**Steps in the tsinfer analysis**

1) **Merging vcf files** by samples

* what are the parameters here?
* what are the checks here?

2) **Phasing**

* what are the parameters here?
* what are the checks here?

3) **Prepare Files for tsinfer**

* we start with compressed and indexed vcf files
* whether we start with one or multiple vcfs –if there is only vcf → we split it into chromosome vcfs
* we decompress the files
* next, we put in the ancestral information
  + the ancestral file need to be standardized
  + for the sites that don’t have an ancestral allele known, we leave a black space in the VCF (make sure this script runs fine and do what expected!)
* - we compress and index the files

4) **Infer the trees**

* prepare the samples file
* what steps do each of us take
* standardize the meta file (to read in the populations)
* infer by chromosome (this is probably true for everyone)

5) **Do statistics on the trees**

* GNN
* Fst
* Tajima’s D
* ...