Mason Lien  
Agro 932 – Biometrical Genetics & Plant Breeding

AGRO-932 Spring 2022 HW1 (100 points): Due Date: Feb. 17th, 2022 at 10 PM

Simulate an NGS dataset for two populations with a small reference genome (< 1 Mb) of your choice, each population with 10 diploid individuals, to address a hypothesis about population differentiation.

1. Establish a version-controlled directory system to host the project (paste the link here). Show your work in the GitHub repository in a user-friendly and reproducible manner.

<https://github.com/Mlien89/PhD/tree/main/Agro932/assignments/hw1>

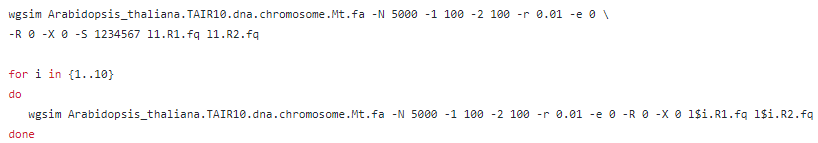
1. Describe your simulation strategy and the hypothesis to test (positive, negative, or neutral selection).

I chose an Arabidopsis dataset for simulating two populations. I considered mitochondrial DNA in the following link <http://ftp.gramene.org/CURRENT_RELEASE/fasta/arabidopsis_thaliana/dna/>. I ran two simulations where simulation A had a mutation rate of 0.01 and simulation B of 0.10.

* Arabidopsis\_thaliana.TAIR10.dna.chromosome.Mt.fa

To simulate sequence reads from a reference genome, wgsim is a tool within SAMtools software package that allows the simulation of FASTQ reads from a FASTA reference. It can simulate diploid genomes with SNP and insertion/deletion, and create reads with uniform substitution sequencing errors. In [hw1\_arab.sh](https://github.com/Mlien89/PhD/blob/main/Agro932/assignments/hw1/lib/hw1_arab.sh), I use wgsim in the following fashion

Simulation A – mutation rate is 0.01



Where the following represents

-N INT number of read pairs [5000]

-1 INT length of the first read [100]

-2 INT length of the second read [100]

-d INT outer distance between the two ends [500]

-r FLOAT rate of mutations [0.01]

-e FLOAT base error rate [0.0]

-R FLOAT fraction of indels [0.0]

-X FLOAT probability an indel is extended [0.0]

-S INT seed for random generator [1234567]

1. Calculate thetas for each population and compute Fst between the two populations.
2. Visualize and briefly interpret your results.

