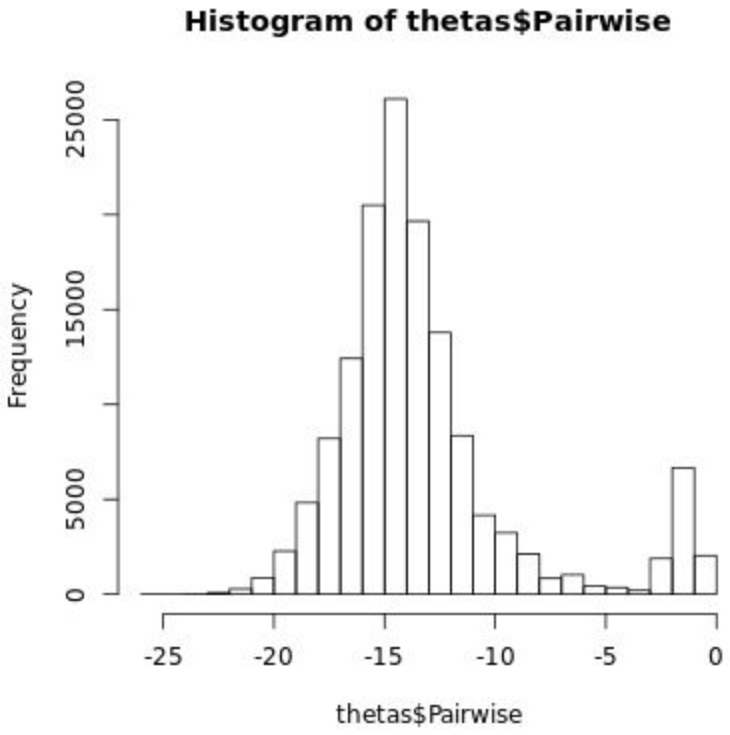
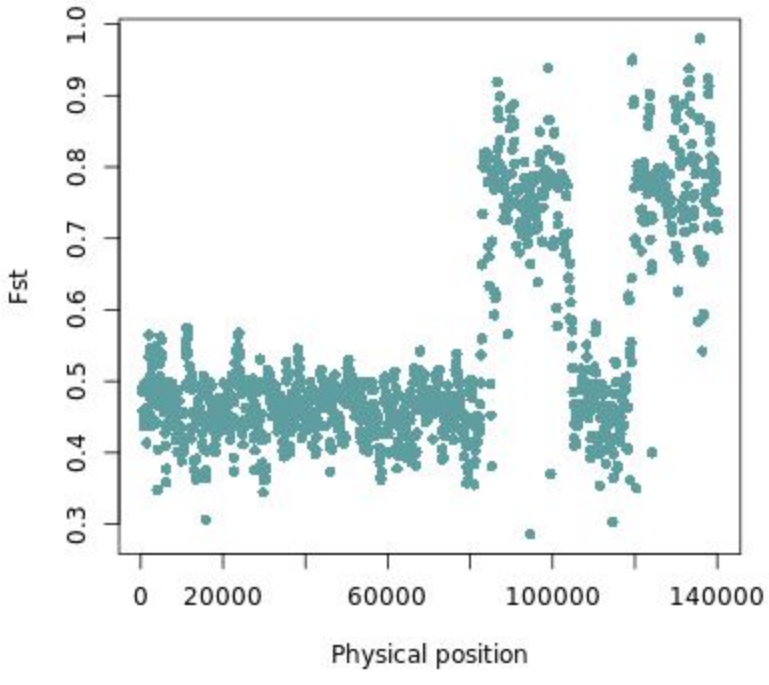
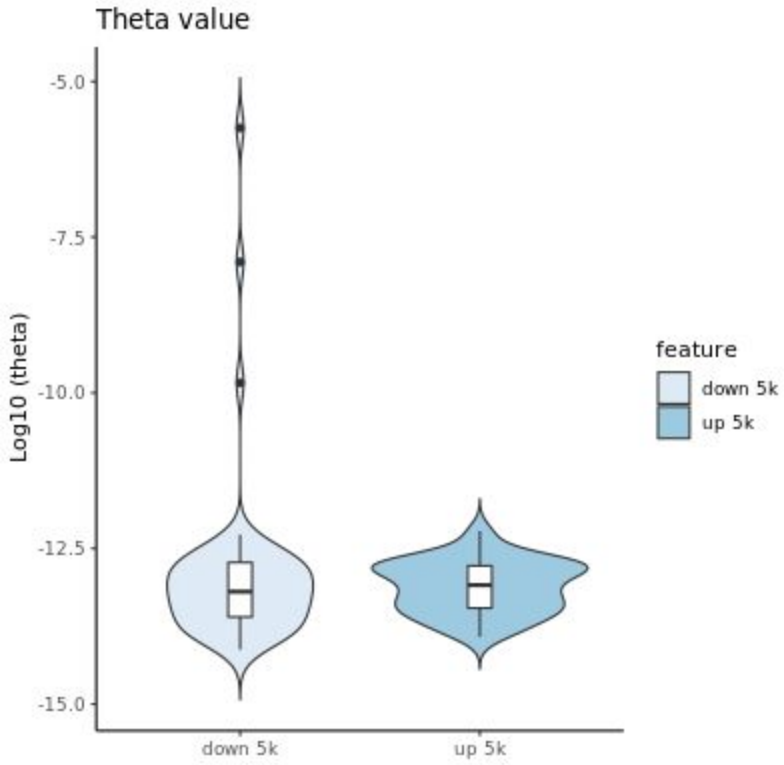
Mary Happ
AGRO 932
2.20.2020
HW #1

1. Version controlled directory system was set up on rhino@unl.edu and maintained at https://github.com/mhapp95/AGRO932/. Git add, commit, and push were used to periodically save work and backup versions.
2. I chose to simulate reads based on the maize (Zea mays) chloroplast genome. It is ~140kb in length. I simulated paired reads, at 150 bp per read and 15X read coverage. Therefore the "-N" parameter of wgsim was 3500. I also simulated an error rate of 1% and a fairly low mutation rate of r=0.2. I would like to look for positive selection.
3. Code used for ANGSD calculations can be found in the github repository in scripts/finished/ANGSD.slurm
4. Code for breakdowns of theta and fst score by genomic feature can be found in the FIGURES.R script
5. This histogram of my pairwise theta values follows a fairly normal distribution, with a small rise towards zero, perhaps indicating some type of differential selection at certain sites between the subpopulations. The scatter plot of the Fst values shows two "spikes" at roughly 90kb and 140kb indicating alleles at those sites were indeed under positive selection and show complete or near complete fixation in one of the subpopulations. Finally, a look at our violin plot shows that these areas under selection were mostly in the downstream portion of genes, indicating by the outliers that stretch the violin plot of theta values towards the top.





Mary Happ
AGRO 932
2.20.2020
HW #1