Solutions

Upload the data

```
# d<-read.table('C:/Users/daniel/Documents/science/teaching/cesky
# krumlov 2016/course.materials/R.files/SNP.mat.txt',
# h=T, check.names=F)
d[1:8, 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
## 3 lib_21_CTGAA stream  A    G     T     C     A     C
## 4 lib_21_CTGAA stream  B    C     N     T     T     N
## 5 lib_21_GAAGC stream  A    C     A     C     T     C
## 6 lib_21_GAAGC stream  B    C     A     C     T     N
## 7 lib_22_GTACA stream  A    G     T     C     T     A
## 8 lib_22_GTACA stream  B    C     A     T     A     C
```
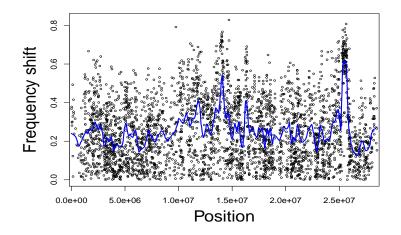
Loop for the calculation of the frequency difference at each SNP

```
diff <- NULL  # empty data container
for (i in 4:length(d[1, ])) {
    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
    cnt.non.N.str <- length(which(loc[1:48] != "N"))  # non-N bases
    mAl.fr.str <- mAl.cnt.str/cnt.non.N.str
    mAl.cnt.lk <- length(which(loc[49:98] == mAl))  # mAl in lake
    cnt.non.N.lk <- length(which(loc[49:98] != "N"))  # non-N bases
    mAl.fr.lk <- mAl.cnt.lk/cnt.non.N.lk
    diff <- c(diff, abs(mAl.fr.lk - mAl.fr.str))  # record freq diff
}
diff[1:8]

## [1] 0.368 0.378 0.281 0.298 0.301 0.493 0.123 0.329
```

Frequency differentiation along chromosome I, with LOESS profile

```
pos <- as.integer(names(d)[4:length(d[1, ])])  # the positions
plot(pos, diff, cex = 0.5)  # dot plot
smo <- predict(loess(diff ~ pos, span = 0.02, degree = 0))  # smooth
lines(pos, smo, col = "blue", lwd = 2)  # add the profile
```