Lecture Notes

Adegenet Package

This package is used to determine genetic clustering based on common genetic markers, or single nucleotide polymorphism using Principle components analysis and Discriminant analysis

library(“adegenet”)

data(dapcIllus)

class(dapcIllus)

names(dapcIllus)

x <- dapcIllus$b

x

grp <- find.clusters(x, max.n.clust = 40)

find.clusters completes a PCA analysis and then runs a k- mean to help determine the groups present. You will need to enter the number of PCs retained after looking at the first report graph. For this we will just be using 200 because we just need to be above the actual value.

Once that has run, you will see a second graph showing the BIC v. number of clusters. You will need to see the number of cluster present based on the location of the valley.

names(grp)

head(grp$Kstat, 8)

grp$stat

grp$size

table(pop(x), grp$grp)

table.value(table(pop(x), grp$grp), col.lab=paste(“inferred”, 1:6), row.lab=paste(“actual”, 1:6))

This will show you if your find cluster analysis did find the correct groups and you will like see a few outliers but that can be expected.

Next, we will be running a Discriminant analysis, this function uses discriminant factors to create genetic clustering through looking for the largest between group variance with the lowest within group variance.

Dapc1 <- dapc(x, grp$grp)

This will first give you a similar PCA curve. For this function you need to input the PCs as close as possible to the point at which it plateaus. Next you will see a bar graph. This bar graph show the number of discriminants being used and you can select the number of them.

scatter(dapc1)

scatter(dapc1, posi.da=”bottomright”, bg=”white”, pch=17:22)

myCol <- c( colors you want to use (6))

scatter(dapc1, posi.da=”bottomright”, bg=white, pch = 17:22, cstar = 0, col= myCol, scree.pca=true, posi.pca=”bottomleft”)

scatter(dapc1, scree.da=false, bg=”white”, pch=20, cell=0, col=myCol, solid=0.4, cex=3, clab=0, leg=true, txt.leg=paste(“Cluster”, 1:6))

The coding above shows several variation that can be completed on the scatter plot in adegenet to allow for inclusion of different relevant information.