Computational Genomics 2025
Homework 1: Population Admixture

PLEASE NOTE THE FOLLOWING INSTRUCTIONS:

1. You are to complete this assignment alone. The assignment is open book, so you are allowed to use any books or information available online, your own notes and your previously constructed code, etc. HOWEVER YOU ARE NOT ALLOWED TO COMMUNICATE OR IN ANY WAY ASK ANYONE FOR ASSISTANCE WITH THIS EXAM IN ANY FORM.

2. Please pay attention to instructions and complete ALL requirements for ALL questions, e.g.
some questions ask for code, plots, AND written answers.

3. A complete answer to this assignment will include a single pdf file named in this format: 'your_name'.compG_HW1.pdf. This file should include all of your commands in linux from installation to the end, your R codes, and your resulting plots.

4. The exam must be ***sent to mehrmohamadi@ut.ac.ir by 11pm on the 5th of Farvardin, 1404.*** *It is your responsibility to make sure that it is received by then and no excuses will be accepted.*

1) Download 3 files for admixture analysis from the following: https://dalexander.github.io/admixture/download.html

- The proper admixture tool
- The sample data: hapmap3-files.tar.gz
- The user manual

2) Install admixture on your system (include steps/commands you used for installation)

3) Follow the instructions in the manual to run admixture on the hapmap3.bed file using K=3. Explain what information your input bed file contains. Also explain what K represents and why you need to provide it as an input to admixture.

4) List the names of files that you get as the resulting output of your admixture run.

5) Follow the instructions in the manual to plot the ancestry fractions (Qs) in R. Include your R commands and your barplot (play with plotting options in R so that your plot has a different look than what is included in the manual!). Interpret your results.

6) How many SNPs were used for this admixture analysis? Explain how you obtained this number.

7) Find the SNPs with the highest and the lowest cross-population variance in allele frequencies. Report population-wise AFs for these two SNPS from your output (P) file. Include your R code.