1.(1 points) Assume you study a particular group of individuals with two hypothetic subpopulations of equal size with the following allele frequencies in a biallelic locus:
subpopulation 1) p = 0.9, q = 0.1
subpopulation 2) p = 0.6, q = 0.4
Calculate F_{ST}. To what extent are the two subgroups isolated?

2. (4 points) Let's turn to genetic clustering algorithms. As mentioned during the lecture, such algorithms may be used not only to model variation in present-day populations, but also to study evolution of closely-related species. We'll study an example from the following paper:
Hailer F, Kutschera VE, Hallström BM, Klassert D, Fain SR, Leonard JA, Arnason U, Janke A. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science*. 2012 Apr 20;336(6079):344-7.
You need to download the structure software using the link below:
https://web.stanford.edu/group/pritchardlab/structure.html
Modify main modeling parameters (mainparams) to match the data (at least INFILE (1216424s2_structure.txt), NUMINDS, NUMLOCI and EXTRACOLS [you should infer this by examining the input file]). Please provide the values you use as part of your submission.
1) Run structure with different values of clusters (-K option) - 3, 4, and 5. Visualize the results using Pophelper (http://pophelper.com/) or R. Which of the plots (for K = 3, 4, or 5) corresponds to the Figure 2 of the original paper?
2) Do the data obtained prove that the white bears are distinct from brown and black ones?

3. (2 points) Let's assume you study a sample of 4 alleles from a particular population. You've made a phylogenetic analysis of these sequences and got the following genealogy together with the coalescence times (t_1, t_2, t_3):
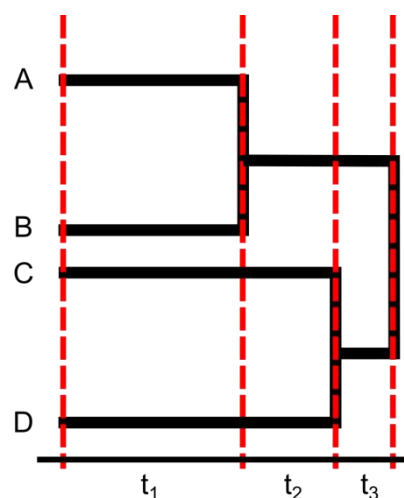
Let t_1 = 6000, t_2 = 5000, t_3 = 3000. Use the logic of the classical skyline method. What is the demographic history of the population? Was the population size increasing or decreasing recently?

4. (5 points) Let's make some Bayesian skyline analysis to model demographic histories. We'll use the data from Atkinson et al., 2007. We'll need BEAST2 tool and the Tracer application to run the analysis and view the results.

Use this tutorial run the analysis (you can find links to software there as well):

https://taming-the-beast.org/tutorials/Skyline-plots/

Important!

- only follow instructions from paragraphs "Setting up the Coalescent Bayesian Skyline analysis" and "Exploring the results of the Coalescent Bayesian Skyline analysis"

- use as input the provided **mtDNA.fasta** instead of the example

- set the parameter of "**Clock rate**" to $1.691 * 10^{-8}$ substitutions/site/year in accordance with the paper

- do not set any value of Gamma when choosing the GTR substitution model

- set "**Chain Length**" to at least 8,000,000

- choose the number of dimensions for "bPopSize" and "bGroupSize" at your own. Please provide the values used as part of the submission. (mind that this parameter affects computation speed)

- when viewing the results with Tracer, specify "**Age of the youngest tip**" between 1979 and 2007.

Questions:

1) What is the (approximately) median log of posterior probability as shown by Tracer? Please provide the plot of the distribution.

2) How does the population size change over time? Were there any bottlenecks? Please provide the skyline plot.

3) Compare your results with plots from the original paper. Which population was used in your analysis? Are there any differences in N_e values?