

# ANT\_Preliminary\_Look

Clare I M Adams

2022-09-21

## R Markdown: Blue antimora

This is a *preliminary* analysis of some Blue antimora (*Antimora rostrata*) data (ANT).

You gotta read in all the info about the stuff

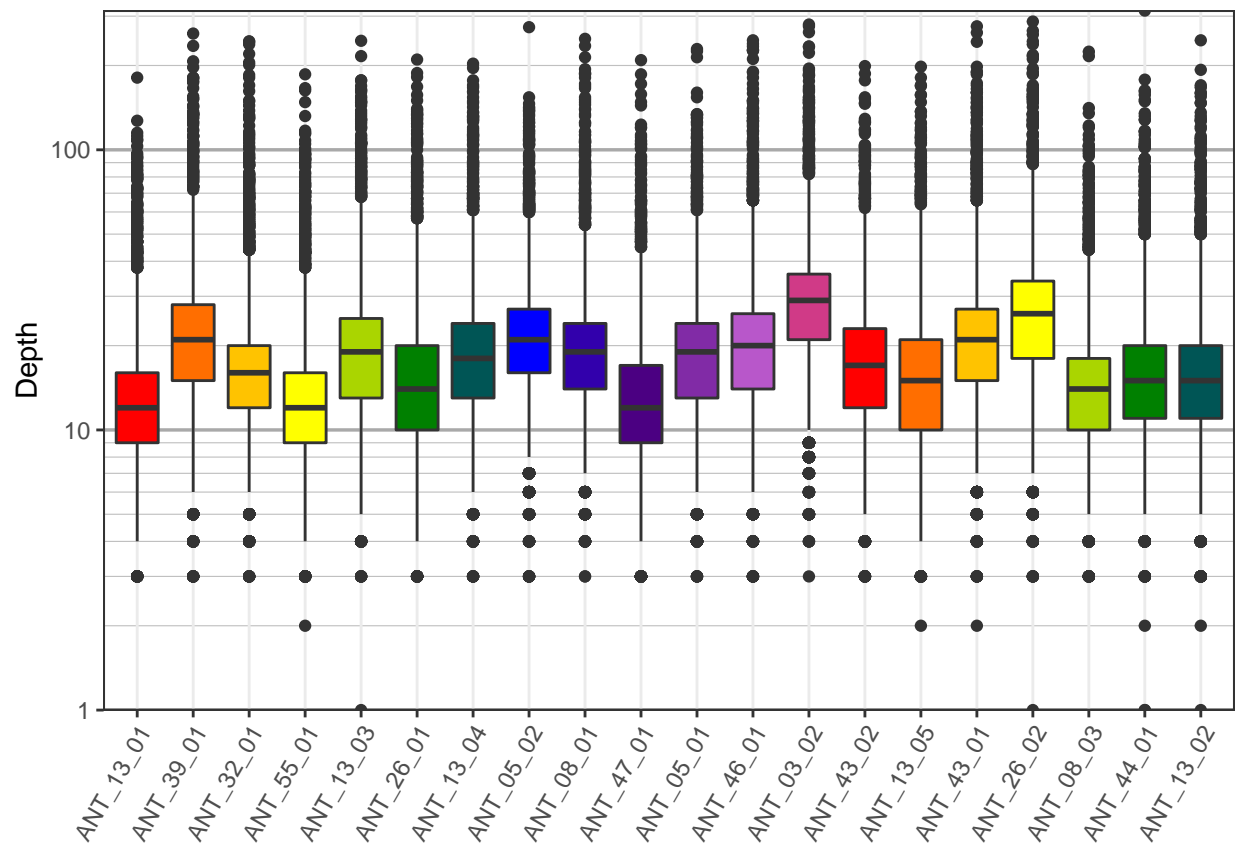
```
## Scanning file to determine attributes.
## File attributes:
##   meta lines: 14
##   header_line: 15
##   variant count: 6525
##   column count: 29
## Meta line 14 read in.
## All meta lines processed.
## gt matrix initialized.
## Character matrix gt created.
##   Character matrix gt rows: 6525
##   Character matrix gt cols: 29
##   skip: 0
##   nrows: 6525
##   row_num: 0
## Processed variant 1000Processed variant 2000Processed variant 3000Processed variant 4000Processed va
## All variants processed
```

```
##      CHROM POS   ID      REF ALT QUAL FILTER
## [1,] "5"    "36" "5:35"  "C" "T" NA   "PASS"
## [2,] "8"    "16" "8:15"  "G" "T" NA   "PASS"
## [3,] "19"   "56" "19:55" "A" "G" NA   "PASS"
## [4,] "20"   "36" "20:35" "G" "A" NA   "PASS"
## [5,] "21"   "48" "21:47" "C" "A" NA   "PASS"
## [6,] "22"   "44" "22:43" "A" "C" NA   "PASS"
```

## Depth Information of SNPs

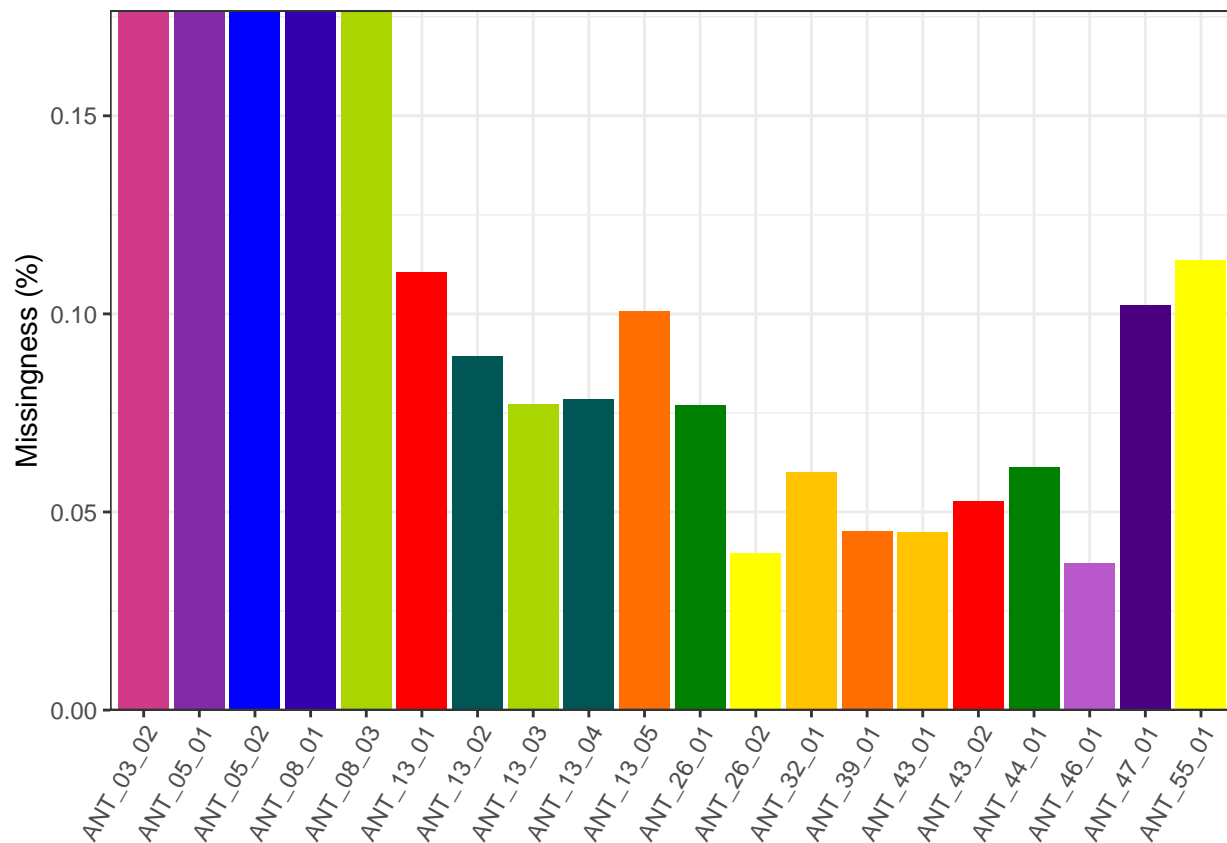
SNP depth info per sample

```
##      ANT_13_01 ANT_39_01 ANT_32_01 ANT_55_01 ANT_13_03 ANT_26_01
## 5:35          7         NA         15          9         14          5
## 8:15         58         72         88         16         84         53
## 19:55         14         33         22          9         28         16
## 20:35         11         32         24         13         21         18
```



## Missingness and relatedness per sample

How many SNP sites are missing per sample; a peek at the first little bit of the dataframe and then a graph per sample

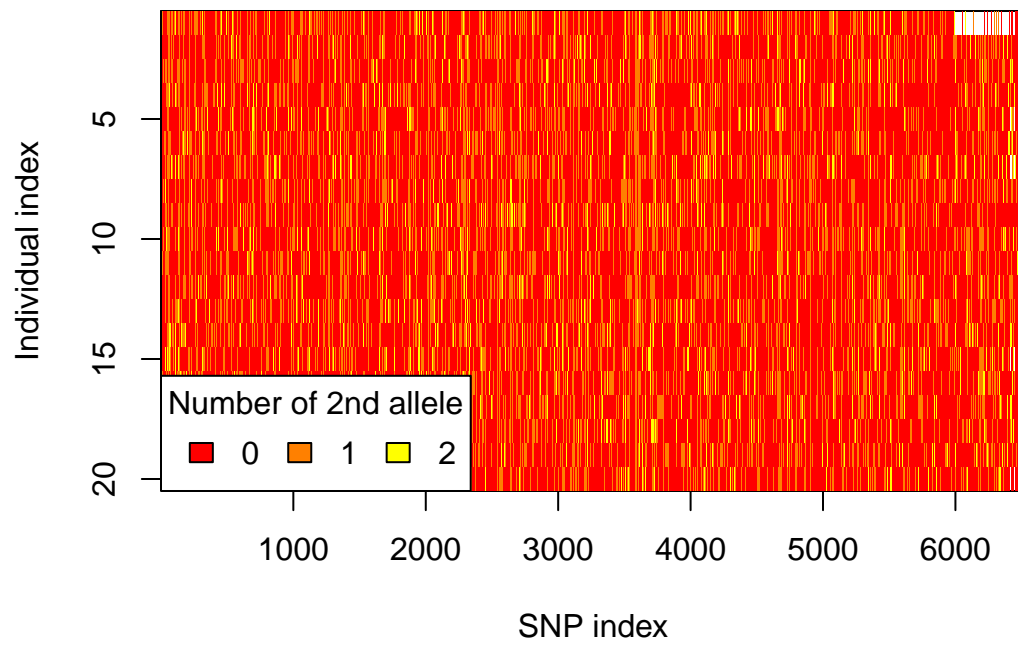


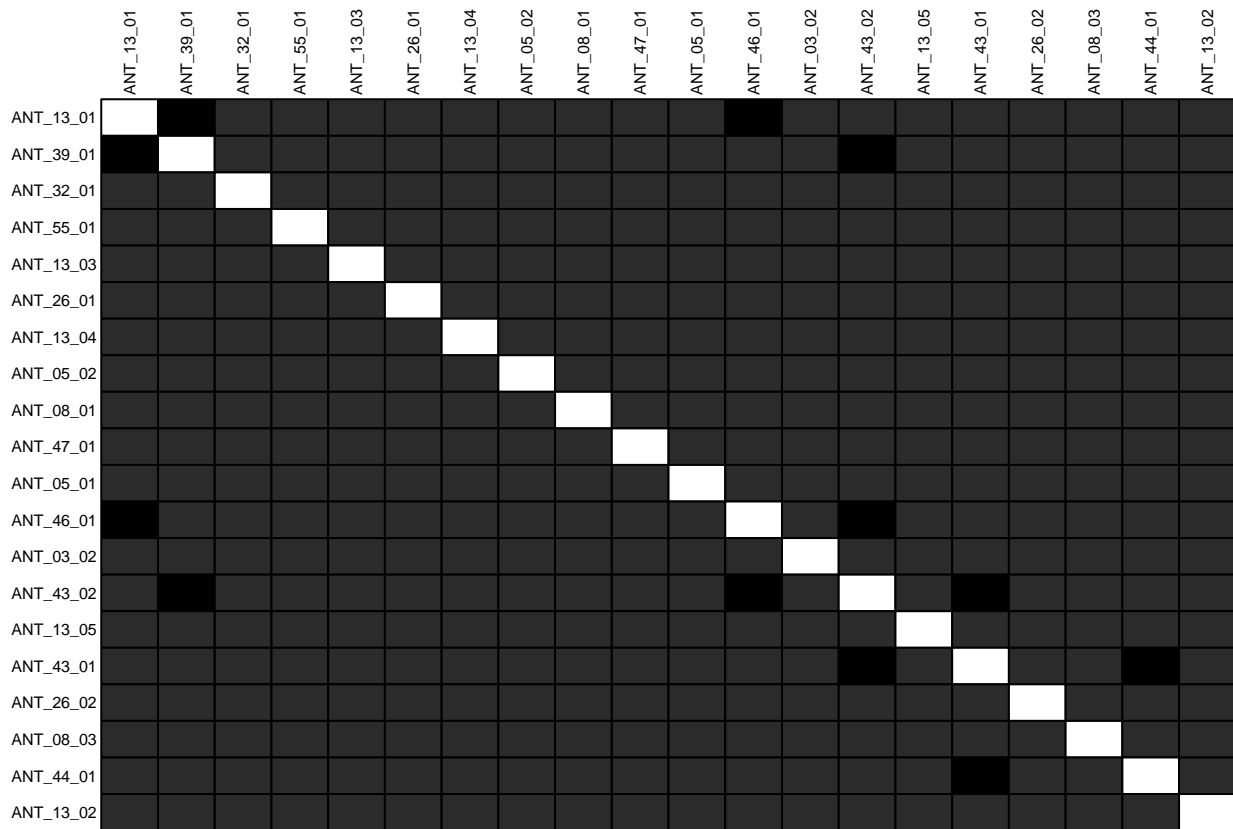
## GenLight/GenInd objects for use

Create some GenLight objects for manipulation; overall looks okay in terms of mostly present alleles. Darker colors in second plot means greater differences between the individuals.

```
## [1] "ANT_13_01" "ANT_39_01" "ANT_32_01" "ANT_55_01" "ANT_13_03" "ANT_26_01"
## [7] "ANT_13_04" "ANT_05_02" "ANT_08_01" "ANT_47_01" "ANT_05_01" "ANT_46_01"
## [13] "ANT_03_02" "ANT_43_02" "ANT_13_05" "ANT_43_01" "ANT_26_02" "ANT_08_03"
## [19] "ANT_44_01" "ANT_13_02"
```

```
## [1] 3 2 2 2 3 2 3 4 4 2 4 2 4 2 3 2 2 4 2 3
## Levels: 2 3 4
```

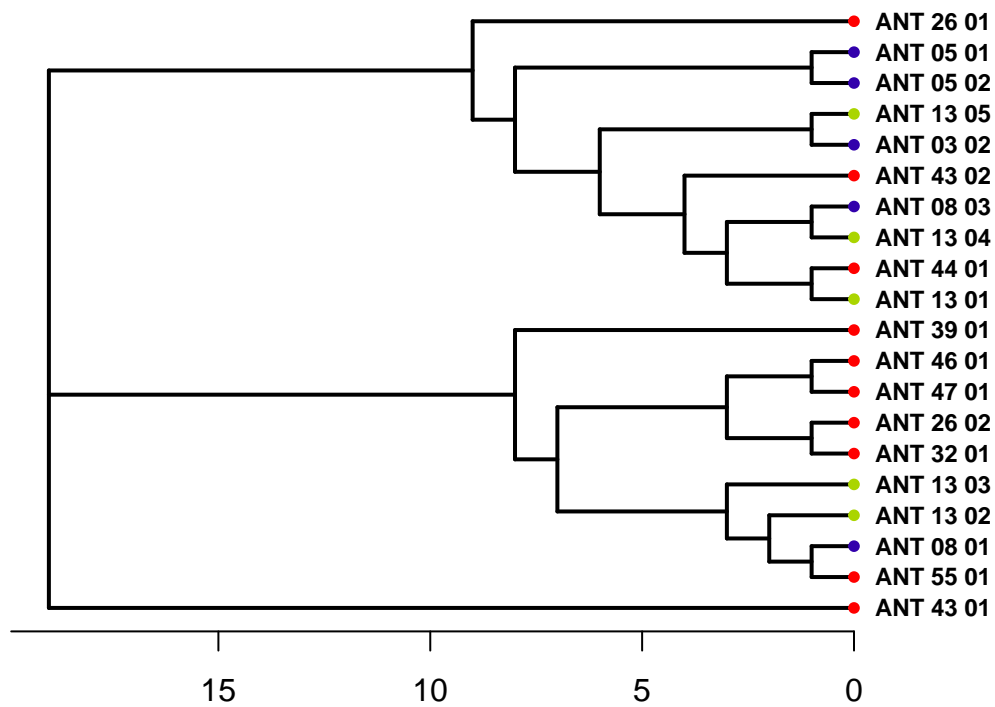




## Neighbor Joining Tree

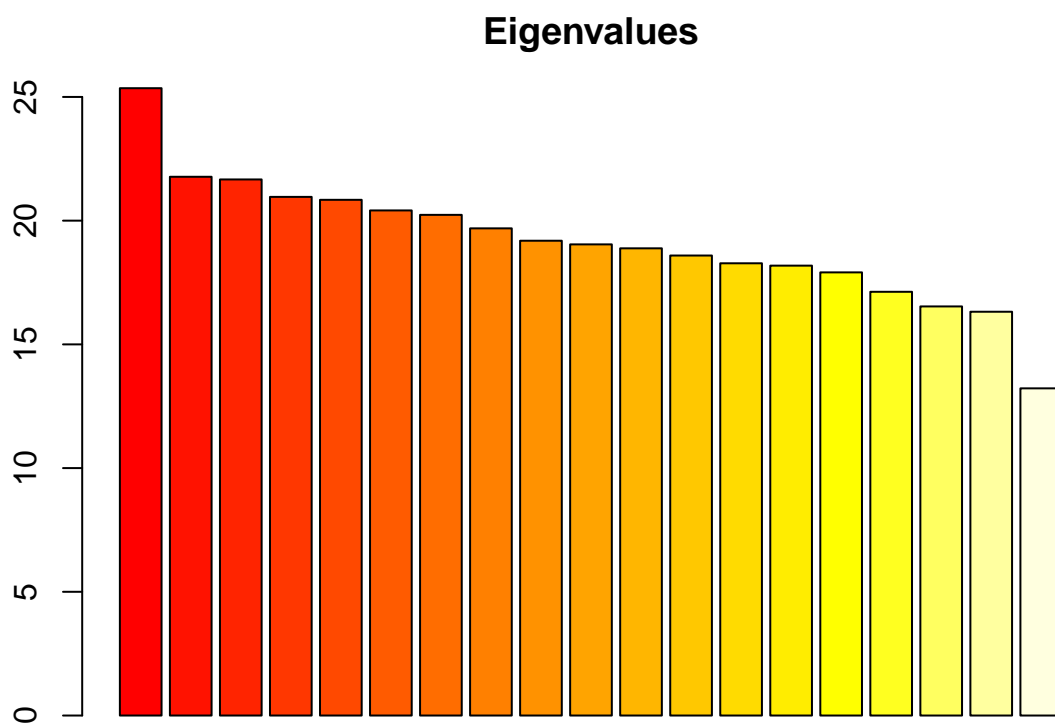
Gotta show how these fish could be related. There are three populations we're drawing from although it looks like a somewhat mixed population.

### Neighbour-joining tree of filtered ANT data

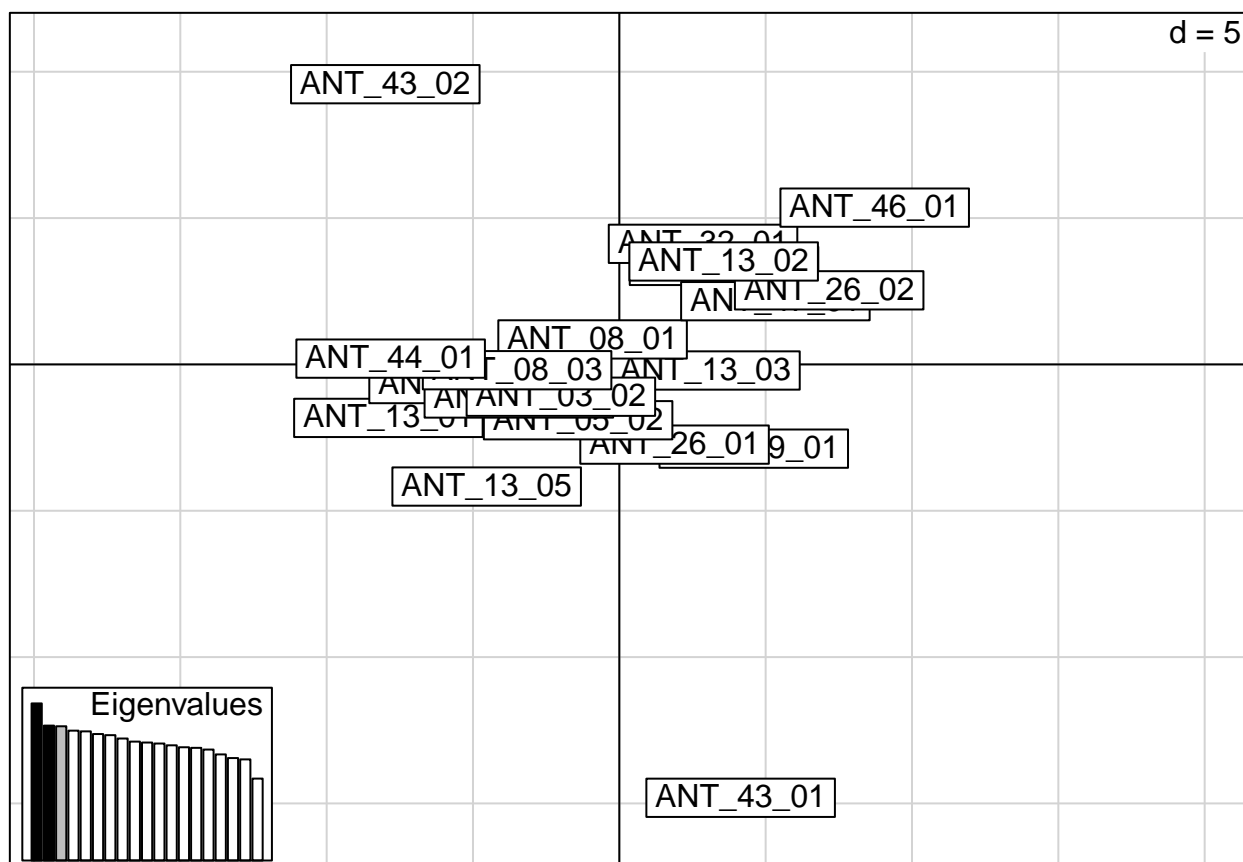


### Some PCA plots

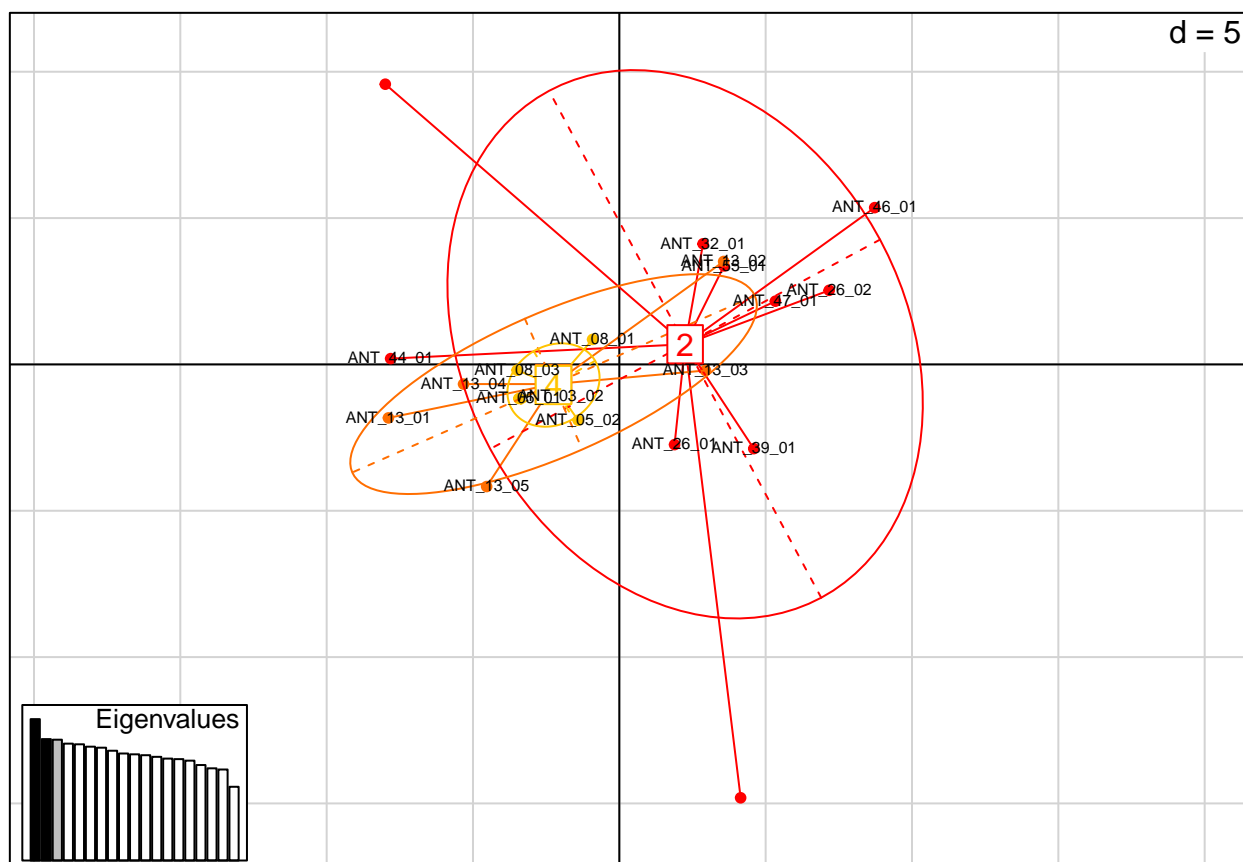
PCA plots are OK for visualization; three axes were chosen but the eigenvalue plot is shown, followed by percentages of what the top axes contributions are. Note that the fourth population sometimes appears within the third population.

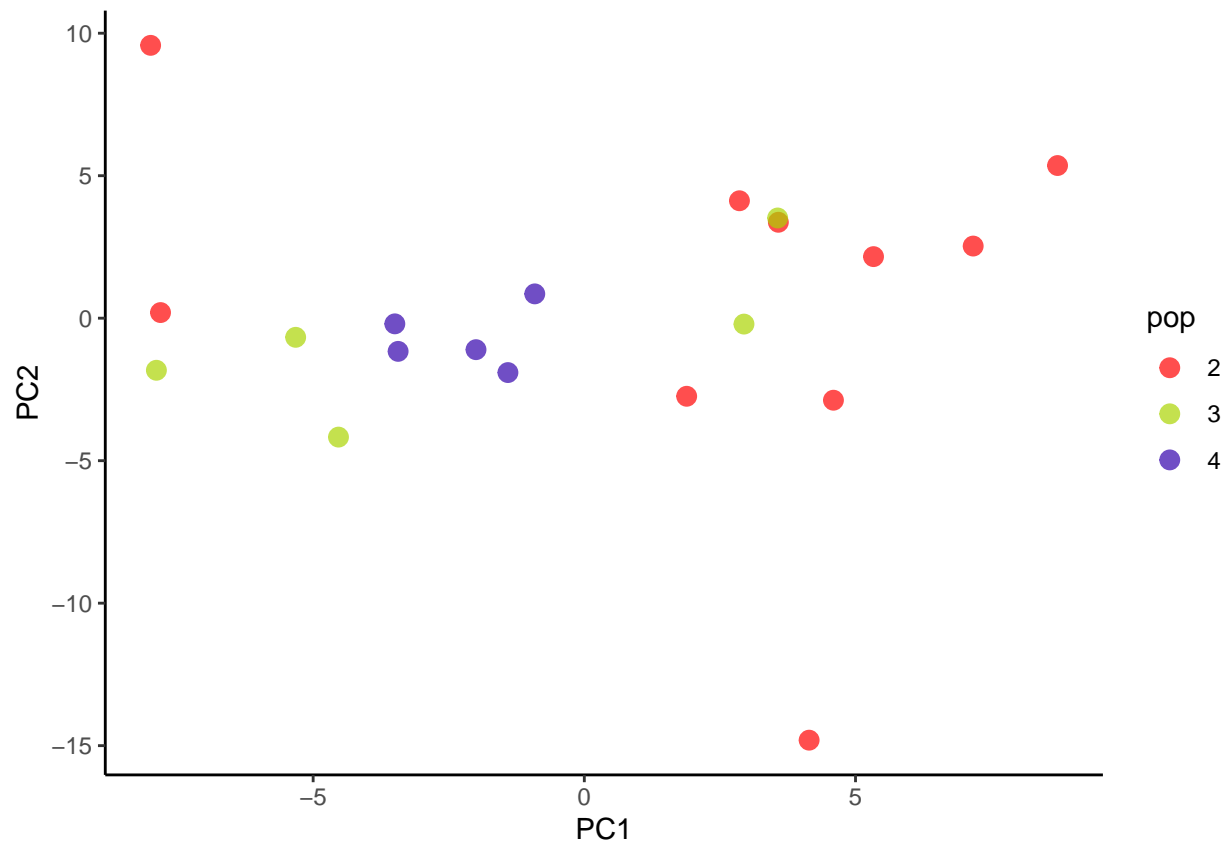


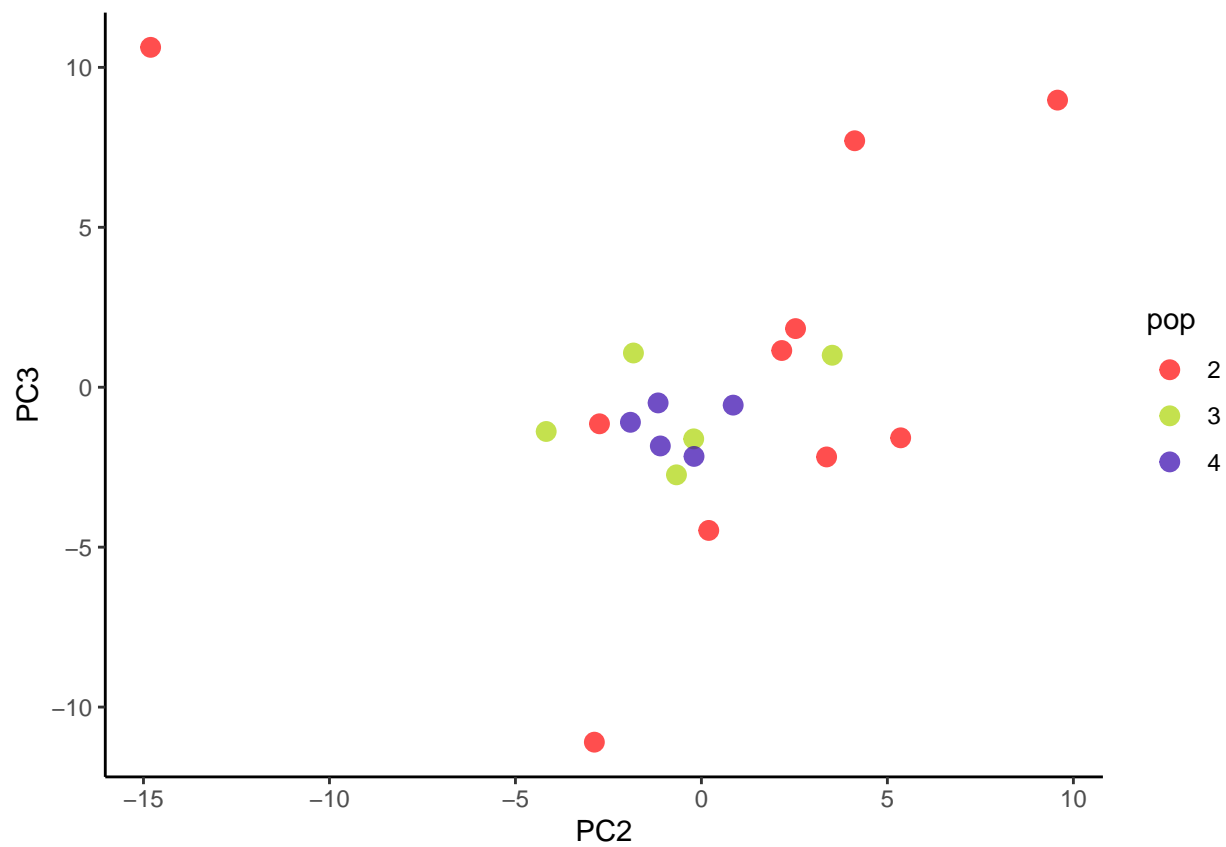
```
## [1] 6.961287 5.978242 5.948675 5.755175 5.722638 5.604775
```

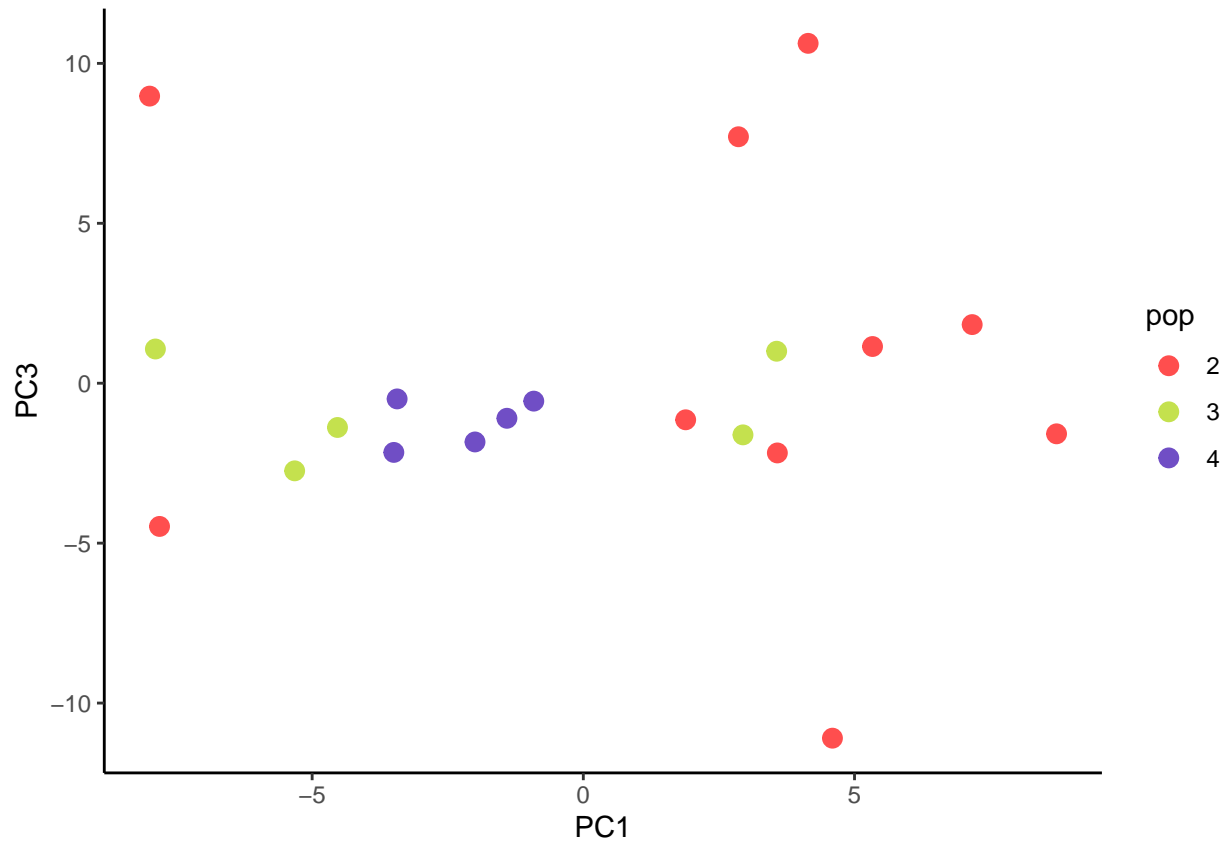








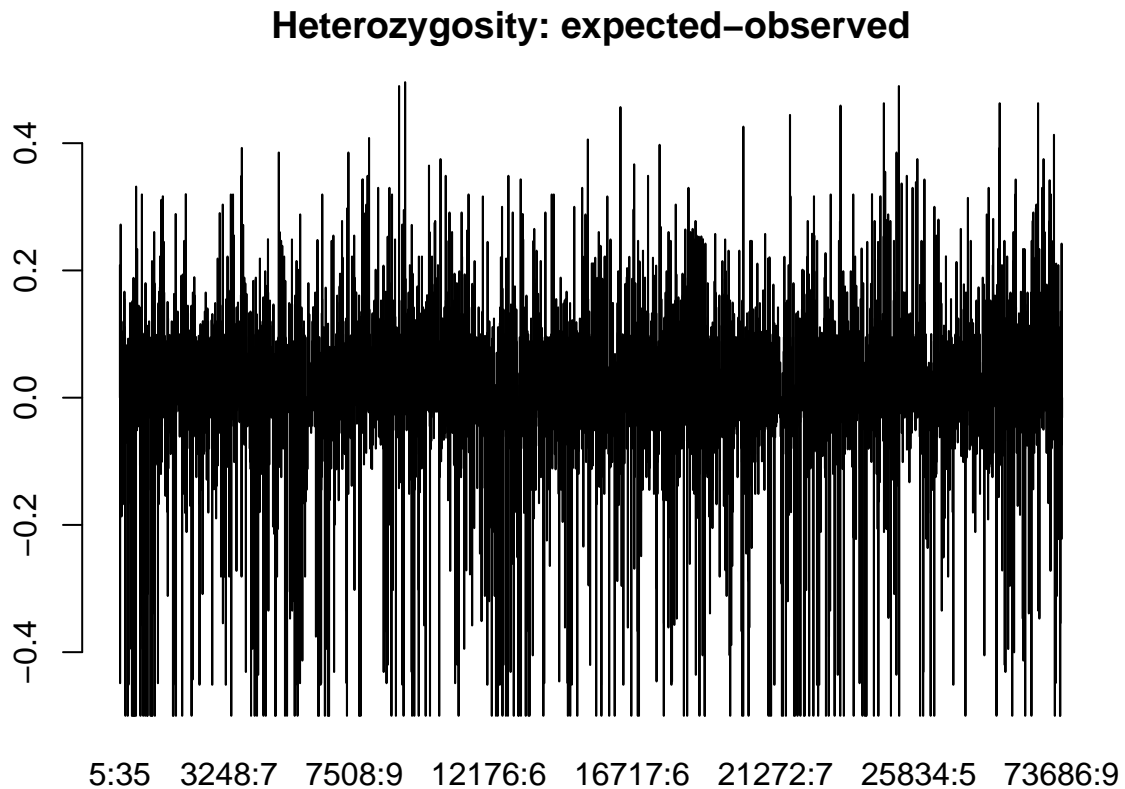




## Some other statistics

Using a GenInd object, we can take a look at heterozygosity of Blue antimora (although I'm not sure I fully understand this plot).

```
## /// GENIND OBJECT ///////////
##
## // 20 individuals; 6,525 loci; 13,050 alleles; size: 4.4 Mb
##
## // Basic content
##   @tab: 20 x 13050 matrix of allele counts
##   @loc.n.all: number of alleles per locus (range: 2-2)
##   @loc.fac: locus factor for the 13050 columns of @tab
##   @all.names: list of allele names for each locus
##   @ploidy: ploidy of each individual (range: 2-2)
##   @type: codon
##   @call: adegenet::df2genind(X = t(x), sep = sep)
##
## // Optional content
##   @pop: population of each individual (group size range: 5-10)
```

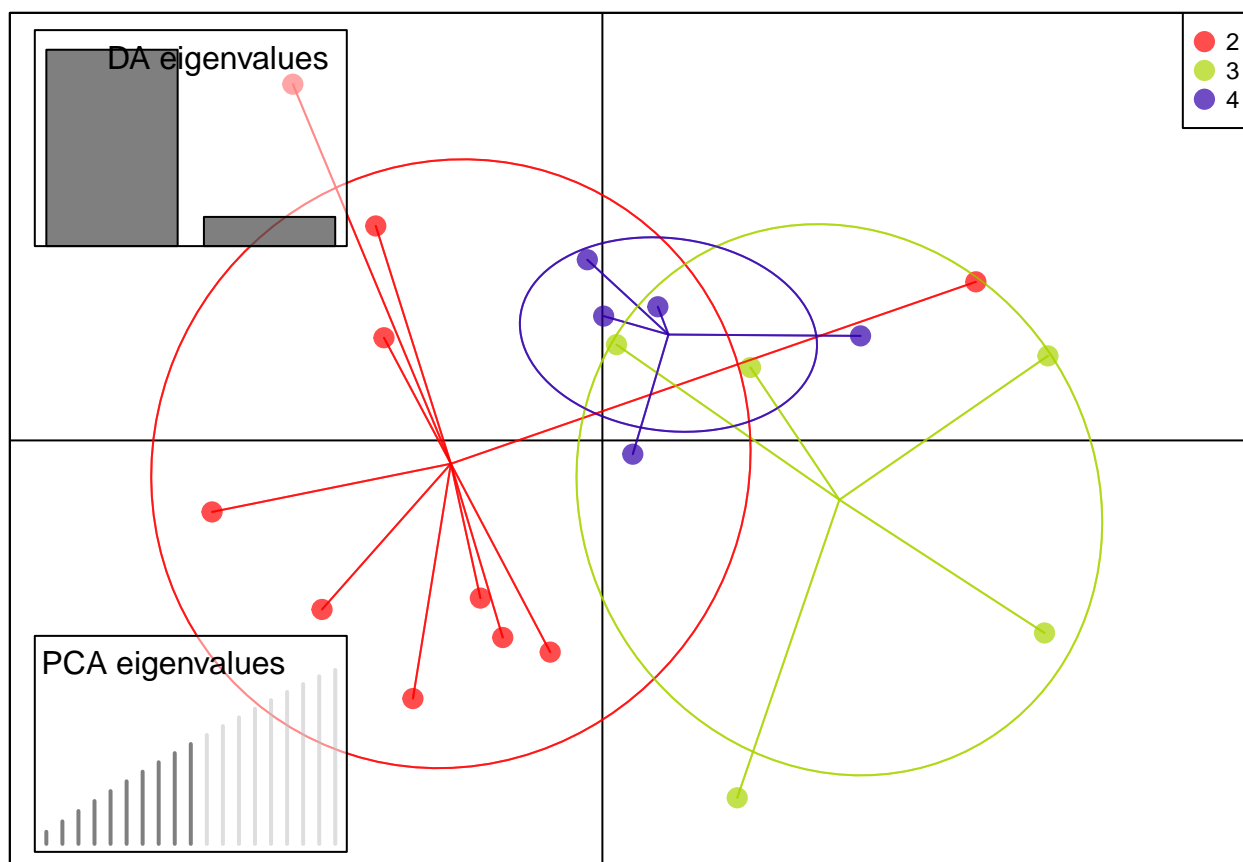


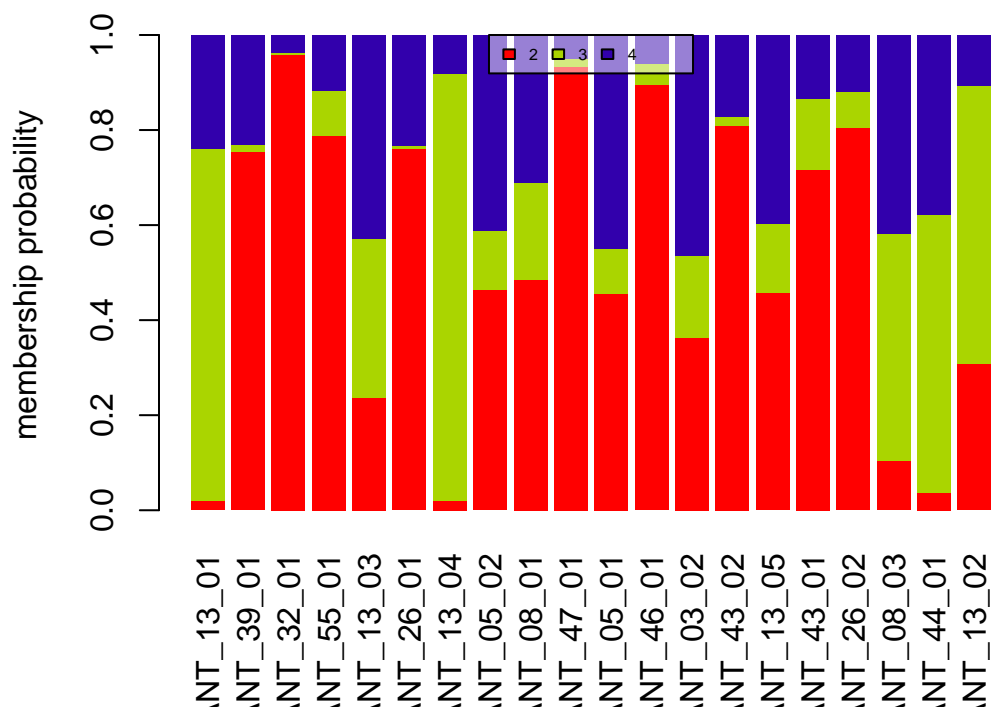
```
##
## Bartlett test of homogeneity of variances
##
## data: list(toto$Hexp, toto$Hobs)
## Bartlett's K-squared = 932.63, df = 1, p-value < 2.2e-16

##
## Paired t-test
##
## data: toto$Hexp and toto$Hobs
## t = 0.84323, df = 6524, p-value = 0.1996
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
## -0.00127238      Inf
## sample estimates:
## mean difference
## 0.001338023
```

## DAPC

I don't trust this DAPC but it looks interesting; running on 10 PCA axes and 2 DA axes





## Using pop, indNames as id variables

