

Find A Gene Project: POMC

Taylor Darby

11/12/2021

[Q7] Generate a sequence identity based heatmap of your aligned sequences using R. If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and “Save as” FASTA format for example). Read this FASTA format alignment into R with the help of functions in the Bio3D package. Calculate a sequence identity matrix (again using a function within the Bio3D package). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.

```
# I will use the function 'seqaln()' which requires a MUSCLE download. Downloaded MUSCLE in the 'Terminal'
# I will also use the bio3d package which is already installed
library(bio3d)
POMC <- read.fasta("POMC_protein.fasta")
```

```
POMC <- read.fasta("POMC_protein.fasta")
POMC_align <- seqaln(POMC)
POMC_align
```

```
##          1          .          .          .          .          .          60
## Deer      -----
## Human     MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPM
## Frog      MFRPLWGCFLAI-LGICIFHIGEVQSQCWESSRCADLSSDGVLCEIKACKTDLSAESPV
## Sheep     MPRLCSSRSGALLLVLLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV
## Cow       MPRLCSSRSAALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV
## Goat      MPRLCSSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV
## Macaque   MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV
## Pig       MPRLCGSRSGALLLTLLLQASMGVRGWCLESSQCQDLSTESNLLACIRACKPDLSAETPV
## Chimpanzee MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPM
## Mouse     MPRFCYSRSGALLLALLLQTSIDVSWWCLESSQCQDLTTESNLLACIRACKLDLSLETPV
## Chicken   -MRGALCHSLPVVLGLLLCHPTTASGPCWENSKCQDLATEAGVLACAKACRAELSAEAPV
## Horse     MPRSCGSRSGALLLALLLQASVEVRGWCLESSQCQDLTTESNLLACIRACKIDLSAETPV
##
##          1          .          .          .          .          .          60
##
##          61          .          .          .          .          .          120
## Deer      -----
## Human     FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNSSS----SGSSGAGQKR--EDVSAGEDCG
## Frog      FPGNGHLQPLSESIRKYVMTHFRWNKFGRNST----GNDGSNTGYKR--EDISSYPVFS
## Sheep     FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SFGAGGAAQKR-EEEVAV-----
## Cow       FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SSGVGAAQKR-EEEVAV-----
## Goat      FPCNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SFGAGGAAQKR-EEEVAV-----
## Macaque   FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNSSS---GSGA---AQKR--EDVAAGEDRG
## Pig       FPGNGDAQPLTENPRKYVMGHFRWDRFGRRNGSSSGGGGGGGGAGQKREEEVAA-----
```

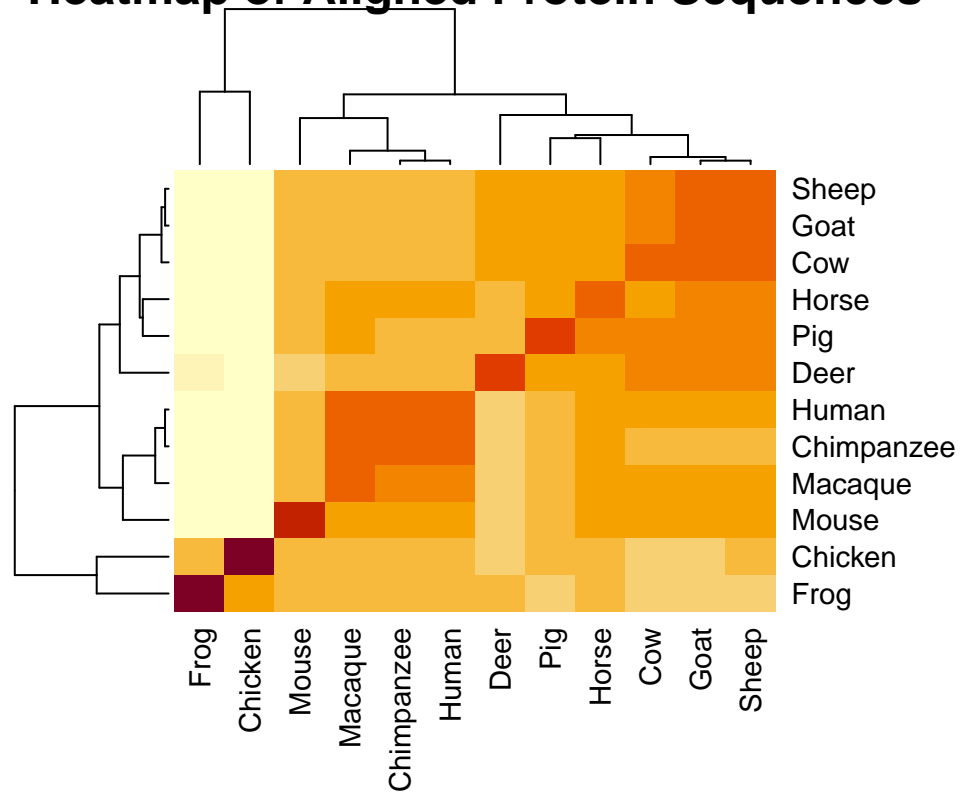


```
##
##          301          .          .  323
## Deer      VDPEGLSSALLPPRSPGEDSPRQ
## Human     -----
## Frog      -----
## Sheep     -----
## Cow       -----
## Goat      -----
## Macaque   -----
## Pig       -----
## Chimpanzee -----
## Mouse     -----
## Chicken   -----
## Horse     -----
##
##          301          .          .  323
##
## Call:
##   seqaln(aln = POMC)
##
## Class:
##   fasta
##
## Alignment dimensions:
##   12 sequence rows; 323 position columns (104 non-gap, 219 gap)
##
## + attr: id, ali, call
```

```
POMC_seqID <- seqidentity(POMC_align)
```

```
POMC_heatmap <- heatmap(POMC_seqID, main = "Heatmap of Aligned Protein Sequences", margins = c(7,2))
```

Heatmap of Aligned Protein Sequences



POMC_heatmap

```
## $rowInd
## [1] 3 11 10 7 9 2 1 8 12 5 6 4
##
## $colInd
## [1] 3 11 10 7 9 2 1 8 12 5 6 4
##
## $Rowv
## NULL
##
## $Colv
## NULL
```