#**1 - Use terminal. Open it.**

#Hereafter, anything typed in the terminal is written in this way:

#command options value input metacharacters output

#**2 - Navigate in the terminal and create a folder**

#cd # change directory

#ls # list

#echo # print on screen

###

cd UNIXworkshop

ls

ls -a

echo "This exercice will be is easy"

#**3 – Modifying content**

#mkdir # make directory

#cp # copy

#rm # remove

#vi # text viewer

###

mkdir Backup

cp MatRead.R Backup/

cp .ScriptBackup\_DO\_NOT\_MODIFY.sh Backup/ScriptBackup\_DO\_NOT\_MODIFY.sh

cp .UNIX.help Backup/UNIX.help

cd Backup

vi UNIX.help

:q

cd ..

#**4 - Extract the name**

#grep # grip

#tr # translate

#sed # stream editor

#' or " # METACHARACTER = input

#. # METACHARACTER = any character

#\* # METACHARACTER = repeat from 0 to any time

#| # METACHARACTER = string together

#< # METACHARACTER = input

#> # METACHARACTER = output

#\n # METACHARACTER = new line (also \r)

###

grep -A 5 'Study reference:' MATRIX.nex > MATRIX.nam.tmp

tr '\n' ' ' < MATRIX.nam.tmp > MATRIX.name.tmp

sed 's/Study reference: //g' MATRIX.name.tmp | sed 's/ TreeBASE Study URI:.\*//g' > MATRIX.name

rm \*.tmp

vi MATRIX.name

:q

#**5 - Extract the number of taxa**

grep 'DIMENSIONS NTAX=' MATRIX.nex > MATRIX.ntax.tmp

sed 's/DIMENSIONS NTAX=//g' MATRIX.ntax.tmp | sed 's/;//g' | sed 's/ //g' > MATRIX.ntax

rm \*.tmp

vi MATRIX.ntax

:q

#**6 - Asigning a variable**

#$ # METACHARACTER = assigning

#let # statement

###

NTAX=$(sed -n '1p' MATRIX.ntax)

let "NTAX += 2"

echo ${NTAX}

#**7 - Extact the matrix values**

grep -A "$NTAX" 'MATRIX' MATRIX.nex > MATRIX.mat.tmp

sed '1,3d' MATRIX.mat.tmp | sed 's/.\* //g' | sed 's/{..}/?/g' | sed 's/{...}/?/g' | sed 's/{....}/?/g' | sed 's/{.....}/?/g' | sed -e 's/\(.\)/\1 /g' > MATRIX.mat

rm \*.tmp

vi MATRIX.mat

:q

#**11 - Extracting the citations**

#for # loop initiator

#in # loop length

#; # METACHARACTER = new line (= start a new prompt)

#do # loop start

#paste # paste

#>> # METACHARACTER = new line

#done # loop end (DON’T FORGET THAT ONE)

###

for f in \*.name ; do paste $f >> citations.txt ; done ;

rm \*.name

vi citations.txt

:q

echo "That was useless"

#**12 – Laaaaaaazzynessssss** *Writing a script*

#sh # execute a shell script

#basename # define a prefix

#\ # METACHARACTER = ignore the metacharacter

## # METACHARACTER = ignore the line (IS NOT IGNORED IN SOME COMMANDS/SCRIPTS)

###

echo "##########################

#Transforming nexus morphological matrices to be read by MatRead.R

##########################

#Generates n matrices to be read by MatRead.R

#Generates a citation file

##########################

#Creating a folder for the output and cleaning the old ones

rm -R matrices

rm citations.txt

mkdir matrices

#Initiating the loop

for f in MatricesSet1/\*.nex

do

prefix=\$(basename \$f .nex)

echo \${prefix}

#Extracting the matrices references

grep -A 5 'Study reference:' MatricesSet1/\${prefix}.nex > \${prefix}.nam.tmp

tr '\n' ' ' < \${prefix}.nam.tmp > \${prefix}.name.tmp

sed 's/Study reference: //g' \${prefix}.name.tmp | sed 's/ TreeBASE Study URI:.\*//g' > \${prefix}.name

#Extracting the number of taxa in order to isolate the matrix

grep 'DIMENSIONS NTAX=' MatricesSet1/\${prefix}.nex > \${prefix}.ntax.tmp

sed 's/DIMENSIONS NTAX=//g' \${prefix}.ntax.tmp | sed 's/;//g' | sed 's/ //g' > \${prefix}.ntax

NTAX=\$(sed -n '1p' \${prefix}.ntax)

let 'NTAX += 2'

#Extracting the matrix, adding a space between each characters in order to read in in R as tab.delim and collapsing the uncertainties (e.g. {01}) into missing data ("?")

grep -A "\$NTAX" 'MATRIX' MatricesSet1/\${prefix}.nex > \${prefix}.mat.tmp

sed '1,3d' \${prefix}.mat.tmp | sed 's/.\* //g' | sed 's/{..}/?/g' | sed 's/{...}/?/g' | sed 's/{....}/?/g' | sed 's/{.....}/?/g' | sed -e 's/\(.\)/\1 /g' > matrices/\${prefix}.mat

done

#Writing the citation file

for f in \*.name ; do paste \$f >> citations.txt ; done ;

#Cleaning temporaries files

rm \*.name.tmp

rm \*.nam.tmp

rm \*.ntax.tmp

rm \*.ntax

rm \*.mat.tmp

rm \*.name

echo 'End'" > MatRead.sh

sh MatRead.sh

#**13 – Using R** *Reading a R script*

R --no-save < MatRead.R

#**14 – Just for the fun, let’s put all together**

echo "#Analysing and ploting the results through R

R --no-save < MatRead.R" >> MatRead.sh

sh MatRead.sh

#**15 – If that’s still not enough…** *Use 100 input matrices*

sed 's/MatricesSet1/MatricesSet2/g' MatRead.sh > MatRead\_2.sh

sh MatRead\_2.sh

#**16 – Going further**

echo "Check the web pages in UNIX.help by using vi command."

echo "Bashing is fun."