

# Population Genomics wk 7

Relate and Demography

Bjarke Meyer Pedersen

# Overview of today

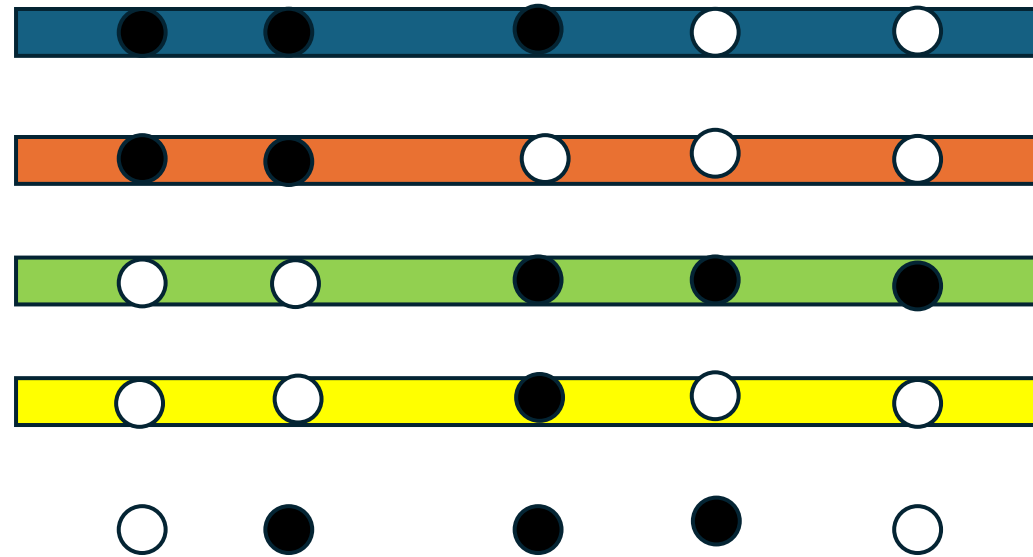
- Run the first couple of commands as one is quite slow
- Introduction to Relate and how it works
- Exercises
- BREAK
- Exercises
- Wrap up in class

# Exercises

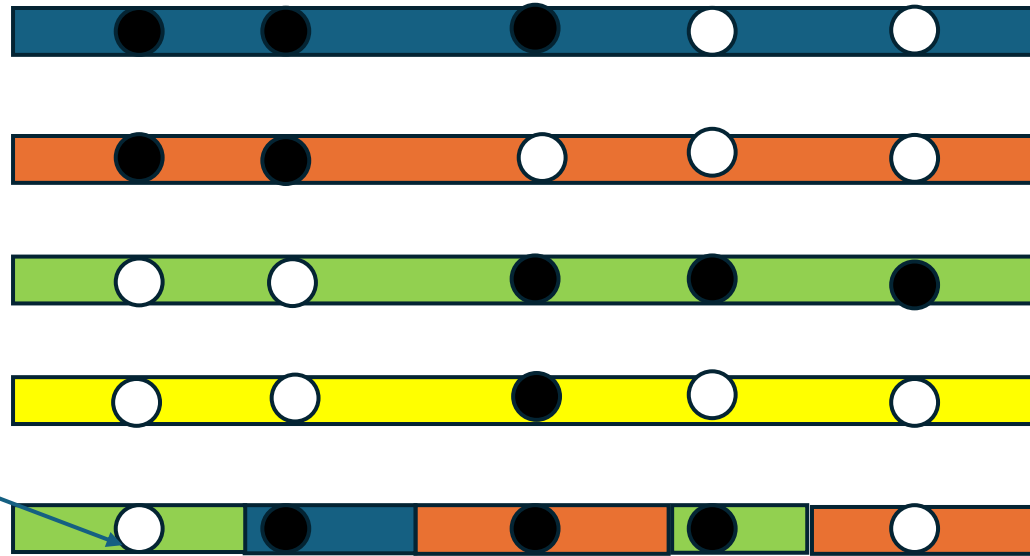
- Conda env create -f env/exercise\_envs/bjarke-relate.yml
- Run upto and including this exercise

```
~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr2_relate -m 1.25e-8 --poplabels 60_inds.txt -o popsize --threshold 0
```

# Relate and ARGs

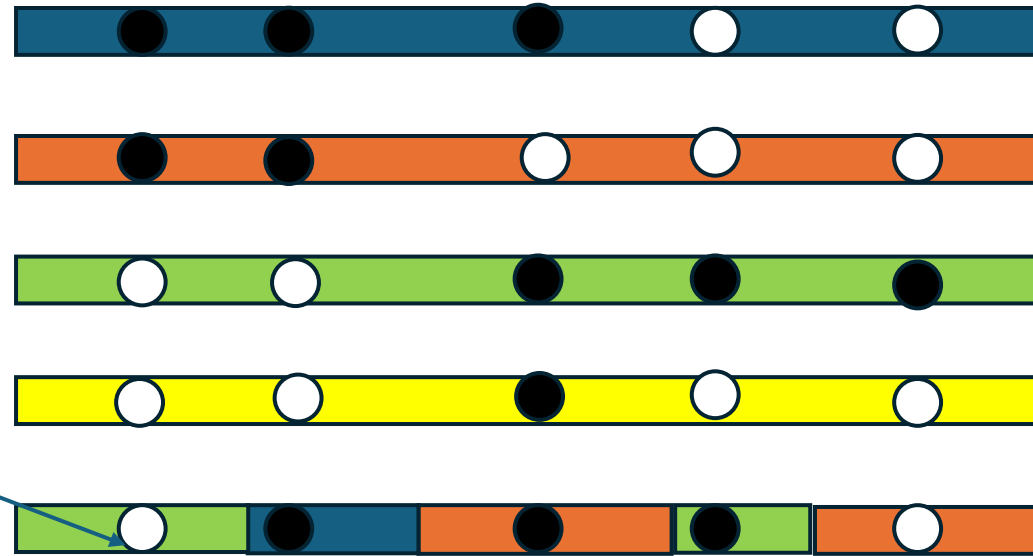


# Relate and ARGs

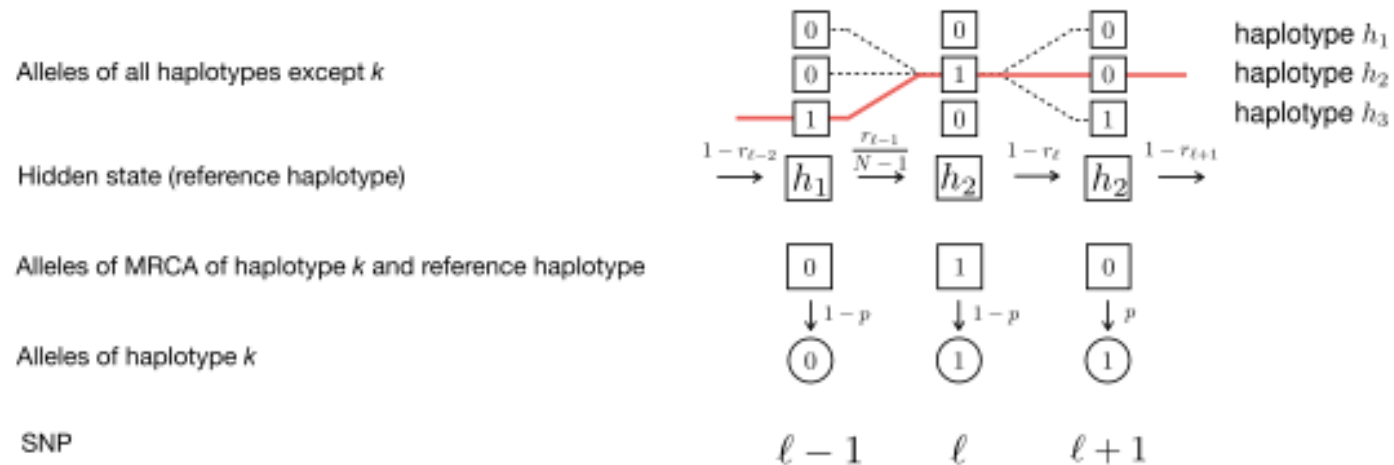


HMM finds the probability of seeing this mutation given the sequences above

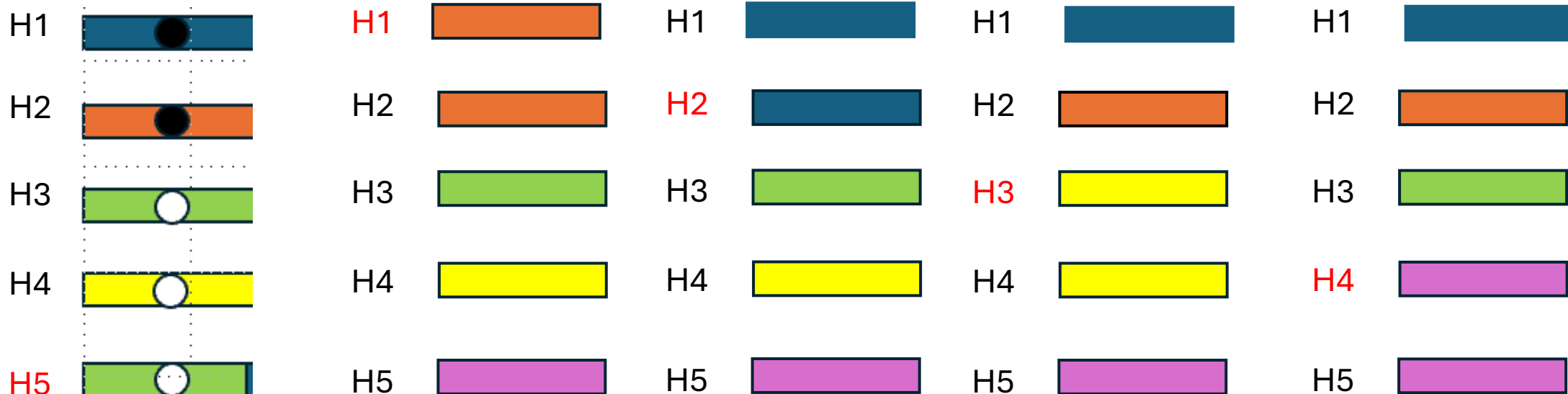
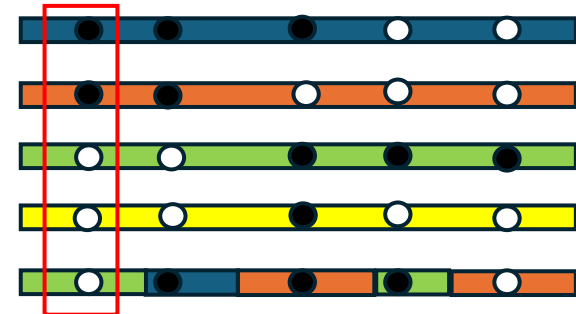
# Relate and ARGs



HMM finds the probability of seeing this mutation given the sequences above



# Relate and ARGs



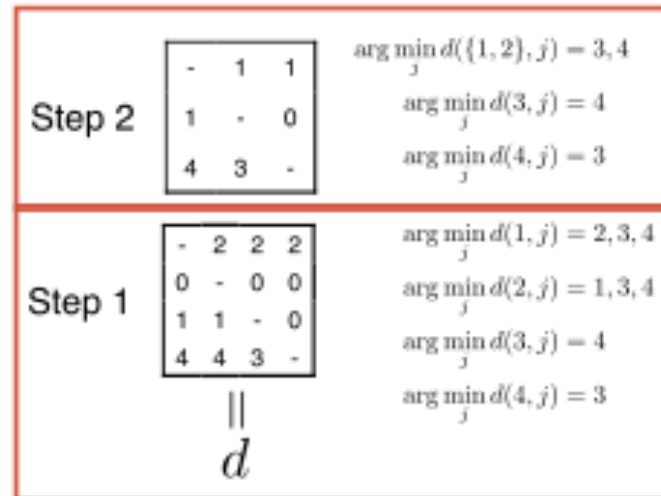
Makes a distance matrix for each position

	H1	H2	H3	H4	H5
H1		1	3	4	1
H2	1		4	3	2
H3	2	3		2	3
H4	3	2	1		4
H5	4	4	2	1	

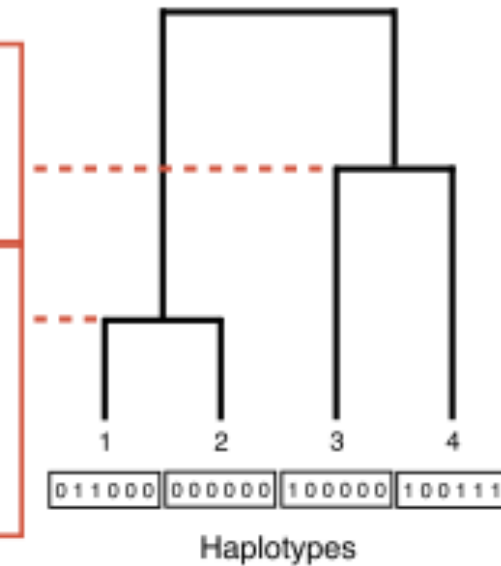
# Relate and ARGs

	H1	H2	H3	H4	H5
H1		1	3	4	1
H2	1		4	3	2
H3	2	3		2	3
H4	3	2	1		4
H5	4	4	2	1	

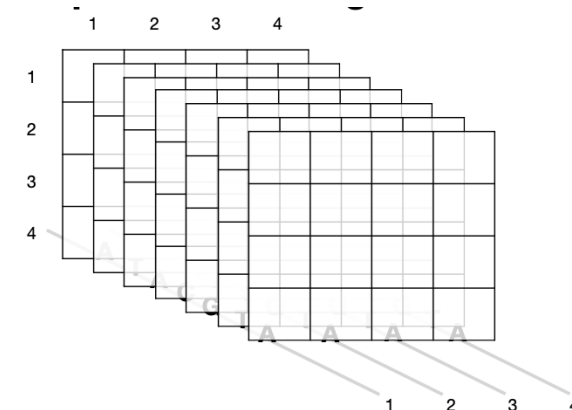
**a** Tree builder



**b**



In each step of the algorithm, we merge the pair of clades that coalesce with each other before coalescing with any other clade, as determined using rows of the distance matrix. If multiple pairs of clades satisfy this condition, we choose the pair with minimum symmetrized score in the distance matrix





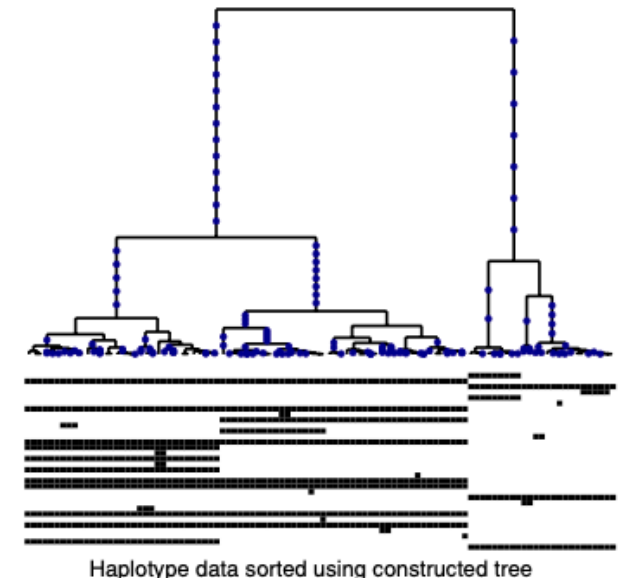
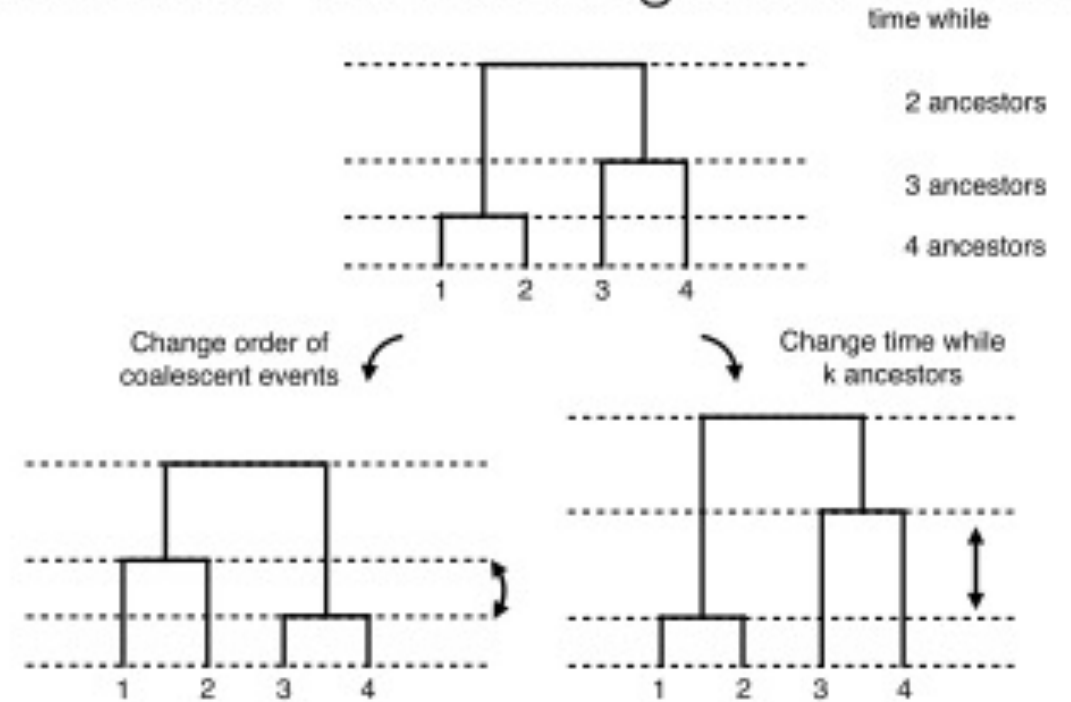
# Relate and ARGs

Now there is a tree for all the positions, the tree length is then determined by

Where possible we map mutations to the (unique) branch that has the identical descendants as the carriers of the derived allele in the data.

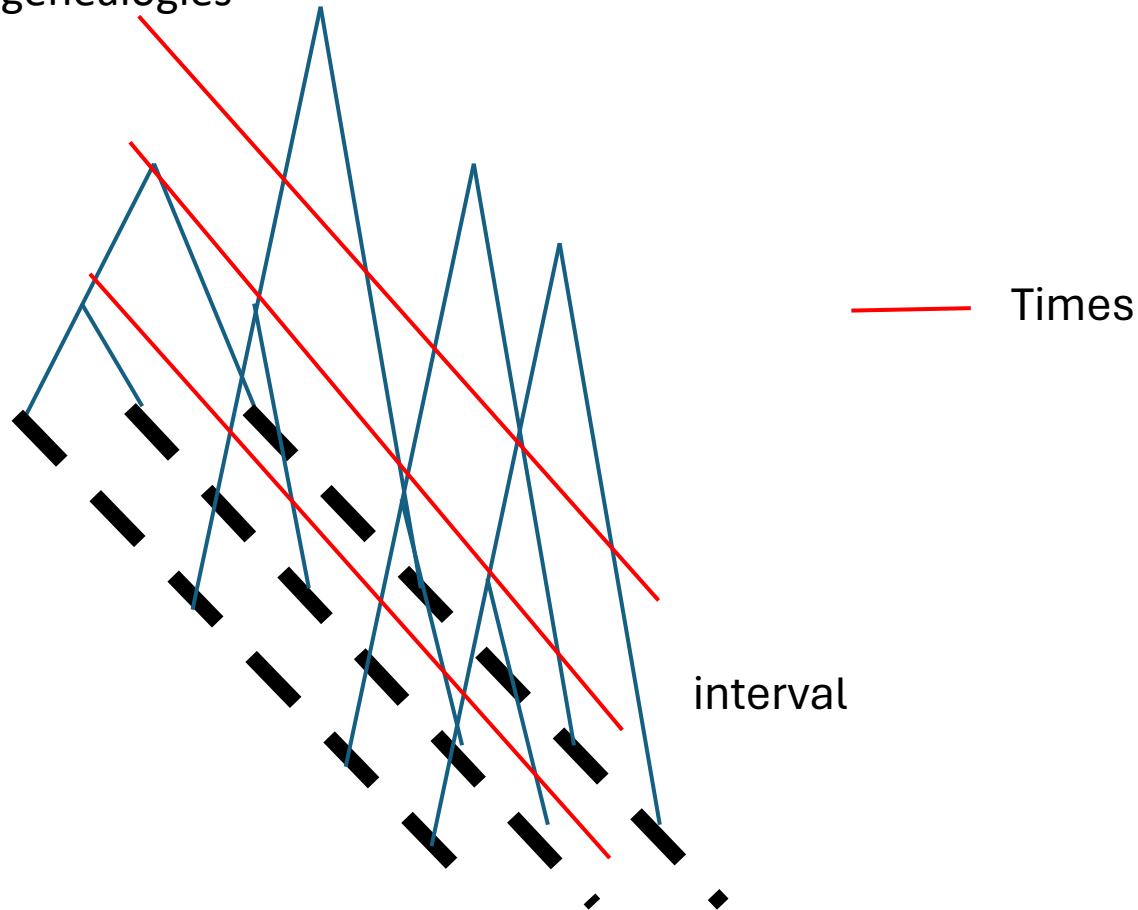
After identifying equivalent branches in adjacent trees along the genome, we apply a Metropolis–Hastings-type MCMC algorithm to estimate branch lengths. The MCMC algorithm has a coalescent prior assuming a single panmictic population

## **b** MCMC for branch length estimation



# Demography from Relate

Estimate a stepwise varying effective population size through time, using the genome-wide collection of estimated genealogies



Coalescence rate at a certain time,  
is inversely related to the population  
size

Infering Selection

**NEXT  
TIME**

# Exercises

# Talk Friday

Dear all

It is a great pleasure to announce our next BiRC talk:

**Speaker:** Haja Kadarmideen, Professor, ANIVET, TECH

**Title:** Applications of Multiomics, Systems Biology and Bioinformatics in Animal & Veterinary Sciences

**Time/place:** Friday, March 15, 11:15-12:00, in room 1870-716

**Abstract:** The talk will focus on applications of multiomics data integration, systems biology/genetics, network biology & bioinformatics within the animal & veterinary sciences including feed efficiency, animal welfare and artificial reproduction technologies and the use of pig models for human obesity & diabetes. Talk will outline statistical omics, AI/ML, and bioinformatics software development for the analyses of (population) genomics, epistasis, epigenomics and transcriptomics datasets. Potential Masters' and PhD thesis topics in Bioinformatics and Systems biology based on available animal multiomics datasets will be presented for those who might be interested.