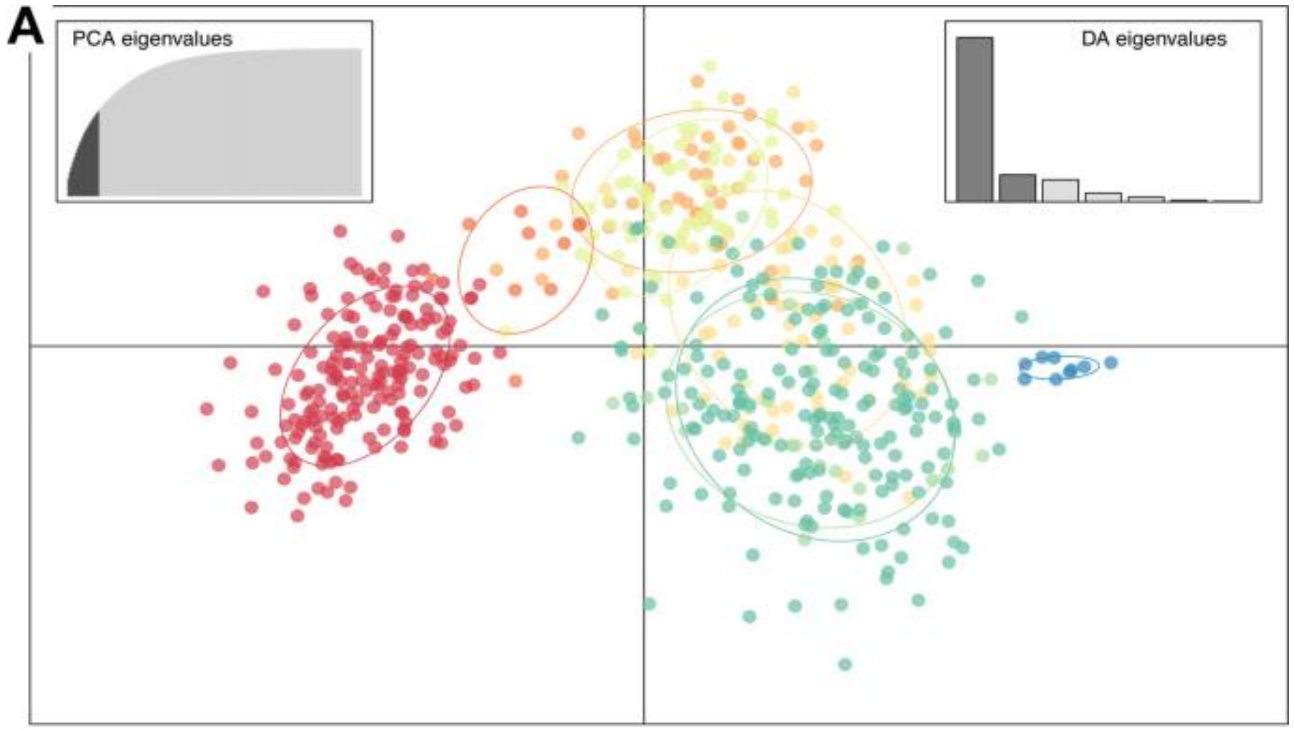
Genetic Structure via Structure Plots

Marine Genomics

5/25/2021

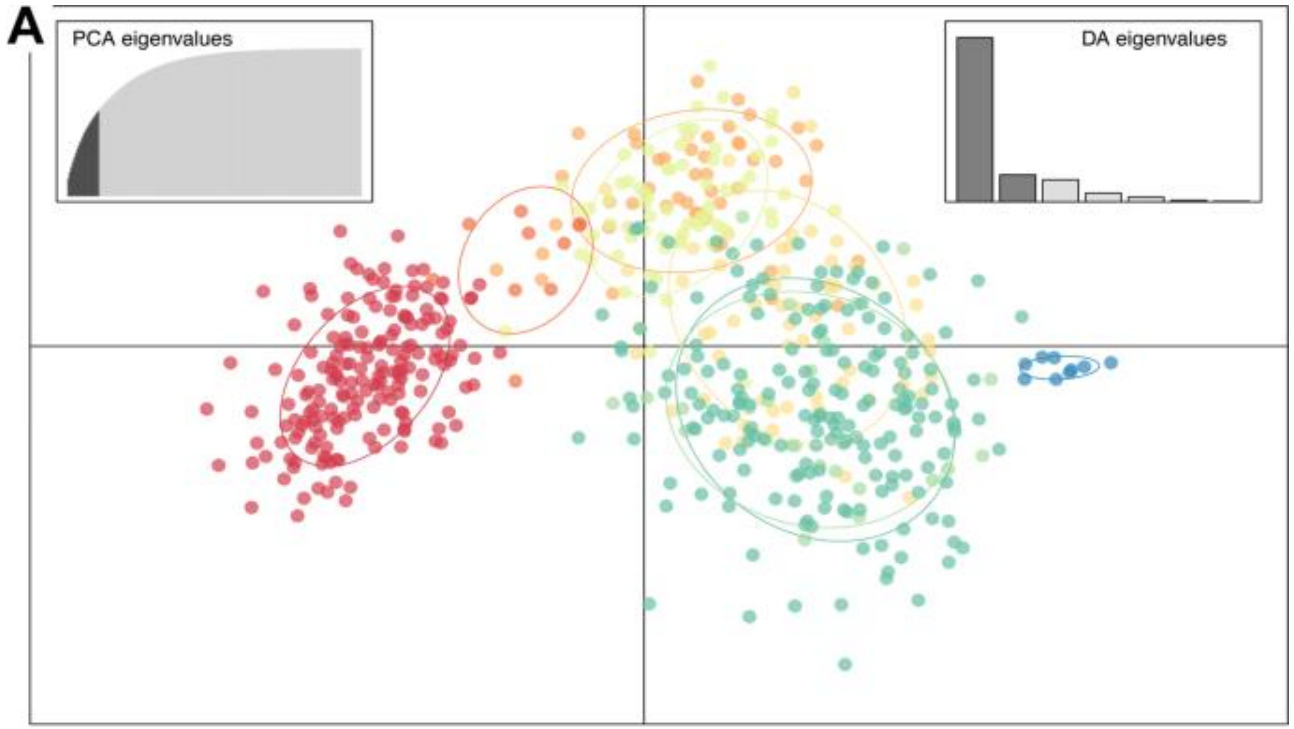
Genetic Structure of Populations via Clustering Algorithms

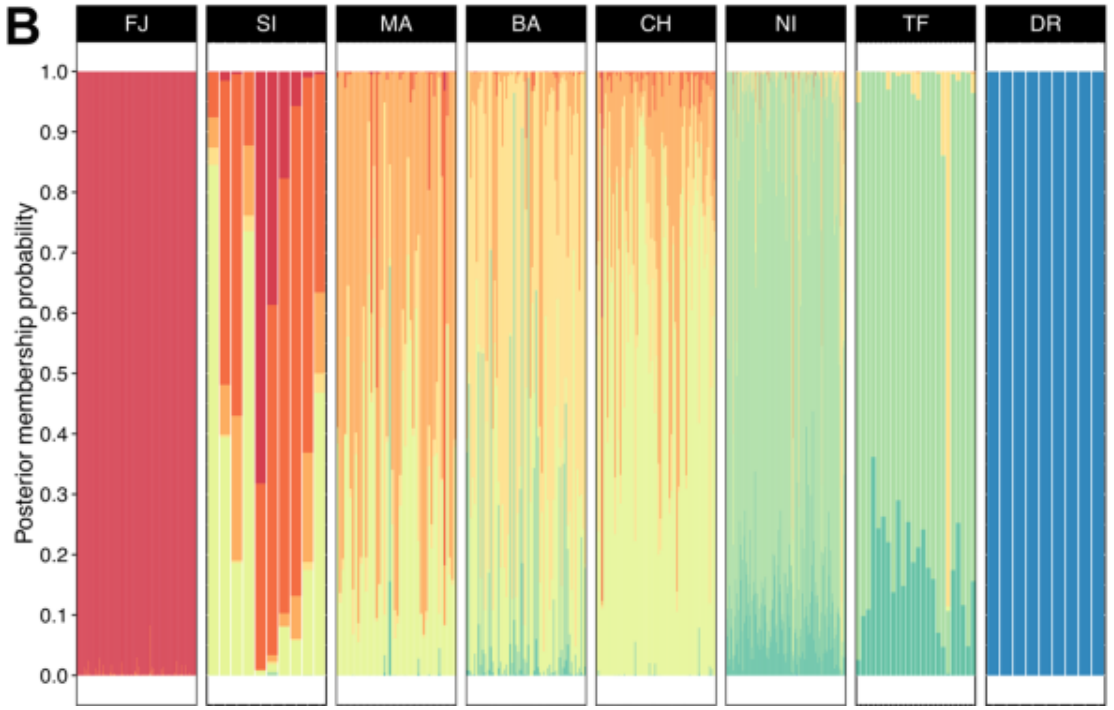


We covered PCA plots in Week 6. Another common way to summarize and graphically display population structure is in a Structure Plot.

https://www.nature.com/articles/s41598-020-66450-7

Genetic Structure of Populations

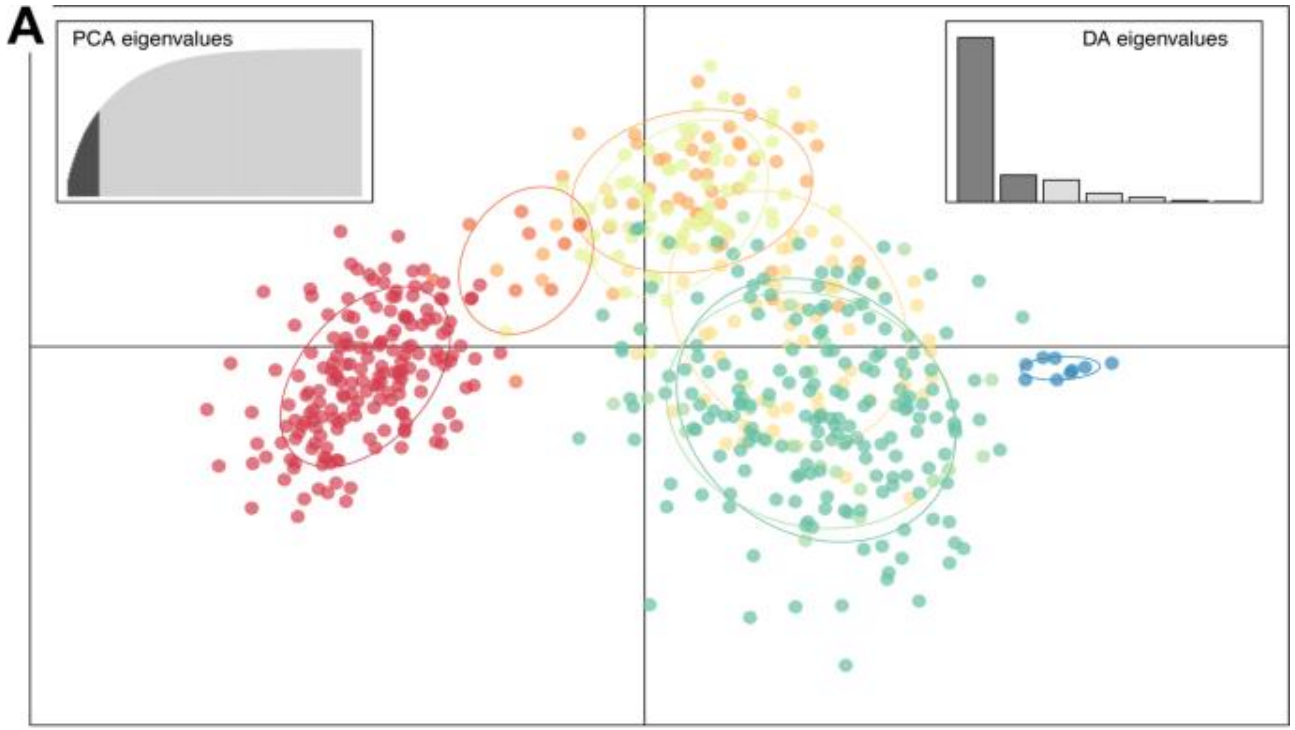


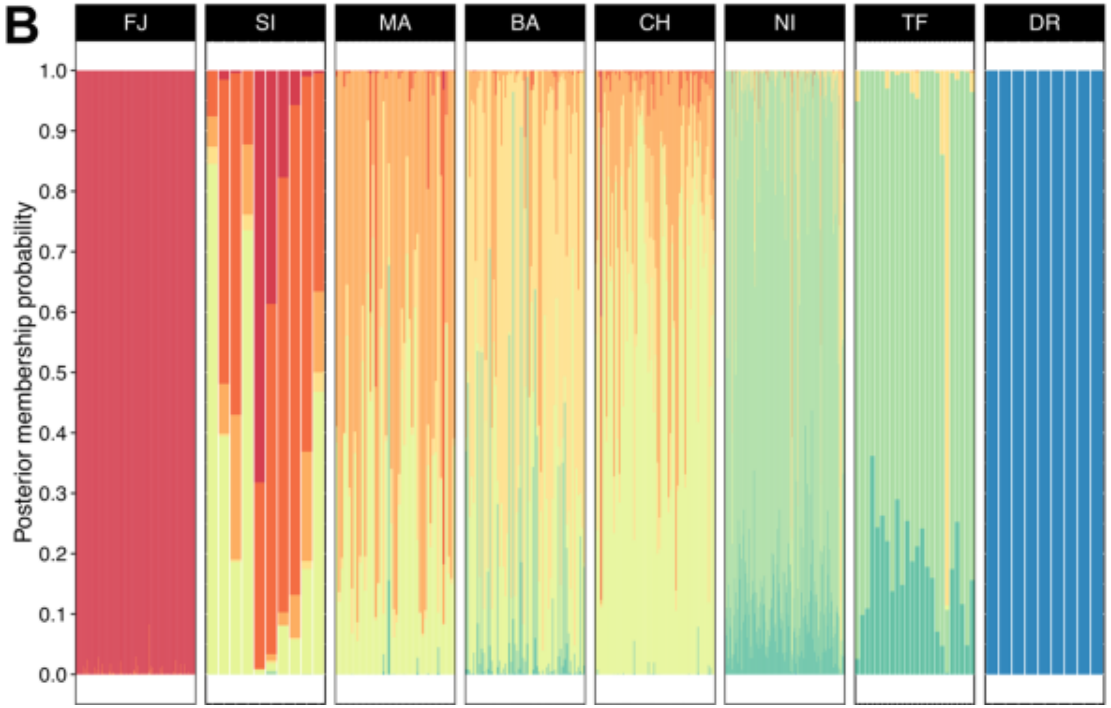


We covered PCA plots in Week 6. Another common way to summarize and graphically display population structure is in a Structure Plot.

https://www.nature.com/articles/s41598-020-66450-7

Structure plots can show patterns of admixture





https://www.nature.com/articles/s41598-020-66450-7

How do we generate a Structure plot?

Pick a program:

STRUCTURE (Pritchard et al. 2000)

NGSadmix in ANGSD

The program we pick will determine the input file type.

NGSadmix will take Genotype Likelihoods, while STRUCTURE takes “hard” genotype calls.

Using NGSadmix

NGSadmix is included in the ANGSD program

Input file need to be in the Beagle format

Run NGSadmix in bash

Do a lot of data wrangling in R

What data are we using this week?

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*)

Amanda Xuereb[1] | Laura Benestan[2] | Éric Normandeau[2] | Rémi M. Daigle[1] | Janelle M. R. Curtis[3] | Louis Bernatchez[2] | Marie-Josée Fortin[1]

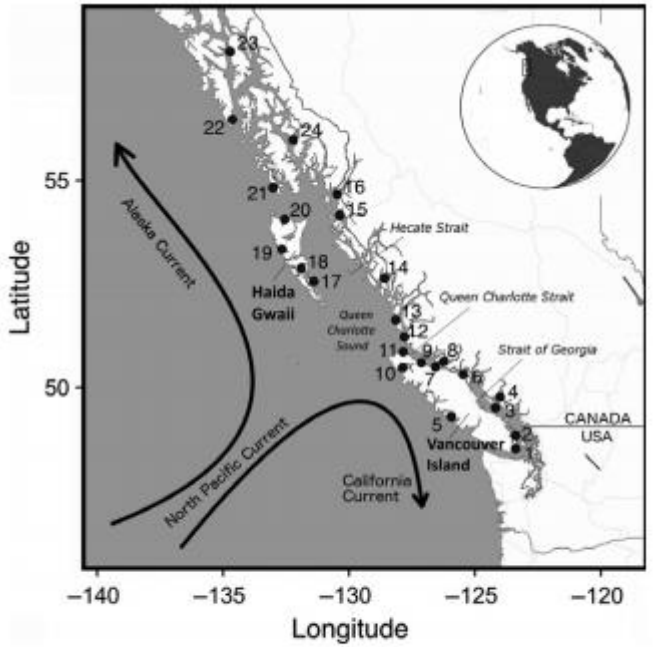
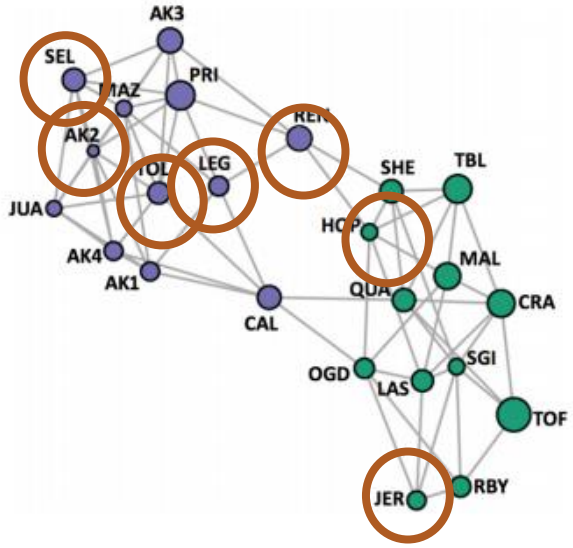


FIGURE 1 Map of sampling locations in coastal British Columbia (1–20) and southeastern Alaska (21–24). Site labels correspond with numbers in Table 1

We're using this subset:

15 individuals from 7 populations



Principal Component Analyses

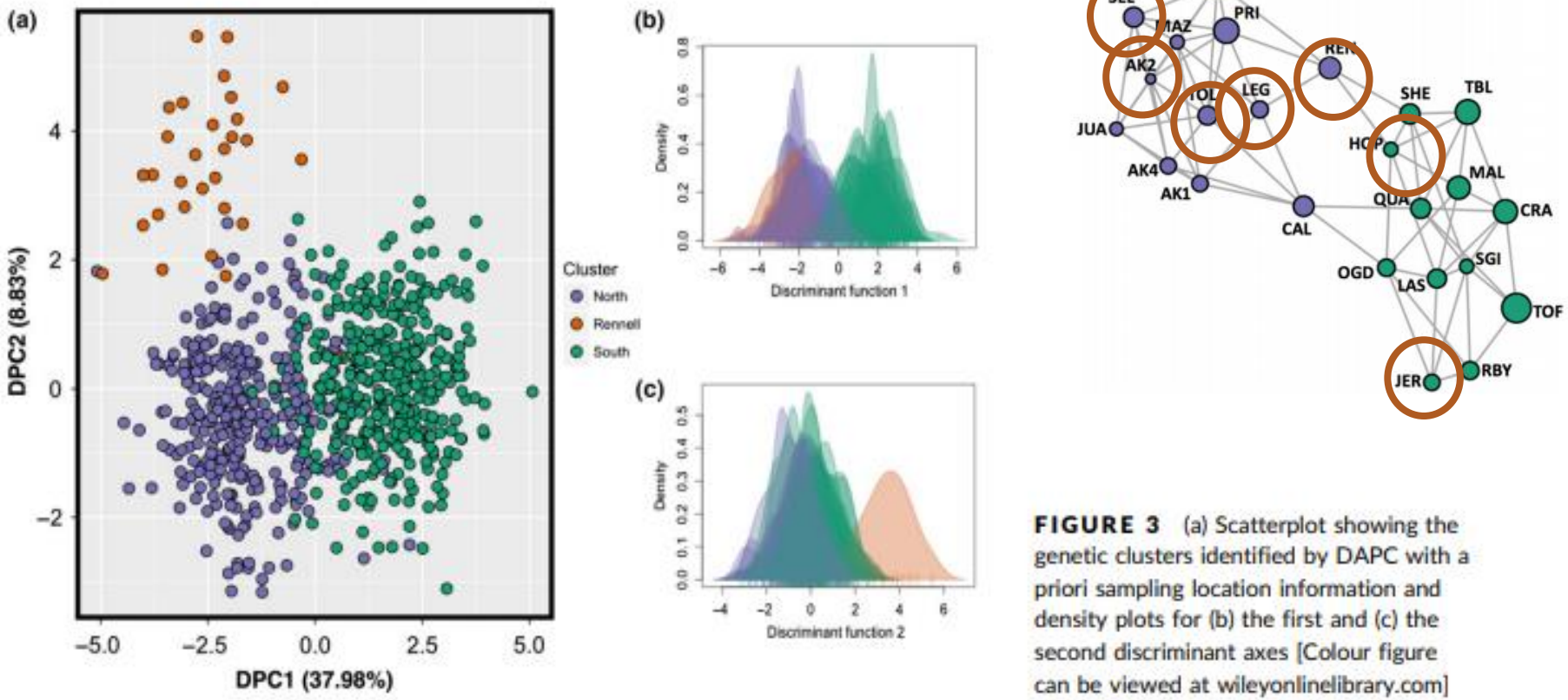
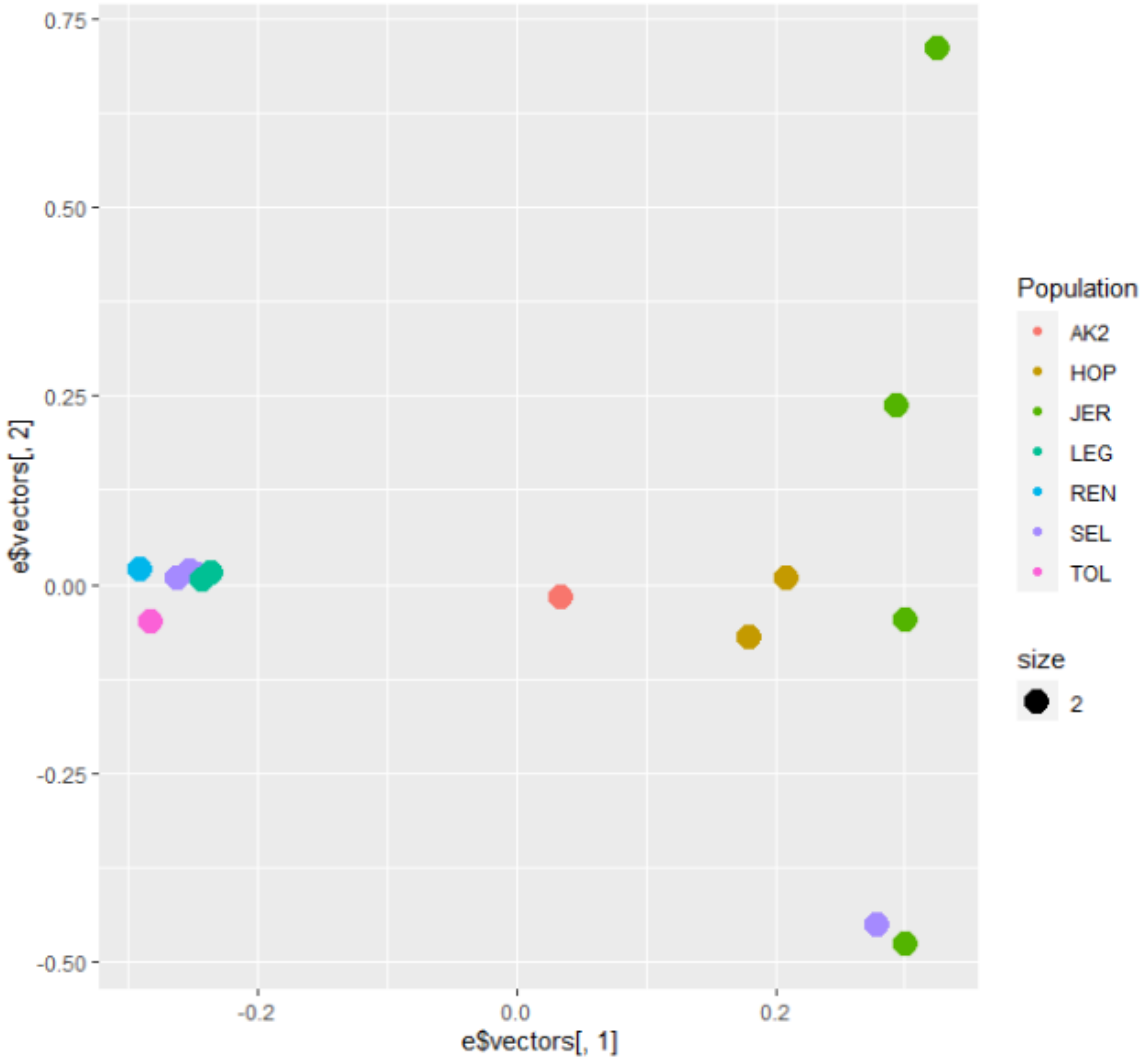
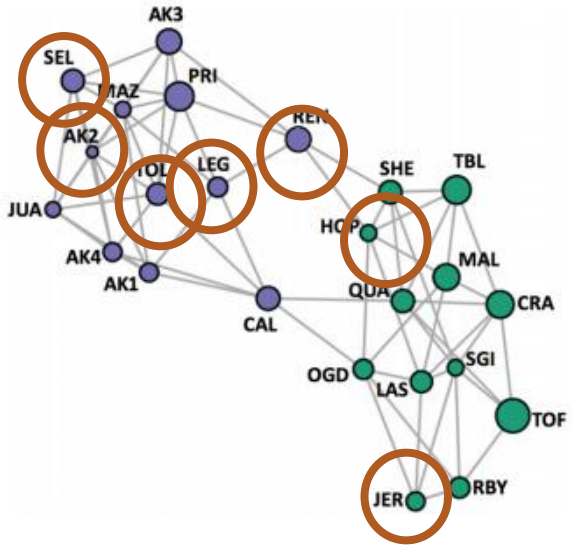
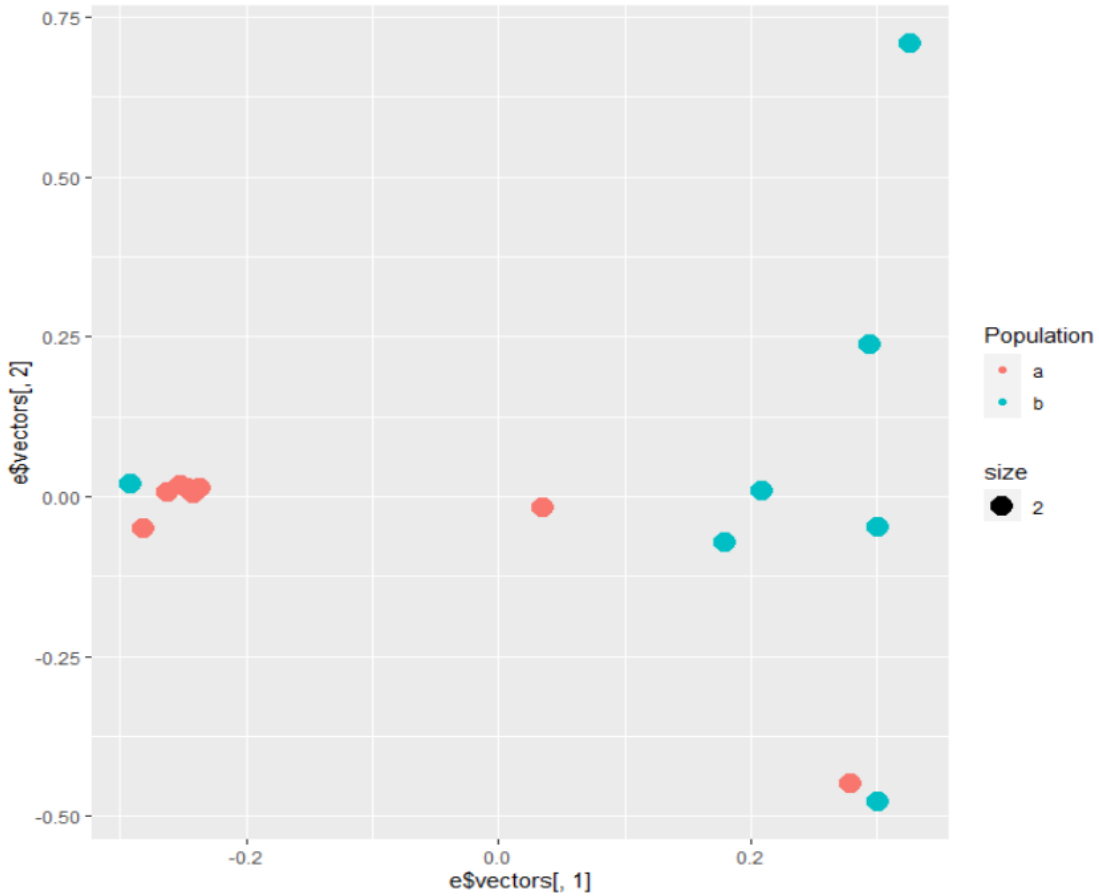


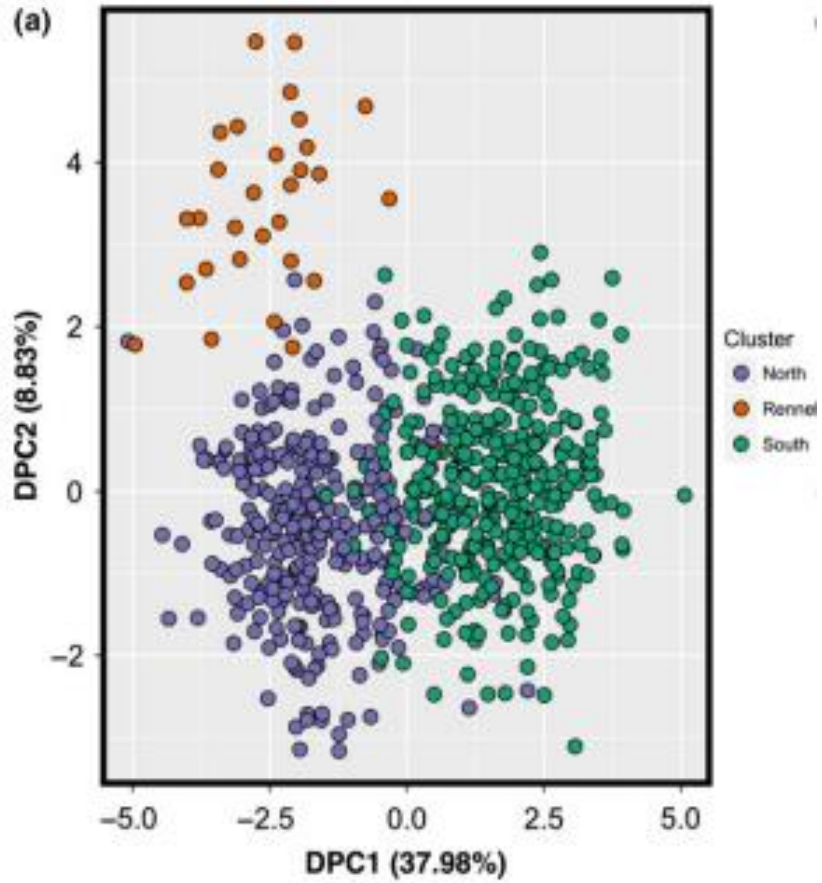
FIGURE 3 (a) Scatterplot showing the genetic clusters identified by DAPC with a priori sampling location information and density plots for (b) the first and (c) the second discriminant axes [Colour figure can be viewed at wileyonlinelibrary.com]

Principal Component Analyses






0.75
0.50
0.25
0.00
-0.25
-0.50
e$vectors[, 2]
-0.2
0.0
0.2
e$vectors[, 1]
Population
a
b
size
2


(a)
4
2
0
−2
DPC2 (8.83%)
−5.0
−2.5
0.0
2.5
5.0
DPC1 (37.98%)
Cluster
North
Rennel
South

Caveats

Our methods differ from those used in this paper, primarily for convenience and time (ours).

What they did prior to making the PCA plot

Many filtering steps

Separated neutral SNPs from “selected” SNPs