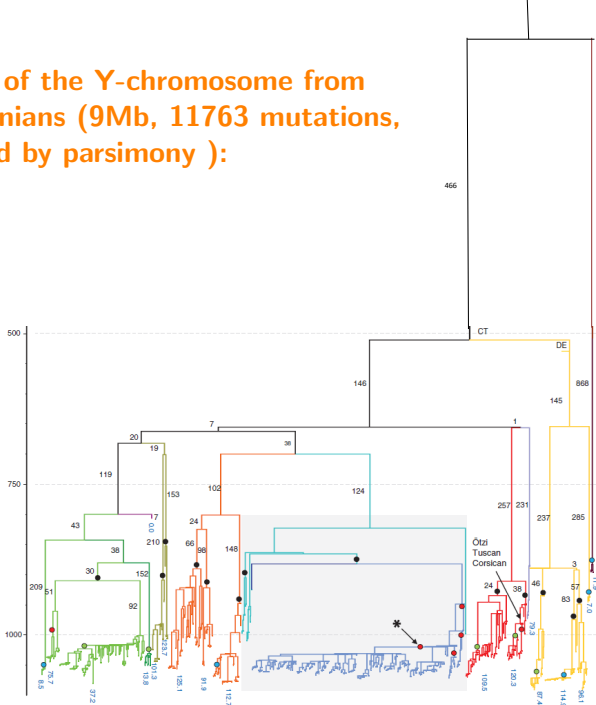


Genealogies, population structure, and selection.

Outline

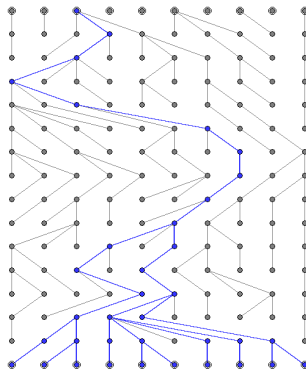
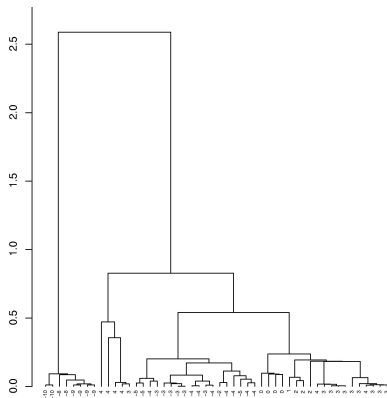
1. Gene genealogies and coalescent theory (this presentation).
2. Practical: Simulating genealogies in R.
3. Natural selection, population structure, detecting selection (next presentation).
4. Practical: Using PCAdapt to detect selection in North American wolves
5. Project: population structure and selection in goat breeds.

Genealogy of the Y-chromosome from
1204 Sardinians (9Mb, 11763 mutations,
constructed by parsimony):



Genealogies and the Wright-Fisher Model

The properties of genealogies can be quite easily understood from an idealised, clonal, model of reproduction where each individual in a fixed population of size $2N$ 'chooses' its parent at random (the Wright-Fisher model).

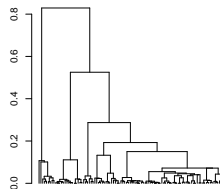
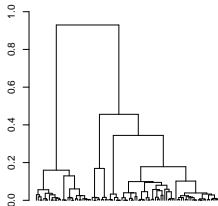
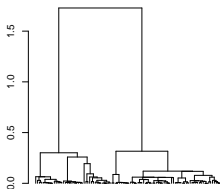
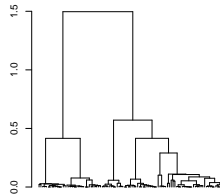
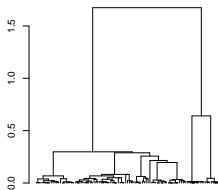
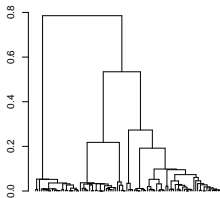


(coalescent.dk)

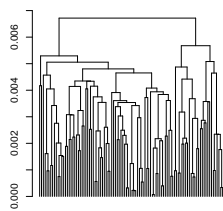
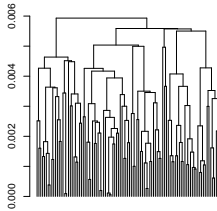
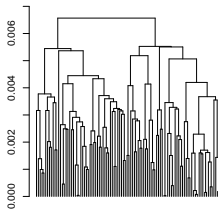
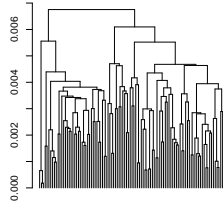
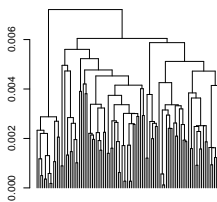
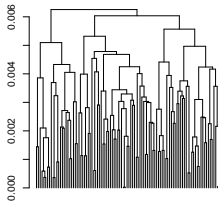
What is the advantage in taking a genealogical approach?

- ▶ Most of the genetic variation is selectively neutral.
- ▶ Neutral mutations have no effect on the shape of a genealogy.
- ▶ Therefore neutral mutations can be considered as adding marks to the DNA as it is copied from generation to generation.
- ▶ All the 'interesting stuff' therefore depends on the genealogy and where in the genealogy a mutation happened.
- ▶ The mutation rate per nucleotide site is typically so low compared to the population size that to a very good approximation in any particular genealogy, for any length of sequence, we can assume that no mutation hits the same site twice (the 'infinite sites approximation').
- ▶ It is orders of magnitude faster to simulate genetic data by simulating from the coalescent rather than from the Wright-Fisher model. Theoretical modelling is much easier too.

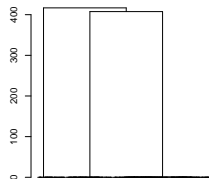
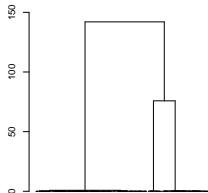
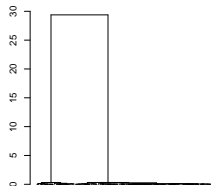
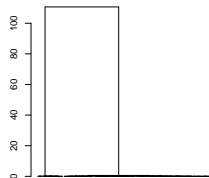
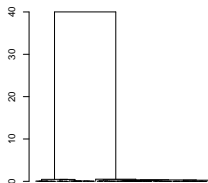
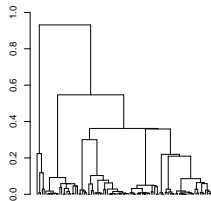
Simulated Genealogies (Constant Population Size)



Simulated Genealogies (Increasing Population Size)

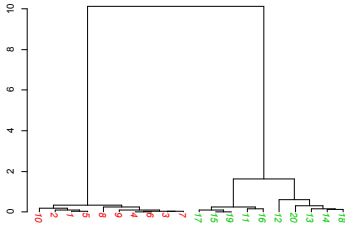


Simulated Genealogies (Decreasing Population Size)

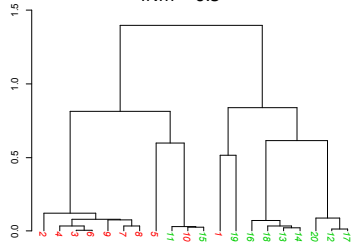


Simulated Genealogies (Two populations, with migration)

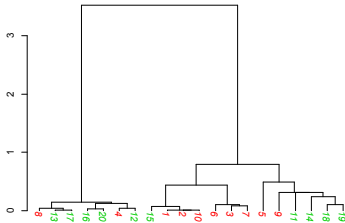
$4Nm = 0.1$



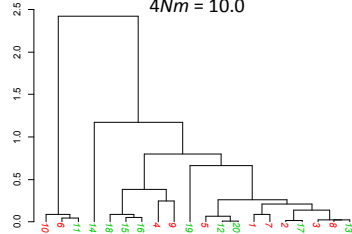
$4Nm = 0.5$



$4Nm = 1.0$



$4Nm = 10.0$



Running Example Simulations in R

- ▶ Aims:
 - ▶ Remind you how to use R.
 - ▶ Allow you to become more familiar with population genetic concepts.
- ▶ We will now open the following GitHub webpage:
`https://ritarasteiro.github.io/FieldCourse2019`
and continue the practical from there ...