README file: Running a pairwise Fst analysis for the *Medicago polymorpha* deep GBS population

Emily Helliwell 9/11/2018

In this experiment, we calculated the pairwise Fst values between all of the deep GBS populations (N > 5 individuals for a population) in the native and invaded ranges. Pairwise Fst values were calculated using Weir and Cockerham’s estimate (1984) in the r package hierfstat. The results were then visualized by constructing a heatmap with a dendrogram, and constructing a histogram based on the clustering patterns seen.

The code for this experiment is all in one file, called ‘pairwise\_Fst\_analysis\_final.R’

There are three input files, as specified in the R code. The first file is “Mpol\_wout\_USDA\_filtered\_Fst.csv’ This file will be used to run the pairwise Fst analysis. The header row designates the first column as the population labels, and all subsequent columns as polymorphic SNPs. Each row designates one individual, but for the purpose of this experiment, the unique ID’s are not specified; only the population of origin.

The second input file, ‘Mpol\_all\_deep\_pops\_corval.csv’ is a transposed matrix of the resulting pairwise Fst values output from the previous step. The population names have been changed from their three-letter abbreviation to the designations used in the paper. This dataset will be used to run the heatmap and bootstrap analysis.

The final input file, ‘Mpol\_pairwise\_Fst\_bycluster.csv’ is an expanded file, which specifies the significance cluster of a particular score as determined by the bootstrap analysis in the previous step. 1v1 and 2v2 refers to the pairwise Fst value between populations that landed in the same bootstrap cluster, while 1v2 refers to pairwise Fst values between populations that landed in different bootstrap clusters. The output of this is a histogram with three peaks, each peak representing a cluster.