Assignment 3

Reading

Korneliussen et al. 2014 [Genotype and SNP calling from next-generation sequencing data](#)

[Site Freq. Spectrum](#)

[nucleotide diversity](#)

[Watterson's estimator](#)

Tajima's D [here](#) and [here](#)

Introduction to [maize domestication](#)

Homework

For the next several classes we will be playing with some example data from maize and teosinte. In order to make analyses feasible given our time and disk space limitations, these data are only for a single 10Mb region of one maize chromosome (0.4% of the genome!), but the compressed data still takes up 1Gb of disk space.

1. **Install ANGSD-wrapper**

Details [here](#). You will need to install using `git clone` then run `angsd-wrapper setup` and `angsd-wrapper setup data` to get the example data. These are bam files and are already aligned. It is not recommended to try these analyses on your own laptop.

1. **Estimate the site frequency spectrum for maize and teosinte**

   Use `angsd-wrapper SFS Site_Frequency_Spectrum_Config`. (Bonus: plot the SFS - - why does it look weird?)

2. **Get summary statistics diversity for maize and teosinte**

   Use `angsd-wrapper Thetas Thetas_Config`. Report the overal diversity values for each taxa.

3. **Explain the differences you see in nucleotide diversity and Tajima's D between maize and teosinte**

Challenge

Run a sliding window analysis of diversity.

1. **Graph the plots in ANGSD-wrapper's shiny (or yourself in R).**

2. **Find windows in the bottom 5% quantile of the ratio of nucleotide diversity between maize and teosinte.**

3. **Find out which, if any, genes are in these intervals using the maize [gene annotation](#) and [bedtools](#).**