

Workshop 3: *Aspergillus*/mold

Population genomic analysis, recombination analysis

Dr Johanna Rhodes

Learning outcomes

- **In practice**
 - Command line arguments
 - Troubleshooting errors

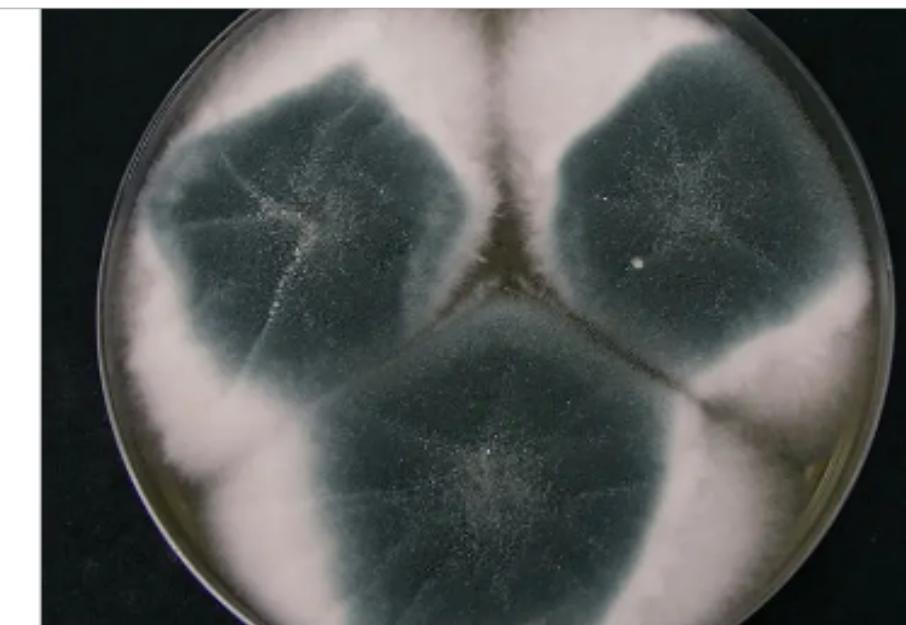
Aspergillus

Diverse genus

- Aspergillus is as diverse as our own phylum, the Vertebrates
 - the ‘very close’ relatives *A. fumigatus* and *Aspergillus fischeri* are as divergent as humans and mice
- Genomes are quite stable, ranging from 28-40 Mb, several similar characteristics
 - It is not what we would call ‘plastic’
 - But why and how are there differences?



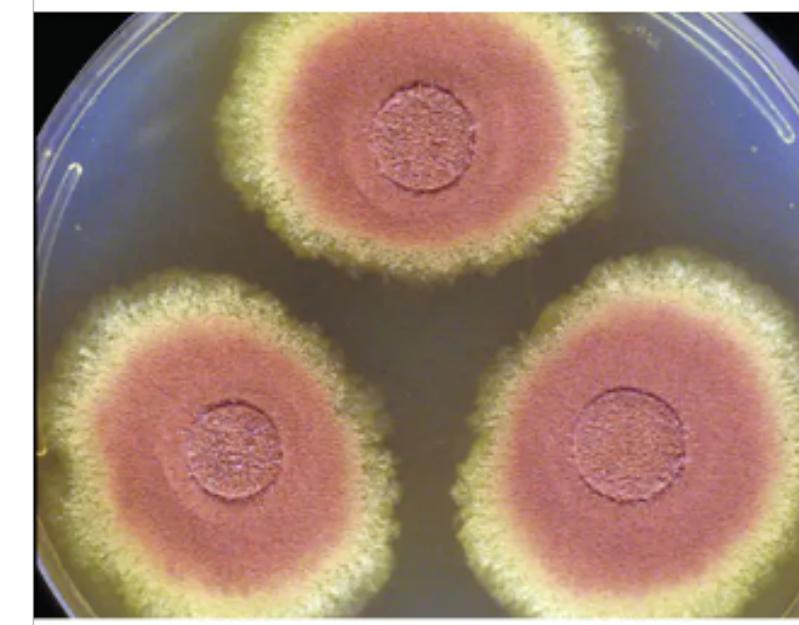
Aspergillus flavus



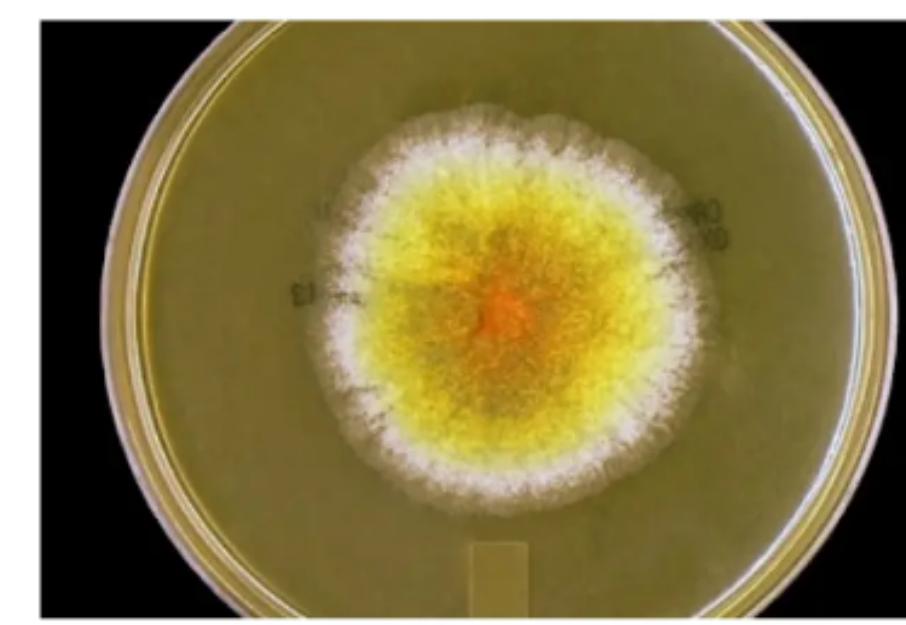
Aspergillus fumigatus



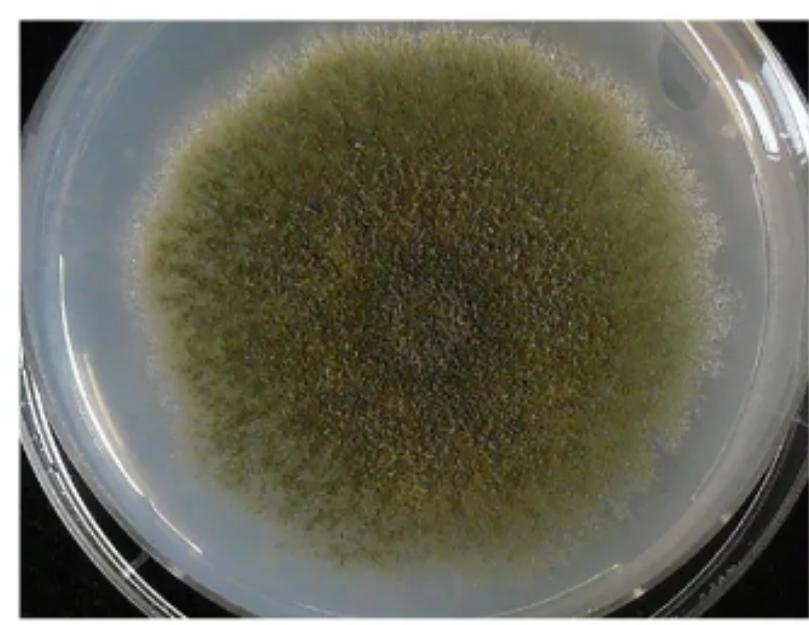
Aspergillus niger



Aspergillus terreus



Aspergillus glaucus

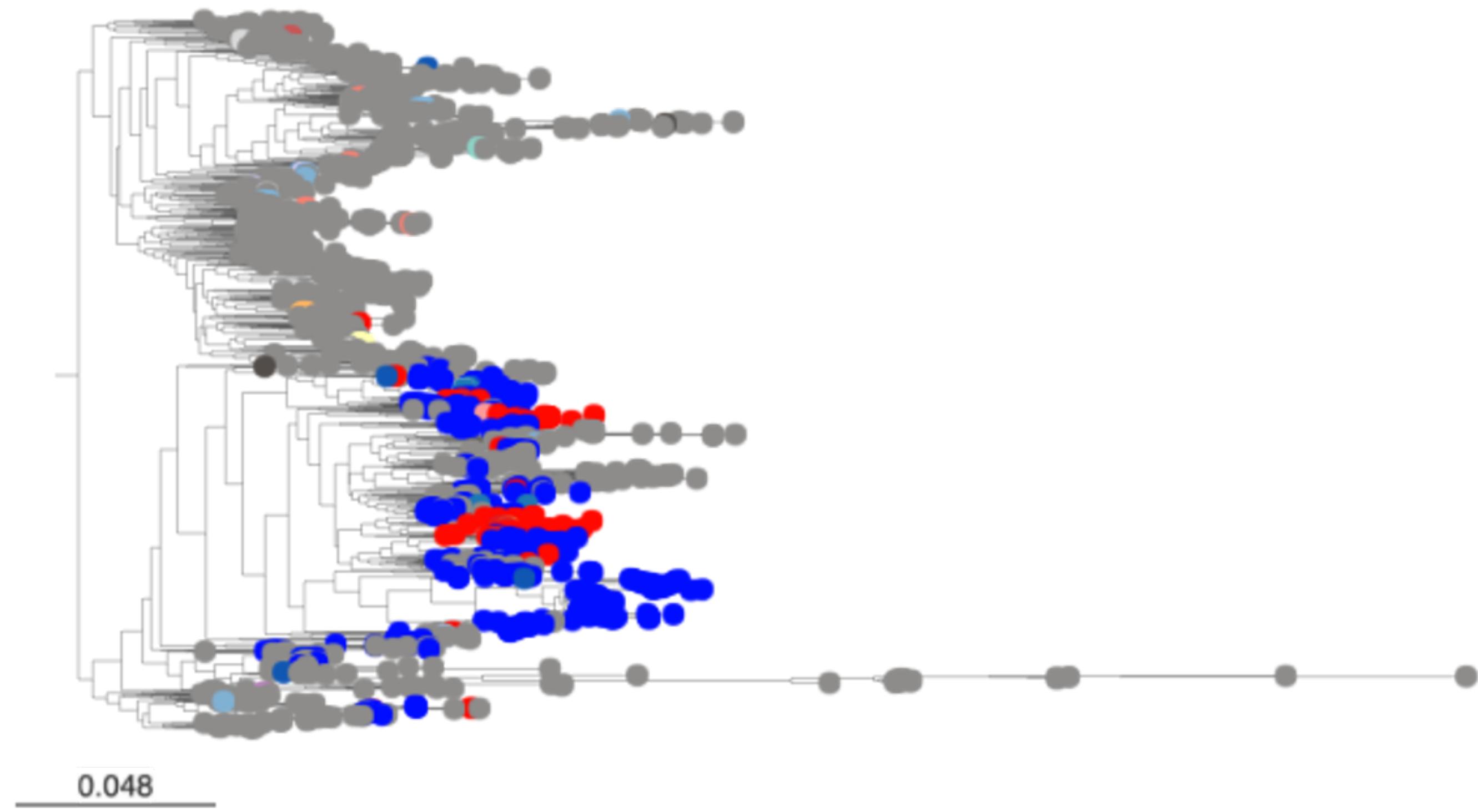


Aspergillus nidulans

A. fumigatus

Extensive recombination

- WHO critical priority fungal pathogen
 - drug resistant infection mortality up to 88%
- Mutations conferring resistance primarily found on one genetic background
 - Can occur occasionally on other genetic background -> recombination/sexual crossing



Data

Basic popgen analysis

- Multi-VCF of *A. fumigatus* isolates
- Metadata associated with these isolates

Basic popgen analysis

R

- Need vcfR, ape and adegenet libraries, a colour library
- Import multi-vcf and metadata

Basic popgen analysis - PCA

R

```
Afum_vcf=read.vcfR("Afum.vcf",verbose=F)
```

```
Afum_genind=vcfR2genind(Afum_vcf)
```

```
metadata=read.csv("metadata.csv",header=T)
```

```
Afum_genind@pop=as.factor(metadata$Source)
```

```
Clean_AfumGenind=tab(Afum_genind,freq=TRUE,NA.method="mean")
```

Basic popgen analysis - PCA

R

- Perform the PCA

```
pca_afum_source<-
dudi.pca(Clean_AfumGenind,scale=FALSE,scannf=FALSE,nf=3)
```

- Visualise

```
colours=brewer.pal(n=3,name="Dark2")
```

```
s.class(pca_afum_source$li,pop(obj),col=transp(col,.6))
```

```
add.scatter.eig(pca1$eig[1:20],nf=3,xax=1,yax=3,posi="topleft")
```

Basic popgen analysis - PCA

R

- What can you say about the *A. fumigatus* population when looking at the isolation source? Can this variable explain the variation?
- There are other columns in the metadata
 - Re-do the PCA using other variables to see if there is a variable that best explains the data
- You can also do PCA using the ‘glPca’ command as well. Is it any different?

Basic popgen analysis - PCA

R

- To publish PCA you usually need to add the percent contribution to each PC, which you can calculate using: `percent=round(pca1$eig/sum(pca1$eig)*100,digits=1)`
- I like to use the ‘mapmixture’ package to then make a nicer PCA plot:
 - `scatter1=scatter_plot(pca1li,objpop,type="points",axes=c(1,2),percent=percent,colours=cols,point_size=2,point_type=21,centroid_size=2,stroke=0.1,plot_title="PCA coloured by...")+`
 - `theme(legend.position="none",axis.title=element_text(size=8),axis.text=element_text(size=6),plot.title=element_text(size=10))`

Basic popgen analysis - DAPC

R

- Can use the same genind object generated (not the cleaned one where NAs were removed)

```
dapc_afum_source<-dapc(Afum_genind)
```

```
scatter(dapc_afum_source,legend=TRUE,col=col)
```

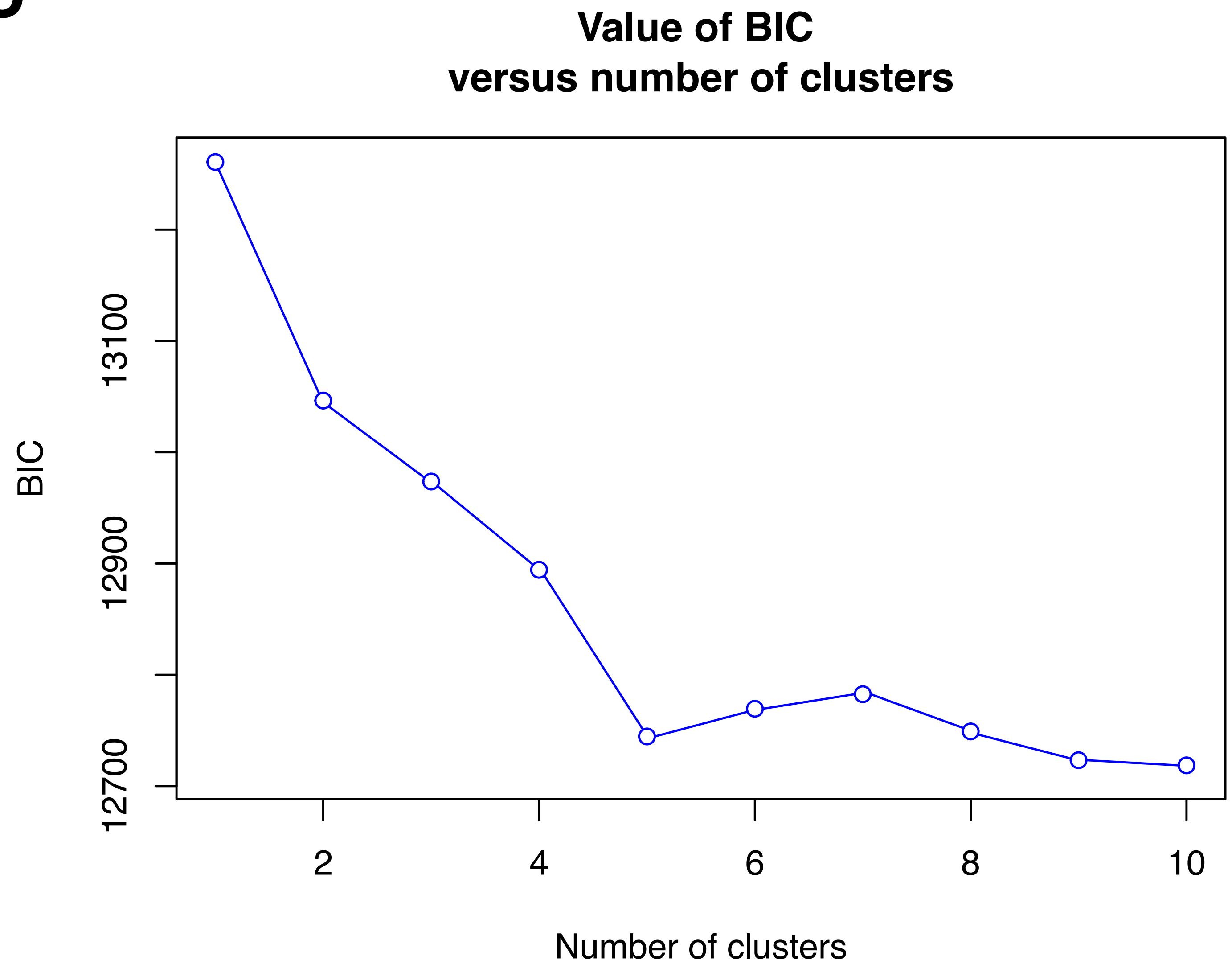
- How does this compare to the PCA with the same variable?
- Can also look at contribution of each individual in terms of populations for a basic look at recombination:

```
compoplot(dapc_afum_source)
```

Basic popgen analysis - DAPC

R

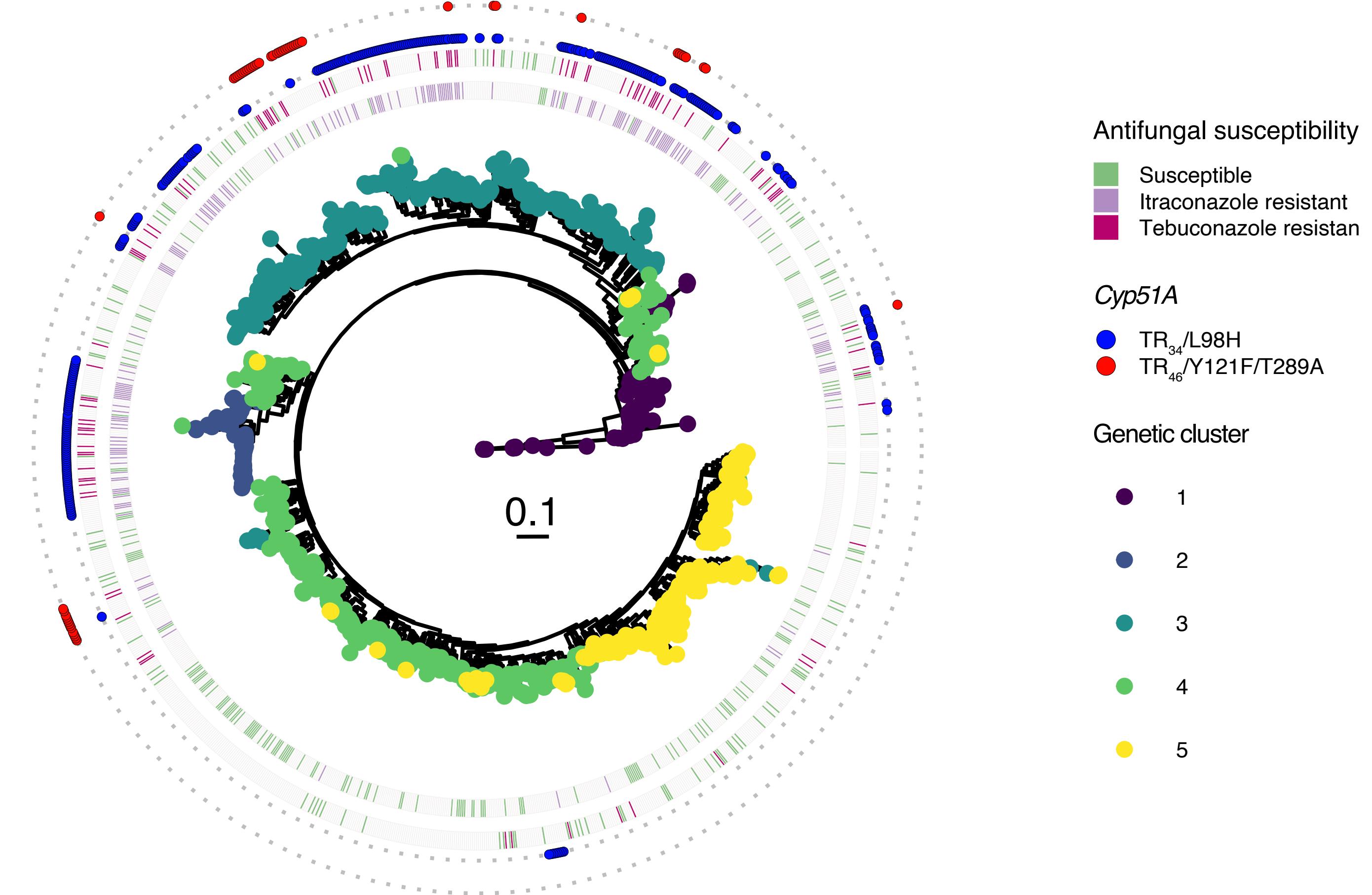
- How many clusters are there within this data?
 - Check for the value of ‘K’
 - This helps investigate which groupings explain the data best
 - `find.clusters(x)`



Basic popgen analysis - DAPC

R

- How many clusters are there within this data?
 - Check for the value of ‘K’
 - This helps investigate which groupings explain the data best
 - `find.clusters(x)`



Infer admixed populations

ADMIXTURE

- Can download ADMIXTURE v1.3 from
- Input is binary PLINK (.bed), ordinary PLINK (.ped) or EIGENSTRAT (.geno)
 - Can use original merged VCF and convert using PLINK

```
plink --vcf all_withRef.vcf --indep-pairwise 50 10 0.1
```

```
plink --vcf all_withRef.vcf --extract plink.prune.in --recode 12 --allow-extra-chr --out pruned_all
```

- First command removes SNPs in linkage disequilibrium
- Second command copies SNPs not removed (and therefore not in LD) to ped file
- Files will be provided so don't worry! This is just an idea of how to generate them yourself in the future

Infer admixed populations

ADMIIXTURE

- It's a good idea to run cross-validation for ADMIXTURE to check for the correct value of K - we had an idea of this with the DAPC analysis
 - Run ADMIXTURE for k=1-10

for K in 1 2 3 4 5 6

do

```
./admixture --cv Afum_LDpruned.bed $K –haploid="" | tee log${K}.out
```

done

Infer admixed populations

Visualisation

- In R - need a colour palette library
- Plot the Q estimates, which are a simple matrix

```
Afum_tbl=read.table("output.Q")
barplot(t(as.matrix(tbl)), col=rainbow(3),
xlab="Individual #", ylab="Ancestry",
border=NA)
```

- I again also like the mapmixture package here, as you can plot the admixture in the context of geography, too.
 - Check out the mapmixture GitHub repo:

