Friday July 13 (Omg! Friday the 13th…)

**Exercise 4: “Introduction to mtDNA analysis”**

1. Look at Figure S2.1 in the Haak et al. (2015) supplement (p. 7). interpret Determine which IDs are associated with which of the above cultures to help.

2. Take the Alberstedt\_LN individual (see IND file). In Table S2.1, look for this individual. Write the haplogroup, list how many SNPs do not match the RSRS. Explain how the notation works.

3. What can you learn about a population from uniparental markers and what can’t you learn about a population? List at least one each.

**4. ADVANCED:** This homework exercise is to reproduce the PCA (on page 7 of Haak et al. 2015 supplement) using the Table S2.2 (page 19). This is exciting! You will see how easy it is to produce a plot on a paper from Nature!

To complete this exercise, we will need to use R. The server has access to R, but currently only on the Terminal. (If we get access to it on jupyter notebook, we will let you know!). If you prefer to do this on a GUI on your computer, please install R.

For Mac users: <https://cran.r-project.org/bin/macosx/>

For PC users: <https://cran.r-project.org/bin/windows/base/>

After you have installed R, you will install an R package called: “ggbiplot”

(you can read about details here, <https://github.com/vqv/ggbiplot>)

To install ggbiplot, first open R, and then at the command line type in:

>library(devtools)

>install\_github("vqv/ggbiplot")

To load the ggbiplot package, just type:

>library(ggbiplot)

Ok, you are ready to go!Now for the hard work, you have to prepare the frequencies table, so that it can to be read into R as a matrix “m” so pass it into the princomp function that performs the PCA analysis (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/princomp.html>)

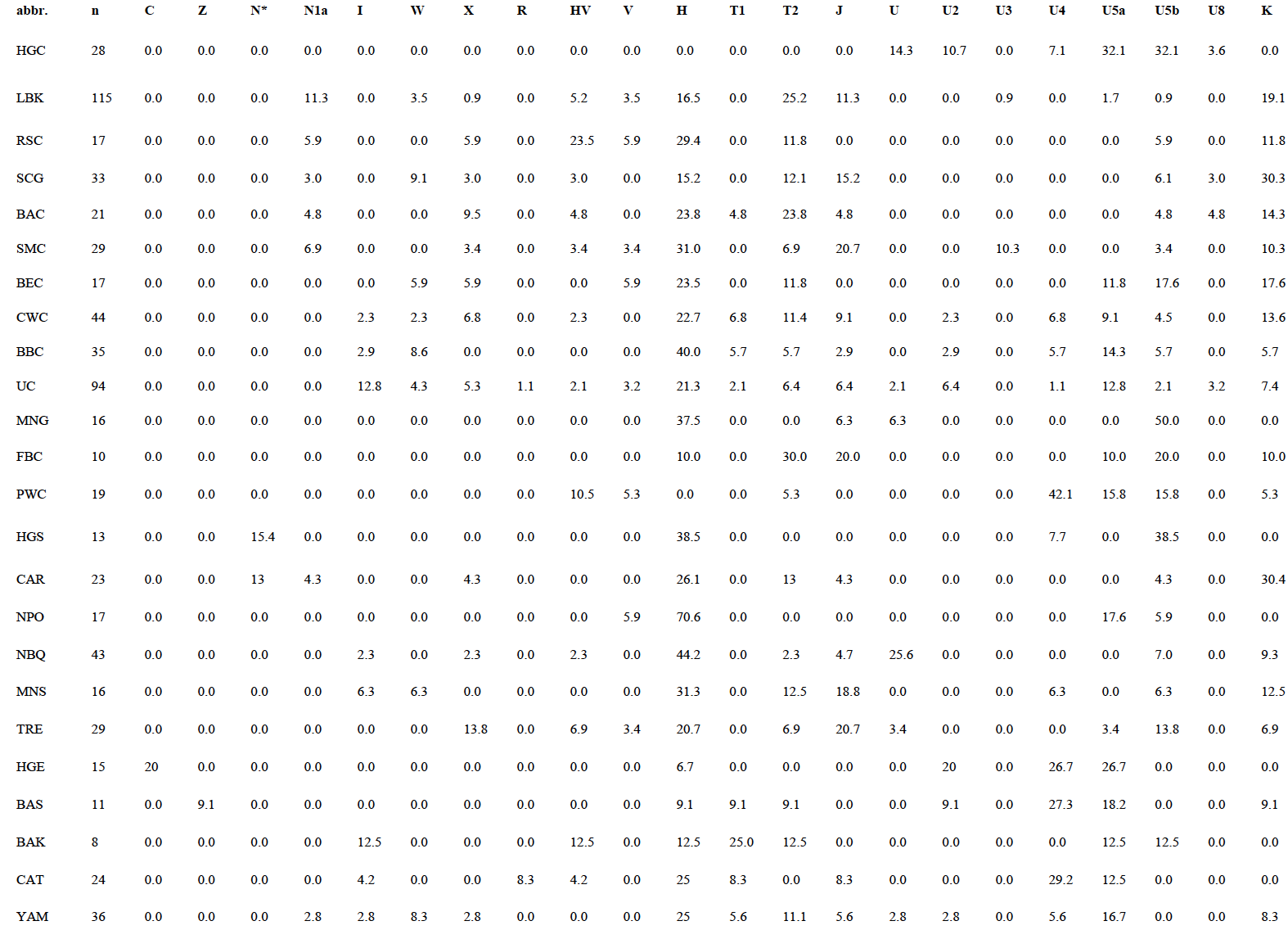
So please find a way to extract the mitochondrial frequencies table of 24 prehistoric populations (Table S2.2 page 19 of the supplement)…

Hint: You can prepare the table (below) in Excel and name it “table.csv” to be read into R. Remember to put the file in the same directory as R, or specify a working directory, for example:

>setwd("~/Downloads/course - haak paper")

Sorry you have to think about how to best extract this table without typing in the numbers, unfortunately this is part of the daily job of being a population geneticist...

>m<-read.csv(“table.csv”, head=T)



This is important, if you read in the table as shown, you will need to make the first column as row names, then delete the first column, so that only frequencies remain, something like:

>row.names(m)<-m[,1]

>m<-m[,-1]

Then at the command line, type:

>fit <- princomp(data.matrix(m), cor=TRUE)

>ggbiplot(fit, labels=rownames(m), scale=0.05, labels.size=3)

Voila! What do you see?

Yes. This concludes the homework exercise. Please let me know if there are any questions. Thank you, see you next time!