Find A Gene Project: POMC

Taylor Darby

11/12/2021

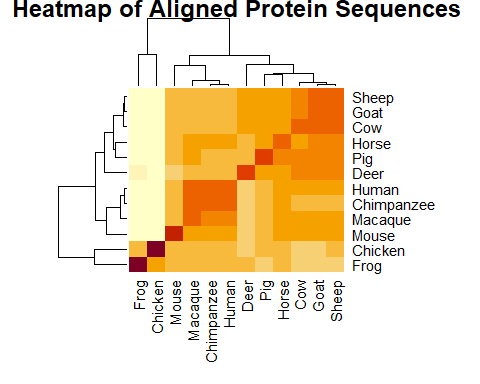
**[Q7]** Generate a sequence identity based heatmap of your aligned sequences using R. If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and “Save as” FASTA format for example). Read this FASTA format alignment into R with the help of functions 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' tab using: curl -o "muscle.exe" "https://www.drive5.com/muscle/downloads3.8.31/muscle3.8.31\_i86win32.exe"  
# 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 MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLECIRACKPDLSAETPM  
## Frog MFRPLWGCFLAI-LGICIFHIGEVQSQCWESSRCADLSSEDGVLECIKACKTDLSAESPV  
## Sheep MPRLCSSRSGALLLVLLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV  
## Cow MPRLCSSRSAALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV  
## Goat MPRLCSSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLACIRACKPDLSAETPV  
## Macaque MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLECIRACKPDLSAETPV  
## Pig MPRLCGSRSGALLLTLLLQASMGVRGWCLESSQCQDLSTESNLLACIRACKPDLSAETPV  
## Chimpanzee MPRSCCSRSGALLLALLLQASMEVRGWCLESSQCQDLTTESNLLECIRACKPDLSAETPM  
## Mouse MPRFCYSRSGALLLALLLQTSIDVWSWCLESSQCQDLTTESNLLACIRACKLDLSLETPV  
## Chicken -MRGALCHSLPVVLGLLLCHPTTASGPCWENSKCQDLATEAGVLACAKACRAELSAEAPV  
## Horse MPRSCGSRSGALLLALLLQASVEVRGWCLESSQCQDLTTESNLLACIRACKIDLSAETPV  
##   
## 1 . . . . . 60   
##   
## 61 . . . . . 120   
## Deer ------------------------------------------------------------  
## Human FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNSSS----SGSSGAGQKR--EDVSAGEDCG  
## Frog FPGNGHLQPLSESIRKYVMTHFRWNKFGRRNST----GNDGSNTGYKR--EDISSYPVFS  
## Sheep FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SFGAGGAAQKR-EEEVAV-----  
## Cow FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SSGVGGAAQKR-EEEVAV-----  
## Goat FPCNGDEQPLTENPRKYVMGHFRWDRFGRRNGSS---SFGAGGAAQKR-EEEVAV-----  
## Macaque FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNSSS---GSGA---AQKR--EDVAAGEDRG  
## Pig FPGNGDAQPLTENPRKYVMGHFRWDRFGRRNGSSSGGGGGGGGAGQKREEEEVAA-----  
## Chimpanzee FPGNGDEQPLTENPRKYVMGHFRWDRFGRRNSSS---SSSGSGAGQKR--EDVSAGEDRG  
## Mouse FPGNGDEQPLTENPRKYVMGHFRWDRFGPRNSSS---AGSA---AQRRAEEEAVW-----  
## Chicken YPGNGHLQPLSESIRKYVMSHFRWNKFGRRNSSS---G------GHKR--EEVAG---LA  
## Horse FPGNGEEQPLTENPRKYVMGHFRWDRFGRRNSSS---GGGA---SQKREEEEVVV-----  
##   
## 61 . . . . . 120   
##   
## 121 . . . . . 180   
## Deer ------------------------------MEHFRWGKPVGKKRRPVKVYPNGAEDESAQ  
## Human PLPEGGPEPR---SDGA-KPGPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAE  
## Frog LFPLSDQNAPGDNMEEE-PLDRQENKRAYSMEHFRWGKPVGRKRRPIKVYPNGVEEESAE  
## Sheep ---GEGPGPR---GDGA-ETGPREDKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAQ  
## Cow ---GEGPGPR---GDDA-ETGPREDKRSYSMEHFPWGKPVGKKRRPVKVYPNGAEDESAQ  
## Goat ---GEGPGPR---GDGA-ETGPREDKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAQ  
## Macaque LLPEGGPEPR---GDGA-EPGPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAE  
## Pig ---GEGPGPR---GDGV-APGPRQDKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDELAE  
## Chimpanzee PLPEGGPEPR---SDGA-KPGPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAE  
## Mouse -------------GDGSPEPSPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNVAENESAE  
## Chicken LPAASPHHPAGEEEDGE-GLEREEGKRSYSMEHFRWGKPVGRKRRPIKVYPNGVDEESAE  
## Horse ---LGGPGPRGDGGDGG-EAGPREGKRSYSMEHFRWGKPVGKKRRPVKVYPNGAEDESAE  
## \*\*\*\* \*\*\*\*\*\*^\*\*\*\*^\*\*\*\*\* ^ \* \*   
## 121 . . . . . 180   
##   
## 181 . . . . . 240   
## Deer AFPLEFKRELTGERLDQARGPE-------AQAESAAARAELEYGLV--AEAEAAEKKDEG  
## Human AFPLEFKRELTGQRLREGDGPD------GPADDGAGAQADLEHSLL-----VAAEKKDEG  
## Frog SYPMELRRELSLEL----DYPE------------IDLDEDIED-----NEVESALTKKNG  
## Sheep AFPLEFKRELTGERLEQARXPE-------AQAESAAARAELEYGLV--AEAEAAEKKDSG  
## Cow AFPLEFKRELTGERLEQARGPE-------AQAESAAARPELEYGLVAEAEAEAAEKKDSG  
## Goat AFPLEFKRELTGERLEQARGPEAQAESAAAQAESAAARAELEYGLV--AEAEAAEKKDSG  
## Macaque AFPLEFKRELTGQRPRAGDGPD------GPADDGAGPRADLEHSLL-----VAAEKKDEG  
## Pig AFPLEFRRELAGAPPEPARDPE-------APAEGAAARAELEYGLV--AEAEAAEKKDEG  
## Chimpanzee AFPLEFKRELTGQRPREGDGPD------GPADDGAGAQADLEHSLL-----VAAEKKDEG  
## Mouse AFPLEFKRELEGER------PL-----------------GLEQVL------ESDAEKDDG  
## Chicken SYPMEFRREMA---------PD-------GDPFGLSEEEEEEE-----EEEGEEEKKDGG  
## Horse AFPLEFKRELAGER------PE-----------GAAARAELGYSLV--AEAEAAEKKDEG  
## ^\*^\* ^\*\*^ \* \* \*   
## 181 . . . . . 240   
##   
## 241 . . . . . 300   
## Deer PYKMEHFRWGSPPKDKRYGGFMTSEKSLTPLVTLFKTPSSRTPTRRASEGAAGRGLSPRK  
## Human PYRMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE------------  
## Frog NYRMHHFRWGSPPKDKRYGGFMTPERSQTPLMTLFKNAIIKNSHKKGQ------------  
## Sheep PYKMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ------------  
## Cow PYKMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ------------  
## Goat PYKMEHFRWGSPPKDKHYGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ------------  
## Macaque PYRMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAYKKGQ------------  
## Pig PYKMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIVKNAHKKGQ------------  
## Chimpanzee PYRMEHFRWGSPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE------------  
## Mouse PYRVEHFRWSNPPKDKRYGGFMTSEKSQTPLVTLFKNAIIKNAHKKGQ------------  
## Chicken SYRMRHFRWHAPLKDKRYGGFMSLEHSQTPLMTLFKNAIVKSAYKKGQ------------  
## Horse PYKMEHFRWGSPRKDKRYGGFMSSEKSQTPLVTLFKNAIIKNAHKKGQ------------  
## \*^^ \*\*\*\* \* \*\*\*^\*\*\*\*\*^ \*^\* \*\*\*^\*\*\*\* ^ ^^^   
## 241 . . . . . 300   
##   
## 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\_heatmap <- heatmap(seqidentity(POMC\_align), main = "Heatmap of Aligned Protein Sequences", margins = c(7,5))



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