

Evolution and genomics, Cesky Krumlov

Daniel Berner

2. February 2016

Genomic analyses using RADseq:

3. Exploring selective sweeps based on allele frequencies

# Demo

Upload and inspect a SNP matrix (a subset of the Roesti et al. Nat. Commun. lake-stream stickleback data, reduced to Chr1, SNPs quality filtered)

```
# d<-read.table('C:/Users/daniel/Documents/science/teaching/cesky  
# krumlov 2016/course.materials/R.files/SNP.mat.txt',  
# h=T, check.names=F)  
dim(d) # call number of rows and columns
```

```
## [1] 98 3462
```

```
d[c(1:2, 97:98), 1:8]
```

```
##           id      pop al 1733 28844 50852 50966 79770  
## 1  lib_21_CGATA stream A    C      A      C      T      A  
## 2  lib_21_CGATA stream B    G      T      T      A      N  
## 97 lib_31_GTCAC  lake  A    G      T      T      T      A  
## 98 lib_31_GTCAC  lake  B    G      N      T      T      N
```

# Calculate lake-stream allele frequency difference at a single SNP

```
i <- 4 # index of the first SNP in the matrix  
loc <- d[, i]  
loc
```

```
## [1] C G G C C C G C C C C C G N G C G N C G G G C C N N  
## [27] C C C N C N N N C C G C G N N N C C N N G C G G G G  
## [53] C G G G C G G G G G G G G G G N N C C G G C G G G  
## [79] G N C G C C C C C G C C G G G G N N G G  
## Levels: C G N
```

```
d[, 2]
```

```
## [1] stream stream stream stream stream stream stream  
## [8] stream stream stream stream stream stream stream  
## [15] stream stream stream stream stream stream stream  
## [22] stream stream stream stream stream stream stream  
## [29] stream stream stream stream stream stream stream  
## [36] stream stream stream stream stream stream stream  
## [43] stream stream stream stream stream stream lake  
## [50] lake lake lake lake lake lake lake  
## [57] lake lake lake lake lake lake lake  
## [64] lake lake lake lake lake lake lake  
## [71] lake lake lake lake lake lake lake  
## [78] lake lake lake lake lake lake lake  
## [85] lake lake lake lake lake lake lake  
## [92] lake lake lake lake lake lake lake  
## Levels: lake stream
```

```
# the major allele at the SNP
mAl <- names(sort(table(loc), decreasing = T))[1]
mAl

## [1] "G"

mAl.cnt.str <- length(which(loc[1:48] == mAl)) # count mAl in stream
mAl.cnt.str

## [1] 12

cnt.non.N.str <- length(which(loc[1:48] != "N")) # count non-N bases
cnt.non.N.str

## [1] 35

mAl.fr.str <- mAl.cnt.str/cnt.non.N.str
mAl.fr.str

## [1] 0.343
```

```
# same for the lake population
mAl.cnt.lk <- length(which(loc[49:98] == mAl)) # count mAl in lake
mAl.cnt.lk

## [1] 32

cnt.non.N.lk <- length(which(loc[49:98] != "N")) # count non-N bases
cnt.non.N.lk

## [1] 45

mAl.fr.lk <- mAl.cnt.lk/cnt.non.N.lk
mAl.fr.lk

## [1] 0.711

# Now can calculate the absolute lake-stream freq
# difference
abs(mAl.fr.lk - mAl.fr.str)

## [1] 0.368
```

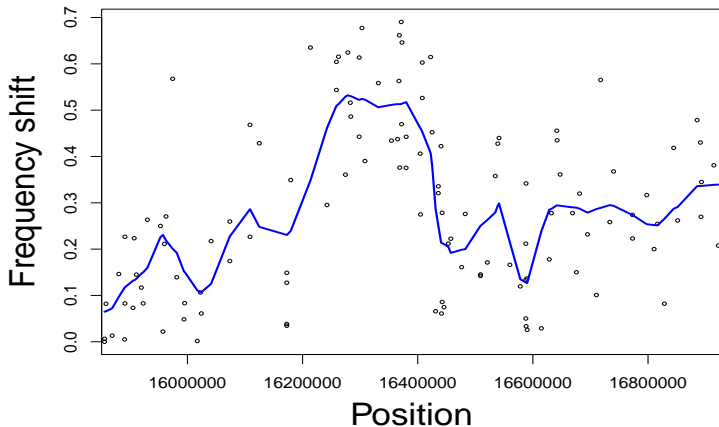
## A minimal loop for the calculation of SNP allele frequency differences

```
diff <- NULL # empty data container
for (i in c(1910:2030)) {
  # SNPs subset, as an example
  loc <- d[, i] # the focal SNP
  mAl <- names(sort(table(loc), decreasing = T))[1] # major allele
  mAl.cnt.str <- length(which(loc[1:48] == mAl)) # mAl in stream
  mAl.fr.str <- mAl.cnt.str/length(which(loc[1:48] != "N"))
  mAl.cnt.lk <- length(which(loc[49:98] == mAl)) # mAl in lake
  mAl.fr.lk <- mAl.cnt.lk/length(which(loc[49:98] != "N"))
  diff <- c(diff, abs(mAl.fr.lk - mAl.fr.str)) # record freq diff
}
diff[1:8]

## [1] 0.00000 0.00633 0.08199 0.01321 0.14622 0.00503
## [7] 0.22666 0.08274
```

## Plotting allele frequency differentiation for the SNPs in target window, with LOESS smoothing

```
pos <- as.integer(names(d)[1910:2030]) # the positions
plot(pos, diff, cex = 0.5) # dot plot
smo <- predict(loess(diff ~ pos, span = 0.1, degree = 0)) # smooth
lines(pos, smo, col = "blue", lwd = 2) # add the profile
```



# Tasks

- ▶ Upload the stickleback SNP data set *SNP.mat.txt*
- ▶ Write a loop calculating the magnitude of lake-stream allele frequency differentiation for every SNP
- ▶ Draw the allele frequency differentiation along the whole chromosome, including a LOESS-smoothed profile