# sNMF analysis - S. pseudococos

```r
# Load library with sNMF function
library(LEA)
# Set input filenames
inputs <- c("ps6_10_ingroup", "ps6_30_ingroup", "ps6_50_ingroup", "ps6_70_ingroup", "ps6_90_ingroup")
```

```r
# Read in sample names from vcf file
library(vcfR)
```

```
## Warning: package 'vcfR' was built under R version 3.6.3
```

```
##
##    *****       ***   vcfR   ***       *****
##    This is vcfR 1.10.0
##      browseVignettes('vcfR') # Documentation
##      citation('vcfR') # Citation
##    *****       *****     *****       *****
```

```r
vcf <- read.vcfR(paste(inputs[1],".vcf",sep=""))
samples <- colnames(vcf@gt)[-1]
```

```r
# Convert vcf file to .geno format for sNMF
for(i in 1:length(inputs)){
  vcf2geno(paste(inputs[i],".vcf",sep=""))
}
```

```
##
##  - number of detected individuals:    52
##  - number of detected loci:       1205
##
## For SNP info, please check ./ps6_10_ingroup.vcfsnp.
##
## 32 line(s) were removed because these are not SNPs.
## Please, check ./ps6_10_ingroup.removed file, for more informations.
##
##
##  - number of detected individuals:    52
##  - number of detected loci:       8259
##
## For SNP info, please check ./ps6_30_ingroup.vcfsnp.
##
## 218 line(s) were removed because these are not SNPs.
## Please, check ./ps6_30_ingroup.removed file, for more informations.
##
##
```

```
##  - number of detected individuals:    52
##  - number of detected loci:      28262
##
## For SNP info, please check ./ps6_50_ingroup.vcfsnp.
##
## 770 line(s) were removed because these are not SNPs.
## Please, check ./ps6_50_ingroup.removed file, for more informations.
##
##
##  - number of detected individuals:    52
##  - number of detected loci:      88944
##
## For SNP info, please check ./ps6_70_ingroup.vcfsnp.
##
## 2156 line(s) were removed because these are not SNPs.
## Please, check ./ps6_70_ingroup.removed file, for more informations.
##
##
##  - number of detected individuals:    52
##  - number of detected loci:      299657
##
## For SNP info, please check ./ps6_90_ingroup.vcfsnp.
##
## 5777 line(s) were removed because these are not SNPs.
## Please, check ./ps6_90_ingroup.removed file, for more informations.
```

```r
#Run sNMF with k from 1 to 1+number of populations
k <- c(1:15)

for(i in 1:length(inputs)){
  snmf(paste(inputs[i],".geno",sep=""), K = k, repetitions = 5, CPU=6,
          project = "new", entropy = TRUE, iterations = 2000, alpha = 1)

  snmf(paste(inputs[i],".geno",sep=""), K = k, repetitions = 5, CPU=6,
          project = "continue", entropy = TRUE, iterations = 2000, alpha = 10)

  snmf(paste(inputs[i],".geno",sep=""), K = k, repetitions = 5, CPU=6,
          project = "continue", entropy = TRUE, iterations = 2000, alpha = 100)
}
```
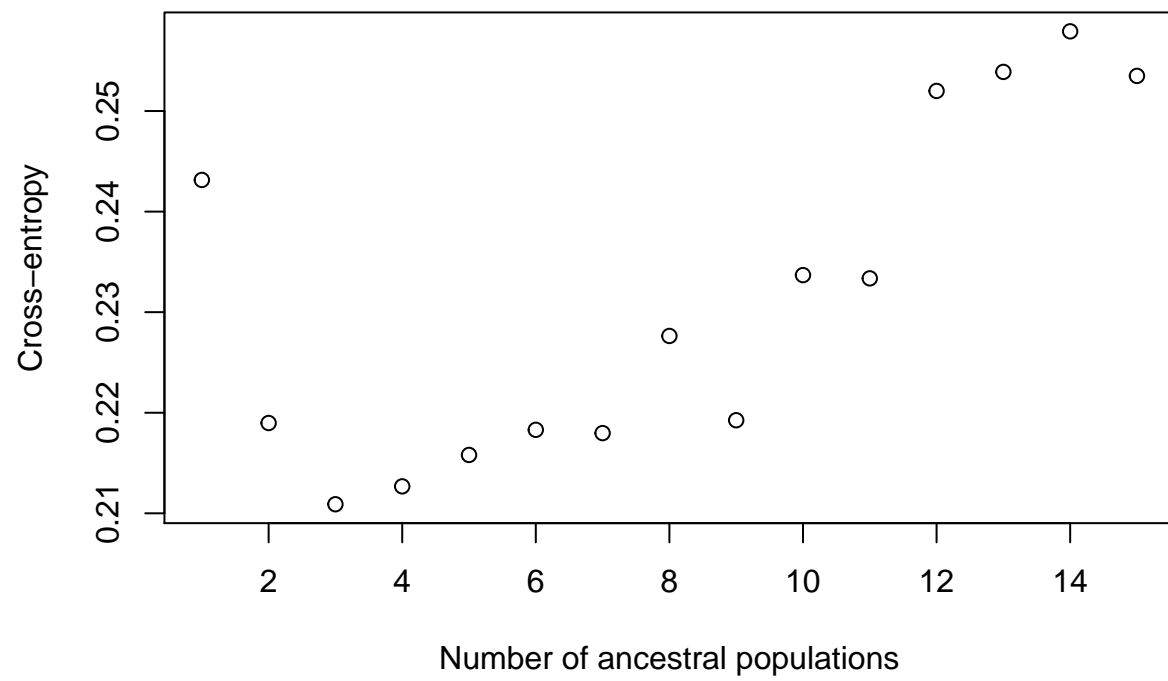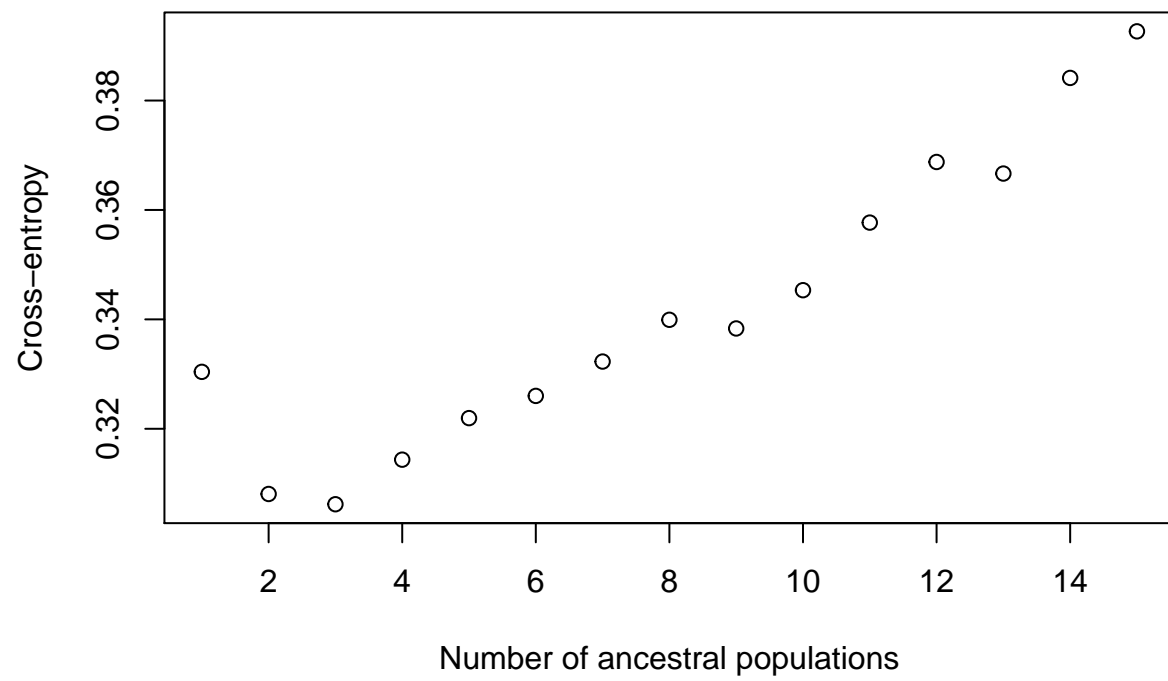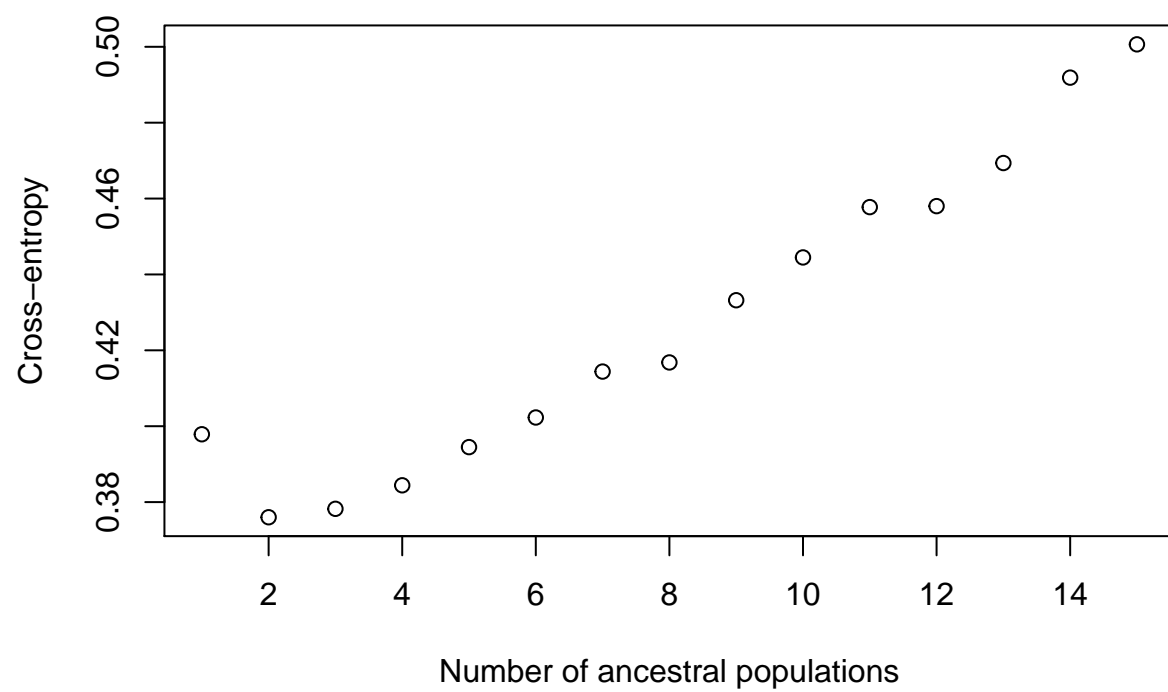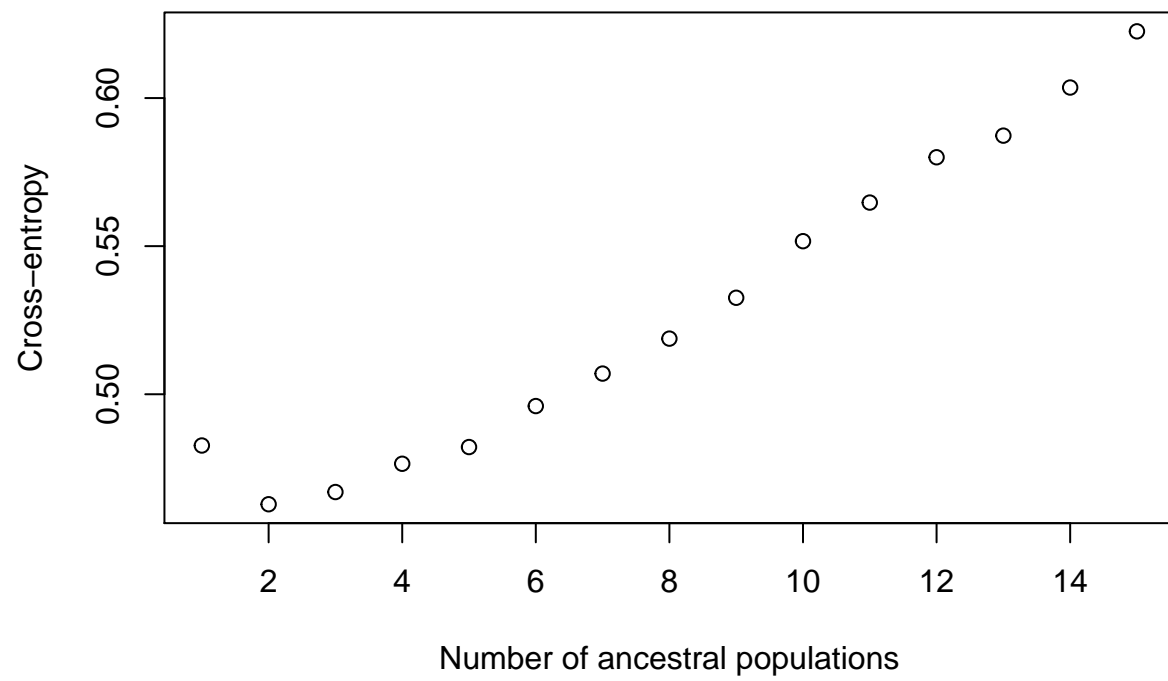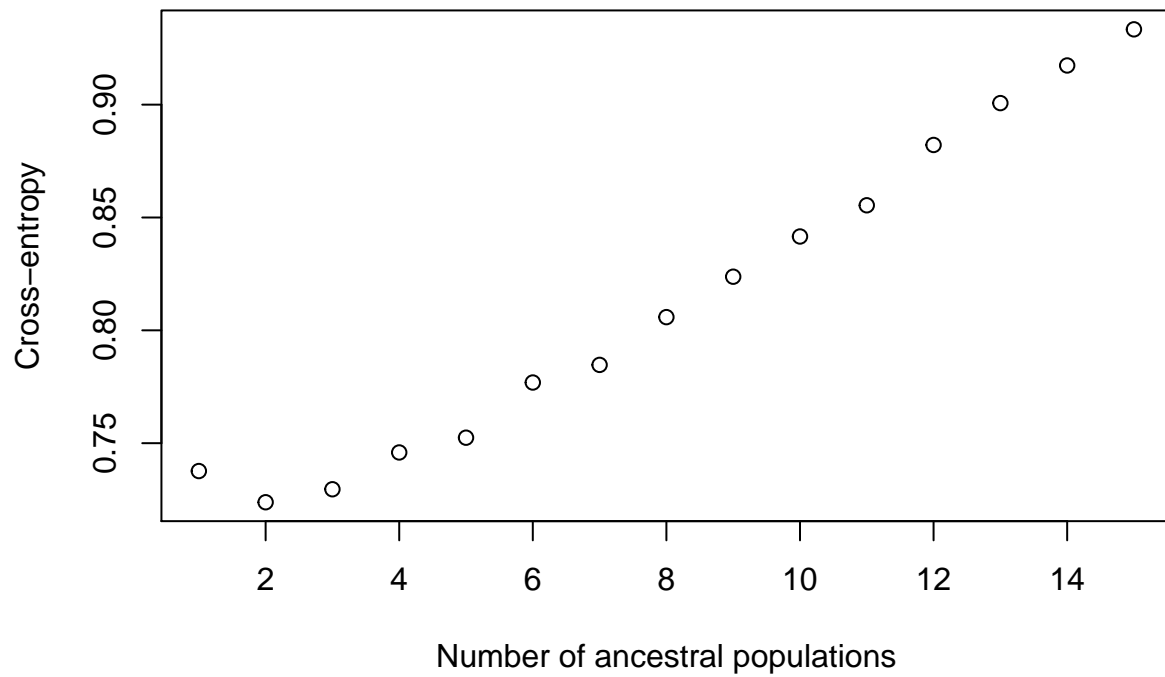
```r
# Plot the cross entropy
for(i in 1:length(inputs)){
  snmf <- load.snmfProject(paste(inputs[i],".snmfProject",sep=""))
  plot(snmf)
  summary(snmf)
}
```

```r
# Plot barplots
for(i in 1:length(inputs)){
  snmf <- load.snmfProject(paste(inputs[i],".snmfProject",sep=""))

  for(j in 2:5){
    ce <- cross.entropy(snmf, K = j)
    best <- which.min(ce)

    qmatrix = Q(snmf, K = j, run = best)

    barplot(t(qmatrix), border = NA, space = 0, xlab = "", ylab = "Admixture coefficients",
            names.arg=samples, las=2)
    mtext("Individuals", side=1, line=4)
  }
}
```