**AGRO-932 HW1**

**Santos Barrera Lemus**

**02-20-2020**

**1.** Establish a version-controlled directory system to host the project.

[**https://github.com/Sybalemus1/Agro932hw.git**](https://github.com/Sybalemus1/Agro932hw.git)

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

**Simulation strategy**

I used the wgsim program to generate the NGS simulated data of two populations with 10 indiviuduals each one (I ran the program twice to generate the 20 individual). Then I download the reference genome (Zea\_mays) using the wget program. The genome used was small with about 500k, 20x coverage, 100bp pair ends reads, only SNP no Indel, and a low mutation rate.

**Hypothesis**

To determinate how different the both simulated populations are, is necessary to determinate the Fst values. Two scenarios 1). If the Fst value is = 0 it indicated that there is not population differentiation which indicates no evolution forces involved. 2) If the Fst value is = 1 it indicated completely differentiation (completely homozygous). Under this concept we can assume that if on average the Fst value is low there is a negative selection.

**3.** Use ANGSD to calculate thetas for each population and Fst between the two populations.

**Distribution of site frequency spectrum of alleles (SFS) and pairwise theta distribution**



Scatter plot of the Fst values:

A screenshot of a cell phone

Description automatically generated

Fst values are low (between 0.02 and 0.006) indicating that there is no population differentiation

**4.** Breakdown the theta ratios and Fst values into different genomic features (i.e., genic and intergenic regions) according to the general feature format (GFF) file for the reference genome of your choice.

