# Load Required Libraries

library(vcfR)

library(adegenet)

library(StAMPP)

library(ggplot2)

library(reshape2)

# Define file paths

vcf\_file <- "data/final.50.recode.vcf"

pop\_file <- "data/Cop.txt"

pca\_output <- "outputs/Cop\_adegenetPCA.txt"

fst\_output <- "outputs/Fst.txt"

fst\_pvalue\_output <- "outputs/Fst\_pvalue.txt"

pca\_plot\_path <- "outputs/PCA\_plot.png"

# Step 1: Load SNP and Population Data

snp\_vcf2 <- read.vcfR(vcf\_file)

pop.data2 <- read.table(pop\_file, header = FALSE)

# Convert VCF to genlight object

gl.snp2 <- vcfR2genlight(snp\_vcf2)

pop(gl.snp2) <- pop.data2$V2

# Step 2: Perform PCA

snp.pca2 <- glPca(gl.snp2, nf = 10)

cat("Number of individuals:", nInd(gl.snp2), "\n")

cat("Number of population labels:", length(pop.data2$V2), "\n")

# Save PCA scores

snp.pca.scores2 <- as.data.frame(snp.pca2$scores)

snp.pca.scores2$pop <- pop(gl.snp2)

write.table(snp.pca.scores2, pca\_output, sep = "\t", row.names = FALSE)

# Eigenvalues and percentages

eig.val <- snp.pca2$eig

eig.perc <- 100 \* eig.val / sum(eig.val)

eigen <- data.frame(Eigenvalue = eig.val, Percentage = eig.perc)

print(eigen)

# Step 3: PCA Plot

data2 <- read.delim(pca\_output)

mycol <- c("#f1c039", "#f37d21", "#51692d", "#56ba32")

pca\_plot <- ggplot(data2, aes(x = PC1, y = PC2, color = pop)) +

geom\_point(size = 2) +

scale\_color\_manual(values = mycol) +

theme\_classic() +

xlab("PC1 (8.75%)") +

ylab("PC2 (5.03%)")

ggsave(pca\_plot\_path, plot = pca\_plot)

# Step 4: Fst Calculation

Qfly\_Fst <- stamppFst(gl.snp2, nboots = 100, percent = 95, nclusters = 6)

write.table(Qfly\_Fst$Fsts, fst\_output, sep = "\t", row.names = FALSE)

write.table(Qfly\_Fst$Pvalues, fst\_pvalue\_output, sep = "\t", row.names = FALSE)

# Step 5: Heatmap of Fst

fst\_matrix <- as.matrix(read.table(fst\_output))

fst\_melted <- melt(fst\_matrix, na.rm = TRUE)

heatmap\_plot <- ggplot(data = fst\_melted, aes(Var2, Var1, fill = value)) +

geom\_tile(color = "white") +

scale\_fill\_gradient2(low = "#ffd60a", high = "#001d3d", mid = "#4e9de6",

midpoint = 0.056, limit = c(0.005, 0.11), space = "Lab",

name = "Pairwise Fst") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, vjust = 1,

size = 12, hjust = 1)) +

coord\_fixed()

##creating input files

library(LEA)

library(pophelper)

vcf2geno(input.file = "final.50.recode.vcf", output.file = "Qff.geno")

##snmf clustering

projectalpha = NULL

projectalpha = snmf("Qff.geno", K = 1:10, repetitions = 50, entropy = T, CPU = 8, project = "new")

# plot cross-entropy criterion for all runs in the snmf project

pdf(file = "./cross\_ent\_alphadefualt.pdf")

plot(projectalpha, col = "Cop", pch = 19, cex = 1.2)

dev.off()

best2 = which.min(cross.entropy(projectalpha, K = 2))

best2

best3 = which.min(cross.entropy(projectalpha, K = 3))

best3

best4 = which.min(cross.entropy(projectalpha, K = 4))

best4

best5 = which.min(cross.entropy(projectalpha, K = 5))

best5

best6 = which.min(cross.entropy(projectalpha, K = 6))

best6

best7 = which.min(cross.entropy(projectalpha, K = 7))

best7

best8 = which.min(cross.entropy(projectalpha, K = 8))

best8

best9 = which.min(cross.entropy(projectalpha, K = 9))

best9

best10 = which.min(cross.entropy(projectalpha, K = 10))

best10

##creating admixture plots. For this, you need to first create a new folder (Qfiles) and move the Q files with "best" entropies from the LEA runs into it.

sfiles <- list.files(path=("./Qfiles"), full.names=T)

slist <- readQ(files=sfiles)

plotQ(qlist = slist[2], imgtype = "pdf",

height = 1.5,

clustercol = c("#51692d", "#f1c039", "#1f78b4", "#33a02c", "#e31a1c",

"#ff7f00", "#6a3d9a", "#b15928", "#a6cee3", "#fb9a99"),

dpi = 1200, exportpath = "./")

plotQ(qlist = slist[3], imgtype = "pdf", height = 1.5,

clustercol = c("#51692d", "#56ba32", "#f1c039", "#f37d21"),

dpi = 1200, exportpath = "./")

plotQ(qlist=slist[4],imgtype = "pdf",

height = 1.5, clustercol = c("#51692d","#56ba32","#f1c039","#f37d21"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[5],imgtype = "pdf",

height = 1.5, clustercol = c("#f37d21","#51692d","#f1c039","#56ba32","#a63838"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[6],imgtype = "pdf",

height = 1.5, clustercol = c("#a63838","#f1c039","#ecbcab","#56ba32","#51692d","#f37d21"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[7],imgtype = "pdf",

height = 1.5, clustercol = c("#51692d","#a63838","#f1c039","#ecbcab","#caf291","#56ba32","#f37d21"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[8],imgtype = "pdf",

height = 1.5, clustercol = c("#3d87db","#a63838","#51692d","#f1c039","#ecbcab","#caf291","#56ba32","#f37d21"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[9],imgtype = "pdf",

height = 1.5, clustercol = c("#f1c039","#a63838","#3d87db","#ecbcab","#caf291","#56ba32","#0000FF","#51692d","#f37d21"), dpi = 1200, exportpath = "./")

plotQ(qlist=slist[1],imgtype = "pdf",

height = 1.5, clustercol = c("#dd00ff","#51692d","#caf291","#3d87db","#ecbcab","#a63838","#56ba32","#0000FF","#f1c039","#f37d21"), dpi = 1200, exportpath = "./")