

Assignment 5

Reading

Holsinger and Weir 2009 [Genetics in geographically structured populations: defining, estimating and interpreting \$F_{ST}\$](#)

Homework

Use the `Fst` module](<https://github.com/mojaveazure/angsd-wrapper/wiki/2D-Site-Frequency-Spectrum-and-Fst>) of ANGSD-WRAPPER to estimate a sliding window of F_{ST} between **mays** and **parviglumis**.

Download the [annotation for maize chromosome 10](#) and use [bedtools](#) to find genes in the top 10% quantile of F_{ST} .

Questions

- Use the mean F_{ST} between **maize** and **parviglumis** to estimate divergence time between them. You will need your estimates of the population mutation rate from HW2 to estimate N_e for this calculation.
- Use the mean F_{ST} between **maize** and **parviglumis** to estimate the migration rate between these two taxa.
- What assumptions are you making in each of the above? Do you think these are valid approaches (why/why not)?
- What % of your high F_{ST} windows are in our near genes? Is this surprising?

Hint You will almost certainly want to filter the F_{ST} outfile to **only** look at SNPs that are variable. You can use 3×10^{-8} as a mutation rate for estimating θ

Challenge

1. Using the data from *mexicana* estimate divergence time between the two teosinte. How does this differ from [previous results](#) and which estimate do you believe?
2. Use [bedtools](#) and [AgriGO](#) to do a GO term enrichment analysis of genes in the top 10% F_{ST} .