

# Population Genomics

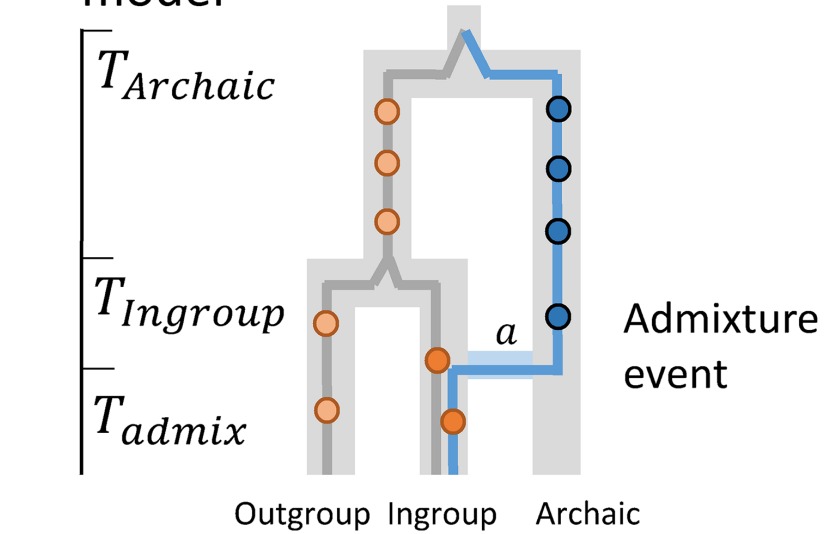
Week 6 Archaic Admixture

Bjarke Meyer Pedersen

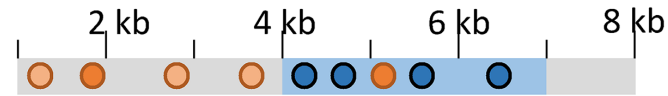
# Overview

- Introduction to the model we will use today
- Group discussion of the model, explain it to each other
- Slurm exercise
- BREAK
- Slurm exercise
- Class wrap up

## a Overview of the model



### Observed variants



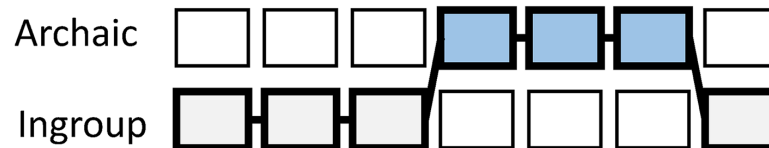
### Remove variants found in outgroup



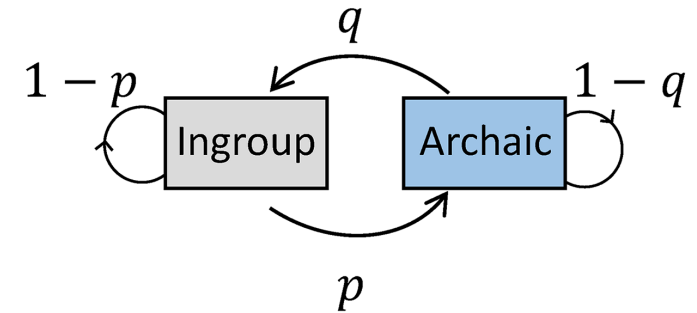
### Count variants in window



### Decode sequence



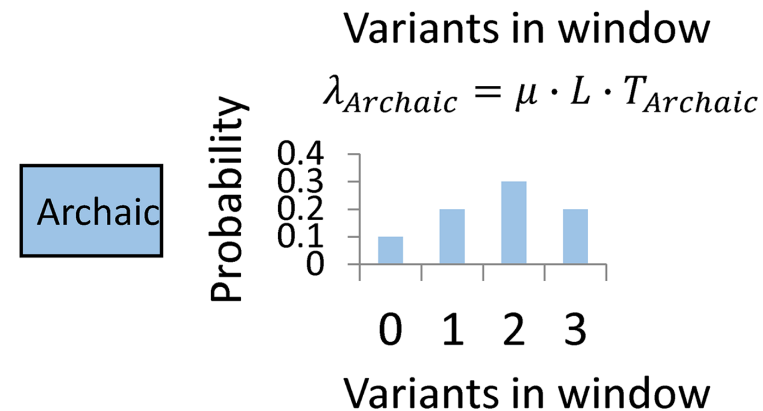
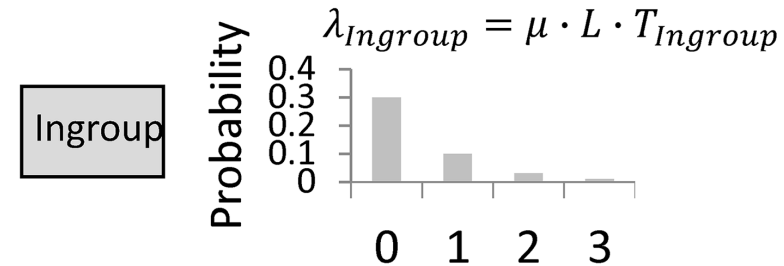
## b Transition probabilities



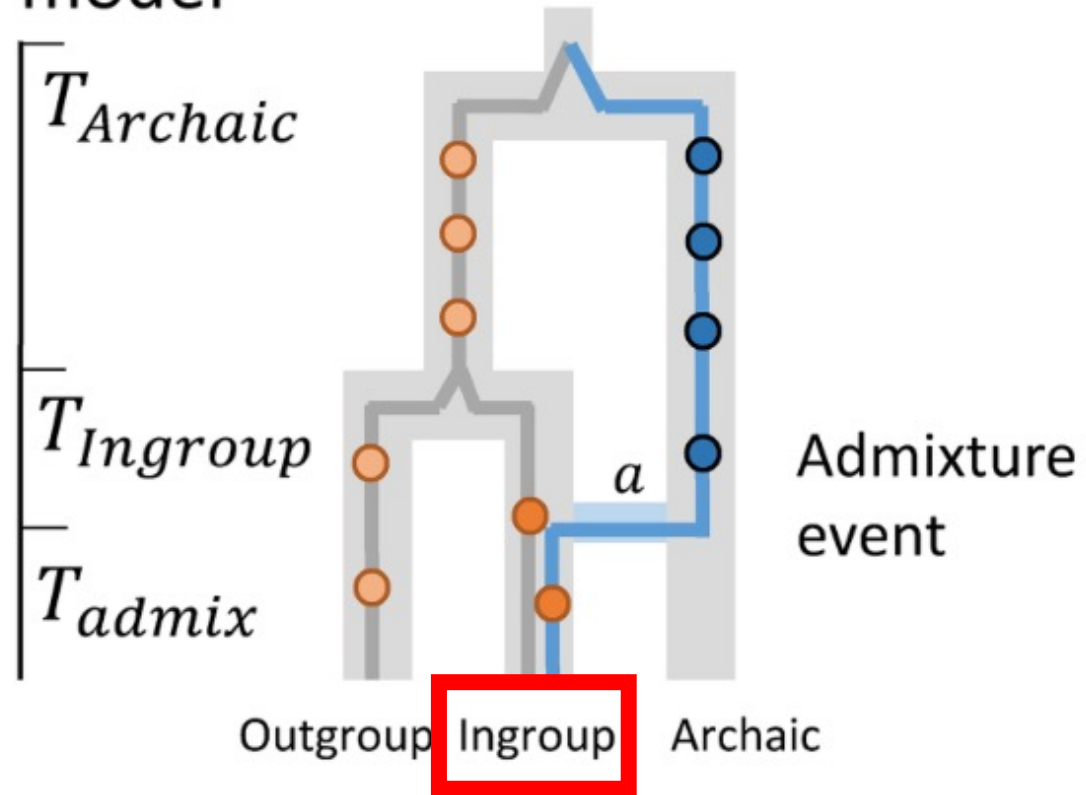
$$q \approx T_{admix} \cdot r \cdot L \cdot (1 - a)$$

$$p \approx T_{admix} \cdot 2 \cdot r \cdot L \cdot a$$

## c Emission probabilities



## Overview of the model

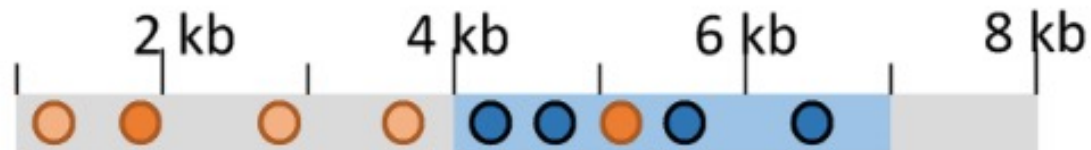


We have samples of the ingroup and outgroup

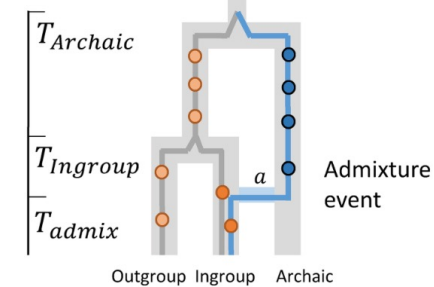
We have samples of the ingroup and outgroup

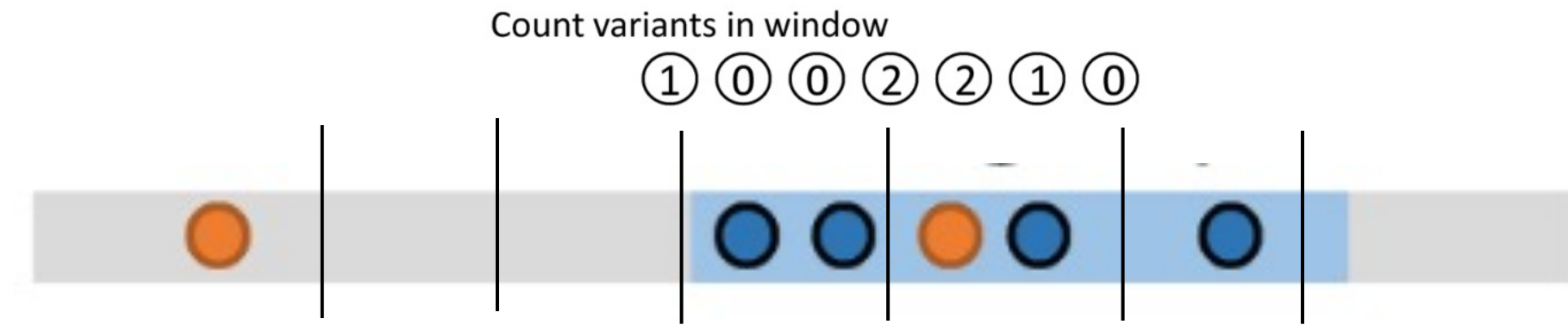
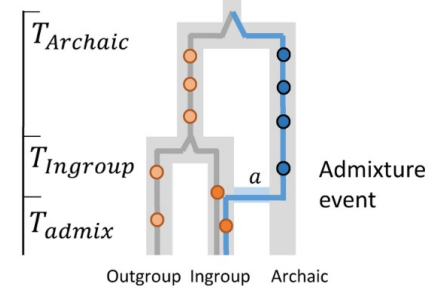


Observed variants

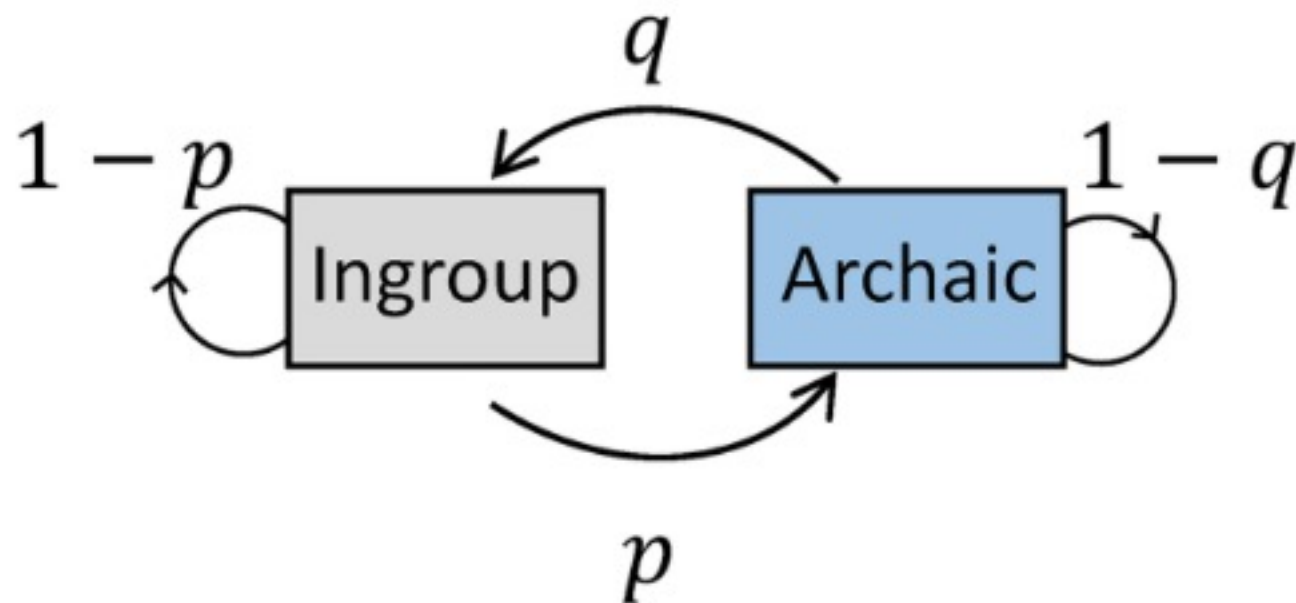


Remove variants found in outgroup



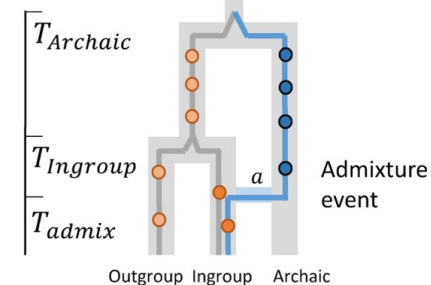


## b Transition probabilities



$$q \approx T_{admix} \cdot r \cdot L \cdot (1 - a)$$

$$p \approx T_{admix} \cdot 2 \cdot r \cdot L \cdot a$$



$L$  = length of a window

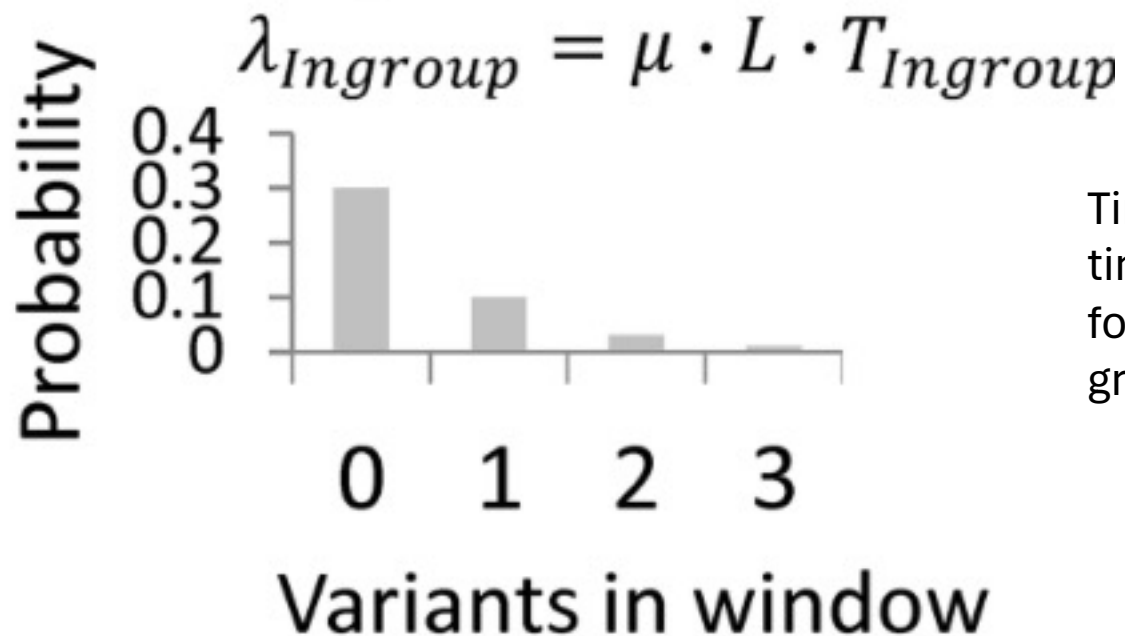
$R$  = recombination rate

$T$  = time to admixture

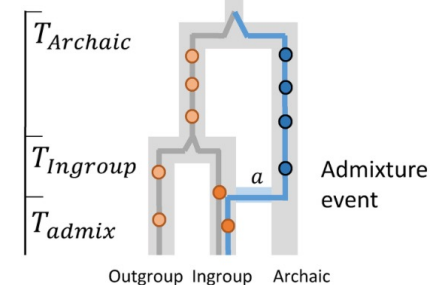
$a$  = admixture proportion

The Emission (is density of variants),  
Which depend on the Time to admixture

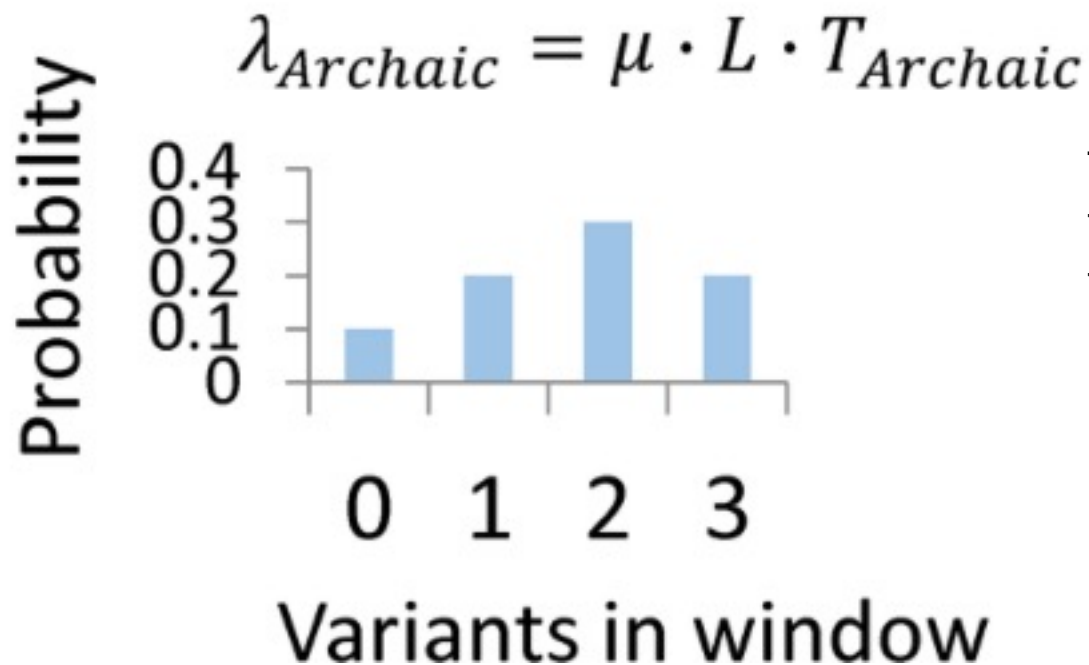
Ingroup



$T_{Ingroup}$  = Mean time to coalescence for the in and out group



Archaic



$T_{Archaic}$  = Mean time to coalescence for the out and archaic group



We find the parameters by optimizing the likelihood given observations for an individual genome

$$q \approx T_{admix} \cdot r \cdot L \cdot (1 - a)$$

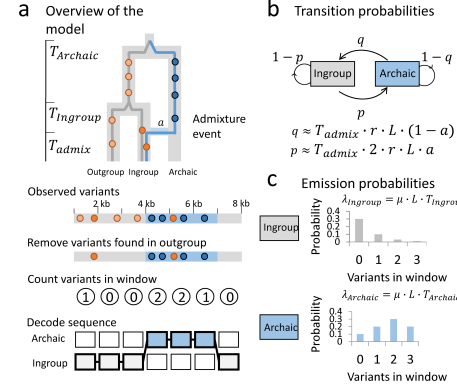
$$p \approx T_{admix} \cdot 2 \cdot r \cdot L \cdot a$$

$$\lambda_{Archaic} = \mu \cdot L \cdot T_{Archaic}$$

$$\lambda_{Ingroup} = \mu \cdot L \cdot T_{Ingroup}$$

What can these tell us?

*if we assume a known mutation rate  $\mu$  and a known recombination rate between windows  $rL$ .*



We find the parameters by optimizing the likelihood given observations for an individual genome

$$q \approx T_{admix} \cdot r \cdot L \cdot (1 - a)$$

$$p \approx T_{admix} \cdot 2 \cdot r \cdot L \cdot a$$

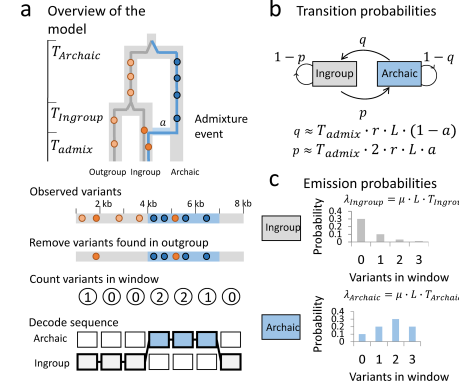
Admixture time and admixture proportions

$$\lambda_{Archaic} = \mu \cdot L \cdot T_{Archaic}$$

Mean time to coalescence between outgroup and archaic

$$\lambda_{Ingroup} = \mu \cdot L \cdot T_{Ingroup}$$

Mean time to coalescence between ingroup and outgroup



Explain the model to each other

# The exercise today

The model has been run on a lot of genomes

And identified segments are in a txt file

Later the segments have been mapped to different reference genomes of an Altai and an Vindija Neanderthal

And against a Denisova



Q3. Combine the Neanderthal fragments for all individuals and plot all the fragments on top of each other along chromosomes (hint use `alpha = 0.1`). Can you find “deserts” of archaic admixture and/or evidence for places where Neanderthal ancestry has reached very high frequency?

```
archaic_df %>%
  arrange(region) %>%
  ggplot() +
    theme_bw() +
    geom_segment(aes(x=start,xend=end,y=0, yend=1, col=closest_to), alpha=0.02) + facet_wrap(~chrom) +
    theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.text.y=element_blank(),axis.ticks=element_blank(),
          xlab("Position"))
```

