PopStructureADE.R

Audrey McCombs

Fri Aug 18 20:23:04 2017

```r
library(adegenet)
```

```
## Warning: package 'adegenet' was built under R version 3.2.5
```

```
## Loading required package: ade4
```

```
## Warning: package 'ade4' was built under R version 3.2.5
```

```
## Warning: replacing previous import by 'expm::expm' when loading 'spdep'
```

```
##
##    /// adegenet 2.0.1 is loaded ////////////
##
##    > overview: '?adegenet'
##    > tutorials/doc/questions: 'adegenetWeb()'
##    > bug reports/feature requests: adegenetIssues()
```

```r
library(pegas)
```

```
## Warning: package 'pegas' was built under R version 3.2.5
```

```
## Loading required package: ape
```

```
## Warning: package 'ape' was built under R version 3.2.5
```

```
##
## Attaching package: 'pegas'
```

```
## The following object is masked from 'package:ape':
##
##     mst
```

```
## The following object is masked from 'package:ade4':
##
##     amova
```

```r
library(hierfstat)
```

```
##
## Attaching package: 'hierfstat'
```

```
## The following objects are masked from 'package:ape':
##
##     pcoa, varcomp
```

```
## The following object is masked from 'package:adegenet':
##
##     read.fstat
```

```r
setwd("D:/Iowa State University/Debinski Lab/Parnassius genetics/ParnassiusGenetics")

### Read in file - uncomment the dataset you want to work with

source("analysis/adegenet_analysis/ReadUnfilImp.R")  #unfiltered-imputed, small file
```

```
## File apparently not yet accessed:
## Scanning file ../ParnassiusGeneticsData/parnassius_clodius_unfiltered_imputed.vcf
##
 6.111896 / 6.111896 Mb
 6.111896 / 6.111896 Mb
## Done.
##
Reading 100 / 1000000 loci
Reading 200 / 1000000 loci
Reading 300 / 1000000 loci
Reading 400 / 1000000 loci
Reading 500 / 1000000 loci
Reading 600 / 1000000 loci
Reading 700 / 1000000 loci
Reading 800 / 1000000 loci
Reading 900 / 1000000 loci
Reading 1000 / 1000000 loci
Reading 1001 / 1001 loci.
## Done.
```

```
#source("analysis/adegenet_analysis/ReadFiltNotimp.r") #filtered-not imputed, medium-sized file
#source("analysis/adegenet_analysis/ReadUnfiltNotimp.r")  #unfiltered, not imputed, large file


### Explore population structure with adegenet

# Find genetic clusters/groups
#grp <- find.clusters(PCdata1, max.n.clusters = 40)  #interactive
grp <- find.clusters(PCdata1, n.pca = 140, n.clust = 3, max.n.clust=40)
names(grp)
```

```
## [1] "Kstat" "stat"  "grp"   "size"
```

```
grp$Kstat
```

```
## NULL
```

```
grp$stat
```

```
## NULL
```

```
grp$grp #group membership for each individual
```

```
## ALM16-001 ALM16-003 ALM16-004 ALM16-005 ALM16-006 ALM16-007 ALM16-008
##         3         3         1         1         2         2         1
## ALM16-009 ALM16-010 ALM16-012 ALM16-014 ALM16-018 ALM16-019 ALM16-020
##         3         3         1         1         2         1         3
## ALM16-021 ALM16-022 ALM16-023 ALM16-024 ALM16-026 ALM16-027 ALM16-029
##         1         2         2         1         3         3         1
## ALM16-030 ALM16-031 ALM16-034 ALM16-035 ALM16-036 ALM16-038 ALM16-043
##         1         1         1         1         2         1         3
## ALM16-044 ALM16-045 ALM16-046 ALM16-048 ALM16-051 ALM16-052 ALM16-053
##         1         1         1         3         1         3         1
## ALM16-056 ALM16-057 ALM16-062 ALM16-069 ALM16-070 ALM16-071 ALM16-074
##         2         1         1         2         2         1         2
## ALM16-075 ALM16-076 ALM16-078 ALM16-079 ALM16-081 ALM16-084 ALM16-088
##         1         1         2         2         1         1         2
## ALM16-089 ALM16-093 ALM16-094 ALM16-095 ALM16-096 ALM16-097 ALM16-098
##         1         2         1         2         1         2         1
## ALM16-099 ALM16-101 ALM16-102 ALM16-104 ALM16-107 ALM16-108 ALM16-111
##         1         1         1         2         2         1         1
## ALM16-112 ALM16-116 ALM16-117 ALM16-118 ALM16-121 ALM16-123 ALM16-124
##         1         1         1         2         1         3         3
## ALM16-125 ALM16-126 ALM16-128 ALM16-129 ALM16-130 ALM16-131 ALM16-134
##         1         1         3         3         1         1         1
## ALM16-135 ALM16-136 ALM16-137 ALM16-138 ALM16-139 ALM16-140 ALM16-141
##         3         2         2         1         3         1         1
## ALM16-142 ALM16-145 ALM16-147 ALM16-148 ALM16-150 ALM16-151 ALM16-152
##         1         1         2         2         2         2         1
## ALM16-155 ALM16-156 ALM16-158 ALM16-159 ALM16-161 ALM16-162 ALM16-163
##         1         2         1         2         1         2         3
## ALM16-164 ALM16-165 ALM16-166 ALM16-167 ALM16-168 ALM16-169 ALM16-170
##         1         2         3         2         2         1         1
## ALM16-171 ALM16-172 ALM16-174 ALM16-175 ALM16-176 ALM16-177 ALM16-180
##         1         1         1         1         1         2         2
## ALM16-181 ALM16-182 ALM16-184 ALM16-185 ALM16-187 ALM16-189 ALM16-191
##         1         2         3         1         2         1         1
## ALM16-192 ALM16-194 ALM16-196 ALM16-198 ALM16-199 ALM16-202 ALM16-204
##         1         2         1         1         1         1         2
## ALM16-205 ALM16-206 ALM16-209 ALM16-210 ALM16-214 ALM16-215 ALM16-217
##         3         2         2         1         2         2         2
## ALM16-218 ALM16-222 ALM16-223 ALM16-224 ALM16-225 ALM16-227 ALM16-229
##         2         1         2         2         1         1         1
## ALM16-230 ALM16-231 ALM16-232 ALM16-237 ALM16-238 ALM16-240
##         1         2         2         1         1         1
## Levels: 1 2 3
```

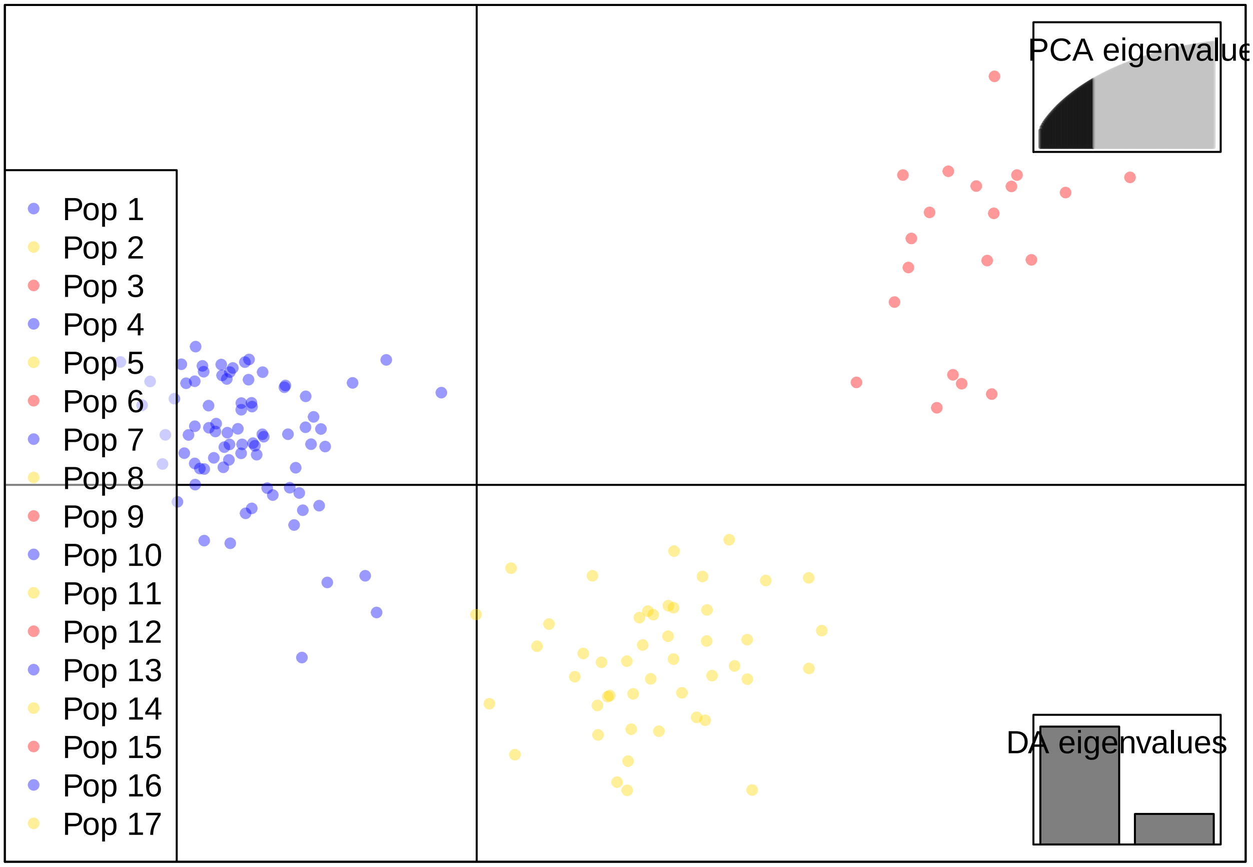
```
grp$size
```

```
## [1] 79 47 20
```

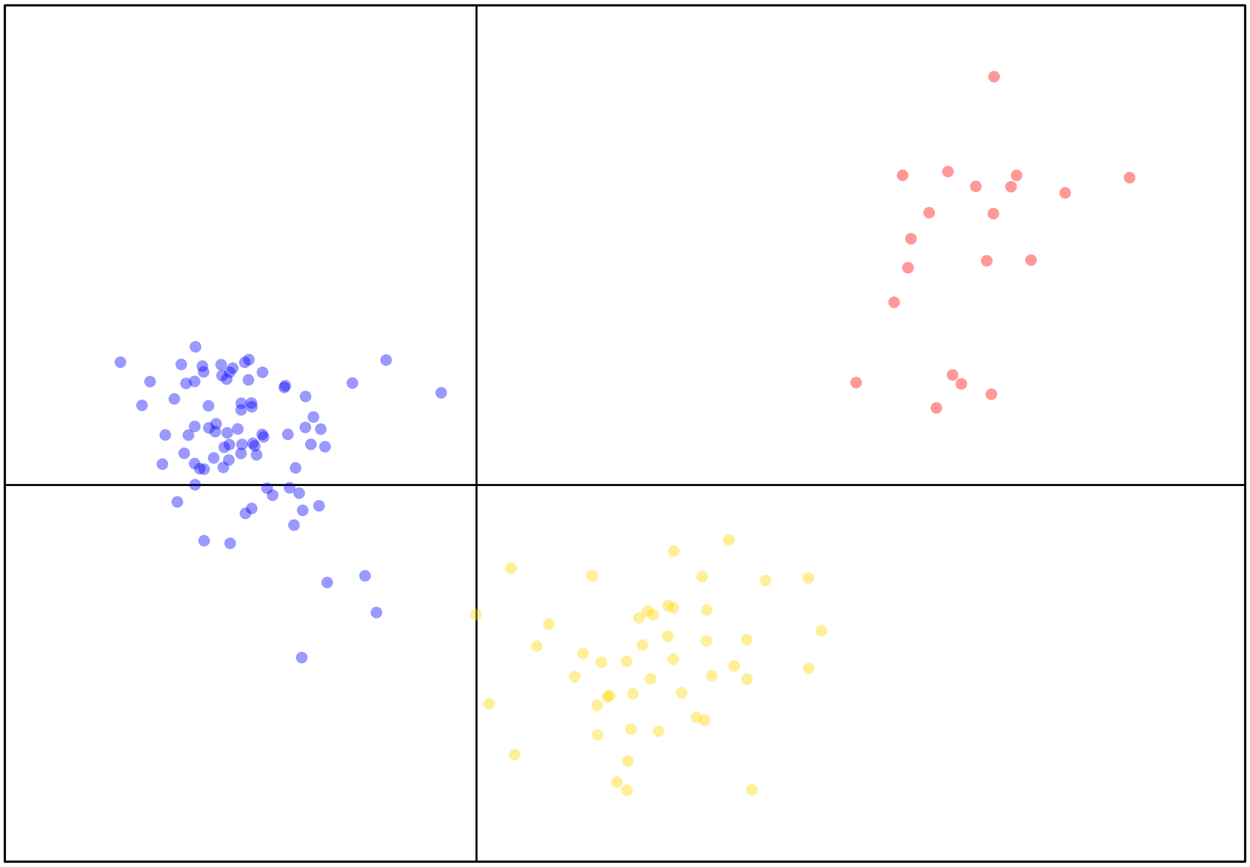
```
# DAPC of data with existing populations
#dapc.pop <- dapc(PCdata1, grp$grp) #interactive
dapc.pop <- dapc(PCdata1, grp$grp,  n.pca = 45, n.da = 4)
dapc.pop
```

```
##  #################################################
##  # Discriminant Analysis of Principal Components #
##  #################################################
## class: dapc
## $call: dapc.genind(x = PCdata1, pop = grp$grp, n.pca = 45, n.da = 4)
##
## $n.pca: 45 first PCs of PCA used
## $n.da: 2 discriminant functions saved
## $var (proportion of conserved variance): 0.644
##
## $eig (eigenvalues): 1155 299.3  vector    length content
## 1 $eig      2      eigenvalues
## 2 $grp      146    prior group assignment
## 3 $prior    3      prior group probabilities
## 4 $assign   146    posterior group assignment
## 5 $pca.cent 2002   centring vector of PCA
## 6 $pca.norm 2002   scaling vector of PCA
## 7 $pca.eig  145    eigenvalues of PCA
##
##   data.frame    nrow ncol
## 1 $tab          146  45
## 2 $means        3    45
## 3 $loadings     45   2
## 4 $ind.coord    146  2
## 5 $grp.coord    3    2
## 6 $posterior    146  3
## 7 $pca.loadings 2002 45
## 8 $var.contr    2002 2
##   content
## 1 retained PCs of PCA
## 2 group means
## 3 loadings of variables
## 4 coordinates of individuals (principal components)
## 5 coordinates of groups
## 6 posterior membership probabilities
## 7 PCA loadings of original variables
## 8 contribution of original variables
```

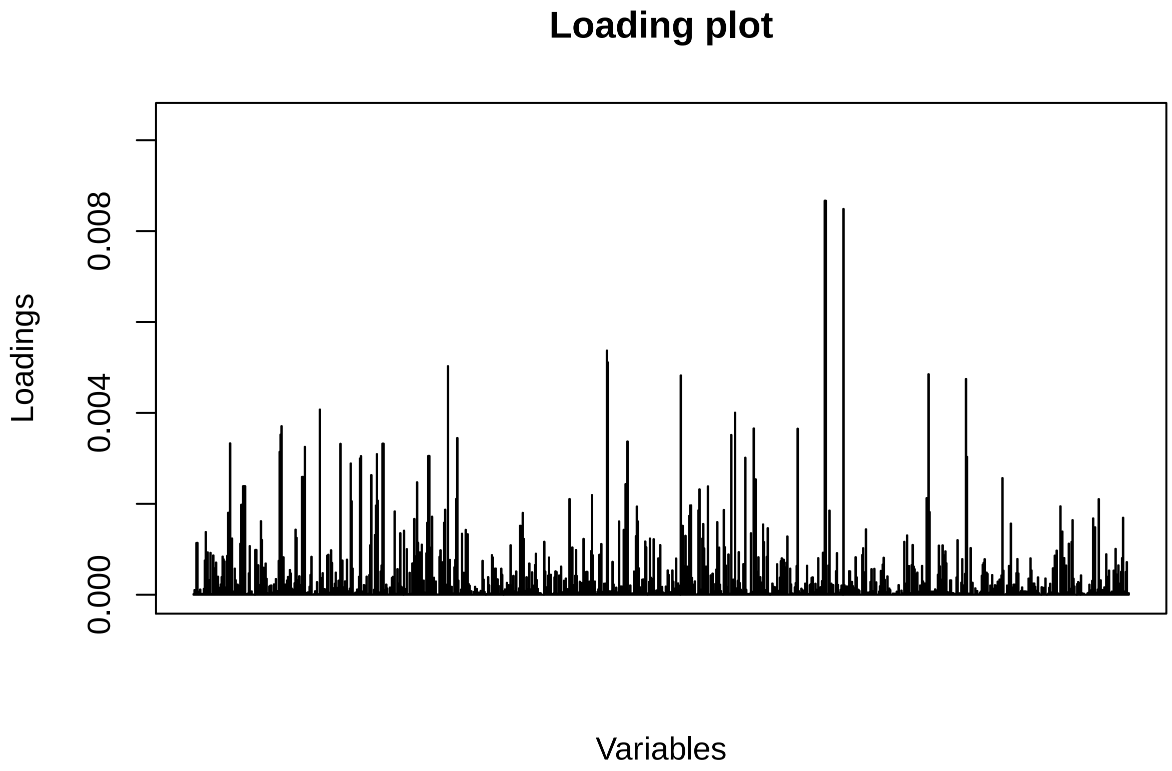
```
# Plots
#scatter(dapc.pop,1,1,bg="white", scree.da=FALSE, legend=TRUE, solid=.4)

scatter(dapc.pop, xax = 1, yax = 2, bg="white", pch=20, cell=0, cstar=0, solid=.4, cex=1, clab=0, scree
```



```
scatter(dapc.pop, xax = 1, yax = 2, bg="white", pch=20, cell=0, cstar=0, solid=.4, cex=1, clab=0,scree.
```



```
# Loadings (i.e., variable contributions)
set.seed(4)
contrib.pop <- loadingplot(dapc.pop$var.contr, axis=2, thres=.07, lab.jitter=1)
```



```
# Probabilities for assignment of individuals into groups
round(head(dapc.pop$posterior),3)
```

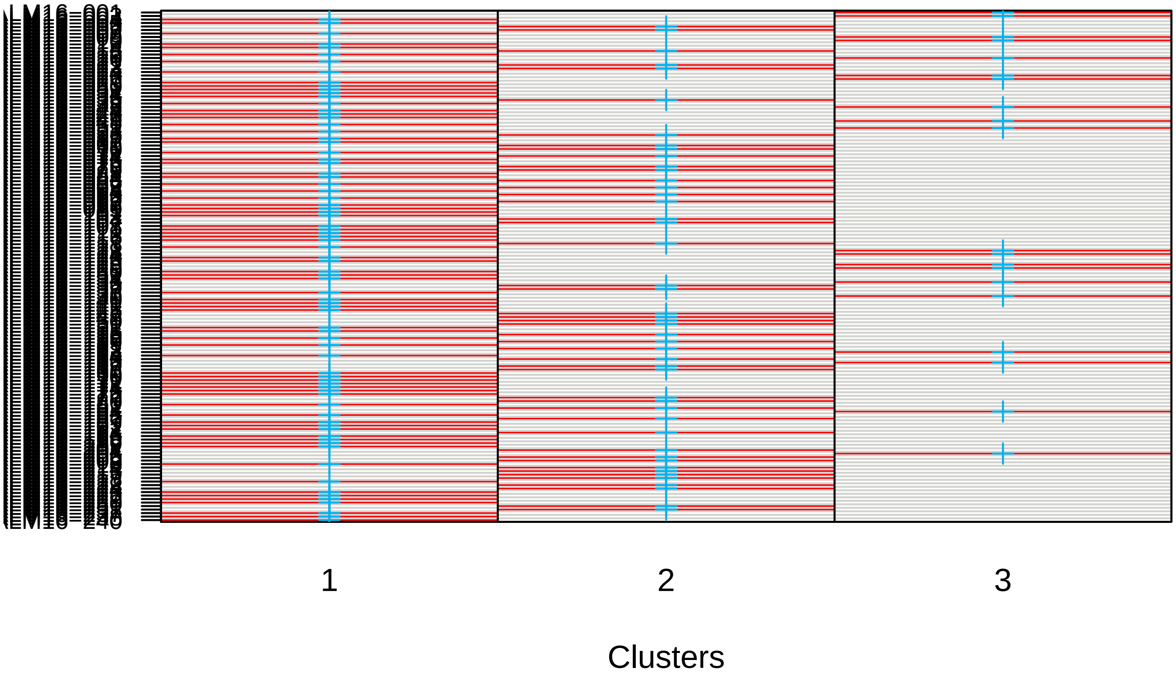
```
##           1 2 3
## ALM16-001 0 0 1
## ALM16-003 0 0 1
## ALM16-004 1 0 0
## ALM16-005 1 0 0
## ALM16-006 0 1 0
## ALM16-007 0 1 0
```

```
#round(dapc.pop$posterior,3)
summary(dapc.pop)  #$assign.per.pop is probabilities of correct assignment into each group
```

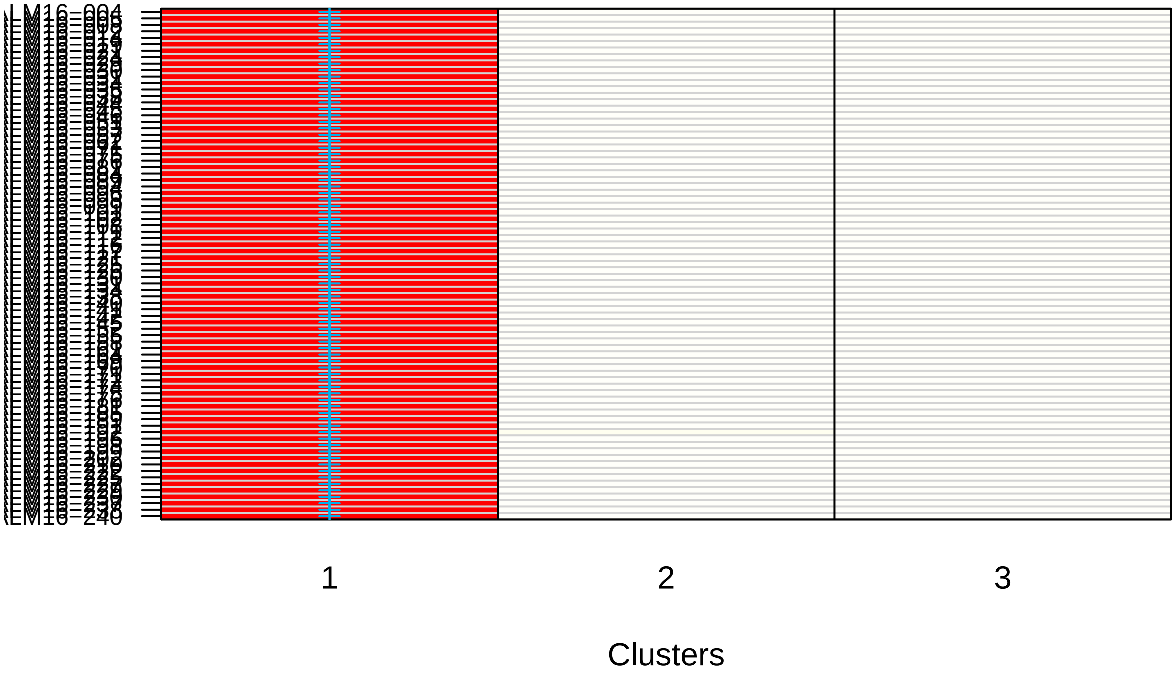
```
## $n.dim
## [1] 2
##
## $n.pop
## [1] 3
##
## $assign.prop
## [1] 1
##
## $assign.per.pop
## 1 2 3
## 1 1 1
##
```

```
## $prior.grp.size
##
##  1  2  3
## 79 47 20
##
## $post.grp.size
##
##  1  2  3
## 79 47 20
```

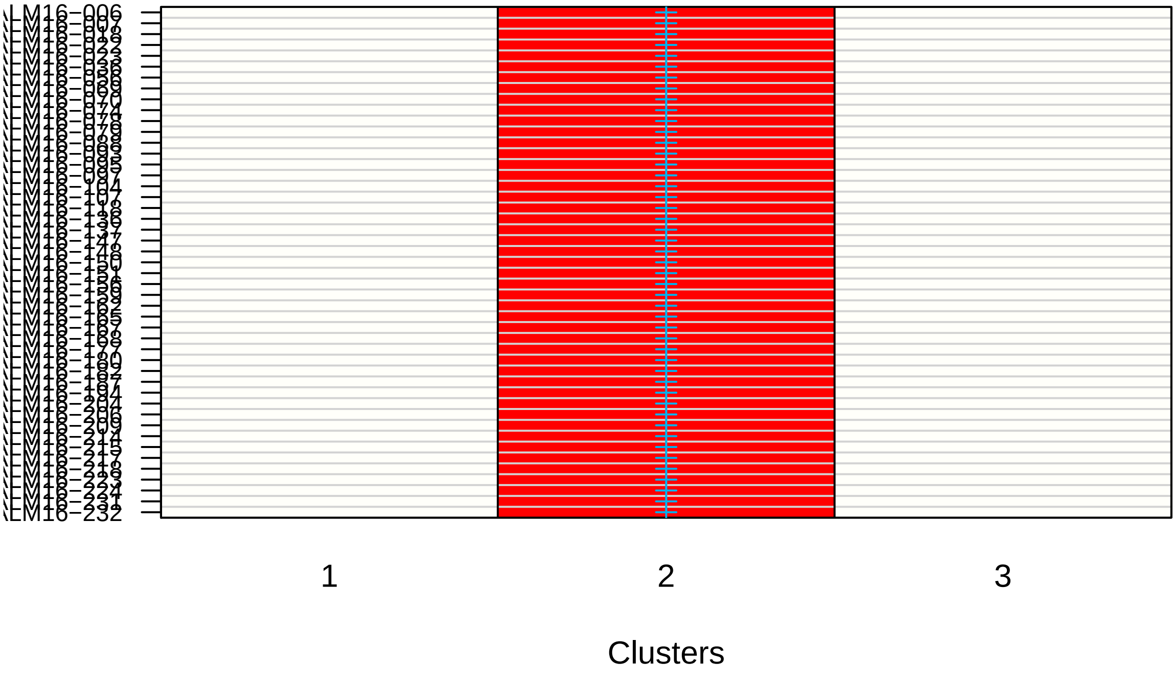
```
# plot probabilities
assignplot(dapc.pop)
```



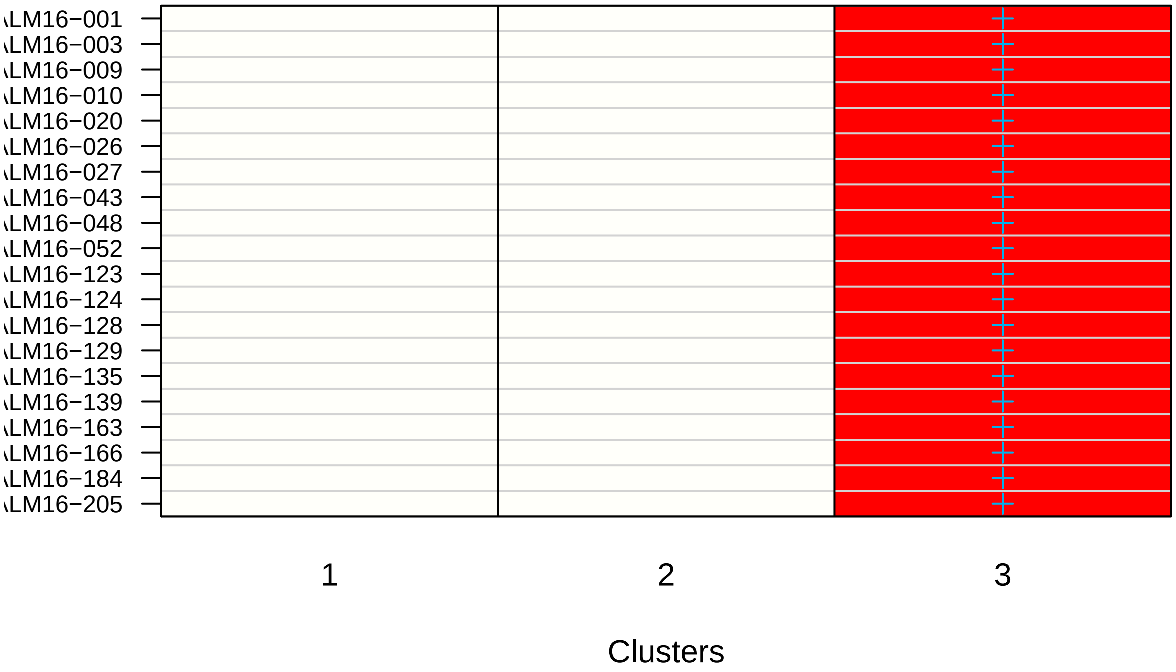
```
assignplot(dapc.pop, only.grp = 1)
```



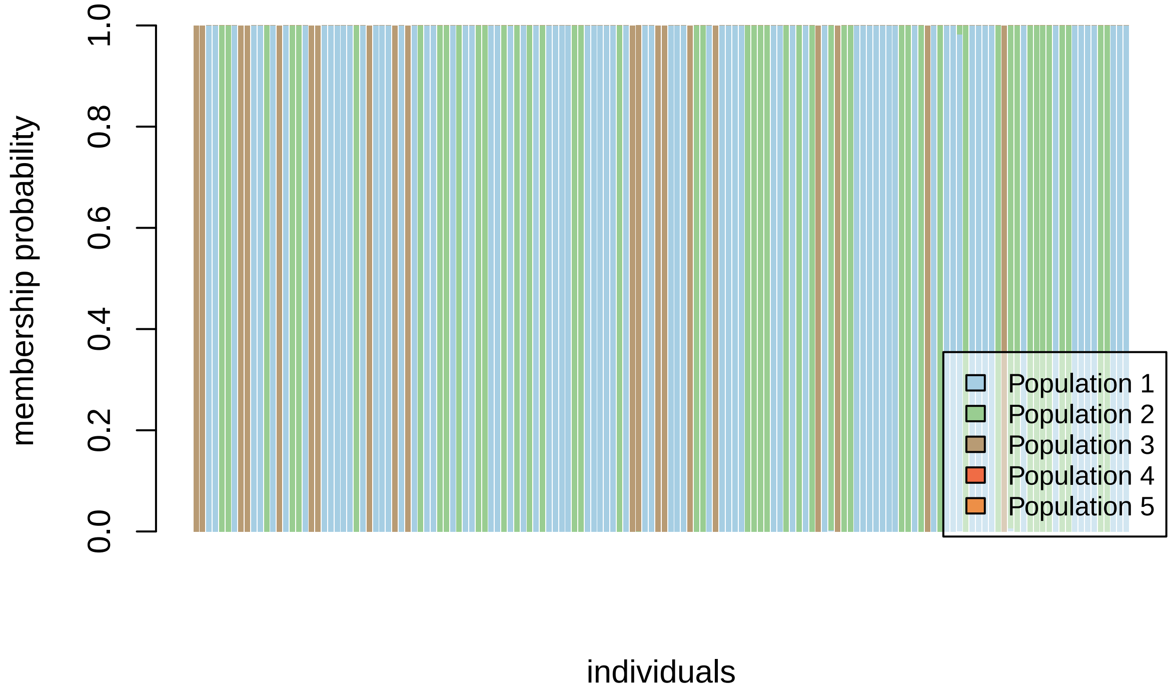
```r
assignplot(dapc.pop, only.grp = 2)
```



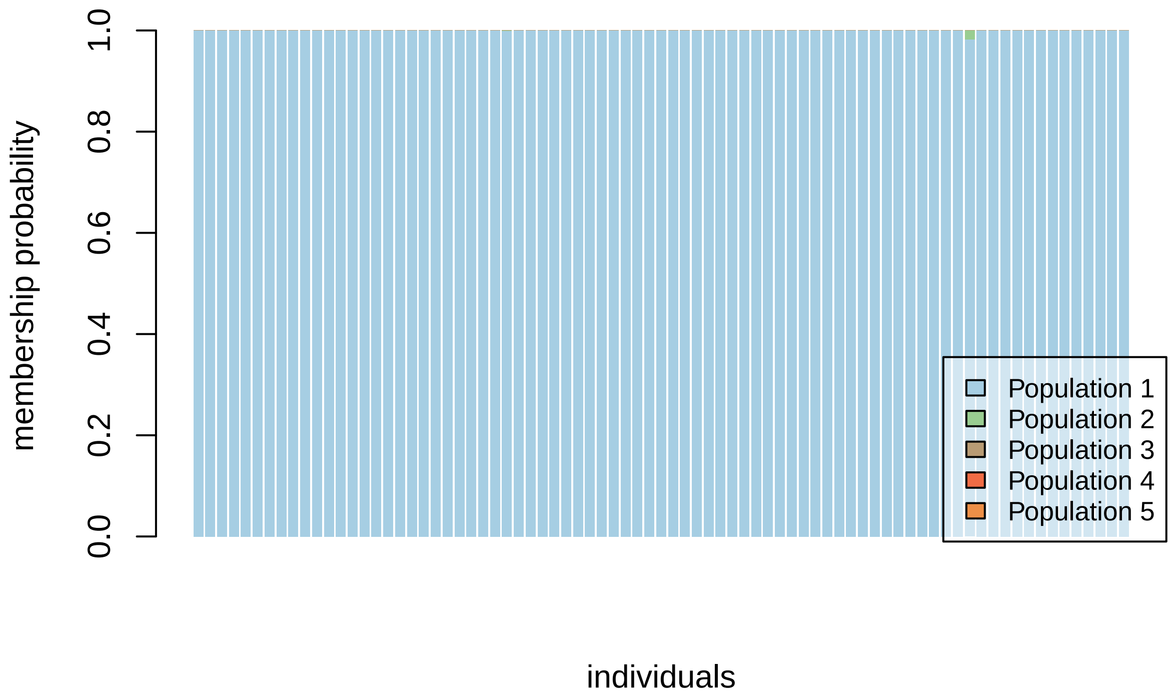
```r
assignplot(dapc.pop, only.grp = 3)
```



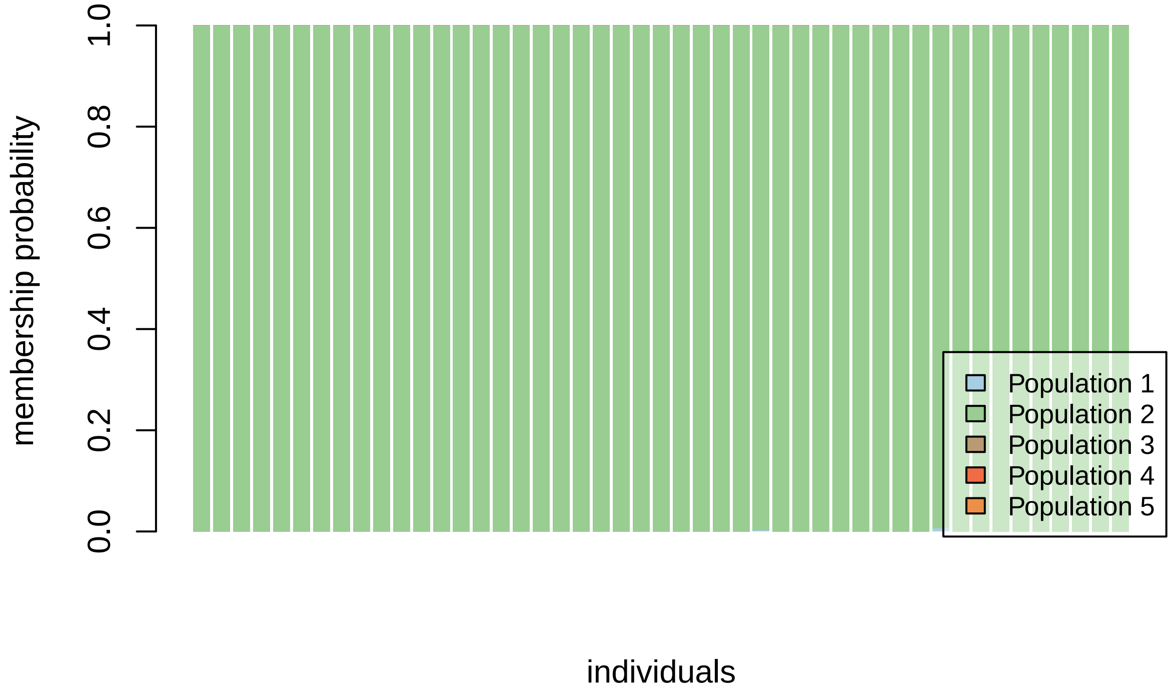
```r
#assignplot(dapc.pop, only.grp = 4)
#assignplot(dapc.pop, only.grp = 5)

compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="individu
```



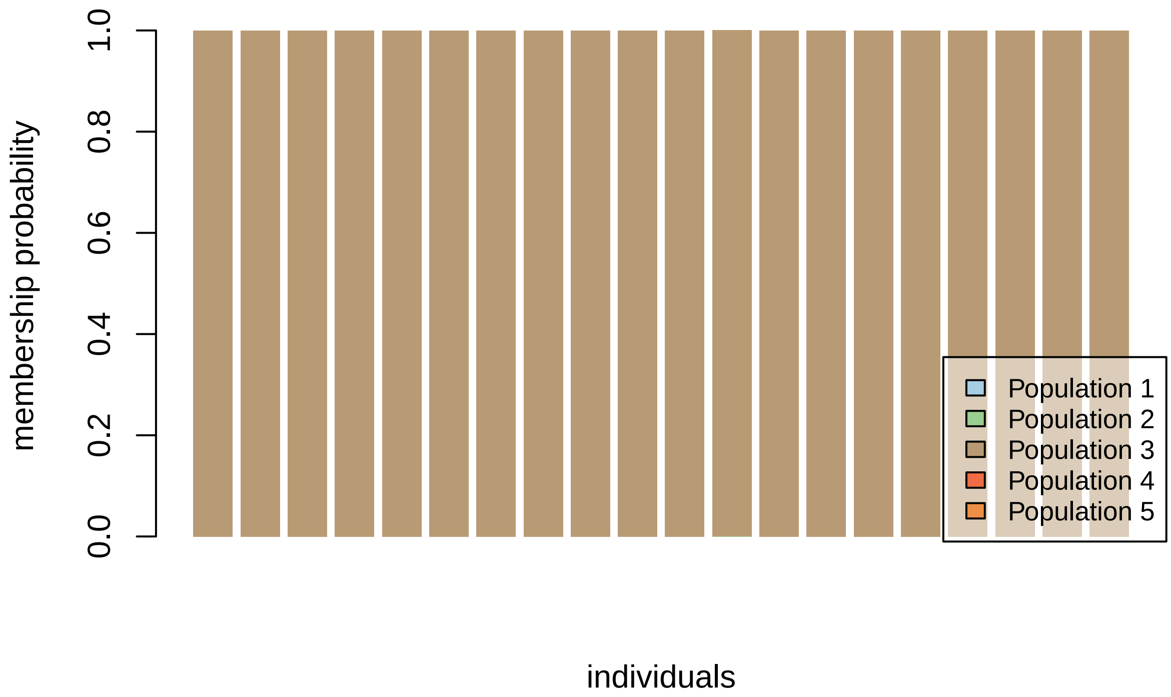
```
compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="individu
```



```
compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="individu
```



```
compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="individu
```



```r
#compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="indivi
#compoplot(dapc.pop, posi="bottomright", txt.leg=paste("Population", 1:5), lab="", ncol=1, xlab="indivi

# Prior vs post memberships
table(pop(PCdata1), grp$grp)
```

```
##
##                               1 2 3
##   Aimees Meadow               5 1 1
##   AMK Ranch                   4 1 1
##   AMK Road                    4 2 1
##   Bear paw lake intersection 1 0 0
##   Bear paw lake trail         5 2 0
##   Buffalo fork                2 0 0
##   Climbers ranch              2 1 3
##   Cygnet Pond                 2 1 2
##   Death canyon phelps lake    4 0 0
##   Death canyon Rangers        5 2 0
##   Death Canyon Trail          0 0 0
##   Dump Road                   0 0 0
##   Grand view 1                0 0 0
##   Grand view 2                1 1 1
##   Grand view parking          5 0 1
##   Hidden Falls Trail          0 0 0
##   Lozier Hill Meadow          2 2 2
##   Lozier Road                 1 0 0
##   Lupine Meadow              0 0 0
##   Mt. Moran turnout          1 3 0
##   North Jenny lake           3 1 0
##   Paintbrush canyon trail    2 3 0
##   Pilgrim Creek              2 8 1
##   Shadow mountain            5 0 0
##   Sound of music             2 3 1
##   String lake parking        1 3 0
##   Surprise 1                 4 1 1
##   Surprise 2                 3 1 1
##   Surprise fall down         0 0 0
##   Taggart Lake trailhead     1 3 1
##   Timbered Island            2 2 0
##   Two Ocean Road 1           0 0 0
##   Two Ocean Road 2           3 0 0
##   Wilderness Rd 1            4 2 0
##   Wilderness Rd 2            2 2 3
##   Wilderness Rd 3            1 2 0
```