

---

# The Wright-Fisher model

- Markov chains are useful because expectations of quantities of interest can be computed via a *first step analysis* [note 4.2].
- This is best illustrated by examples:
  - Example 1: Probability of fixation [**Note 3.1**]
  - Example 2: Time to absorption. [**Note 3.2**]
  - Example 3: Decay in heterozygosity. [**Note 4.4**]

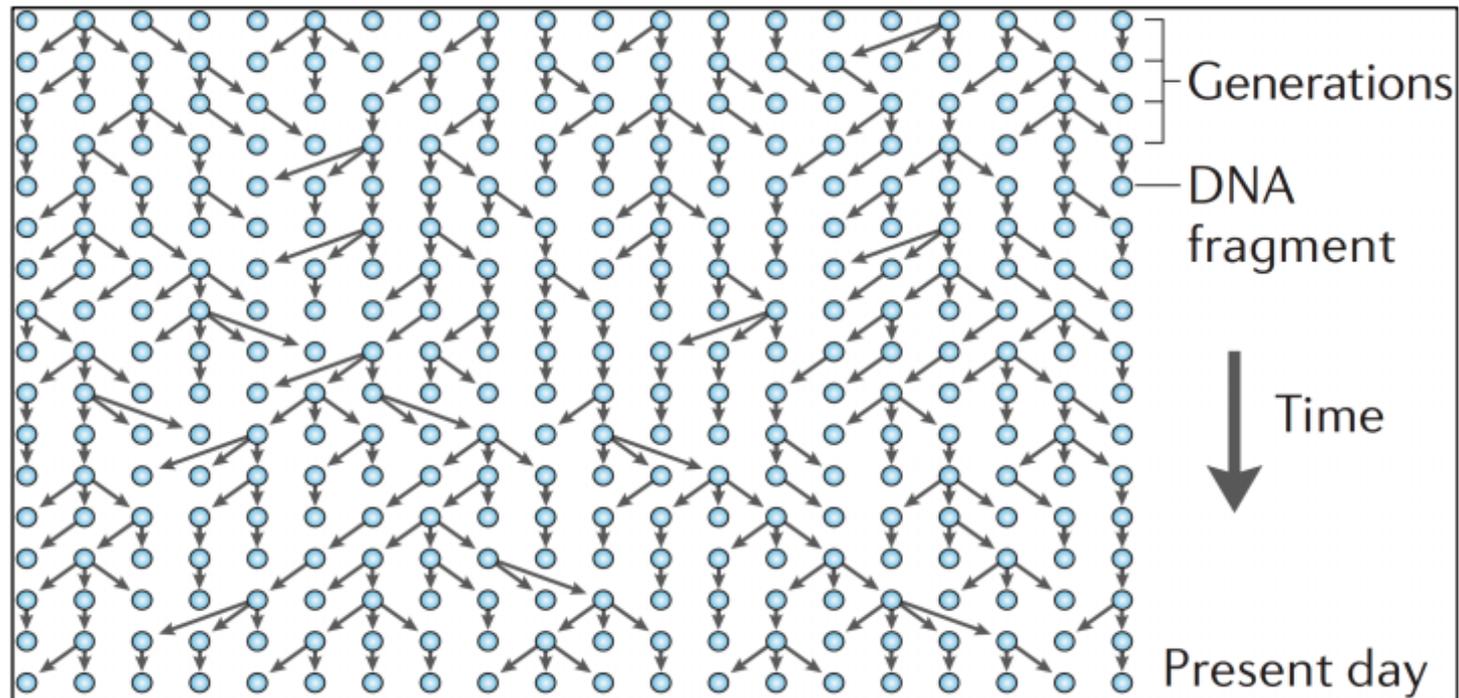
---

## 3. Scaling limits and Kingman's coalescent.

---

# 5. Approximating the Wright-Fisher model

- Recap: We met the Wright-Fisher model,



and computed three quantities of interest under it, exploiting the fact that it is a ***time-homogeneous Markov chain***.

# Approximating the Wright-Fisher model

- We found:
  - The exact fixation probability:

$$\text{Prob}(\text{fixation given initial frequency } p) = p.$$

- The approximate expected time to absorption:

$$E[T | p_0 = p] = -2N [p \log(p) + (1-p) \log(1-p)].$$

- The mean heterozygosity:

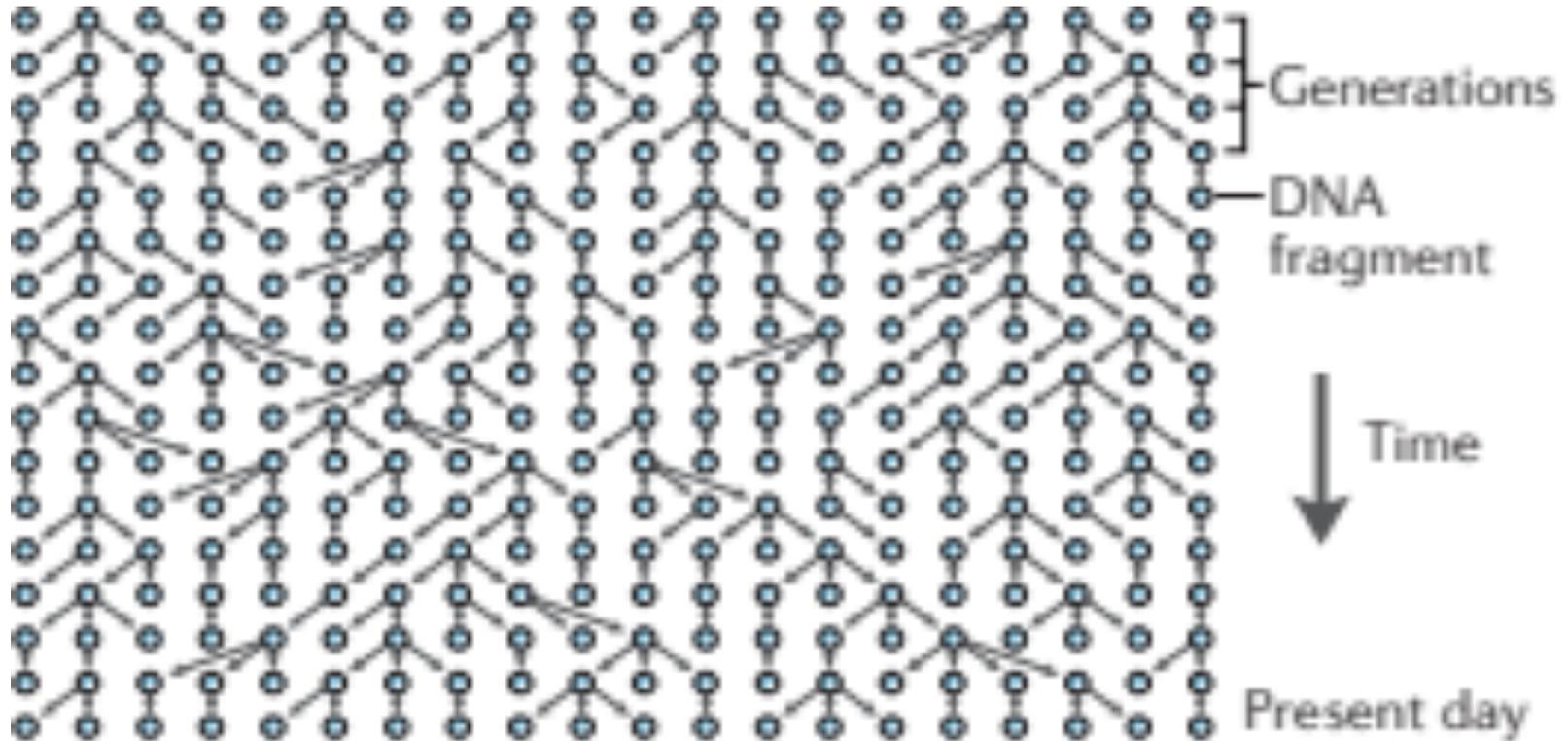
$$E[H_k | H_0] = H_0 \left(1 - \frac{1}{N}\right)^k$$

- The last two results tell us that genetic drift *slows down* in larger populations.

# Rescaling the Wright-Fisher model

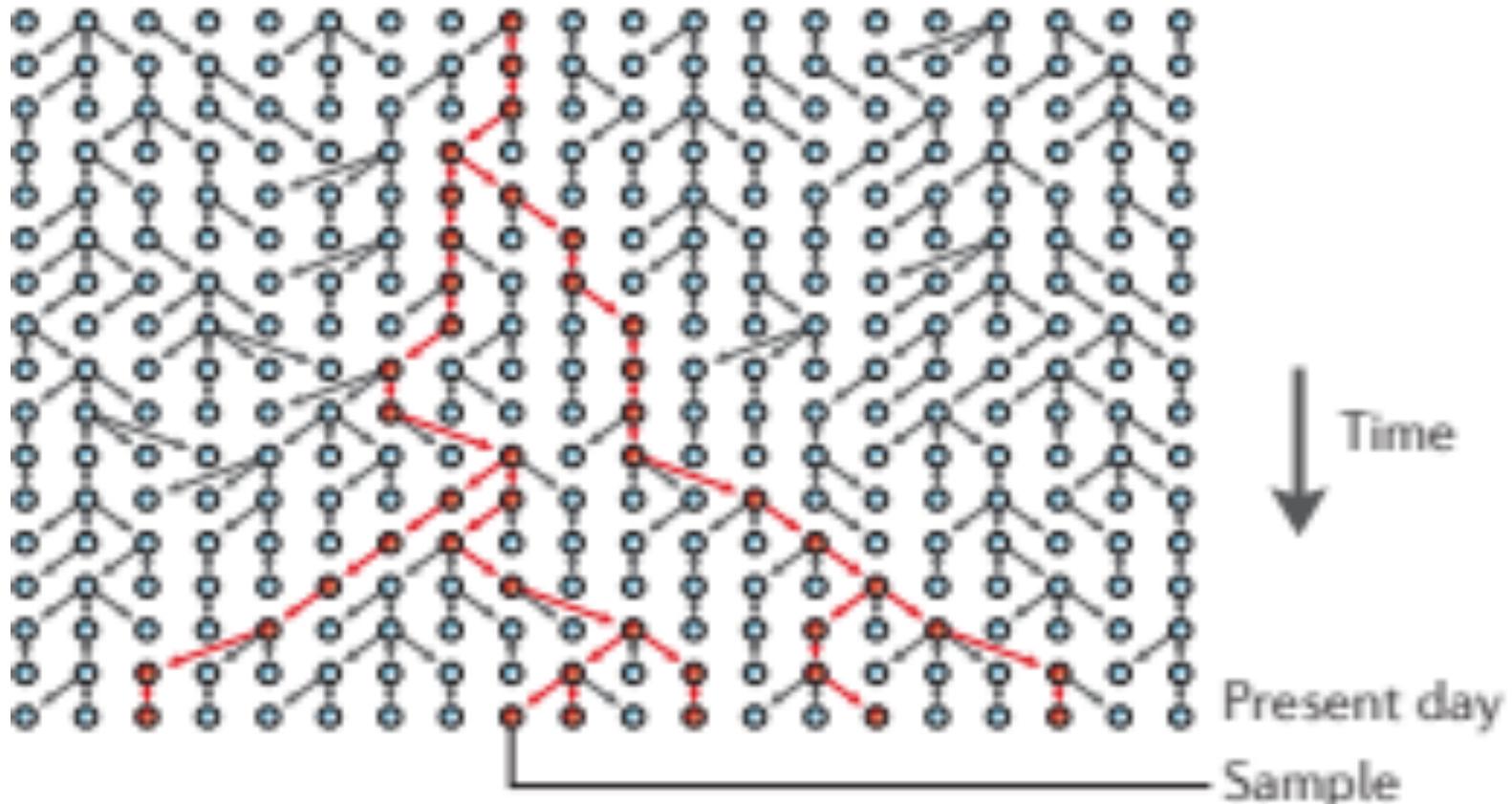
- For large  $N$ , the time for ‘something to happen’ (say, absorption) grows linearly with  $N$ , but in *units of  $N$  generations* there is a non-trivial limiting distribution for the absorption time as  $N$  goes to infinity.
- Under the same rescaling, the decay in heterozygosity is well-behaved too **[Note 5.1]**.
- Something is going on:
  - If we just let  $N$  go to infinity, genetic drift has rate 0.
  - But if we let  $N$  go to infinity *and rescale time appropriately* we get an interesting limit.
- In fact, the whole Wright-Fisher model converges to something. Our goal is to find its limit.

# Rescaling the Wright-Fisher model



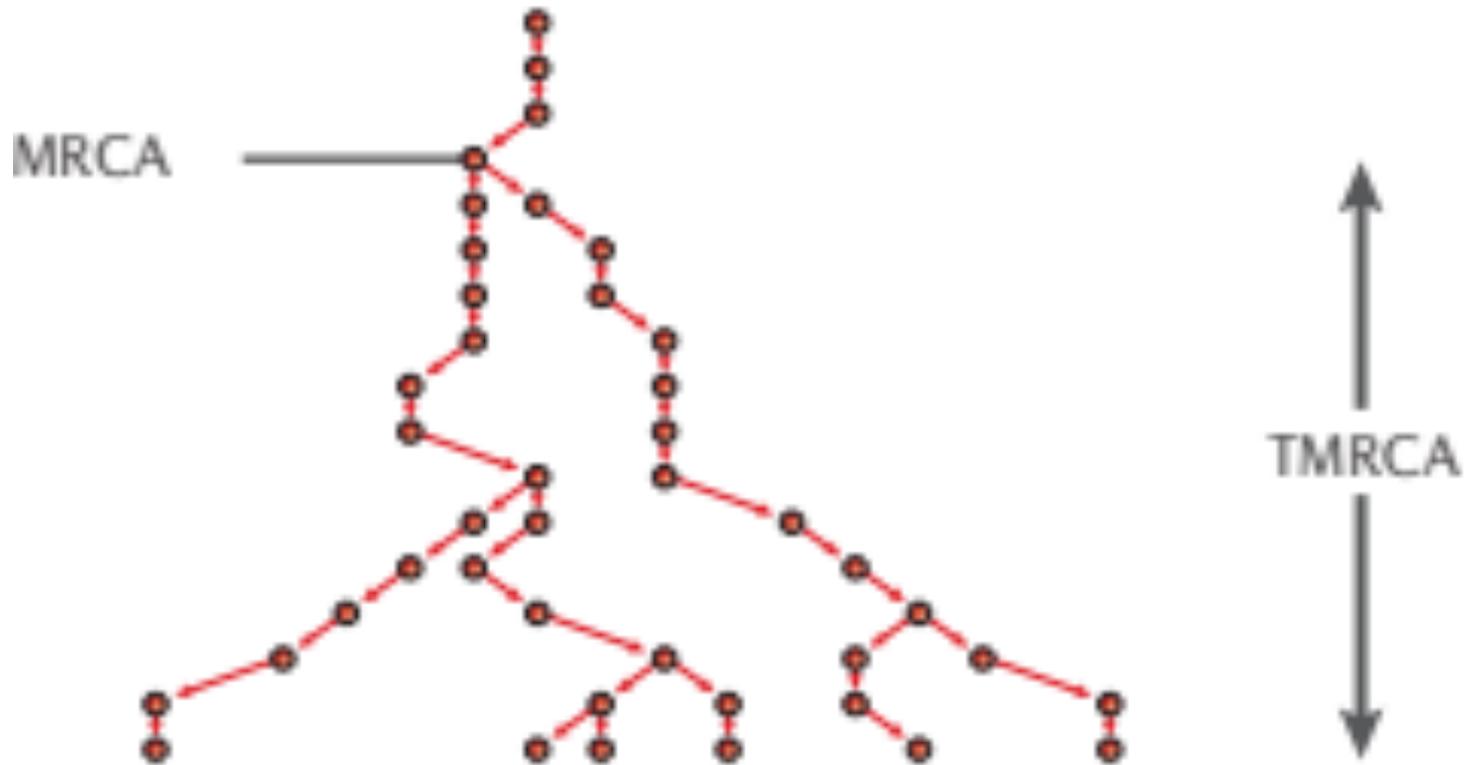
- What do the genealogies of the Wright-Fisher model look like?

# Rescaling the Wright-Fisher model



- What do the genealogies of the Wright-Fisher model look like?

# Rescaling the Wright-Fisher model



- What do the genealogies of the Wright-Fisher model look like?

# Rescaling the Wright-Fisher model

- Let's start with a simpler question: What is the probability that two individuals share their most recent common ancestor  $k$  generations ago? **[Note 5.2]**
- The answer is complicated – already we are in trouble!
- Mathematics thrives on approximating complicated quantities by simpler ones. **[Note 5.3]**

*If we take a sample of  $n$  lineages from a Wright-Fisher model with a large population size  $N$ , then the waiting time until two or more of them find a most recent common ancestor (MRCA) is approximately exponentially distributed with parameter*

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

*in units of  $N$  generations. When a common ancestor is found, it is between precisely two (and not three, four, ...) lineages.*

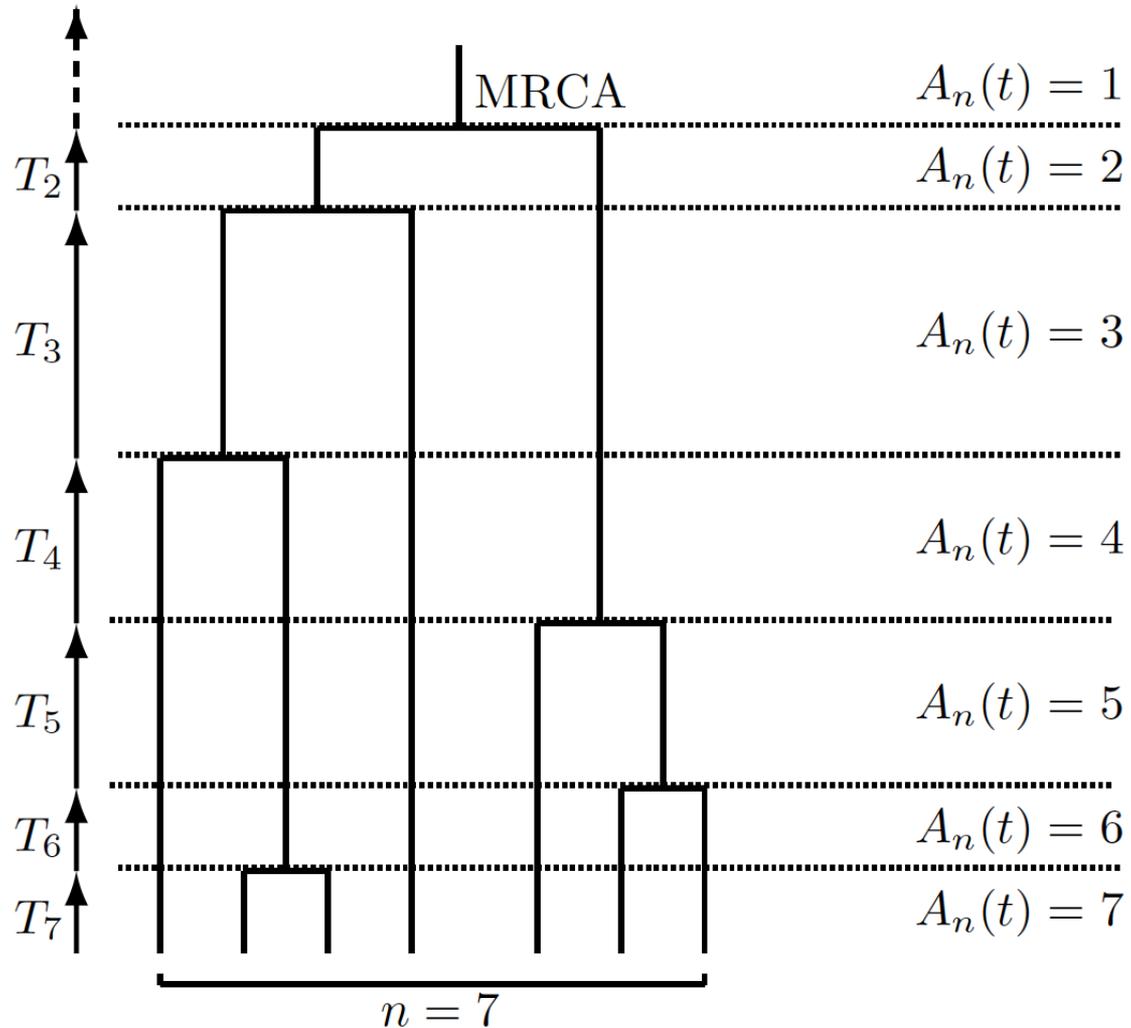
# Rescaling the Wright-Fisher model

- When an MRCA is found, the number of lineages drops to  $n - 1$  and we continue as before, with the remaining lineages coalescing after a time

$$T_{n-1} \sim \text{Exponential} \left( \binom{n-1}{2} \right).$$

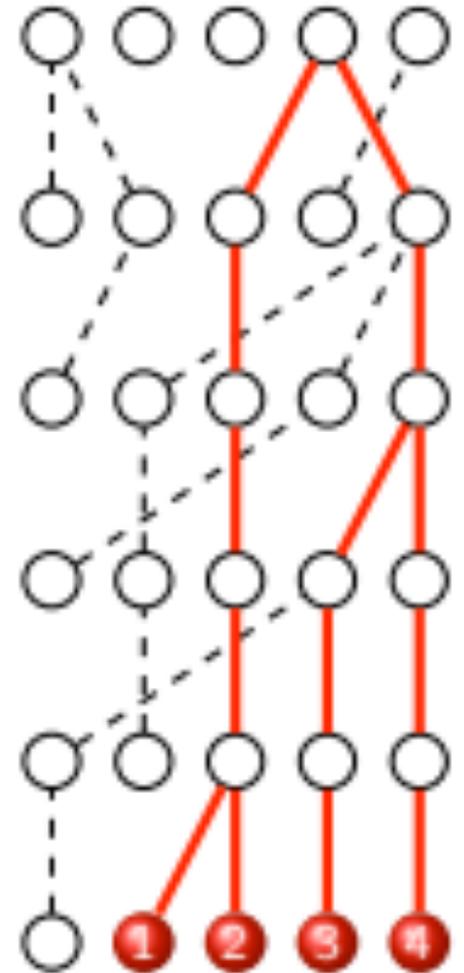
- And so on – we can continue until we reach a MRCA of the whole sample.
- This argument lets us define the ***ancestral process*** [Note 5.4].

# The ancestral process



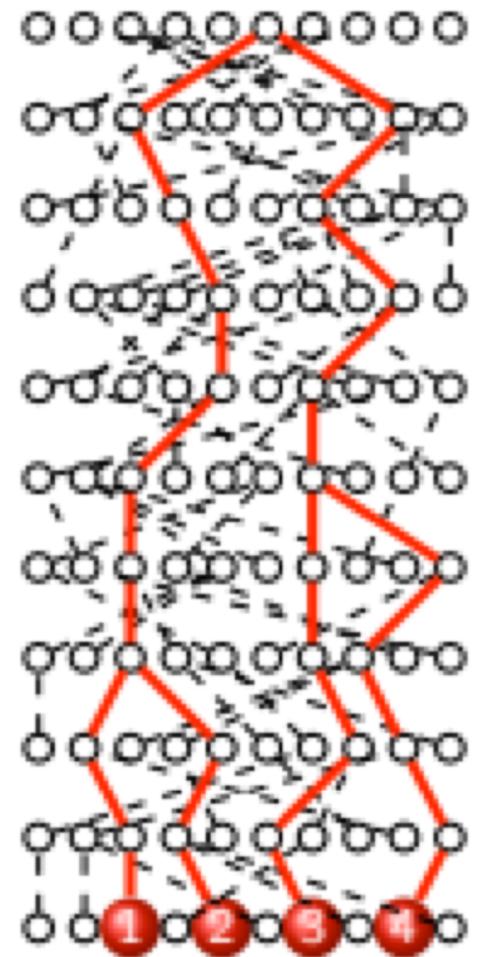
# The ancestral process: summary

- As the population get large, the timescale also gets big.
- Our limit was obtained by ‘zooming out’ from the Wright-Fisher model.



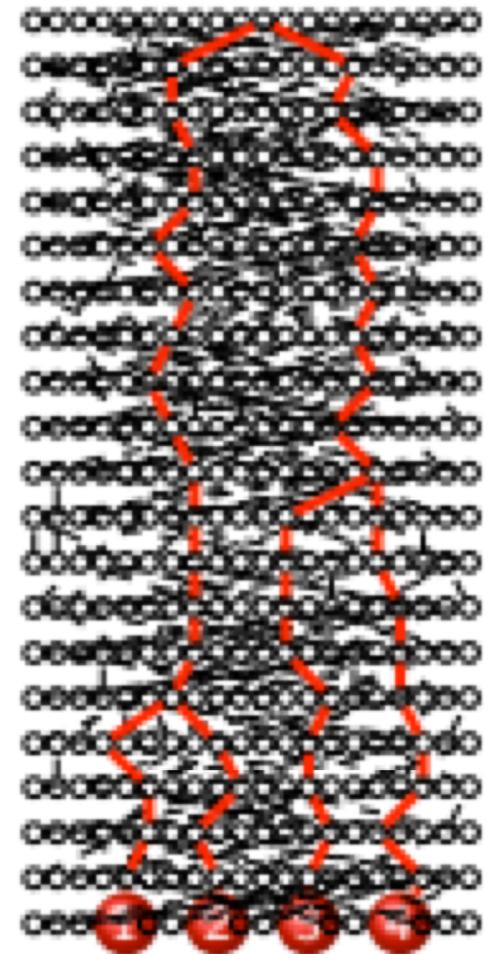
# The ancestral process: summary

- As the population get large, the timescale also gets big.
- Our limit was obtained by ‘zooming out’ from the Wright-Fisher model.



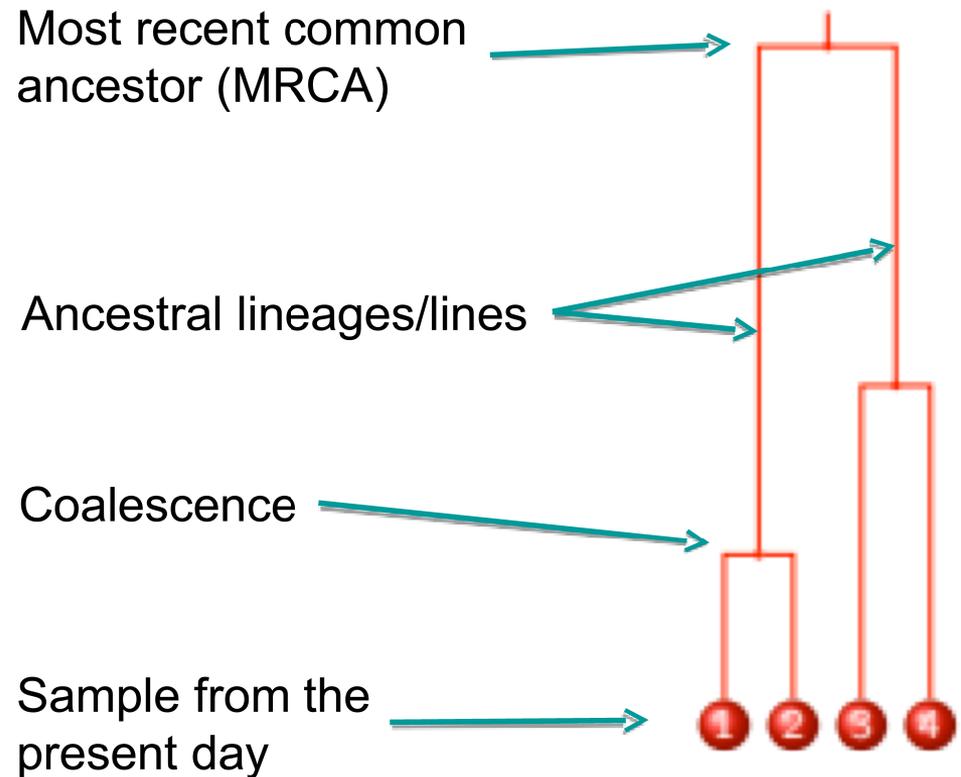
# The ancestral process: summary

- As the population get large, the timescale also gets big.
- Our limit was obtained by ‘zooming out’ from the Wright-Fisher model.



# The ancestral process: summary

- As the population get large, the timescale also gets big.
- Our limit was obtained by 'zooming out' from the Wright-Fisher model.

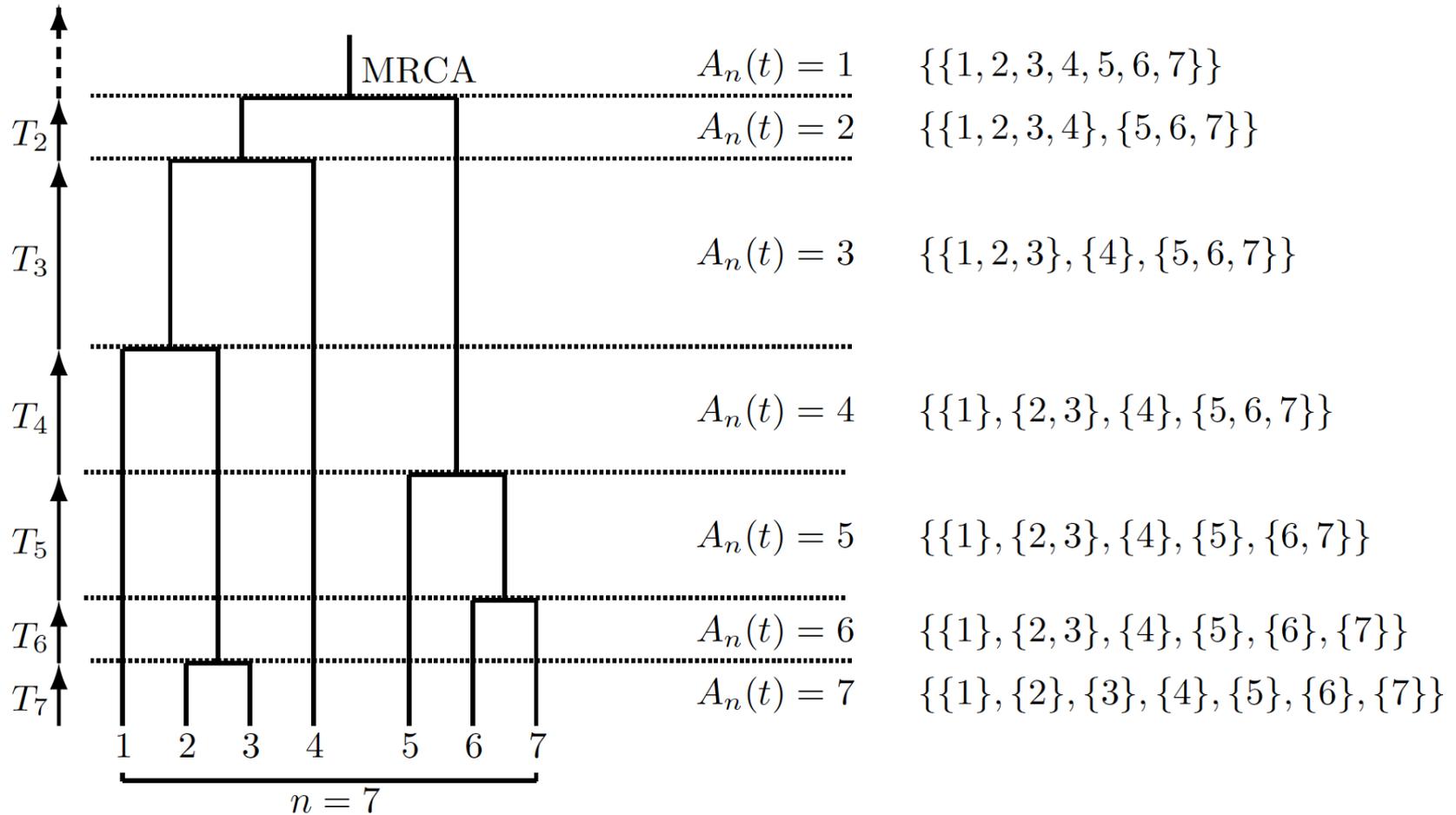


---

## 6. The coalescent

- We can't recover the coalescent tree from the ancestral process.
- In fact, there is a neat way to define a Markov process from which we *can* recover the tree.
- This is how the coalescent was originally defined by Kingman in 1982 [**Note 6.1**].
- The Wright-Fisher model is not the only discrete model that converge to the coalescent; many others do too (see exercise sheet).
- In that sense, the coalescent is ***robust***.

# The coalescent



---

# The coalescent

- The coalescent process hasn't given us much new beyond the ancestral process. All it really says additionally is:

*At the times of the jumps of the ancestral process, pick a pair uniformly at random to coalesce.*

- We can make this more precise [**Note 6.2**].

# Simulating coalescent trees

- The previous analysis suggests a simple algorithm for simulating a coalescent tree:
  1. Start with a partition  $\{\{1\}, \{2\}, \dots, \{n\}\}$  where every label is in its own block, and set  $k = n$ .
  2. Simulate time  $T_k$  from an exponential distribution with rate  $k$  choose 2. At this time, merge two blocks chosen uniformly at random. Decrement  $k$  by 1.
  3. If  $k = 1$  then stop, otherwise go to step 2.
- The output is a coalescent realisation, from which a tree can be reconstructed.