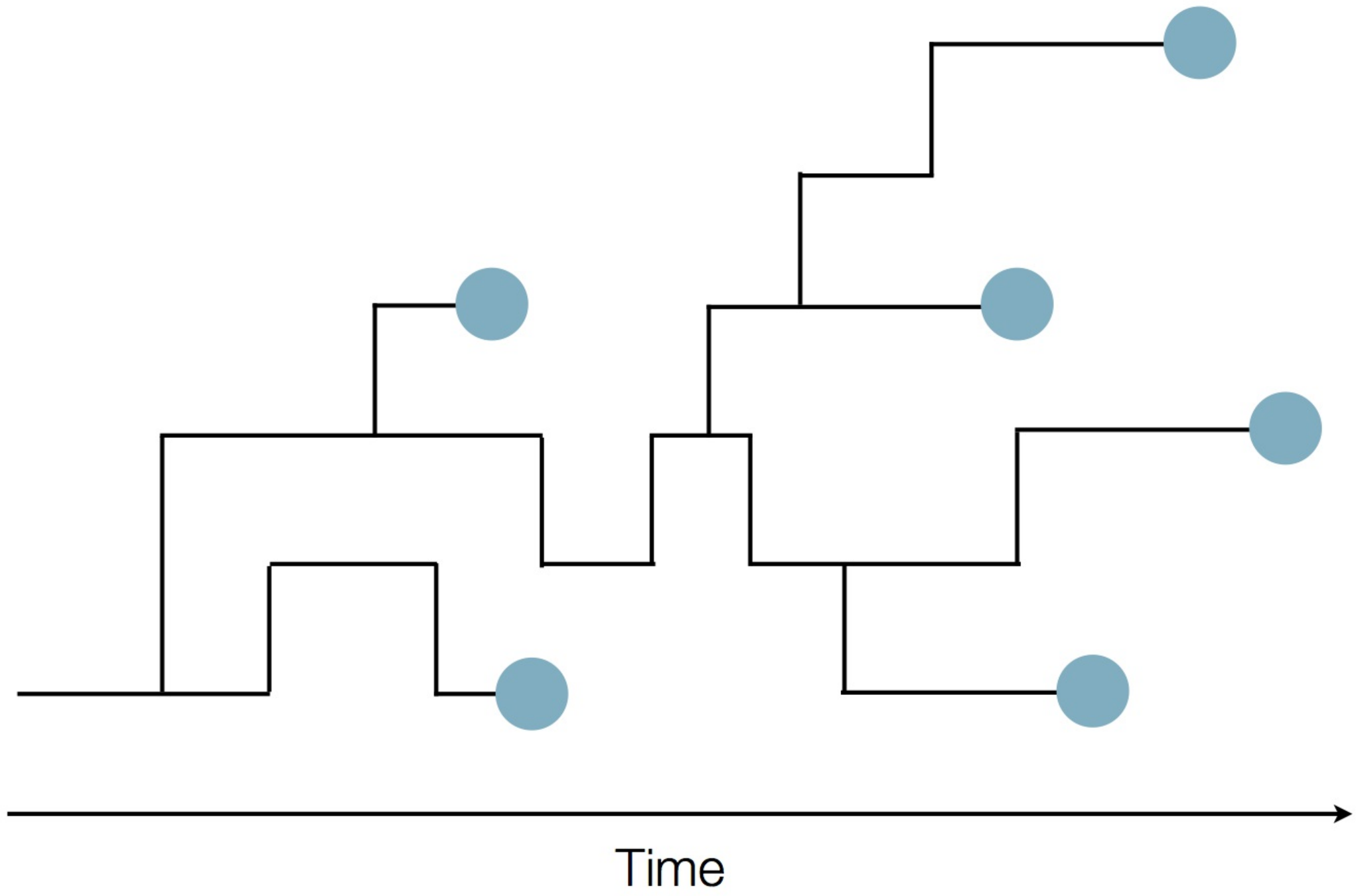
Epidemic process



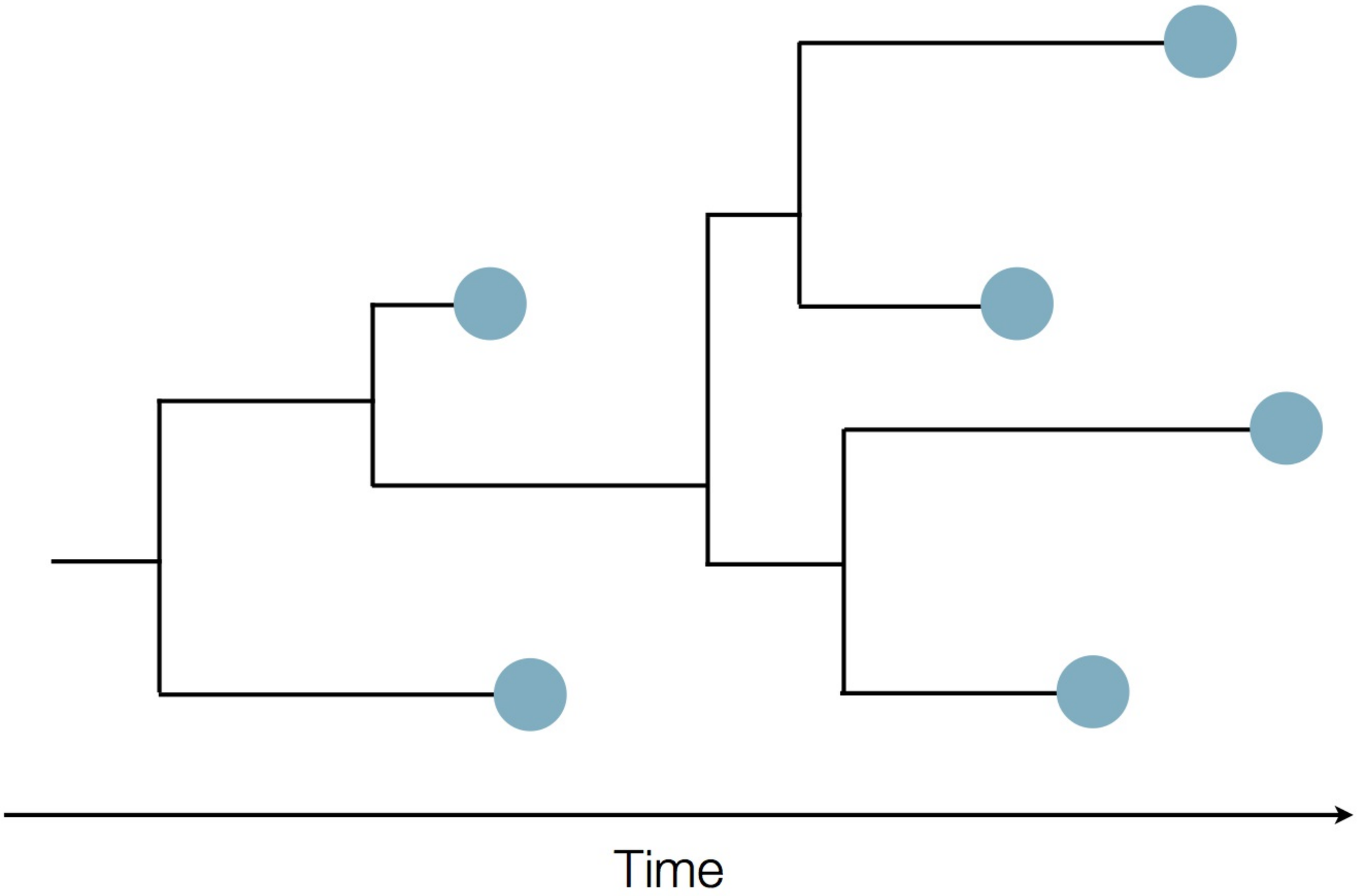
Sample some individuals



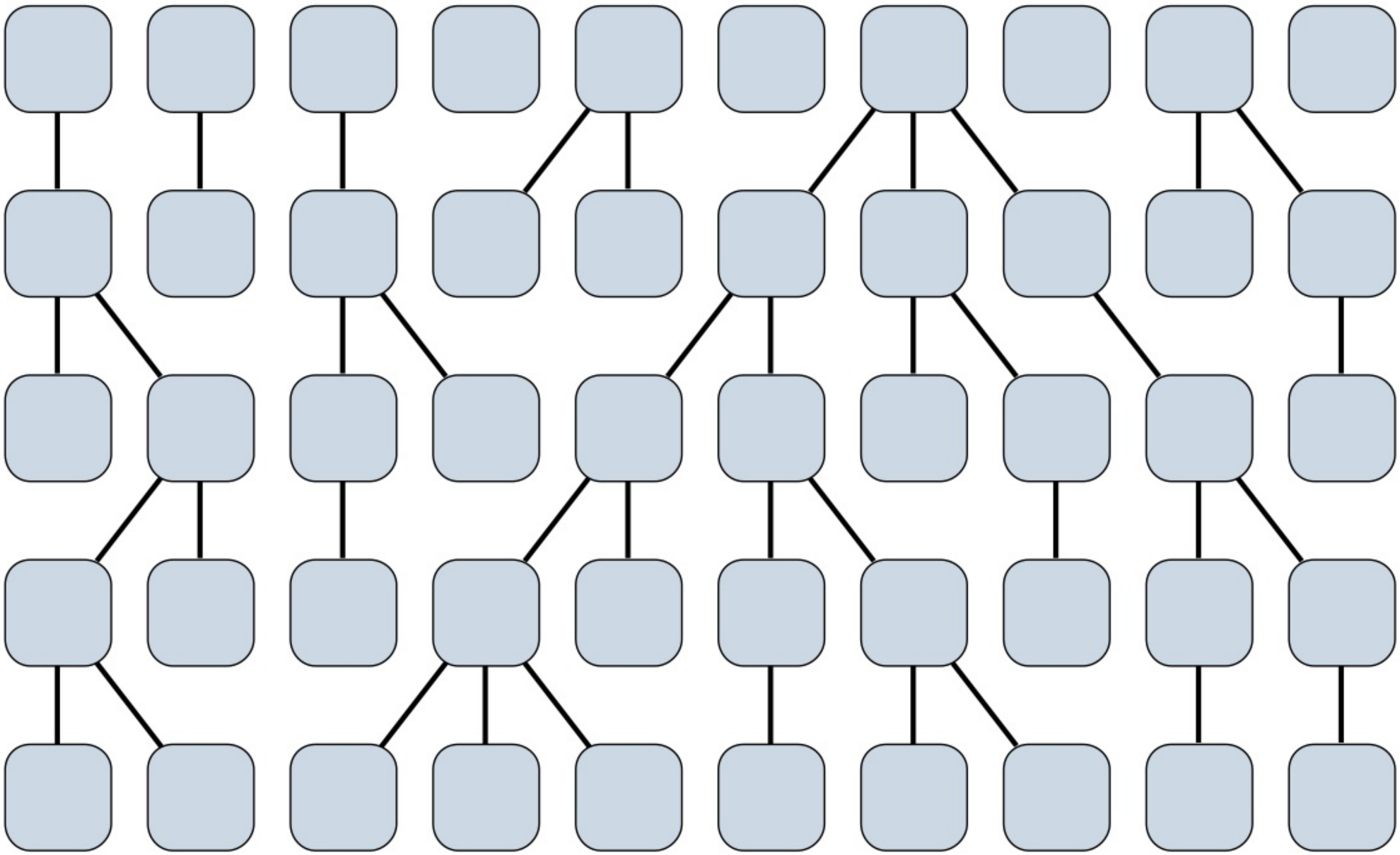
Sequence and determine phylogeny



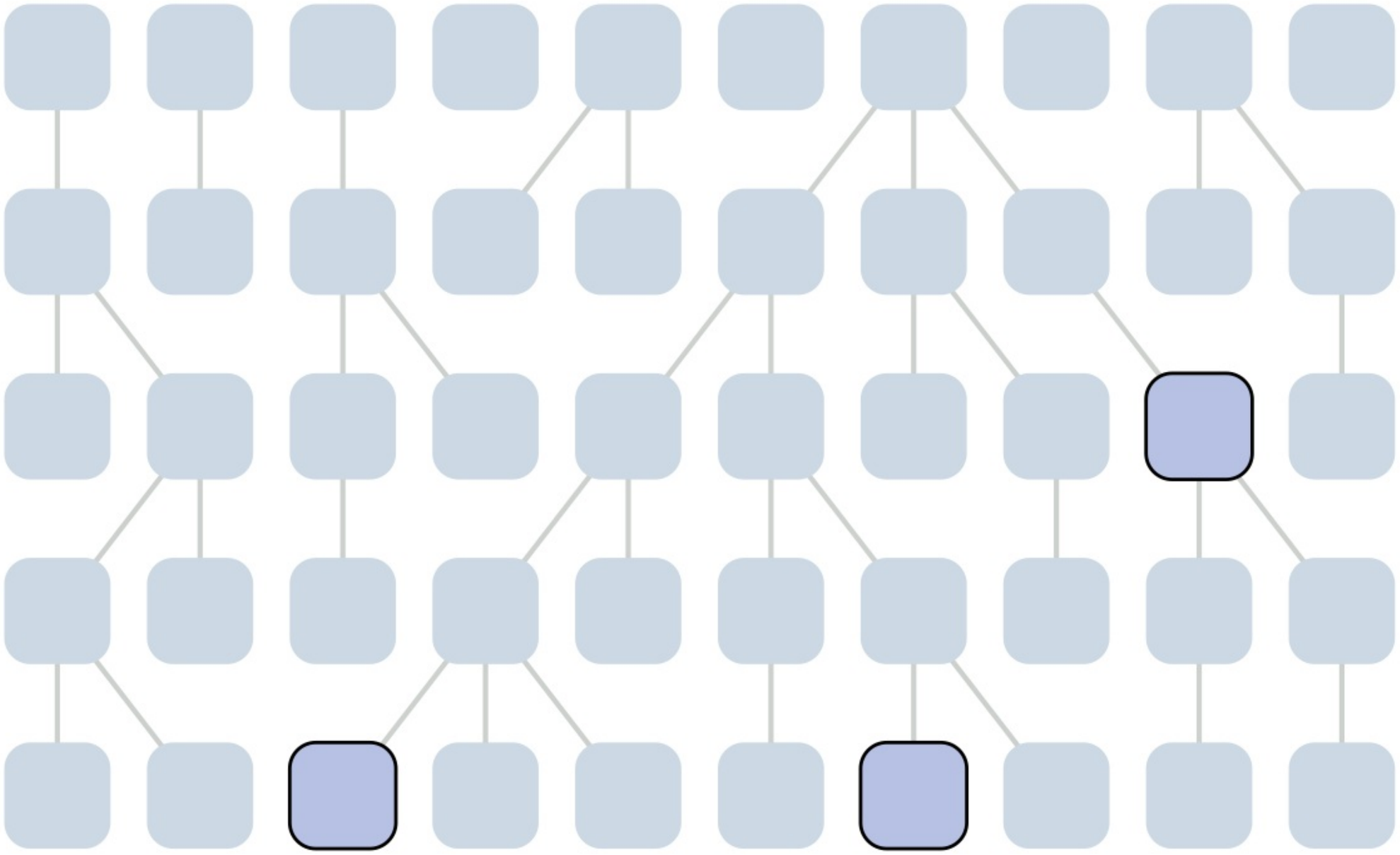
Sequence and determine phylogeny



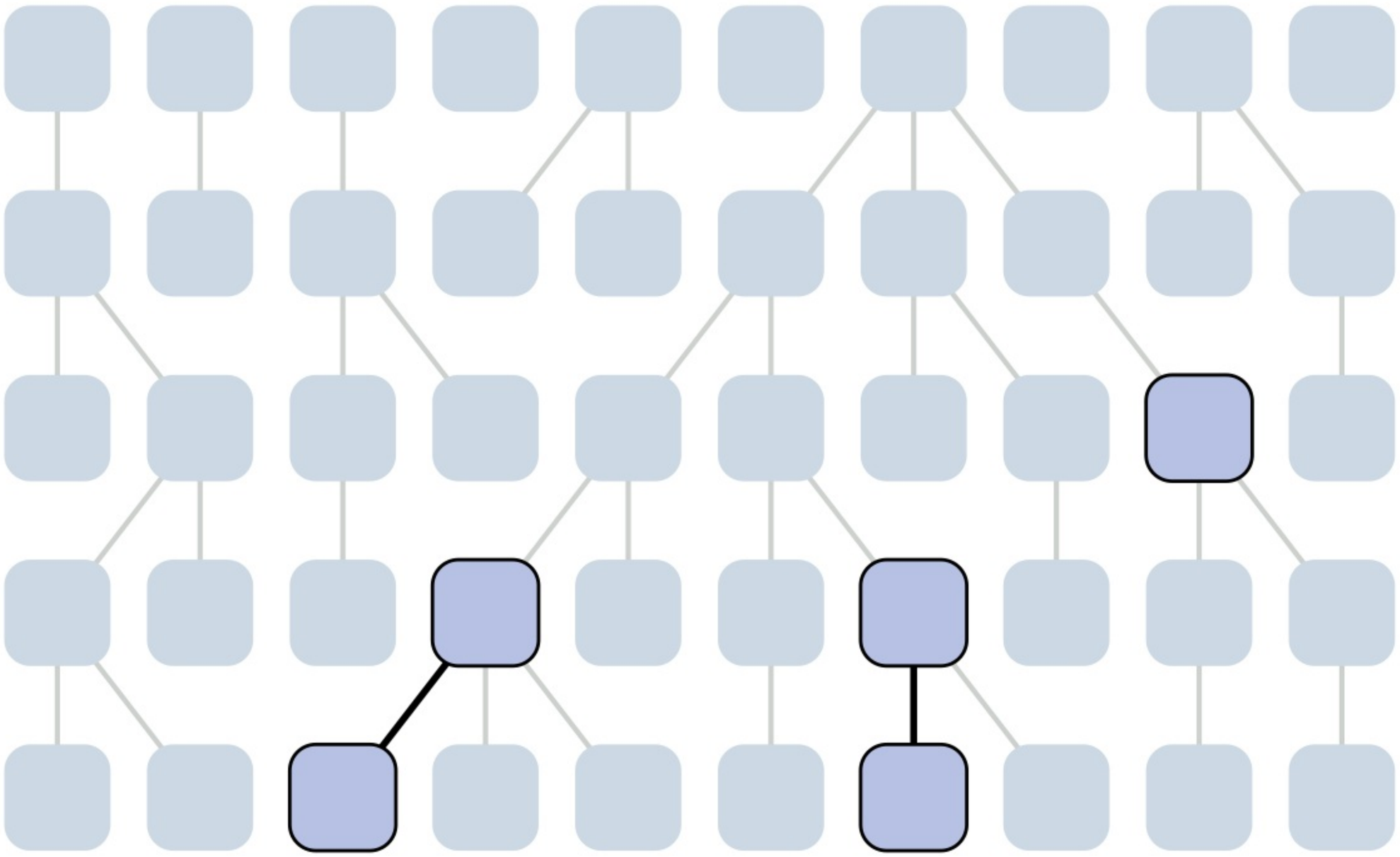
Wright-Fisher process



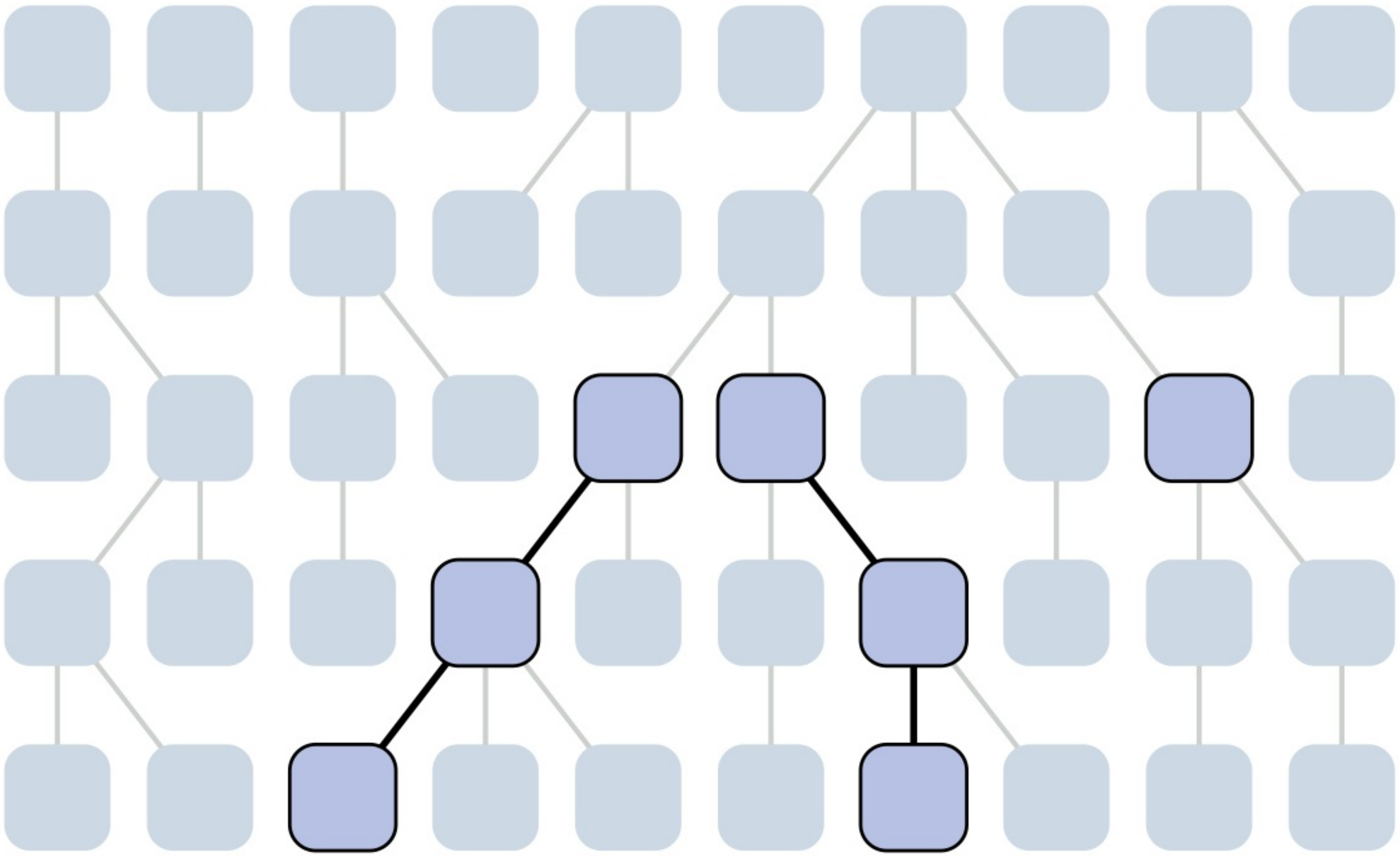
Sample some individuals



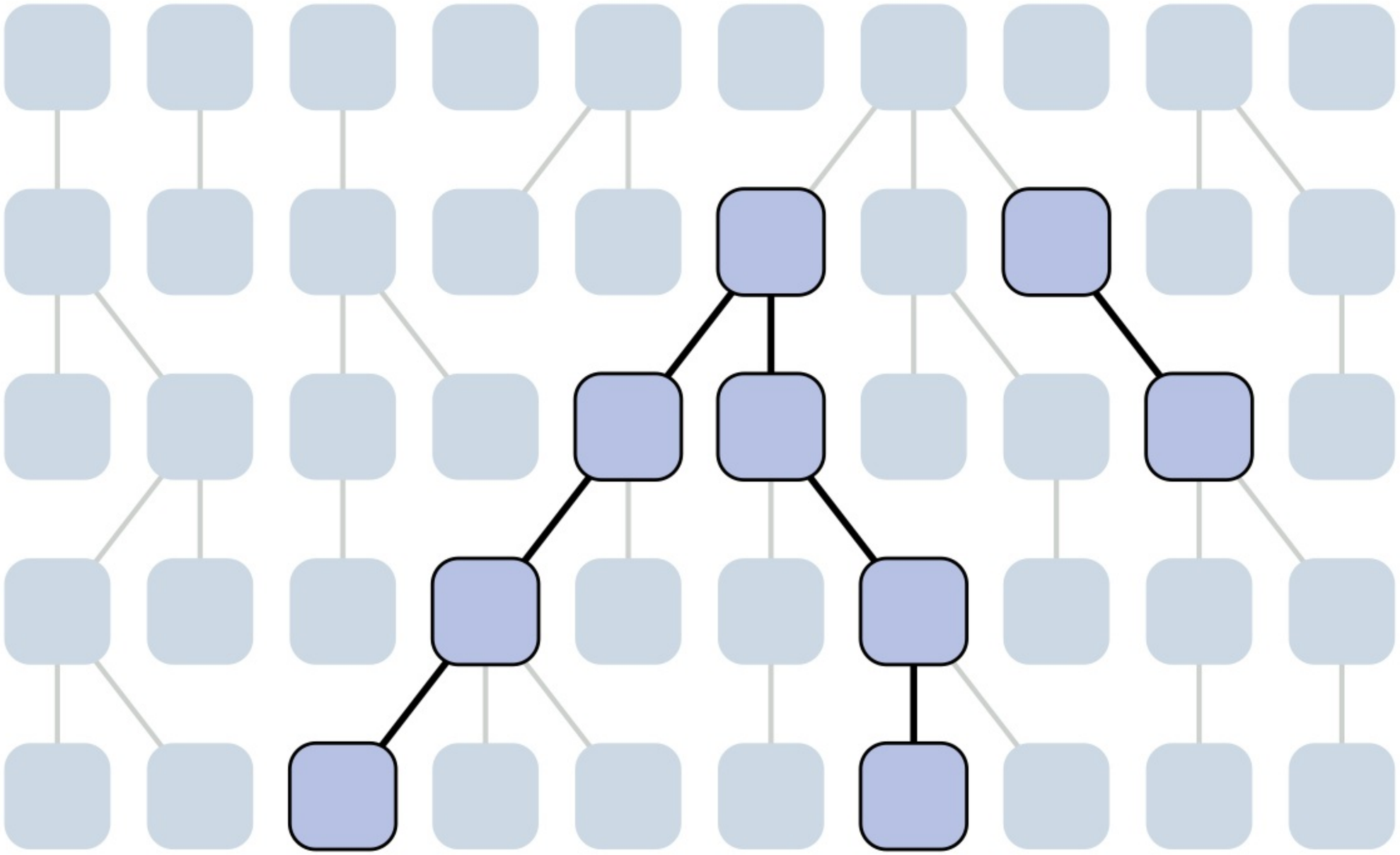
Each generation coalescence may occur



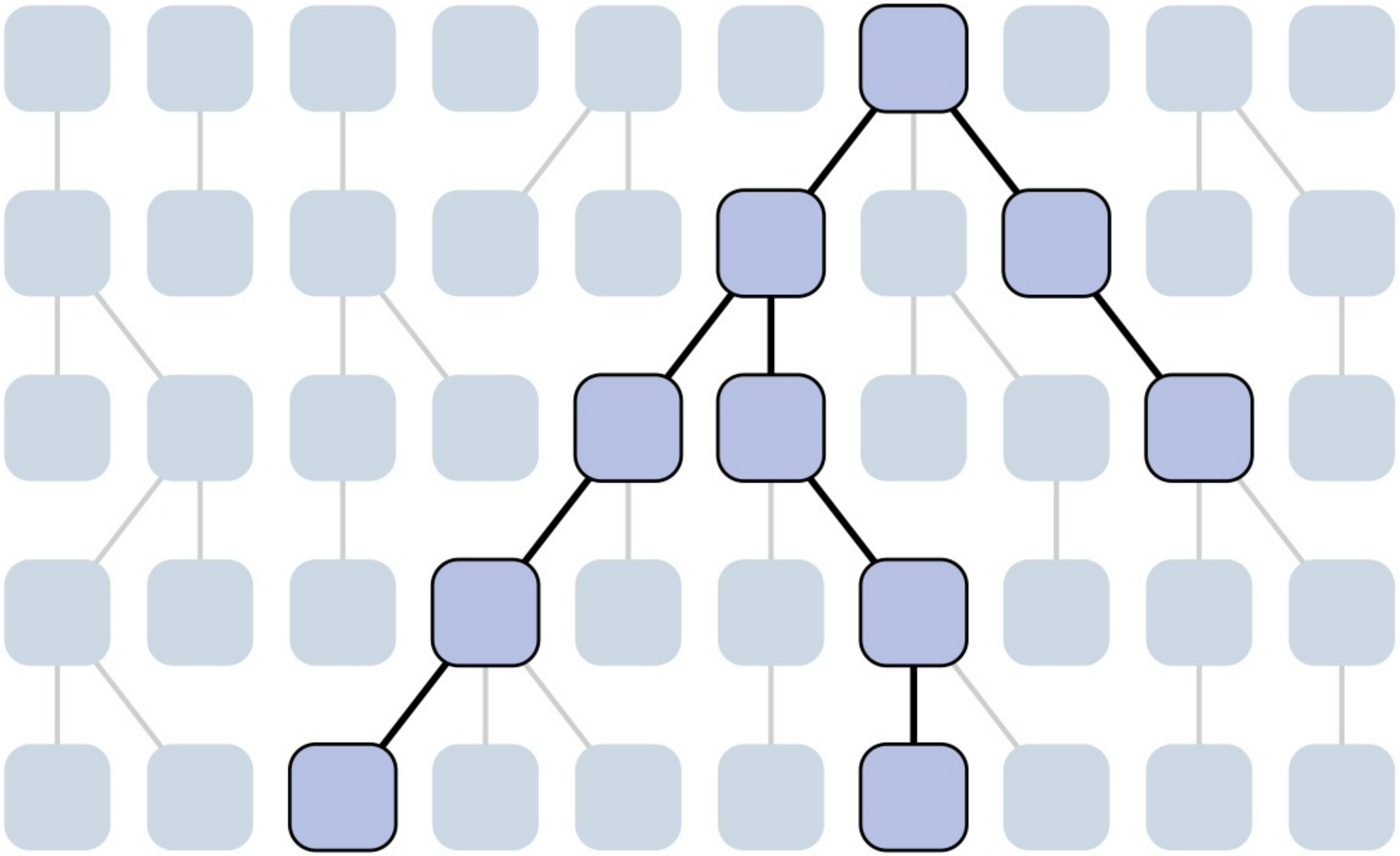
Each generation coalescence may occur

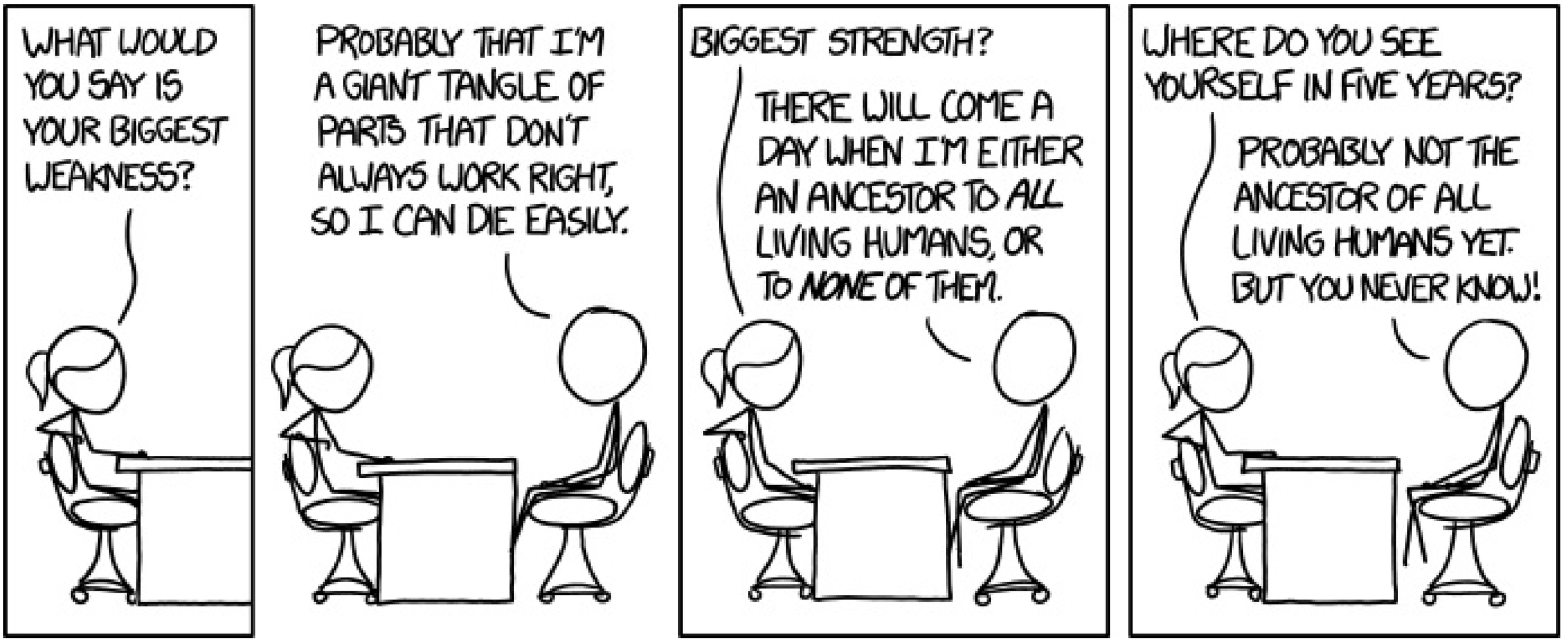


Each generation coalescence may occur

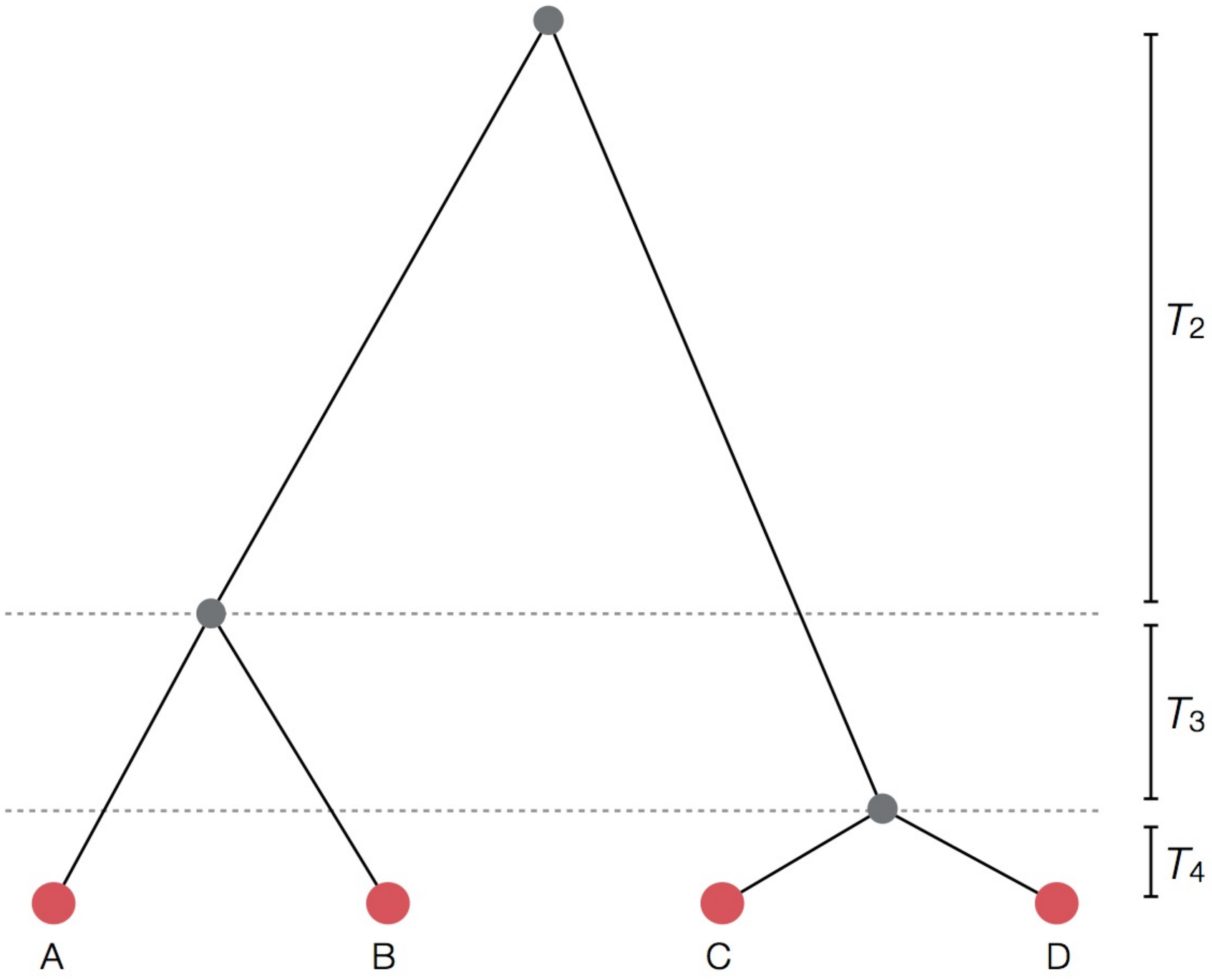


Each generation coalescence may occur




WHAT WOULD YOU SAY IS YOUR BIGGEST WEAKNESS?
PROBABLY THAT I'M A GIANT TANGLE OF PARTS THAT DON'T ALWAYS WORK RIGHT, SO I CAN DIE EASILY.
BIGGEST STRENGTH?
THERE WILL COME A DAY WHEN I'M EITHER AN ANCESTOR TO ALL LIVING HUMANS, OR TO NONE OF THEM.
WHERE DO YOU SEE YOURSELF IN FIVE YEARS?
PROBABLY NOT THE ANCESTOR OF ALL LIVING HUMANS YET. BUT YOU NEVER KNOW!

Set of coalescent intervals



Per-generation probability of coalescence

$$\mathbf{Pr}(\mathbf{coal} \,|\, i = 2) = \frac{1}{N}$$

Probability of first lineage picking an arbitrary parent is 1, while the probability of the 2nd lineage picking the same parent is $\frac{1}{N}$.

Probability of coalescence scales inversely with population size.

Per-generation probability of coalescence

$$\mathbf{Pr(coal)} = \binom{i}{2}\frac{1}{N} = \frac{i(i-1)}{2N}$$

There are $\binom{i}{2}$ ways pairs of lineages can pick the same parent.

Probability of coalescence scales quadratically with lineage count.

Expected waiting time to coalescence

$$\mathbf{E}[T_i] = \frac{2N}{i(i-1)}$$

This is a geometric distribution. If each generation there is a $\frac{1}{x}$ probability of an event occurring, we expect to wait x generations for the event to occur.

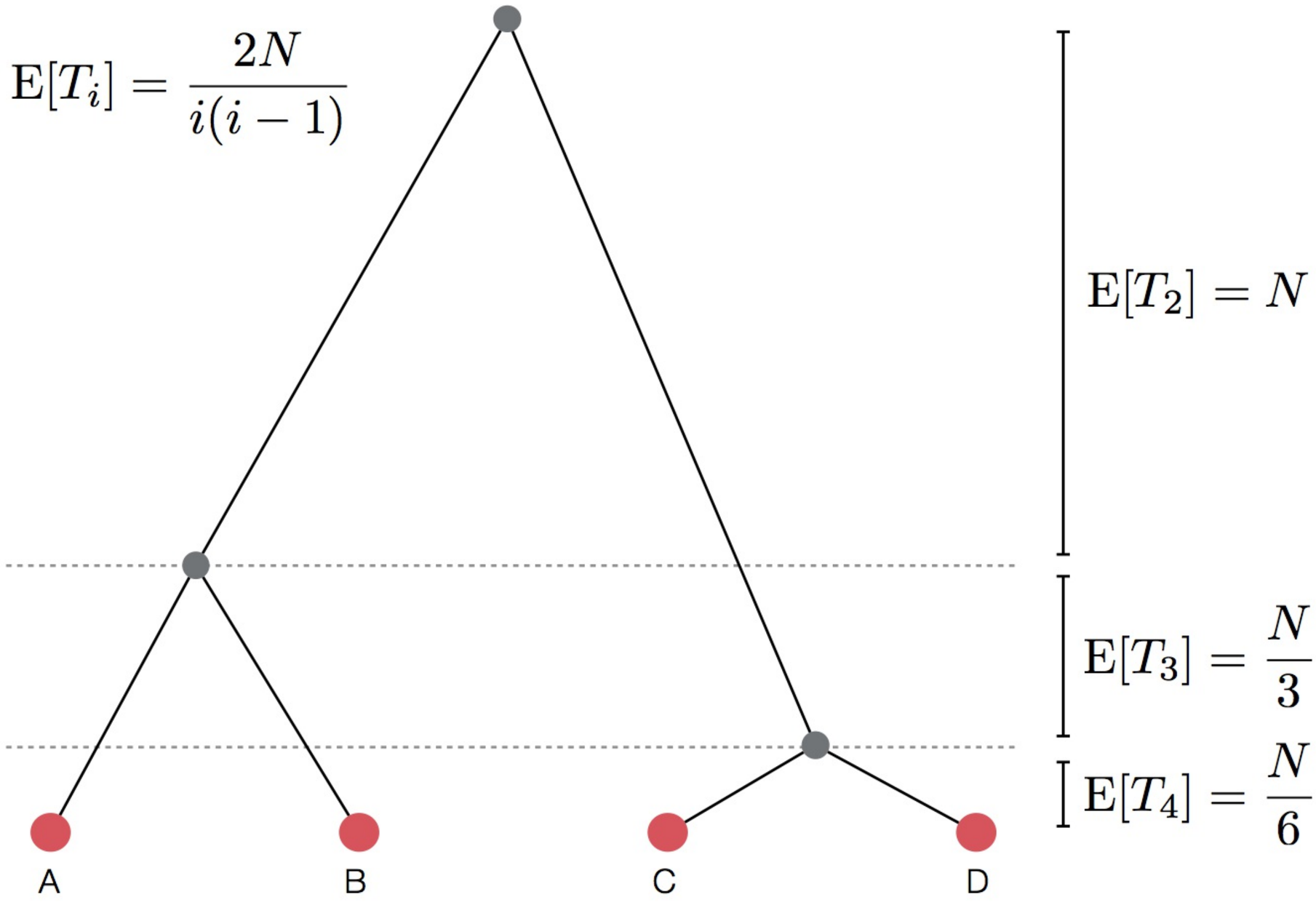
Continuous time limit

With per-generation probability of an event $\frac{1}{x}$ small, but many generations, then the discrete time geometric distribution approximates to a continuous time exponential distribution. Thus, we assume T_i to be exponentially distributed with mean

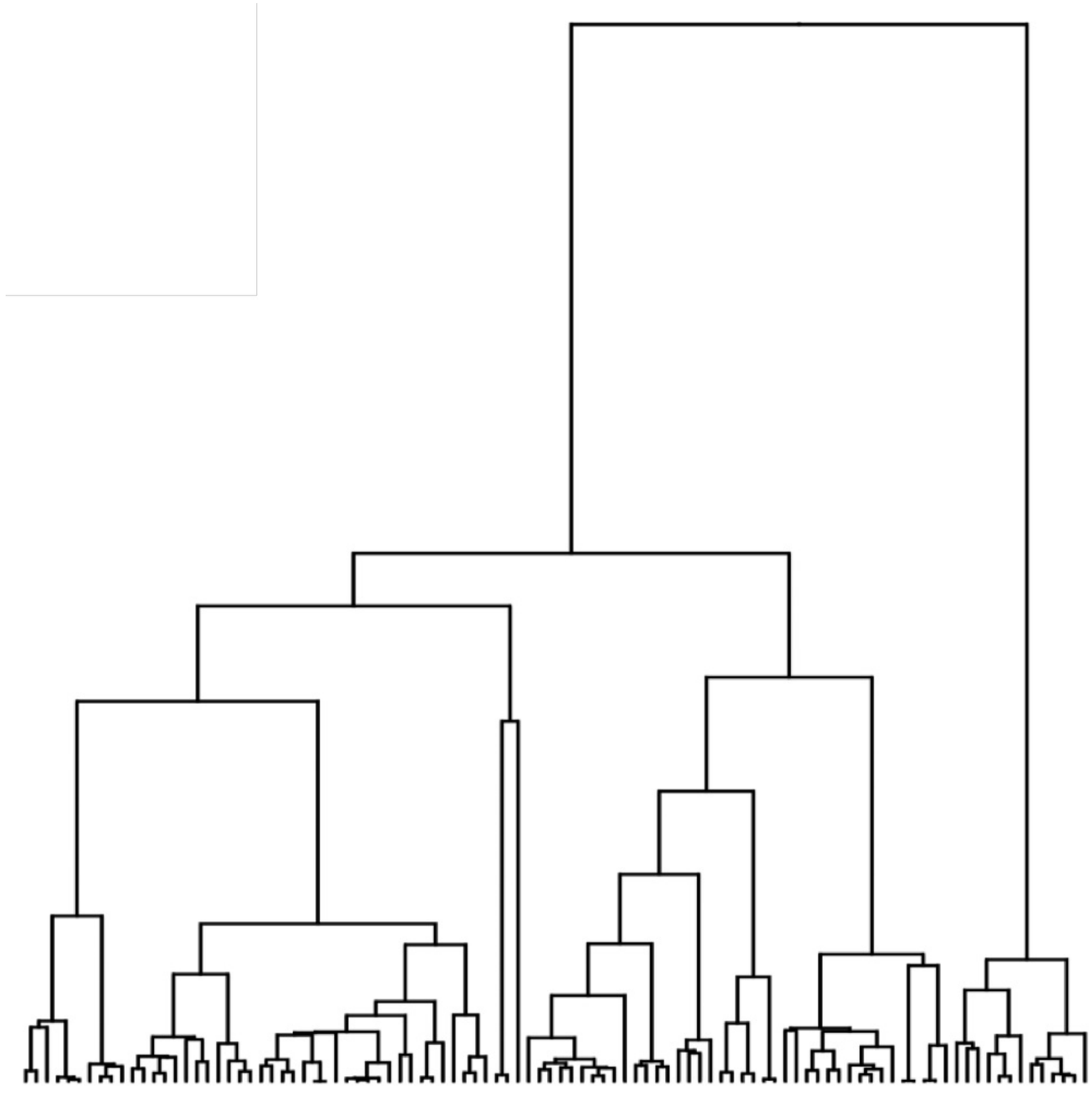
$$\mathbf{E}[T_i] = \frac{2N}{i(i-1)}.$$

Set of coalescent intervals with waiting times

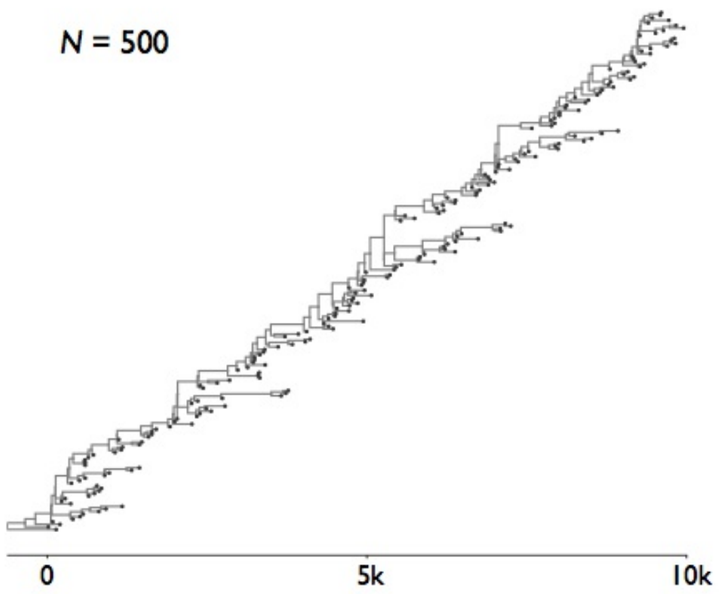
$$\mathrm{E}[T_i] = \frac{2N}{i(i-1)}$$

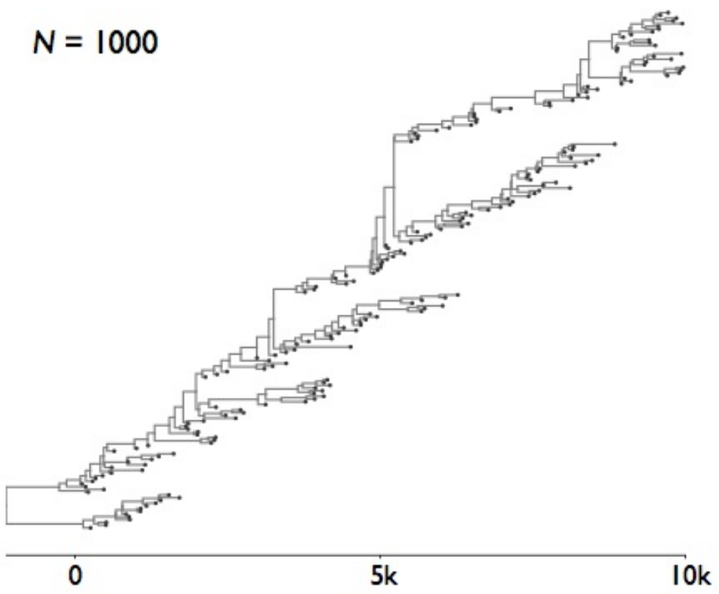


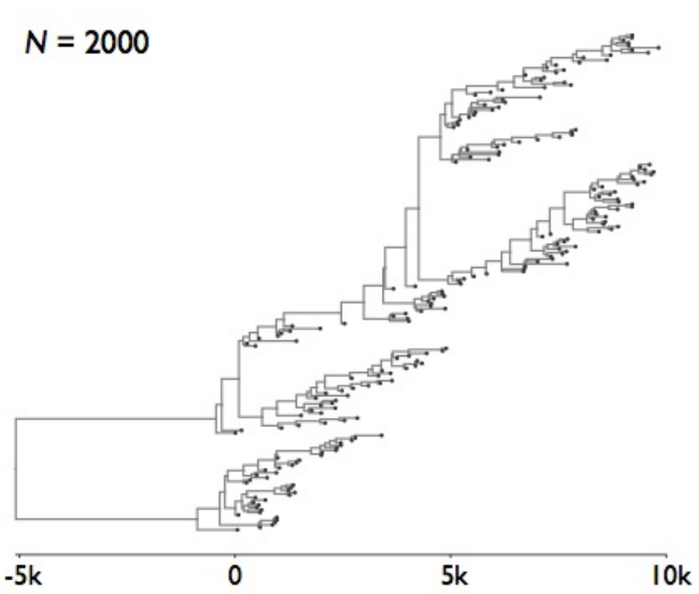
Gives coalescent trees their distinctive shape

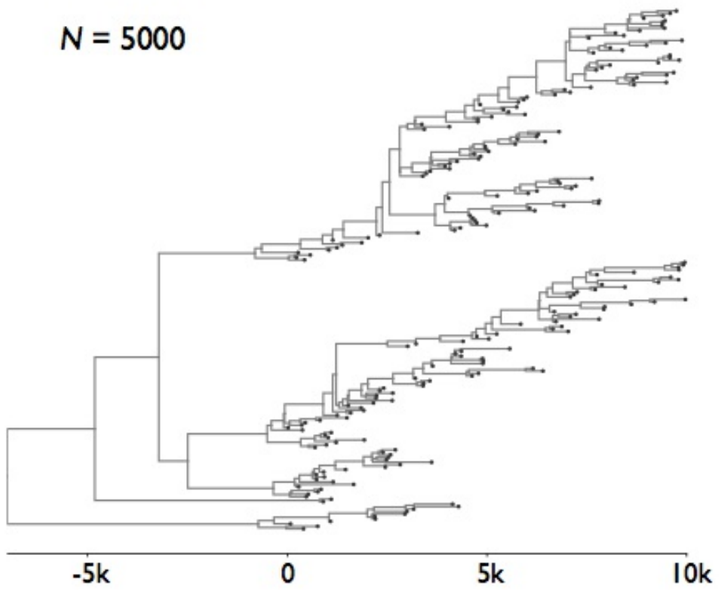


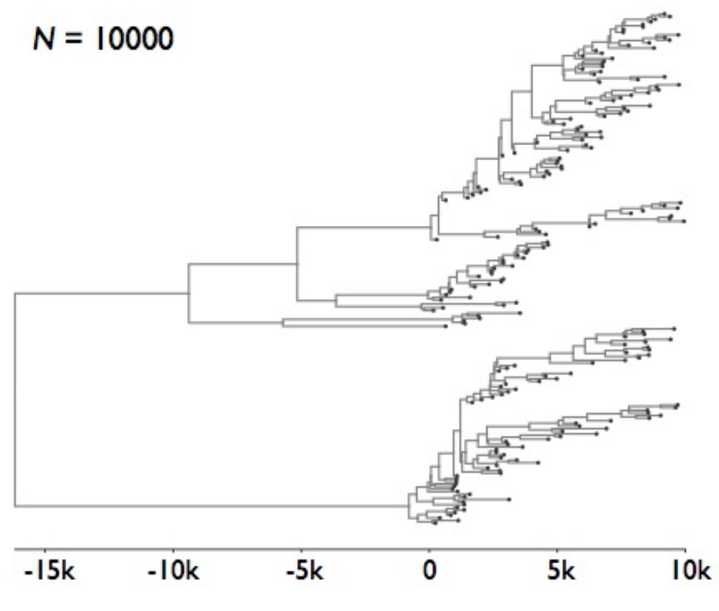
Rate of coalescence scales with population size N

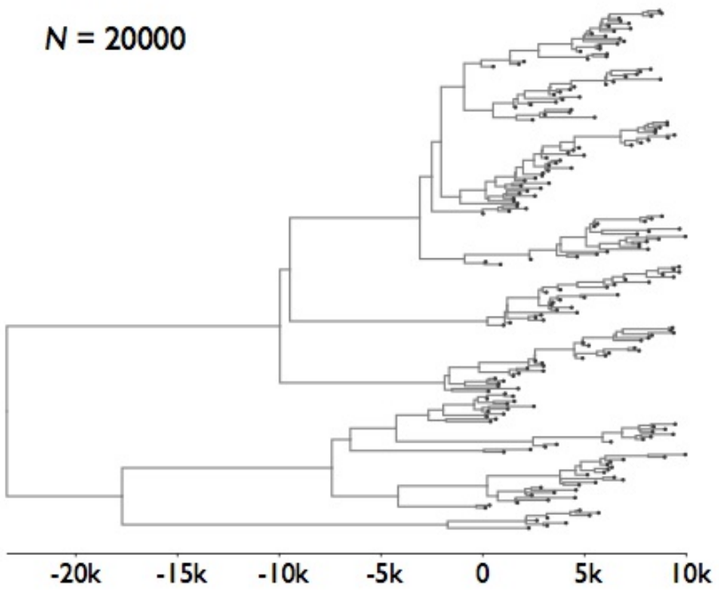








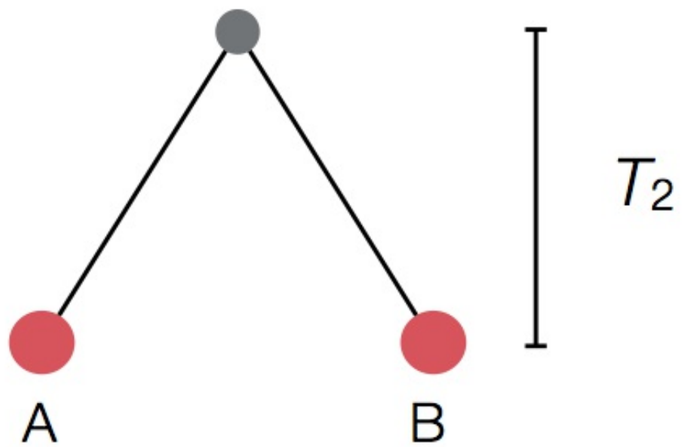




Visualization of the coalescent process

Pairwise genetic diversity

$$\mathrm{E}[T_2] = N$$



Gives 2*N* generations for mutations to occur

π is number of substitutions per site between two random individuals

$\mathrm{E}[\pi] = 2N\mu$ where μ is mutations per site per generation

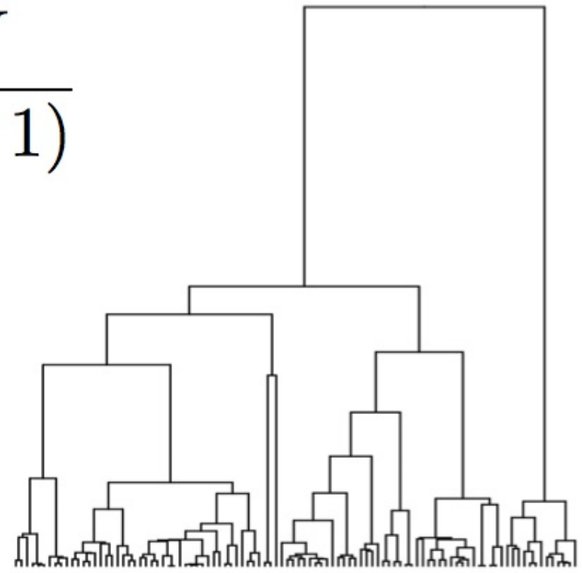
$\mathrm{E}[\pi] = 2N\tau\mu$ where μ is mutations per site per year and

τ is years per generation

Time to the most recent common ancestor (TMRCA)

$$T_{\mathrm{MRCA}} = \sum_{i=2}^{n} T_i \qquad \mathrm{E}[T_i] = \frac{2N}{i(i-1)}$$

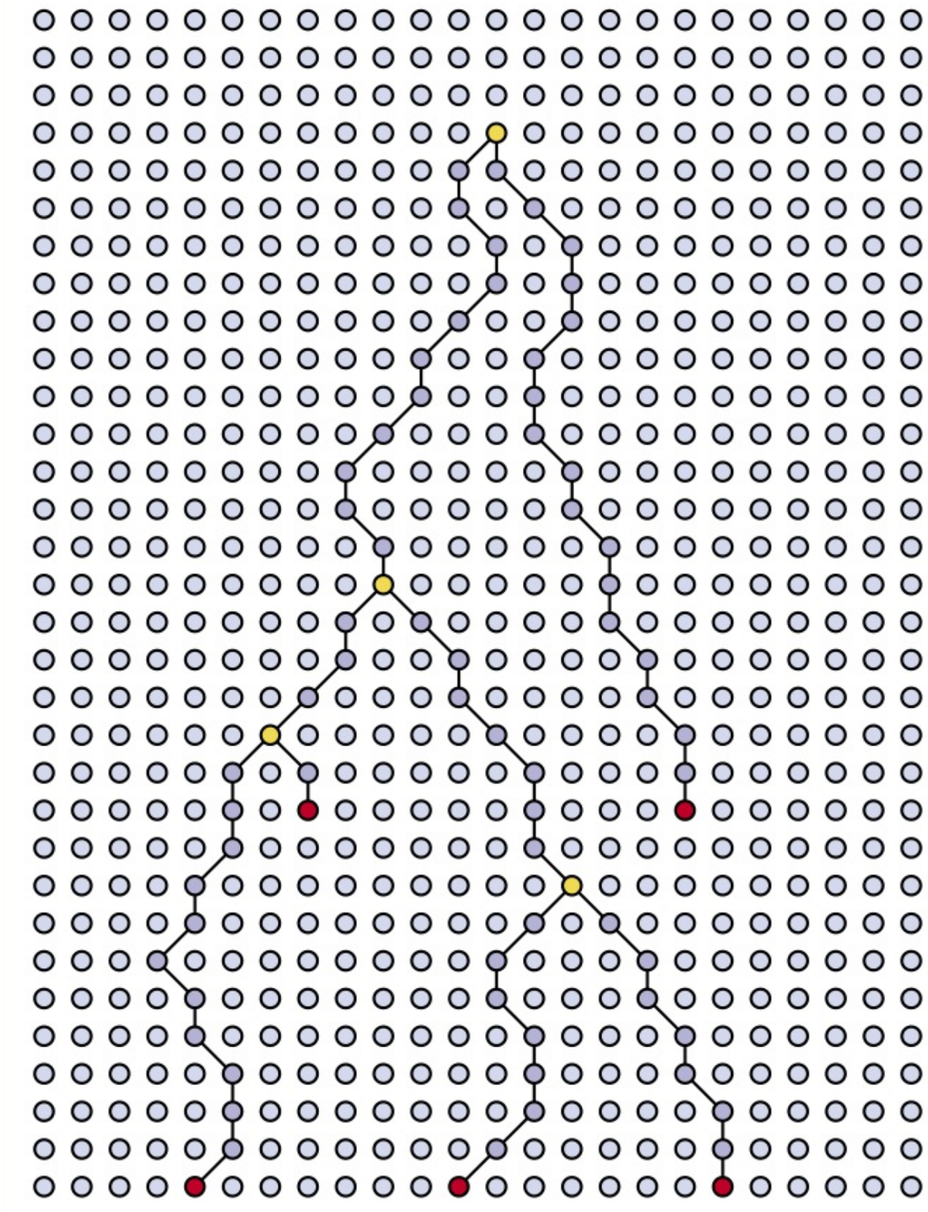
$$\begin{aligned}
\mathrm{E}[T_{\mathrm{MRCA}}] &= \sum_{i-2}^{n} \frac{2N}{i(i-1)} \\
&= 2N \sum_{i-2}^{n} \left(\frac{1}{i-1} - \frac{1}{i} \right) \\
&= 2N \left(1 - \frac{1}{2} + \frac{1}{2} - \cdots - \frac{1}{n-1} + \frac{1}{n-1} - \frac{1}{n} \right) \\
&= 2N \left(1 - \frac{1}{n} \right)
\end{aligned}$$



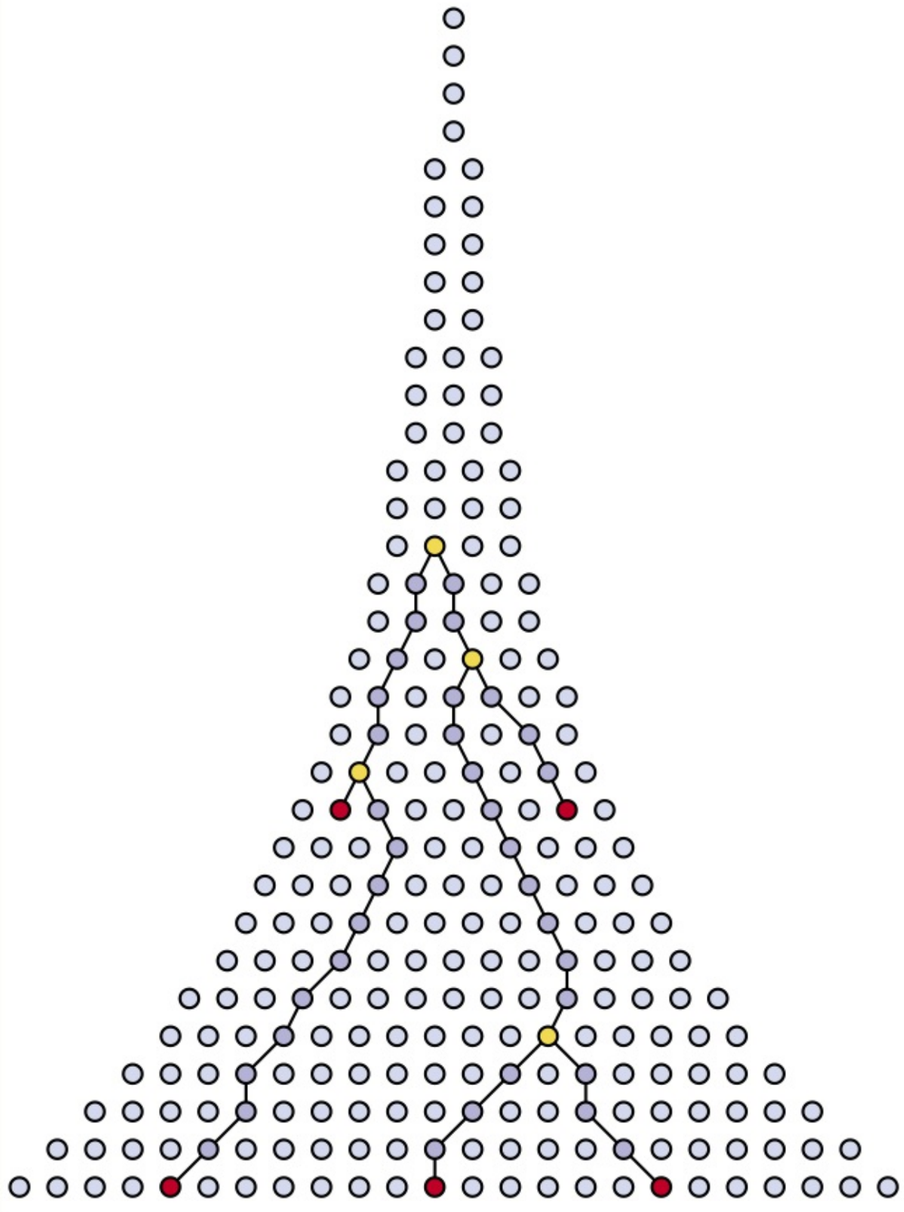
$$\mathrm{E}[T_{\mathrm{MRCA}}] = \lim_{n \to \infty} 2N$$

Coalescence patterns can estimate population growth/decline

Changing population size alters coalescent rate

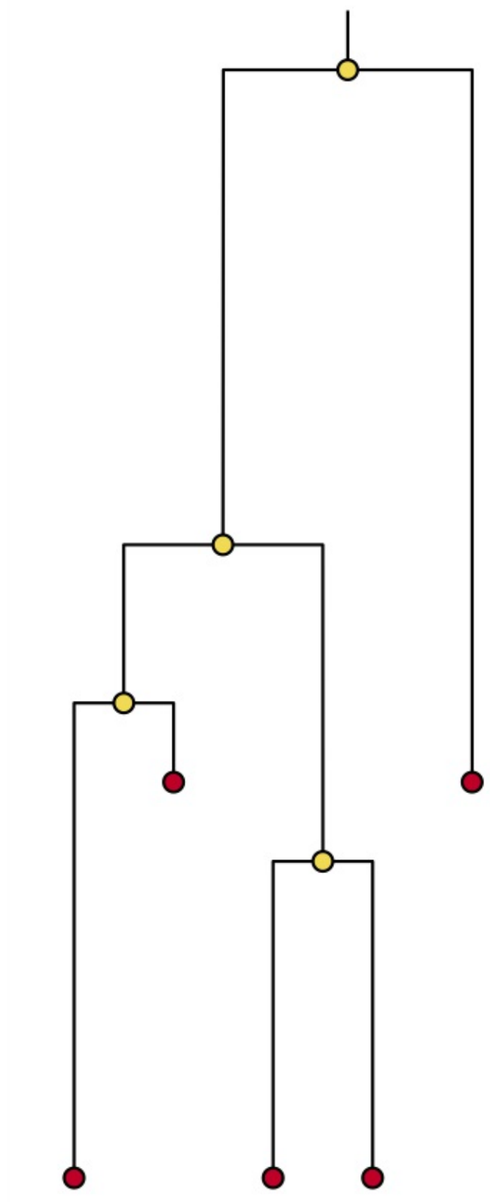


Constant size

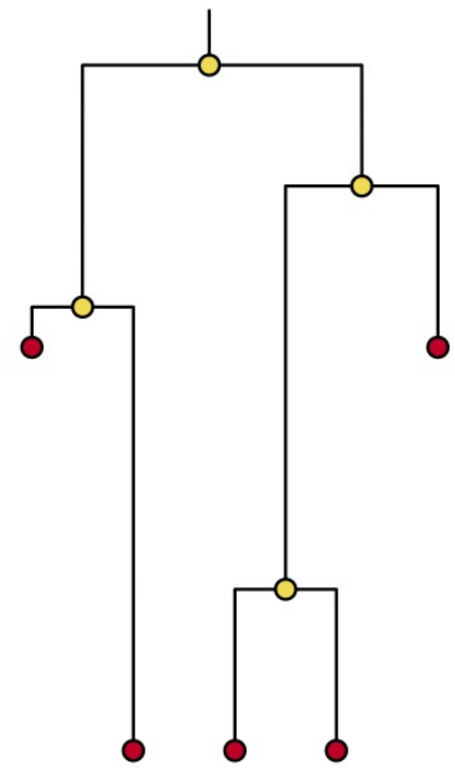


Growing population

Changing population size alters coalescent rate

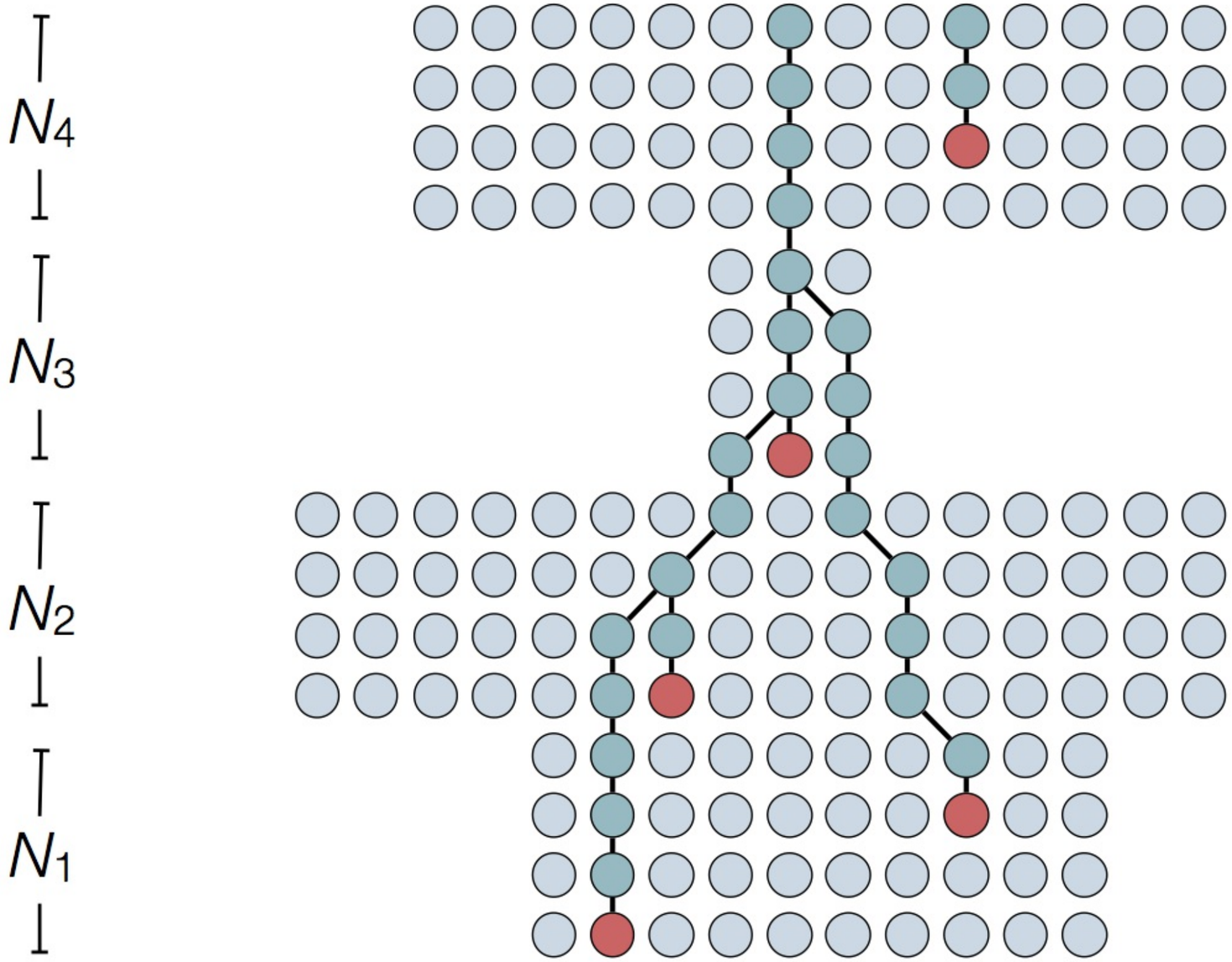


Constant size



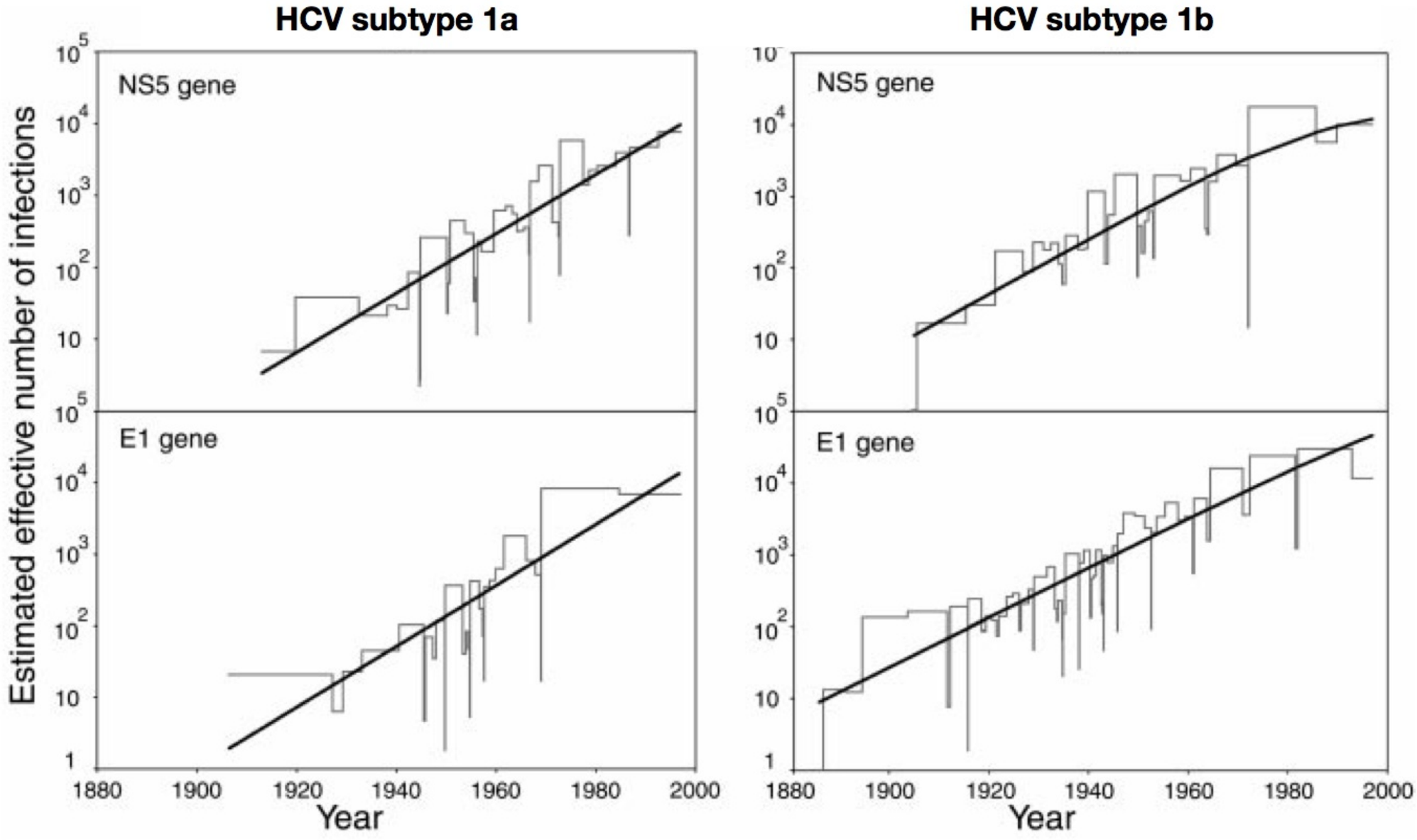
Growing population

'Skyline' is flexible demographic model that estimates windows of coalescent rate



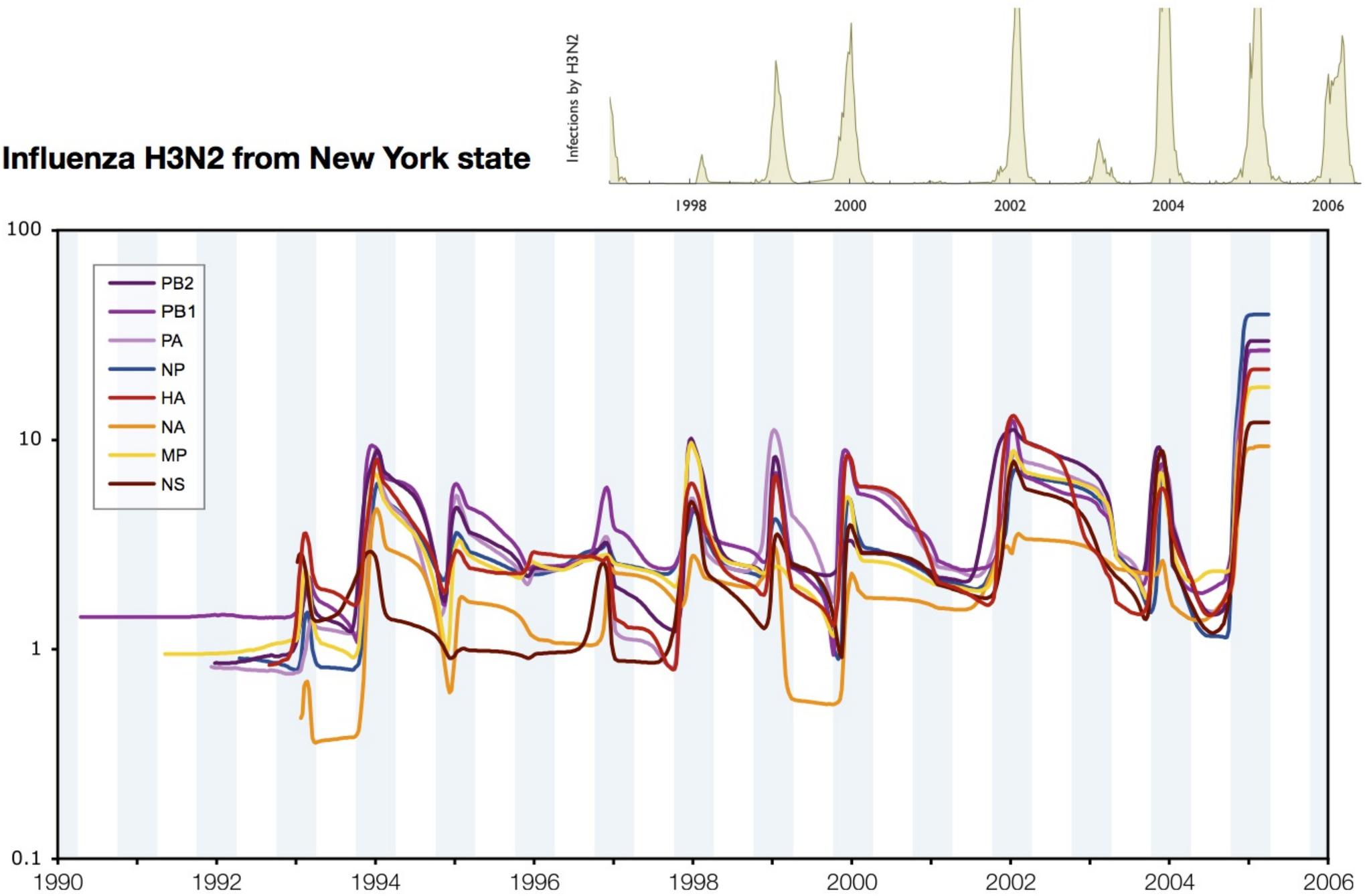
Drummond et al. 2005

Skyline model shows population growth in HCV



Pybus et al. 2001

Skyline model shows seasonality in flu



Rambaut et al. 2008

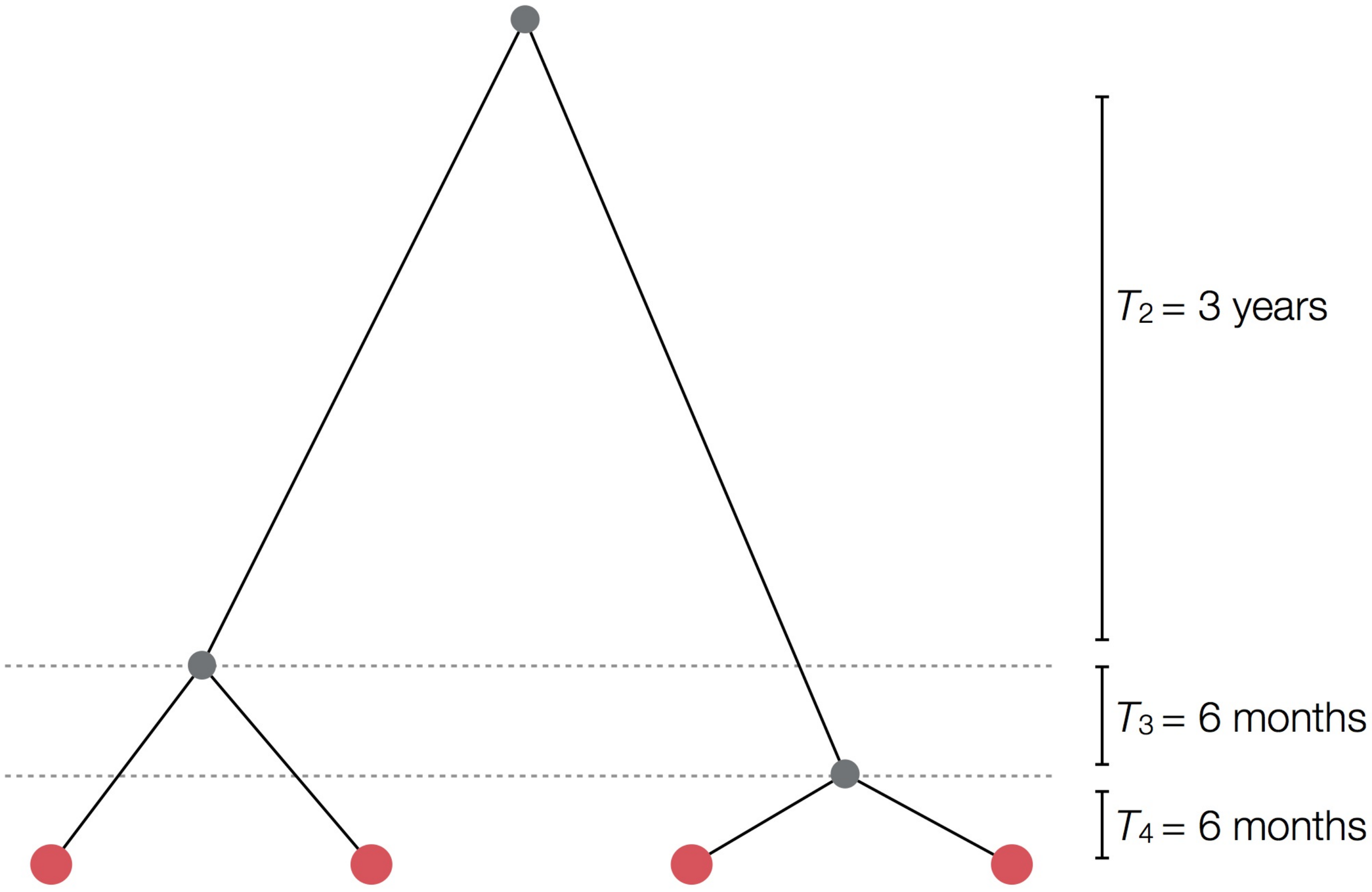
Interpreting rate of coalescence

These approaches directly estimate the pairwise rate of coalescence λ, which is measured in terms of events per year. Thus, the timescale of coalescence $\frac{1}{\lambda}$ is measured as the expected waiting time in years for two lineages to find a common ancestor.

Interpreting rate of coalescence

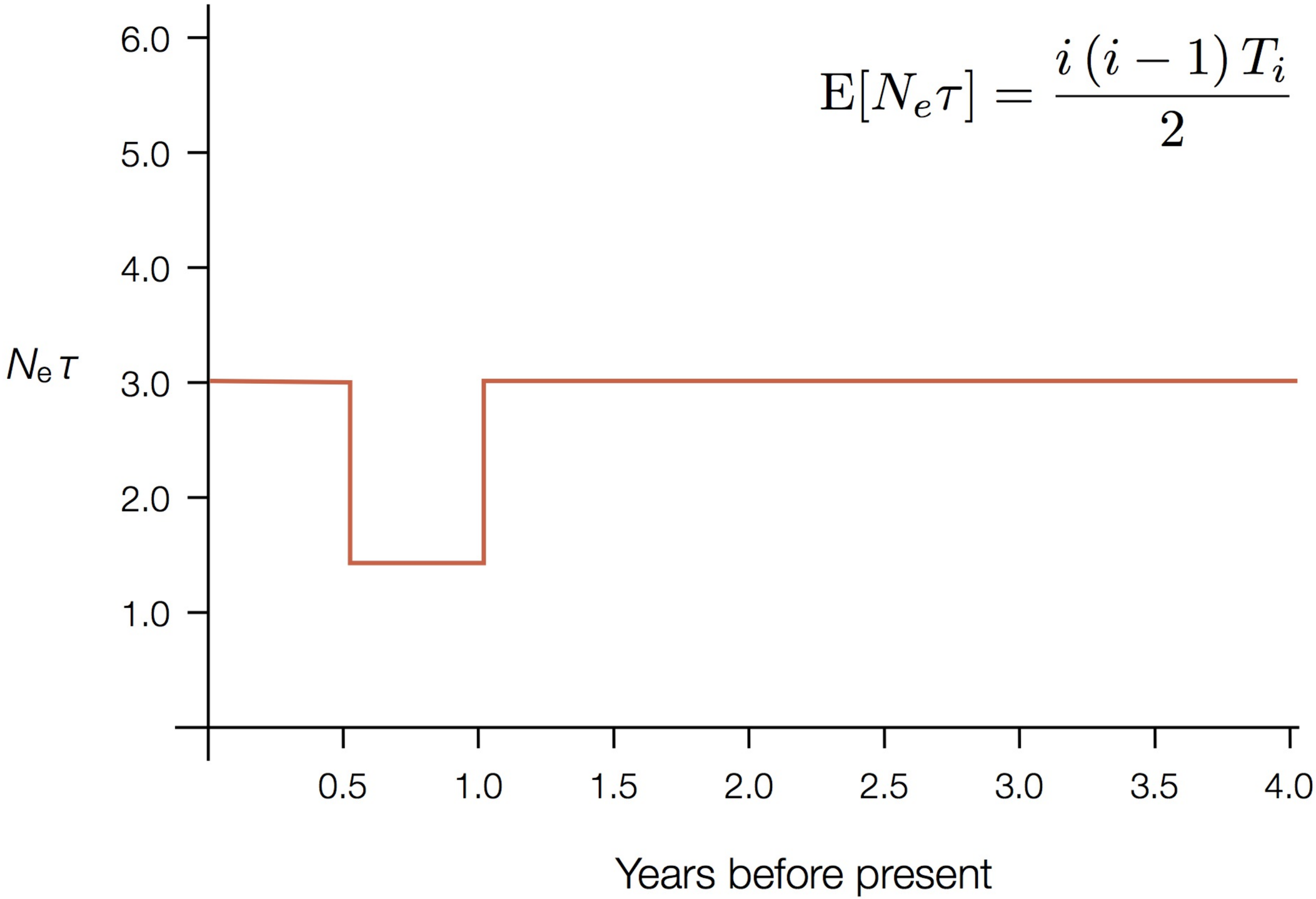
The timescale of coalescence $\frac{1}{\lambda}$ is equal to $N_e\tau$, where N_e is measured in generations and τ is measured in years per generation. τ acts to rescale time from generations to years.

Example tree

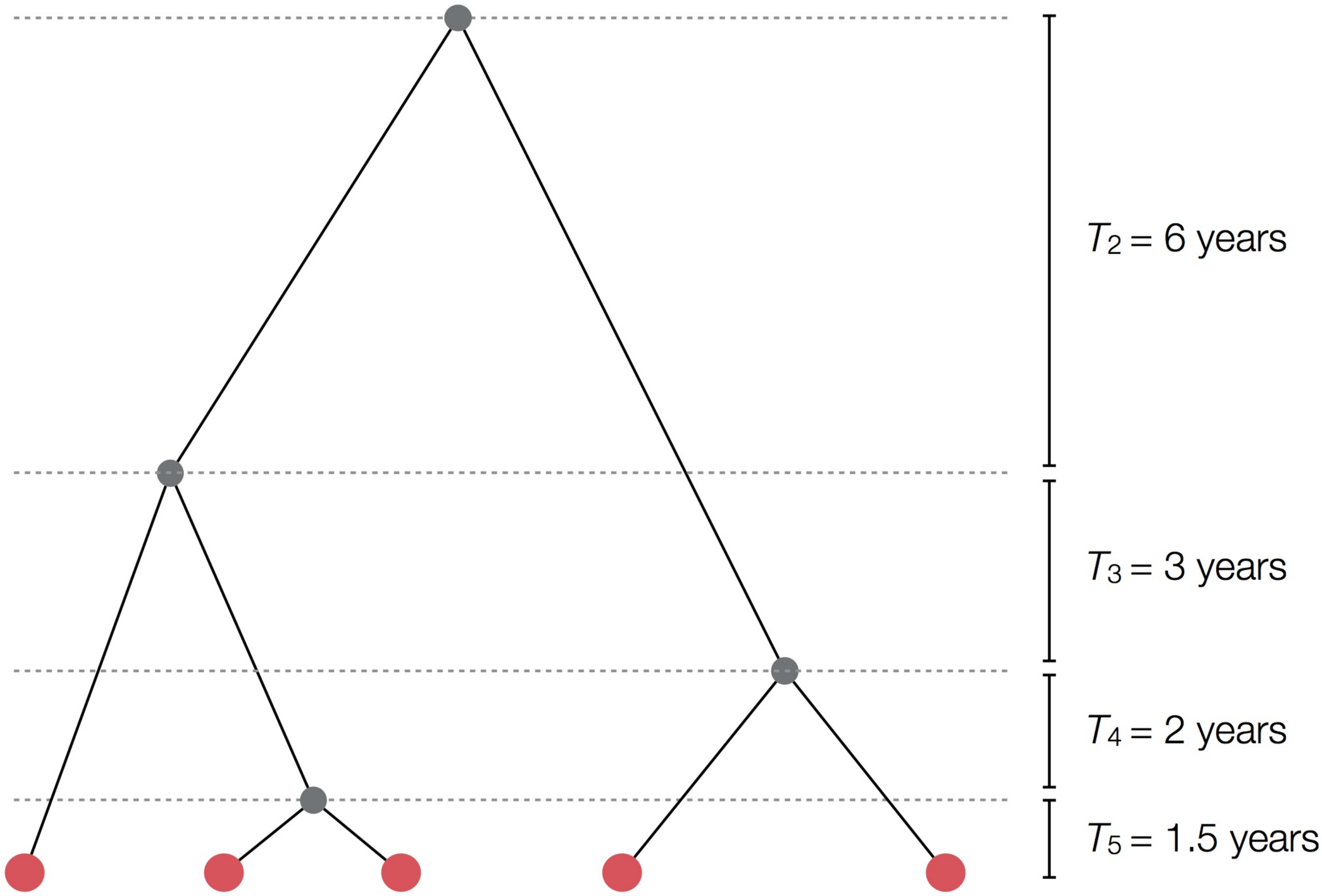


Example skyline plot

$$\mathrm{E}[N_e\tau] = \frac{i\,(i-1)\,T_i}{2}$$

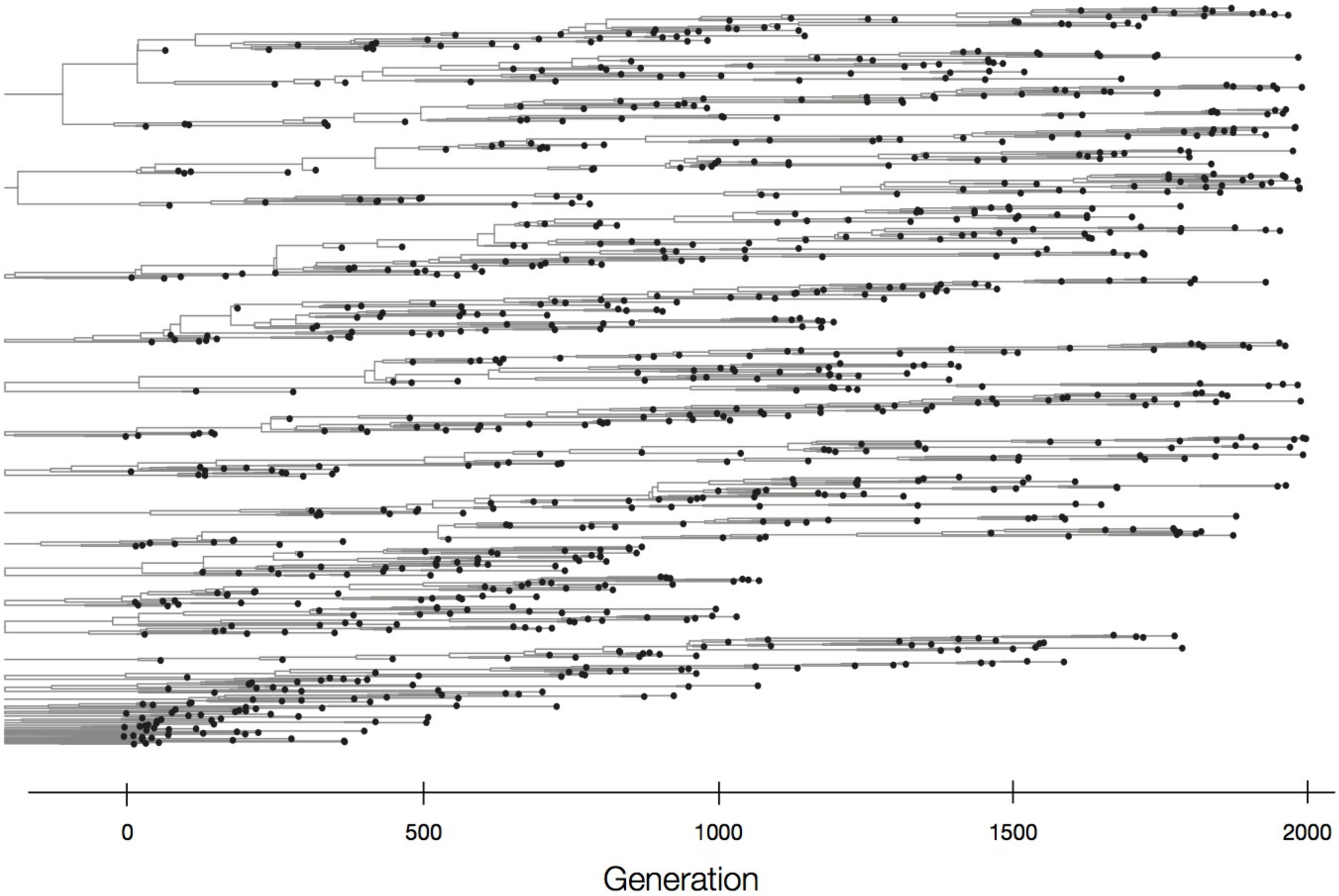


Exercise: plot skyline from tree

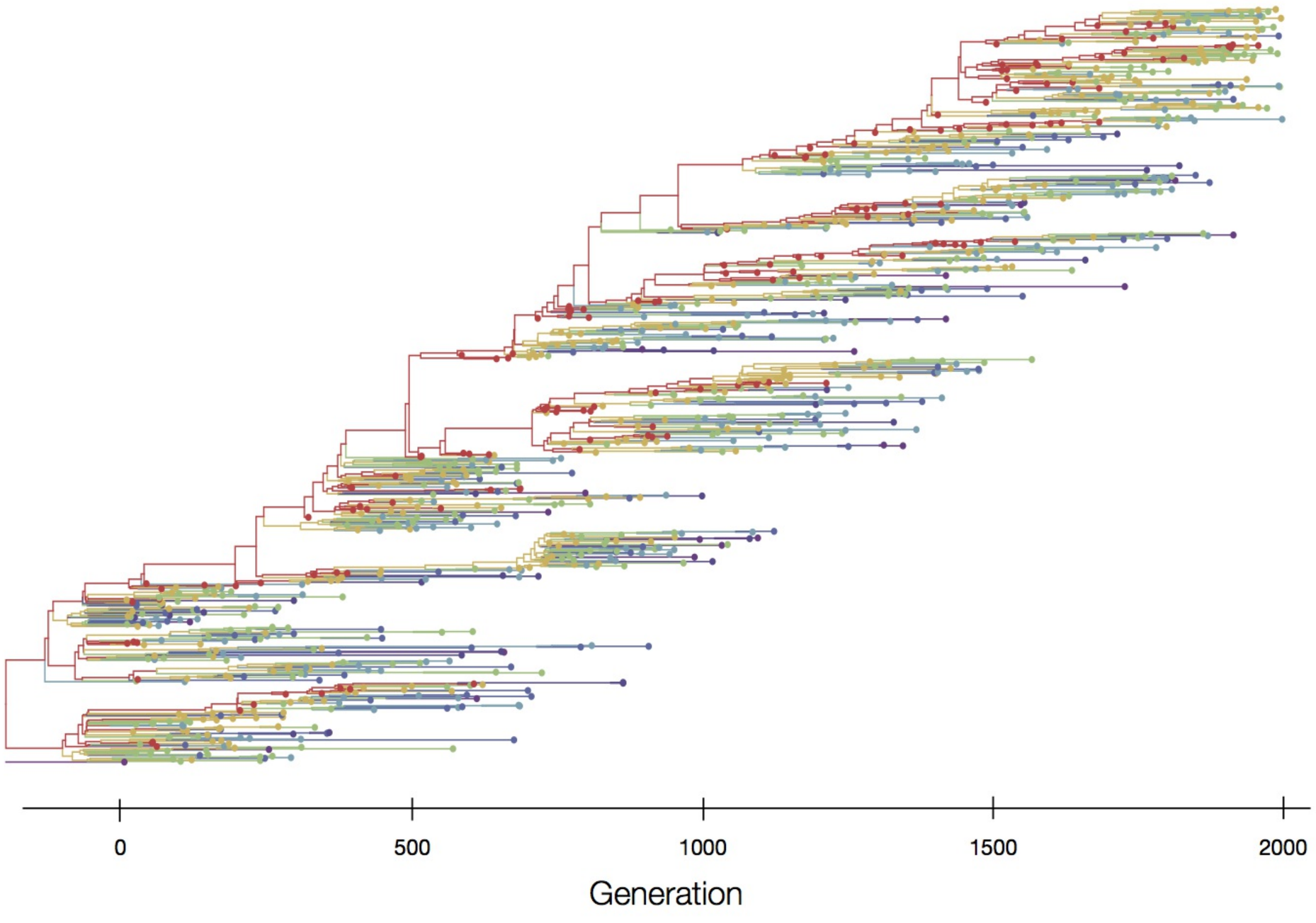


Selection distorts coalescence patterns

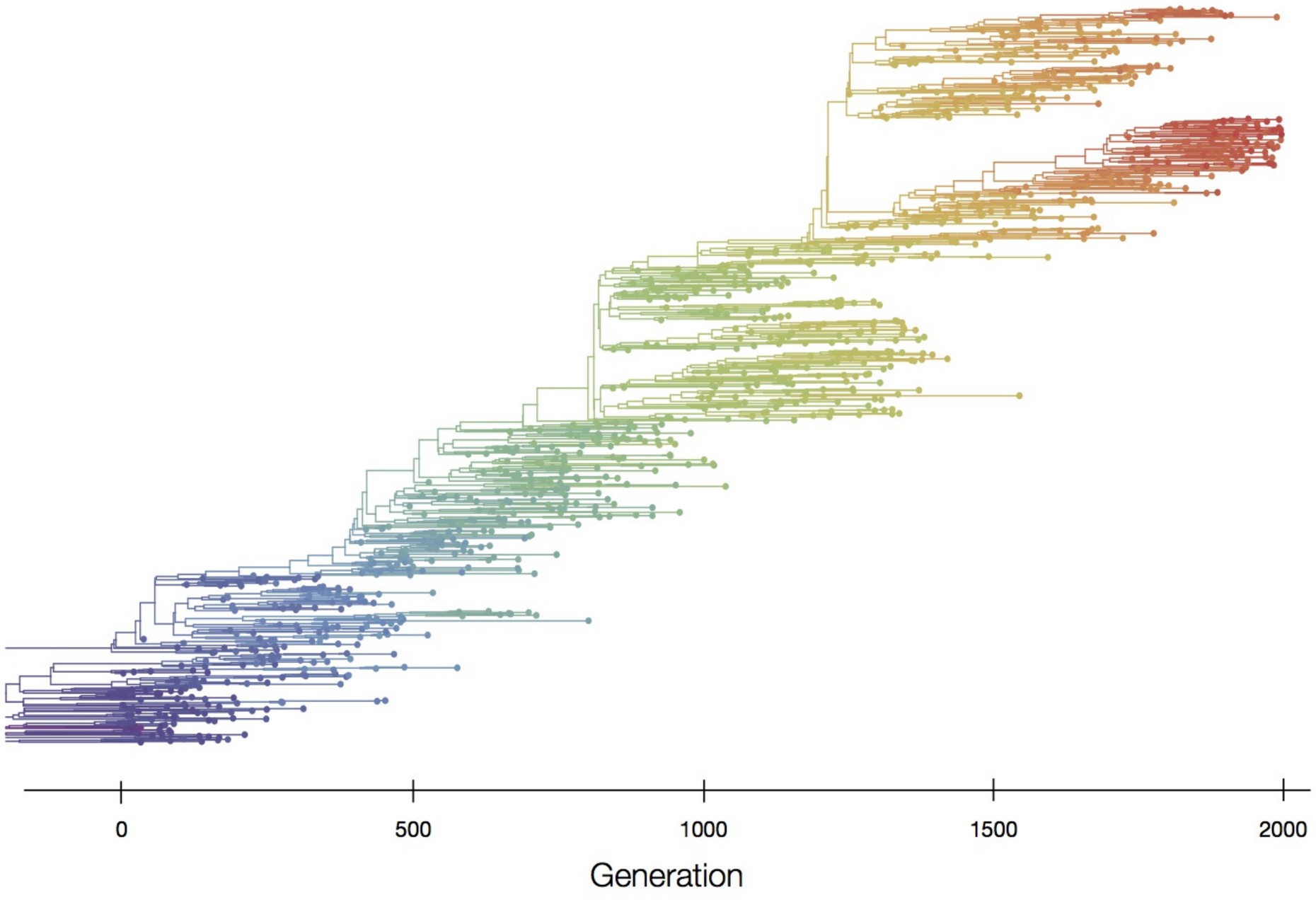
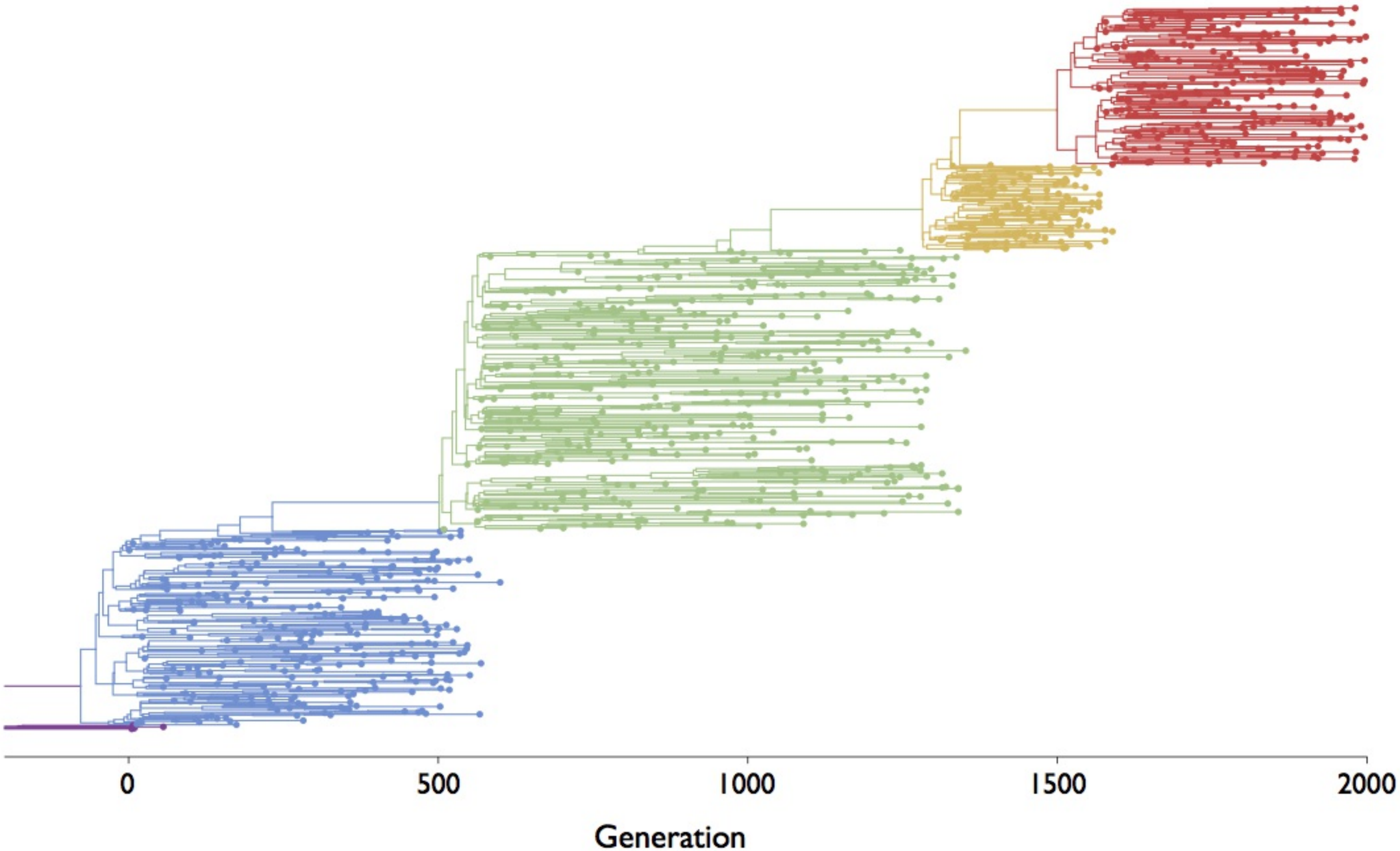
Neutral population dynamics



Purifying selection reduces genetic diversity and leads to population stasis



Positive selection reduces genetic diversity leads to population turnover



Episodic positive selection shows selective sweeps