Session 9 Exercise: Calculating Fst divergence measures for HapMap populations

The PLINK file "plink_hapmap_CEU_YRI_ASW_CHB.bed" (as well as the corresponding ".bim" and ".fam" files) is available on the course website. The file contains genotype data at autosomal SNPs for four HapMap populations: Yoruba in Ibadan, Nigeria (YRI); Han Chinese in Beijing, China (CHB); Utah residents with ancestry from Northern and Western Europe (CEU), and African Americans from the south-western United States (ASW).

1. Use PLINK to obtain allele frequencies for each of the four HapMap Populations. Files with family and individual IDs for the populations are available on the course website: "hapmap-CEU.txt", "hapmap-YRI.txt", "hapmap-CHB.txt", and "hapmap-ASW.txt". You can use the following PLINK command to obtain allele frequencies for only the CEU individuals.

```
./plink --bfile  plink_hapmap_CEU_YRI_ASW_CHB --keep
hapmap-CEU.txt --freq --out CEU
```

2. Read the allele frequency files into R. Obtain Fst measures for all pairs of HapMap populations. (Hint: Use the R-script that is available on the course website)

3. Which of the HapMap population pairs are the most divergent? Which populations are the most similar based on Fst?