MONASH University

Science

A quick introduction to the Coalescent

Anders Gonçalves da Silva

Population and landscape genomics workshop – Day 2

25 March 2014

CBA – ANU

Outline

- Define the coalescent
- Introduce a reproductive model
- Use the reproductive model to turn-back time
- Conceptualize a simple coalescent model
- Try it out in *R*
- Introduce the idea of a *skyline plot*
- Modify our model to accommodate population growth
- Test it with *skyline plots*

Question?

- What is the likelihood that at least one pair of genes in this room shares a common ancestor one generation in the past?
- How many generations do we need to go back before the probability is larger than 50%?

What is the Coalescent?

What is the Coalescent?

- A statistical model:

$$P(\text{Genealogy}|\boldsymbol{\theta})$$

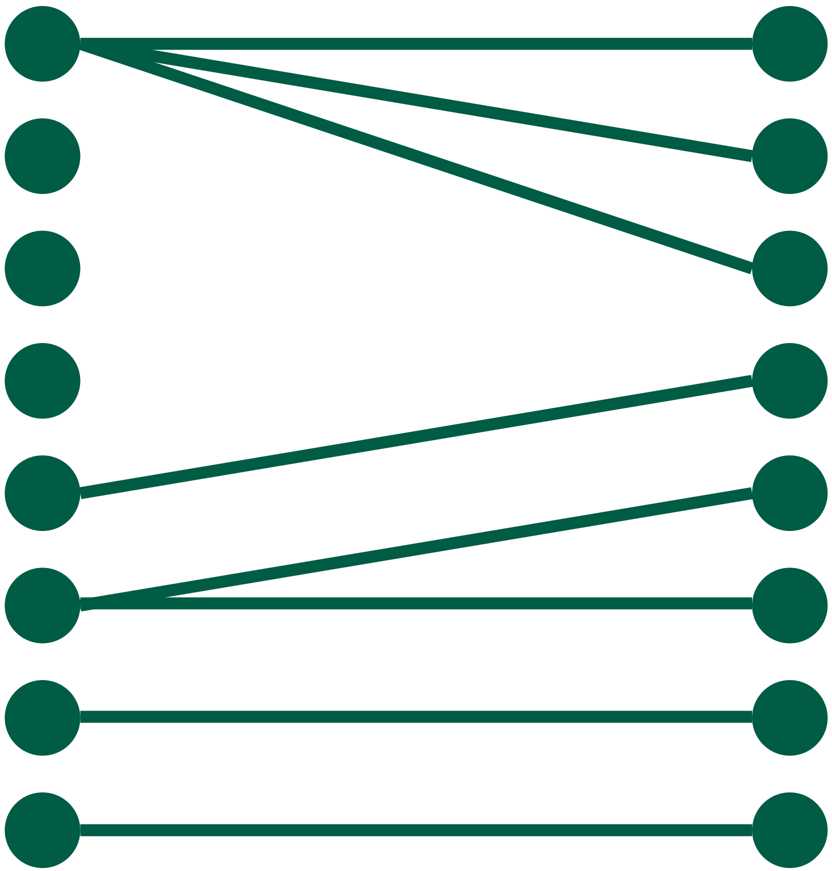
Reproductive model: Wright-Fisher

2N = 8 – number of haploid genomes

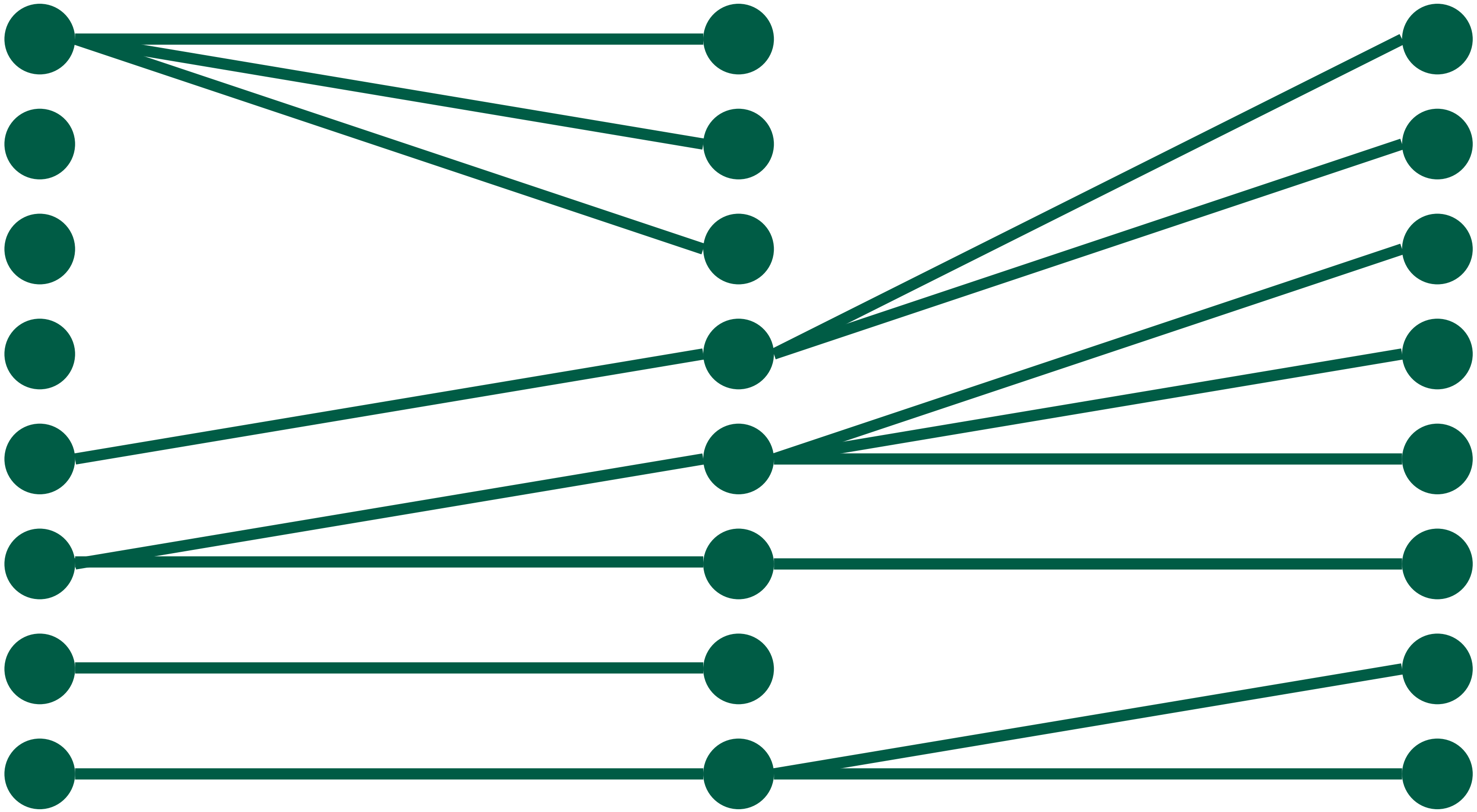
Each genes is equally represented in the gamete pool.

2N gametes are chosen at random to form the next generation.

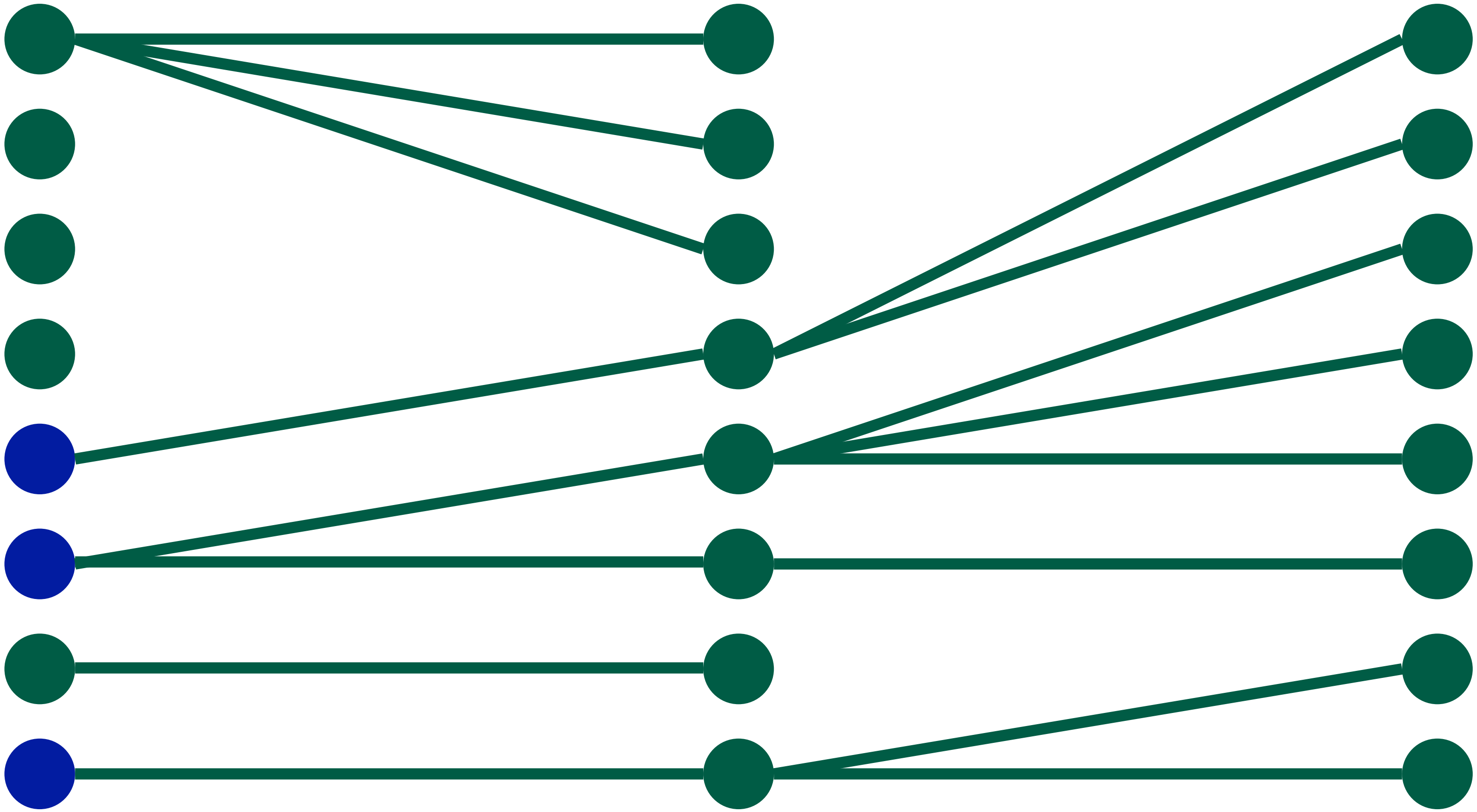
Reproductive model: Wright-Fisher



Reproductive model: Wright-Fisher



Reproductive model: Wright-Fisher



The coalescent

How long before two lineages merge?

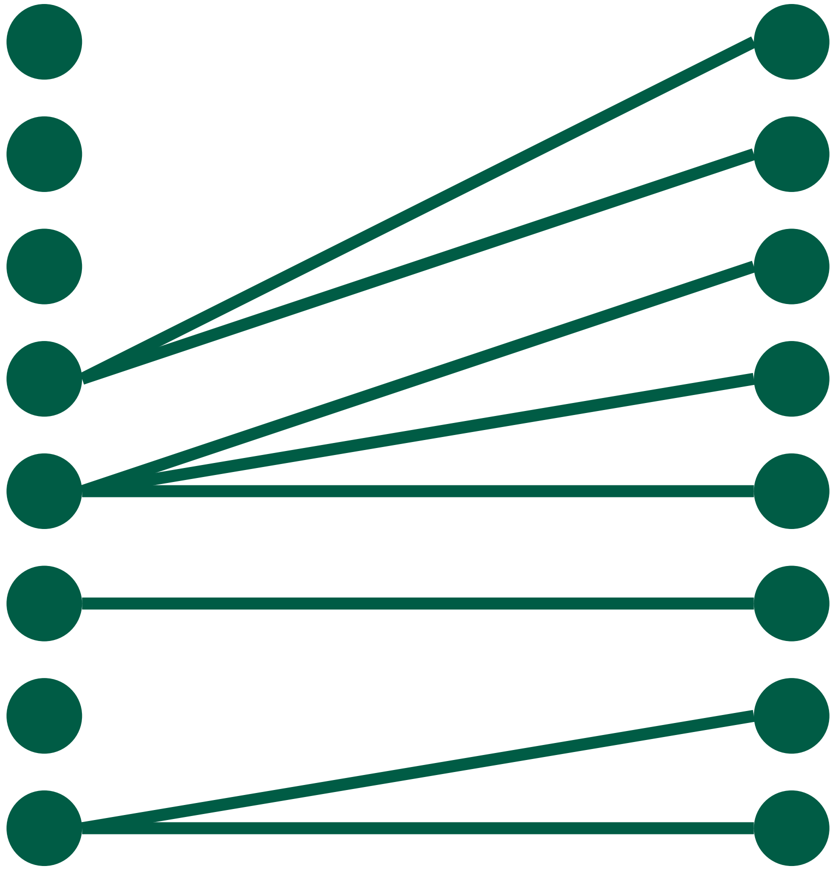
How long before all lineages merge?

How long is measured in generations!

Merging is a coalescent event!

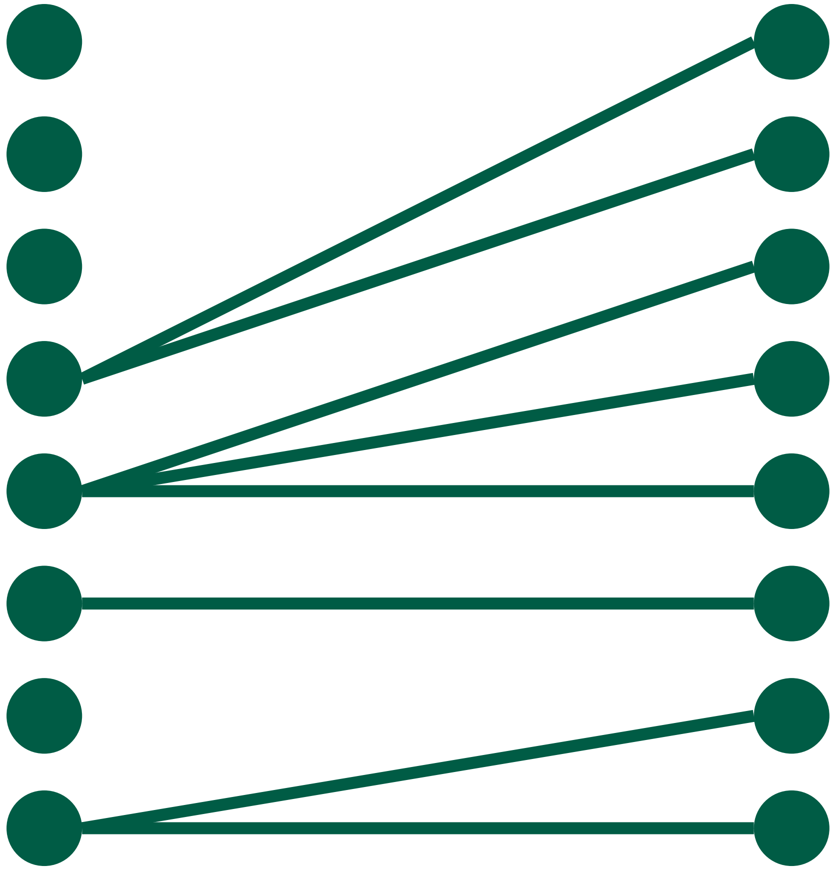
The *n*-coalescent

What is the probability that two lineages coalesce in the previous generation?



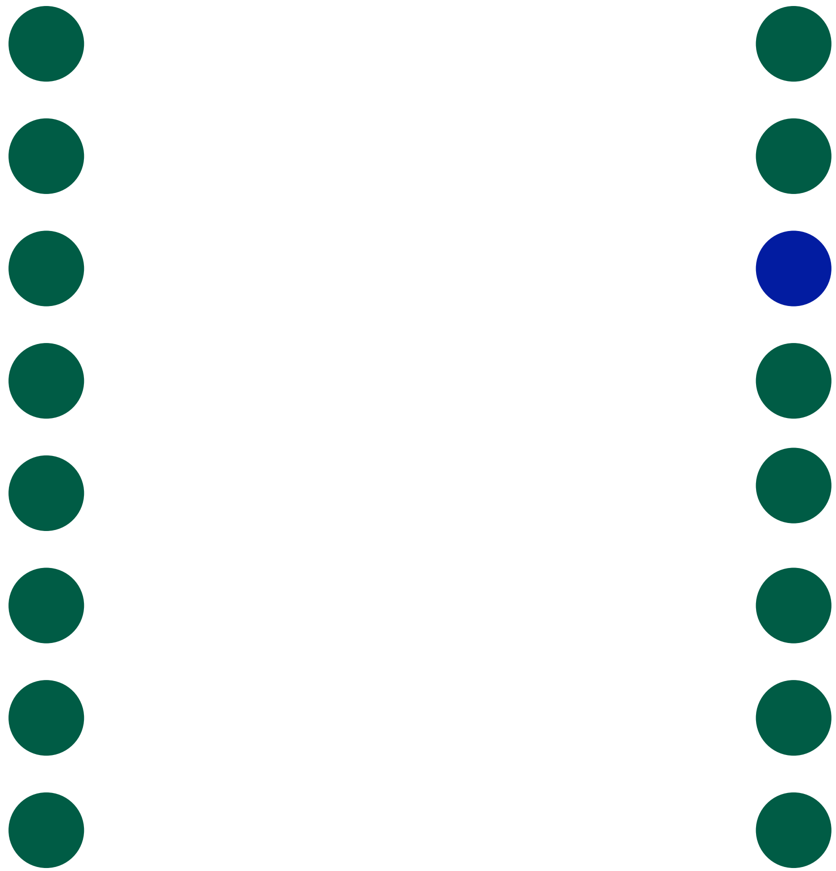
The *n*-coalescent

$P(T_2=1)$



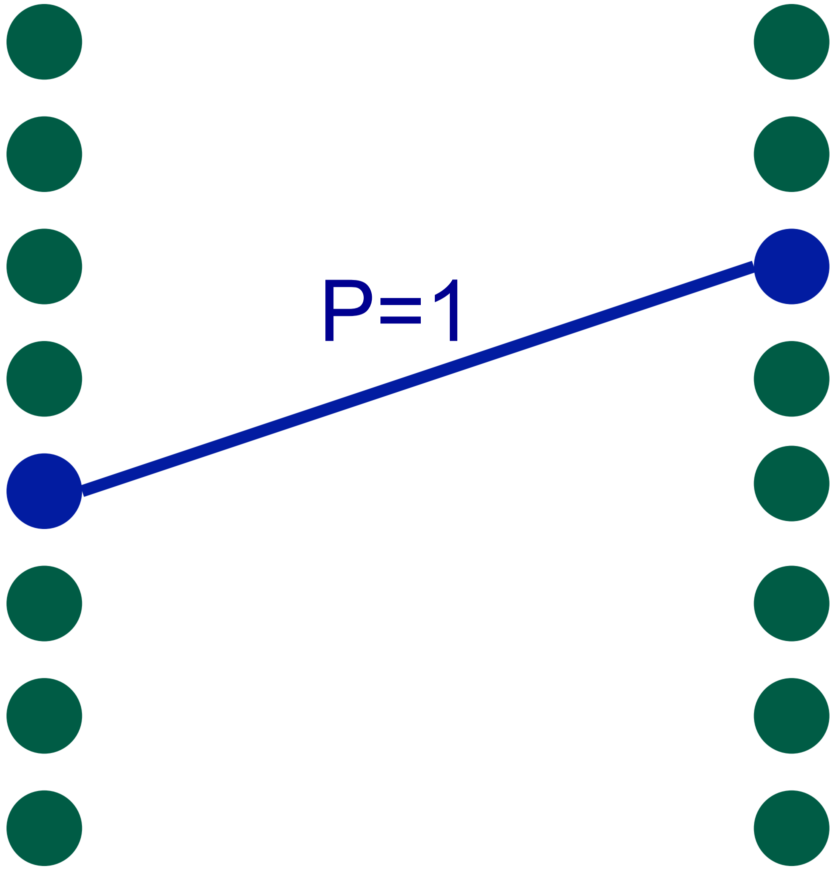
The *n*-coalescent: Wright's insight

What is the probability that a random individual has a parent in the previous generation?



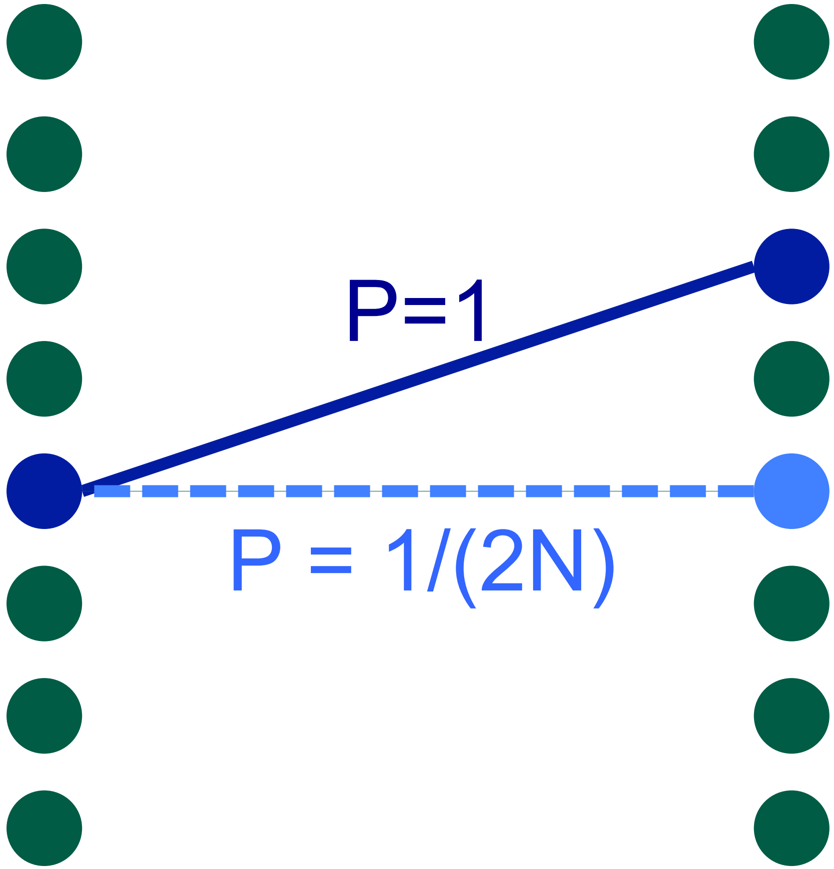
The *n*-coalescent: Wright's insight

What is the probability that a second individual shares the same parent?



The *n*-coalescent: Wright's insight

What is the probability that a second individual shares the same parent?



The *n*-coalescent: Wright's insight

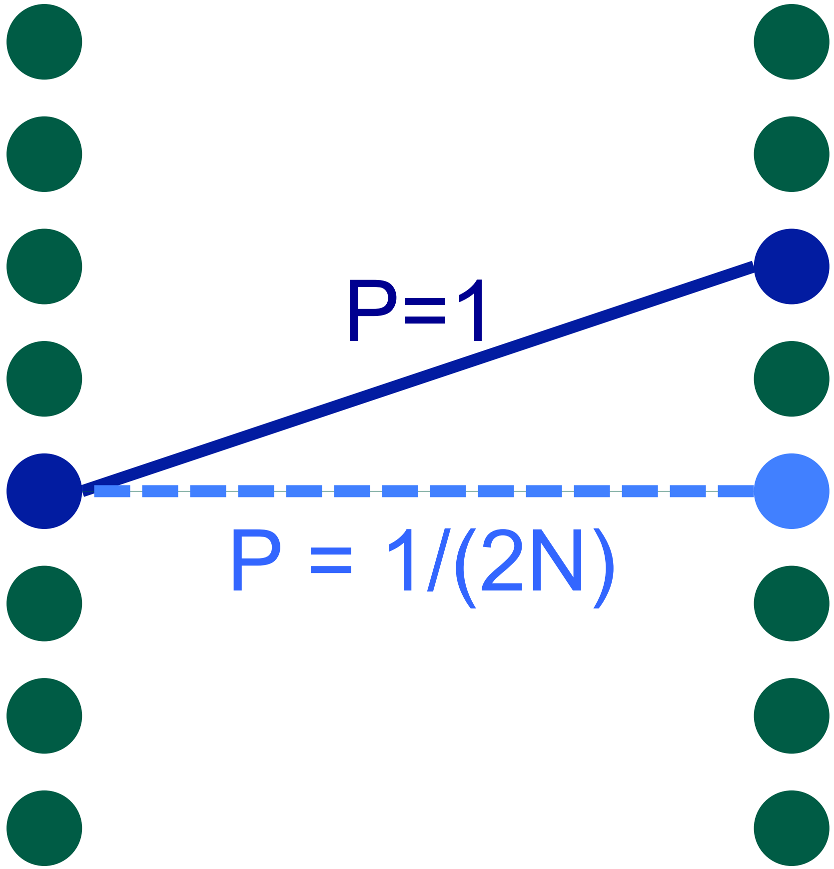
$P(T_2=1)$

Pr(Individual has parent)

X

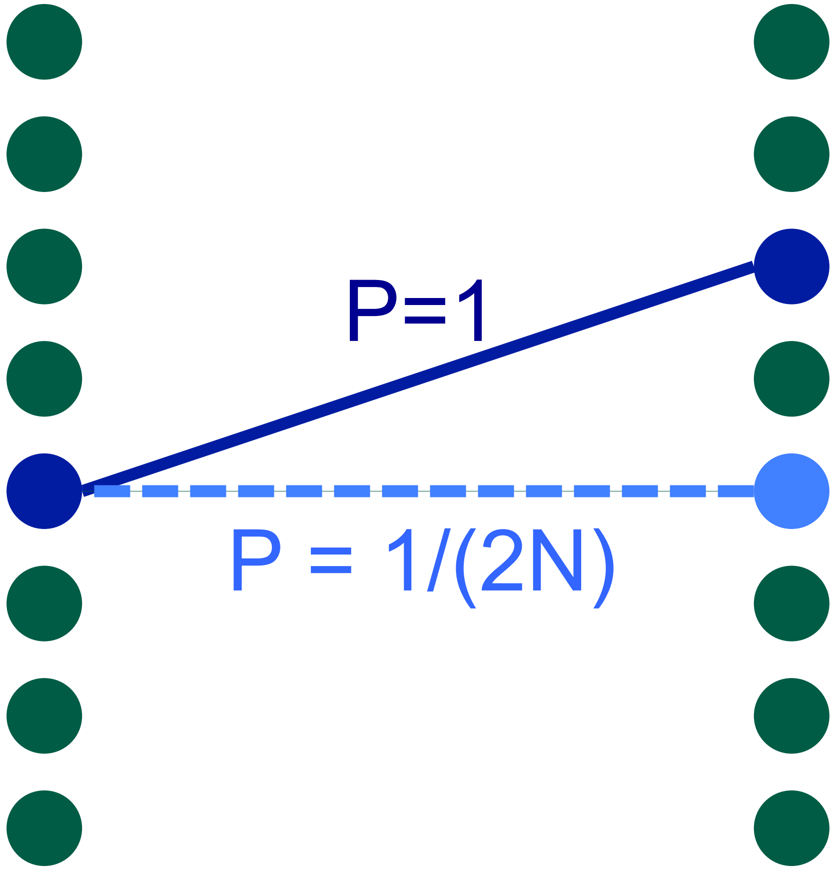
Pr(Individual

shares parent)



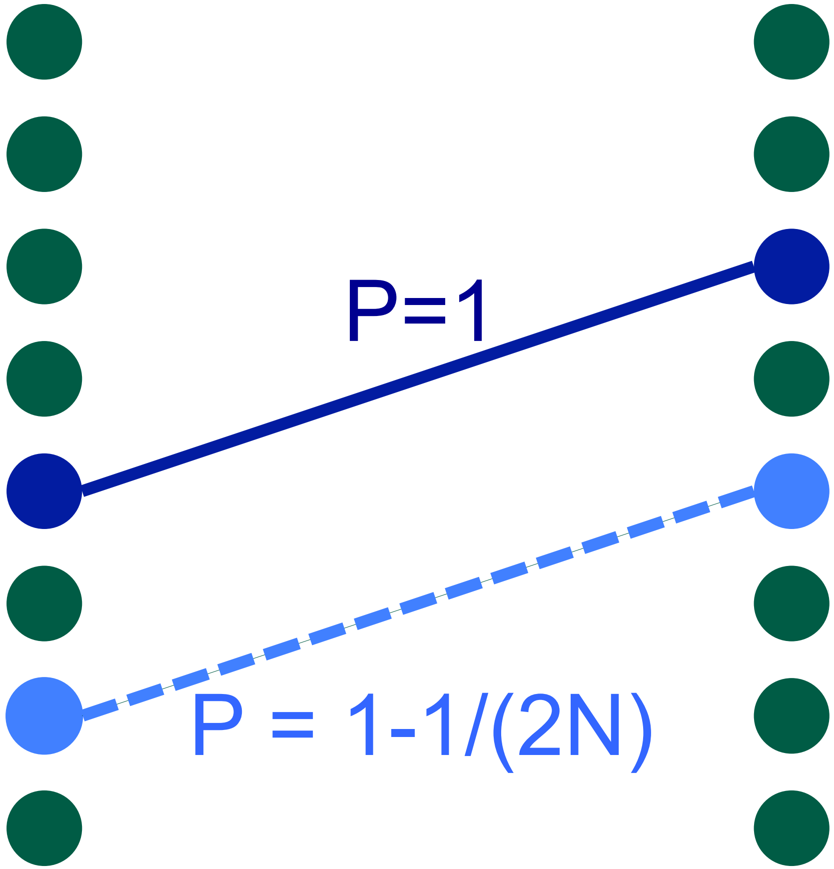
The *n*-coalescent: Wright's insight

$P(T_2=1) = 1/(2N)$



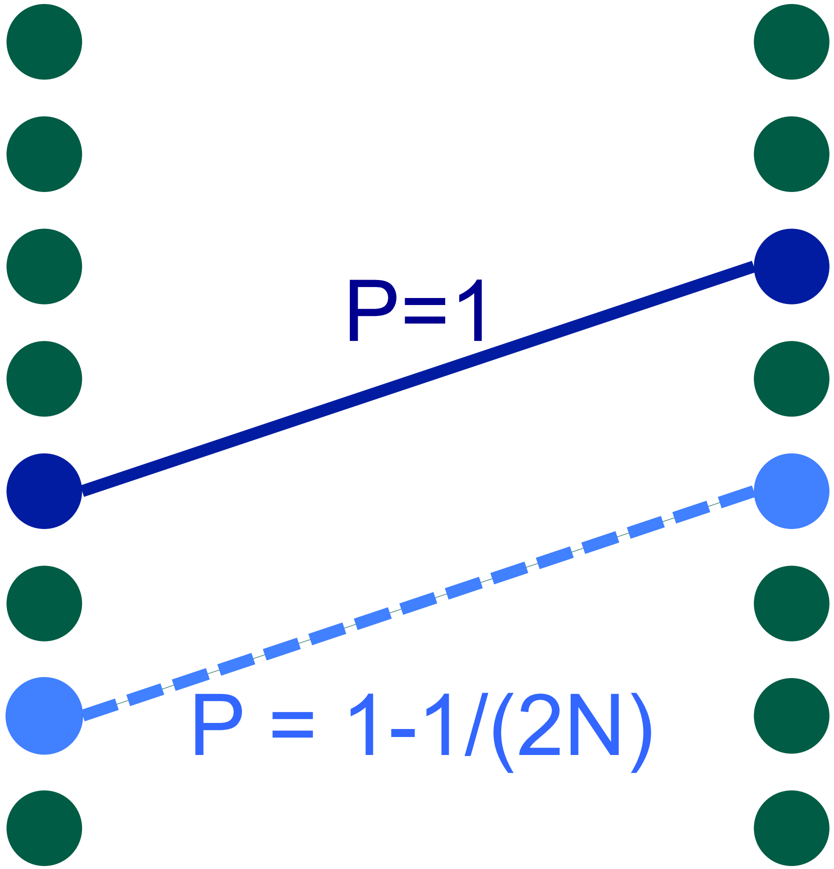
The *n*-coalescent: Wright's insight

$P(T_2 \neq 1)$

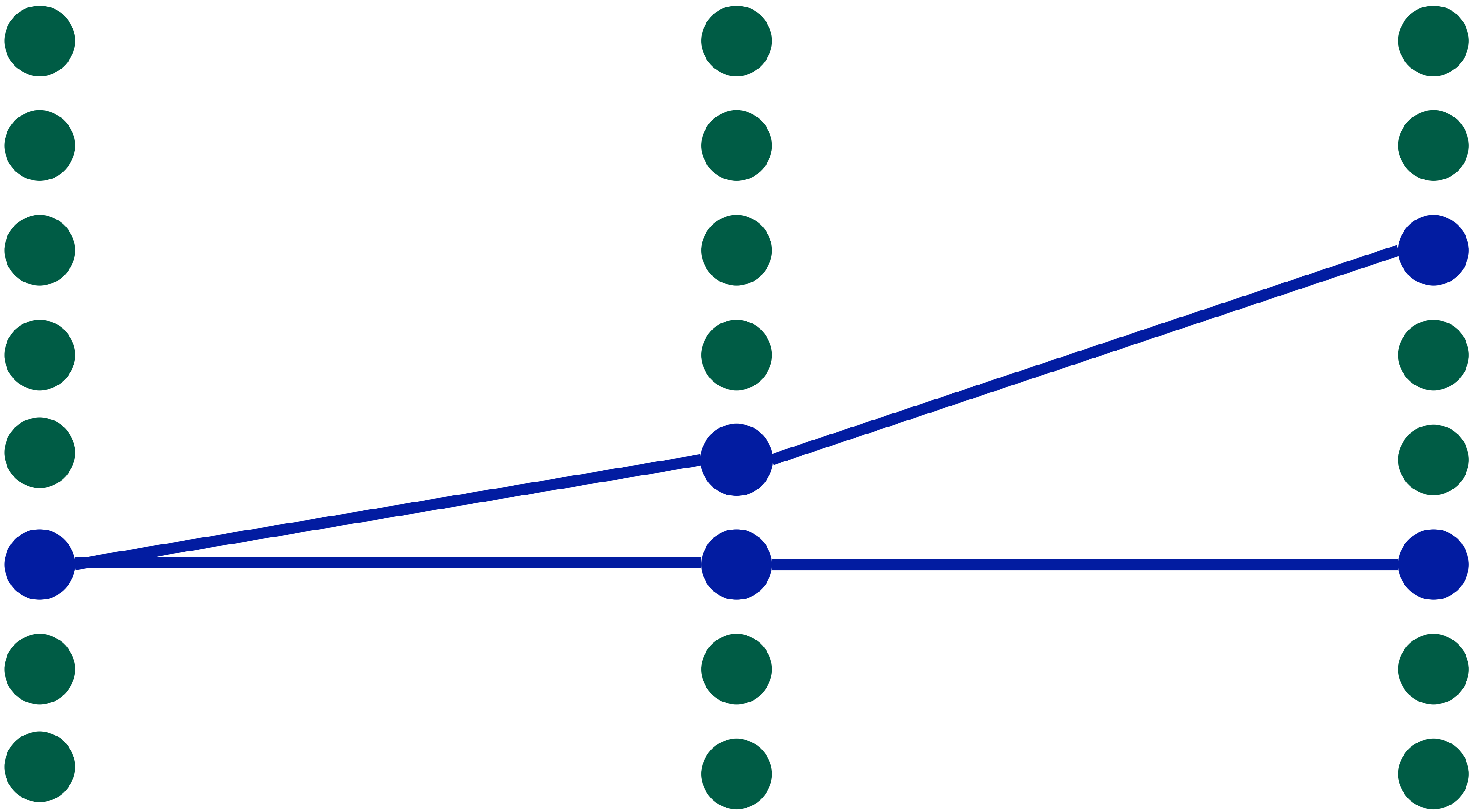


The *n*-coalescent: Wright's insight

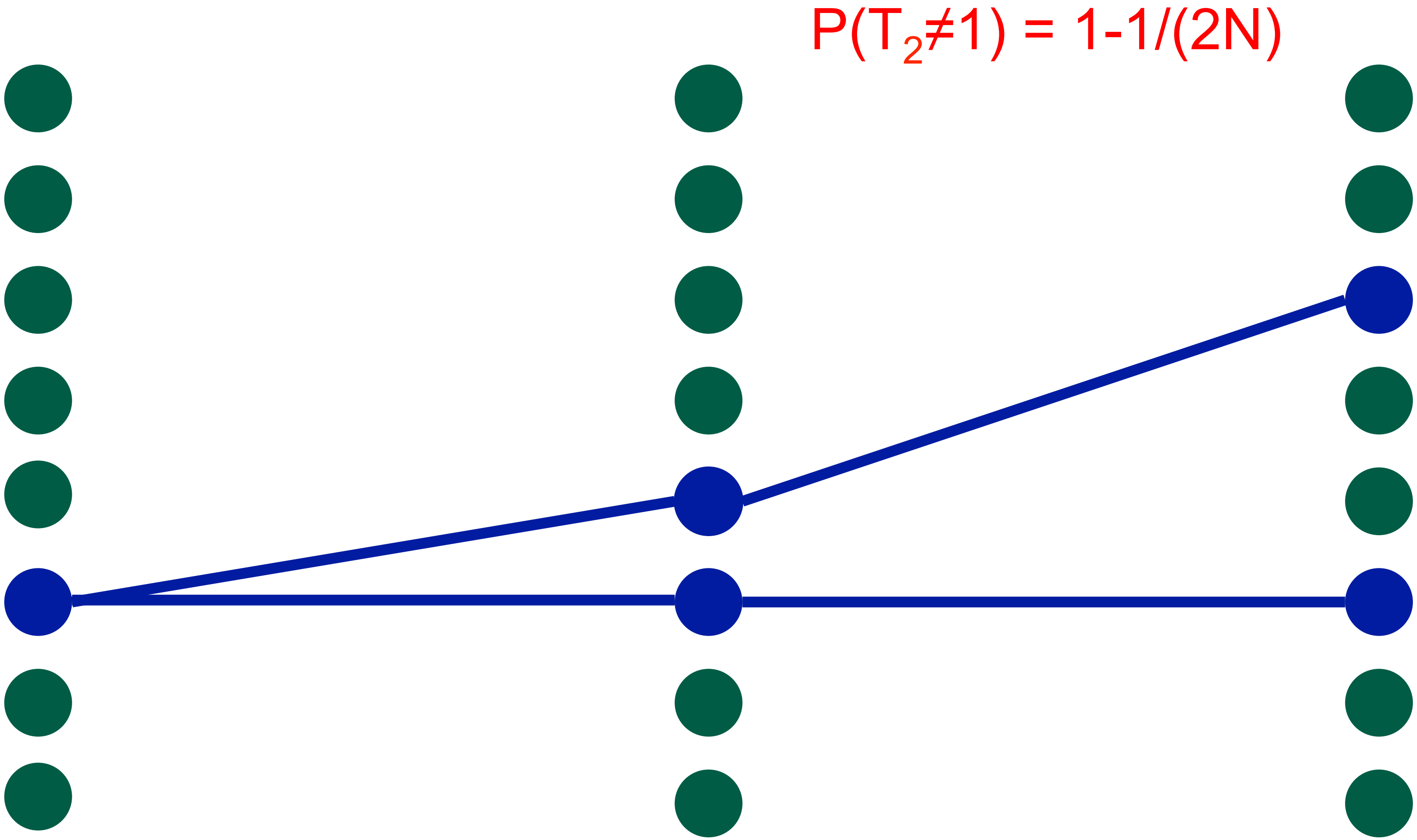
$P(T_2 \neq 1) = 1-1/(2N)$



$P(T_2=2)$?



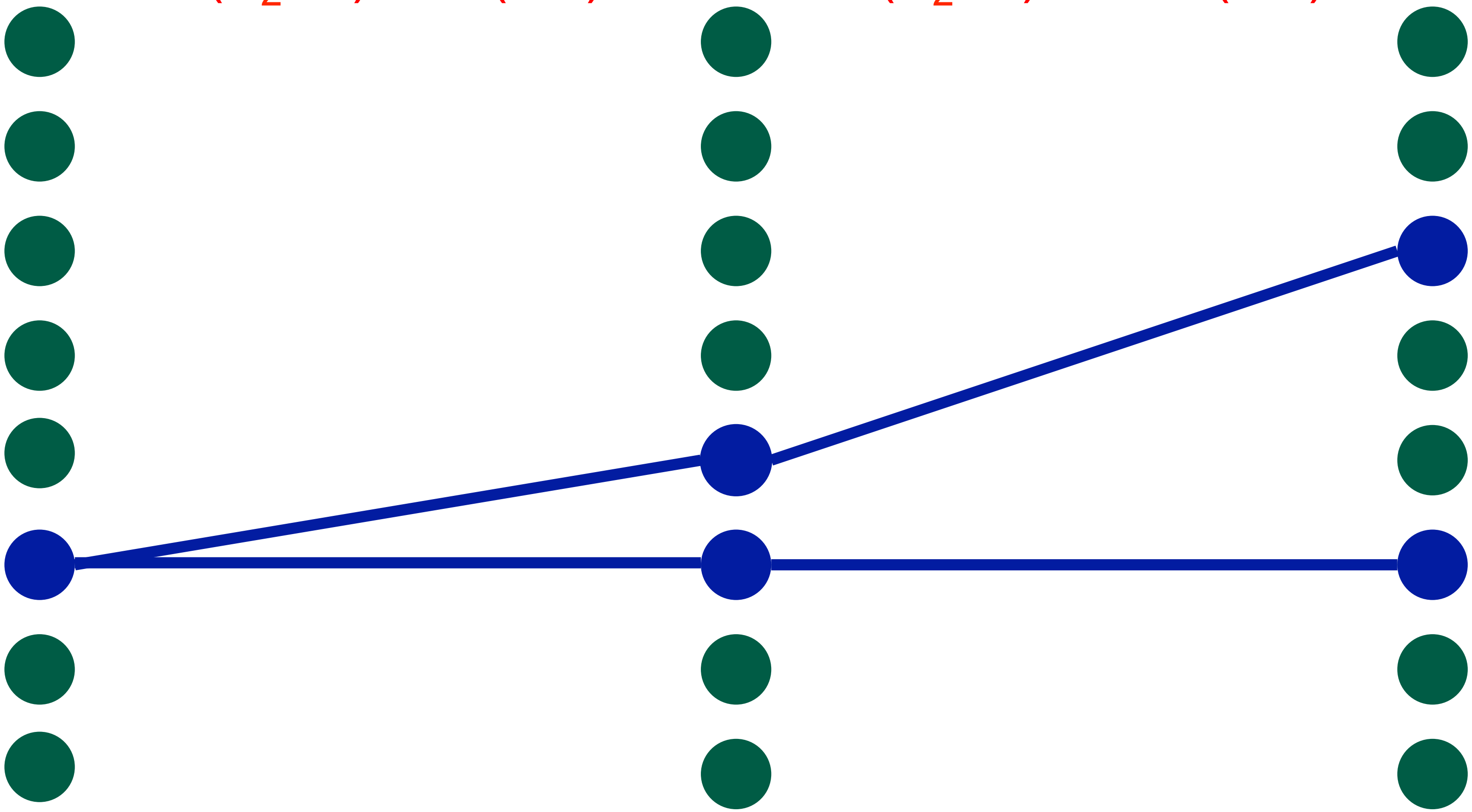
$P(T_2=2)$?



$P(T_2=2)$?

$P(T_2=1) = 1/(2N)$

$P(T_2 \neq 1) = 1-1/(2N)$



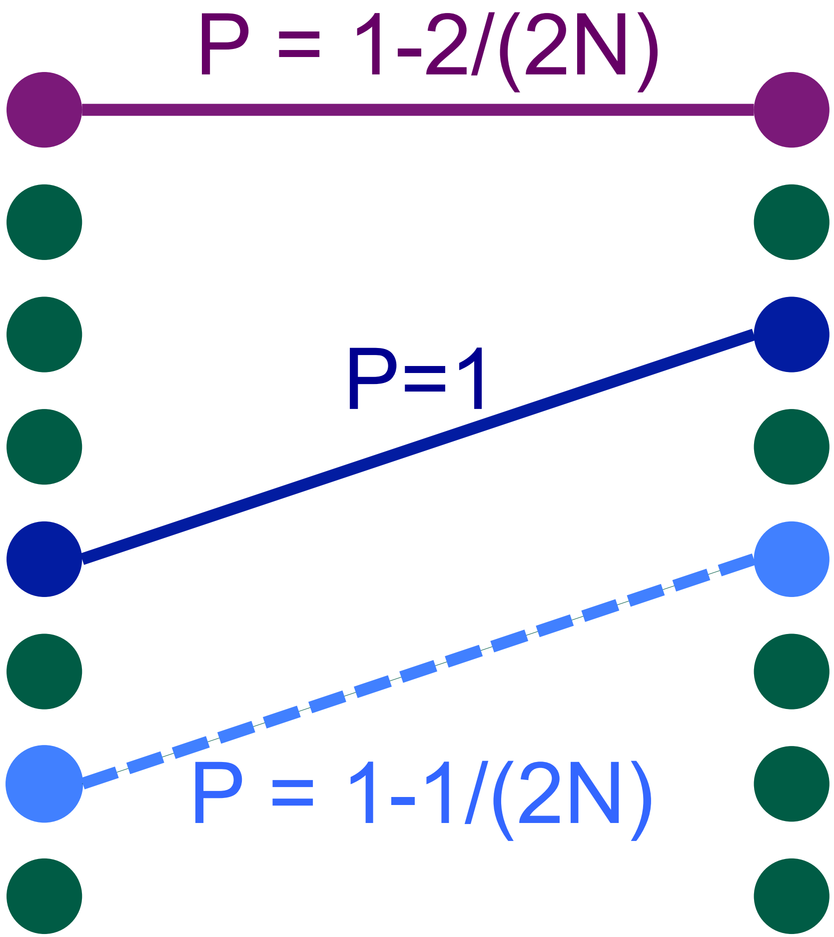
$P(T_2=j)$?

- $P(T_2=2) = P(T_2\neq 1) \times P(T_2=1)$
- $P(T_2=3) = P(T_2\neq 1) \times P(T_2\neq 1) \times P(T_2=1)$
- $P(T_2=4) = P(T_2\neq 1) \times P(T_2\neq 1) \times P(T_2\neq 1) \times P(T_2=1)$

- $P(T_2=j) = P(T_2\neq 1)^{(j-1)} \times P(T_2=1)$
- $P(T_2=j) = (1-(1/(2N))^{(j-1)} \times (1/(2N))$

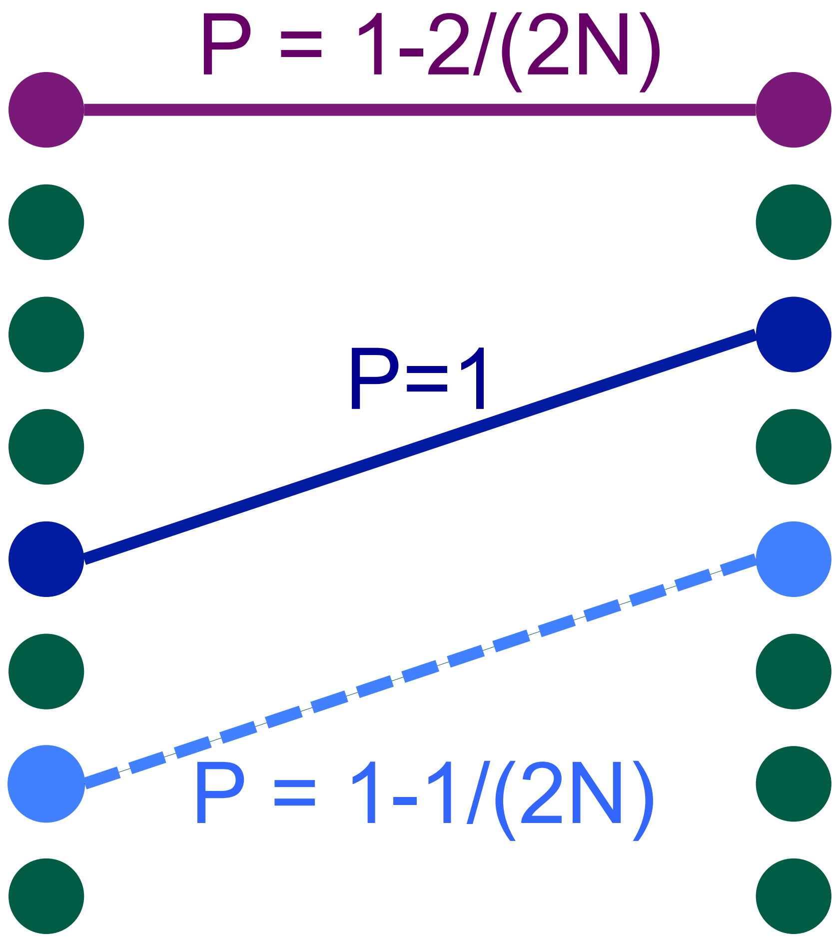
The *n*-coalescent: *k* > 2

$P(T_3 \neq 1)$



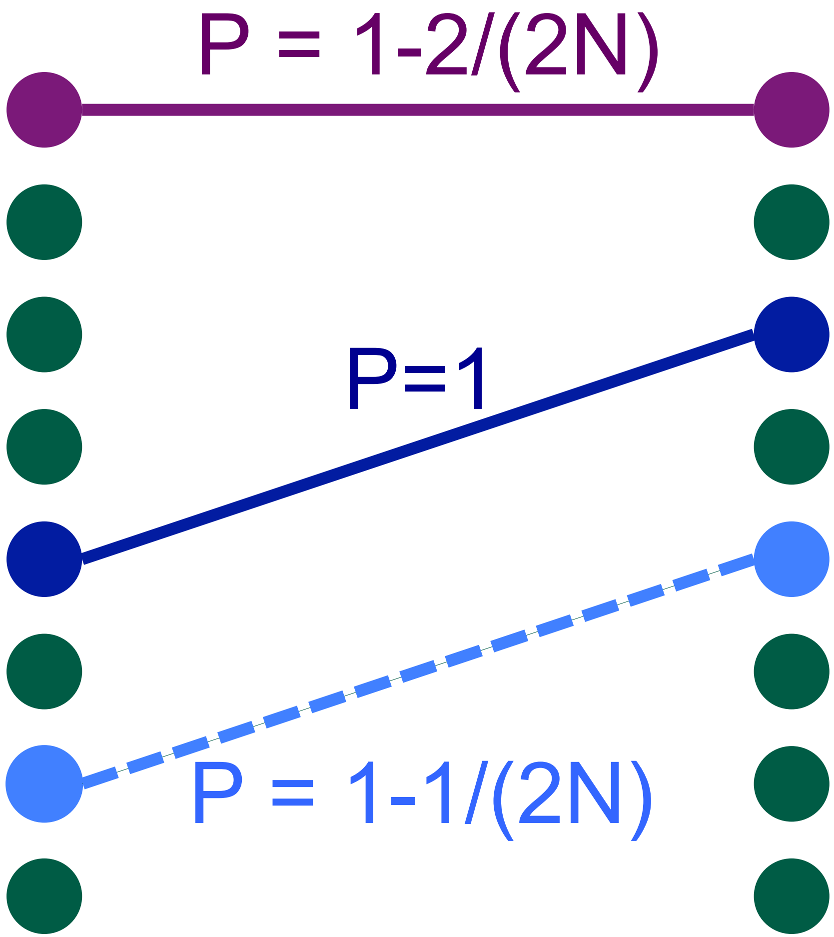
The *n*-coalescent: *k* > 2

$$P(T_3 \neq 1) = 1-(3/(2N))$$



The *n*-coalescent: *k* > 2

$$P(T_k \neq 1) = 1 - \binom{k}{2}\frac{1}{2N}$$



$P(T_k = j)$

$$P(T_k = j) = \left\{1 - \binom{k}{2}\frac{1}{2N}\right\}^{j-1}\binom{k}{2}\frac{1}{2N}$$

Thus…

$$T_k \sim Geom(p)$$

$$P(T_k = j) = (1-p)^{j-1}p$$

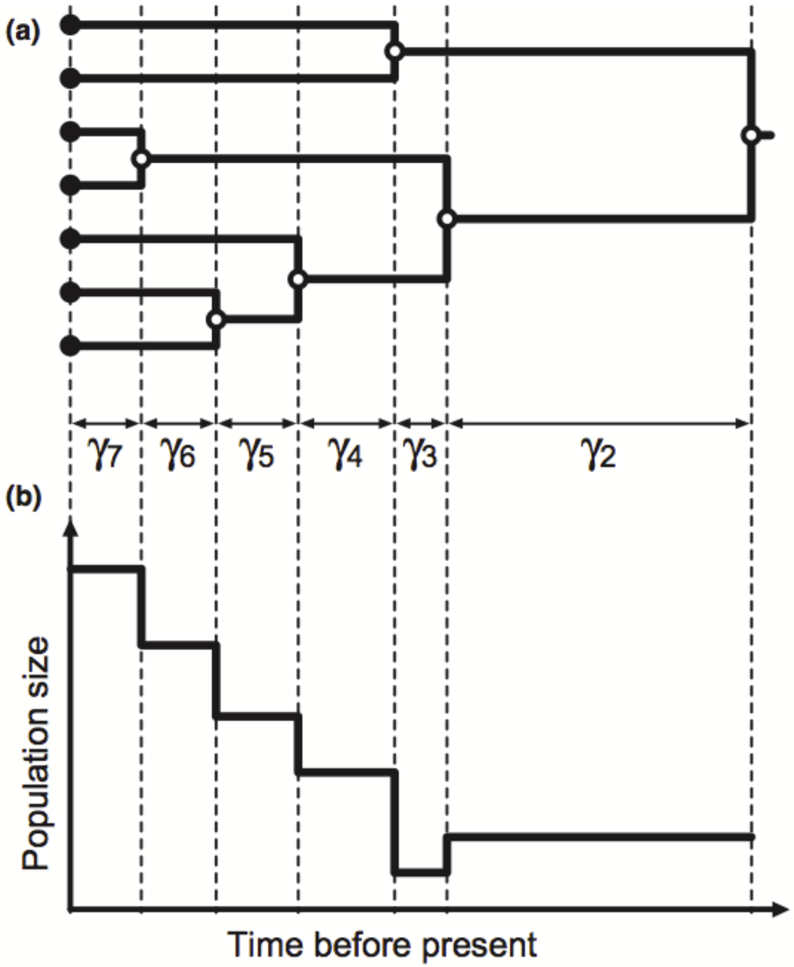
$$p = \binom{k}{2}\frac{1}{2N}$$

LET’S TRY IT IN *R*…

The point of the coalescent…

- Given a topology (genealogy), we can estimate:
  - Effective population size
  - Migration rates
  - Population size changes
  - Mutation rates
  - Age of mutations
  - Recombination rates
  - Time of divergence

Estimating population growth…



$$N_i = \gamma_i \frac{i(i-1)}{2}$$

Ho & Shapiro (2011) *MER* 11: 423

LET'S ESTIMATE POPULATION GROWTH IN *R*...

Further reading

- Hein, Schierup and Wiuf (2010). Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press.

- Wakeley (2009). Coalescent theory: an introduction. Roberts & Company.