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 'Termin
# I will also use the bio3d package which is already installed
library(bio3d)
POMC <- read.fasta("POMC_protein.fasta")

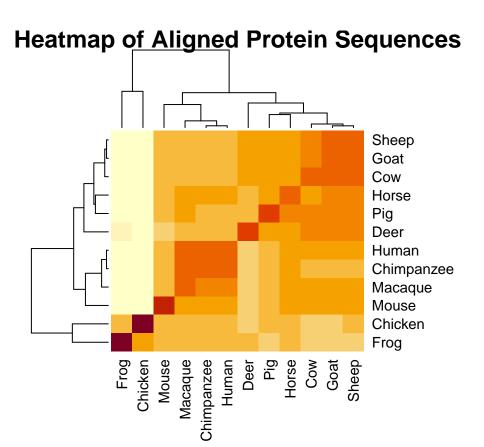
POMC <- read.fasta("POMC_protein.fasta")
POMC_align <- seqaln(POMC)
POMC_align</pre>
```

##		1						60
##	Deer							
##	Human	MPRSCCS	RSGALLLAI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLE	CIRACKPDLS	SAETPM
##	Frog	MFRPLWG	CFLAI-LG	ICIFHIGEV	QSQCWESSRCA	DLSSEDGVLE	CIKACKTDLS	SAESPV
##	Sheep	MPRLCSS	RSGALLLVI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLA	CIRACKPDLS	SAETPV
##	Cow	MPRLCSS	RSAALLLAI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLA	CIRACKPDLS	SAETPV
##	Goat	MPRLCSS	RSGALLLAI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLA	CIRACKPDLS	SAETPV
##	Macaque	MPRSCCS	RSGALLLAI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLE	CIRACKPDLS	SAETPV
##	Pig	MPRLCGS	RSGALLLTI	LLLQASMGV	RGWCLESSQCQ	DLSTESNLLA	CIRACKPDLS	SAETPV
##	Chimpanzee	MPRSCCS	RSGALLLAI	LLLQASMEV	RGWCLESSQCQ	DLTTESNLLE	CIRACKPDLS	SAETPM
##	Mouse	MPRFCYS	RSGALLLAI	LLLQTSIDV	WSWCLESSQCQ	DLTTESNLLA	CIRACKLDLS	SLETPV
##	Chicken	-MRGALC	HSLPVVLG	LLLCHPTTA	SGPCWENSKCQ	DLATEAGVLA	CAKACRAELS	SAEAPV
##	Horse	MPRSCGS:	RSGALLLAI	LLLQASVEV	RGWCLESSQCQ	DLTTESNLLA	CIRACKIDLS	SAETPV
##								
##		1		•			•	60
##								
##		61					•	120
##	Deer							
##	Human	FPGNGDE	QPLTENPRI	KYVMGHFRWI	DRFGRRNSSS-	SGSSGAG	QKREDVS	AGEDCG
##	Frog	FPGNGHL	QPLSESIR	KYVMTHFRWI	NKFGRRNST	GNDGSNTG	YKREDISS	SYPVFS
##	Sheep	FPGNGDE	QPLTENPRI	KYVMGHFRWI	DRFGRRNGSS-	SFGAGGAA	QKR-EEEVAV	J
##	Cow	FPGNGDE	QPLTENPRI	KYVMGHFRWI	DRFGRRNGSS-	SSGVGGAA	QKR-EEEVAV	J
##	Goat	FPCNGDE	QPLTENPRI	KYVMGHFRWI	DRFGRRNGSS-	SFGAGGAA	QKR-EEEVAV	J
##	Macaque	FPGNGDE	QPLTENPRI	KYVMGHFRWI	DRFGRRNSSS-	GSGAA	QKREDVA	AGEDRG
##	Pig	FPGNGDA	QPLTENPRI	KYVMGHFRWI	DRFGRRNGSSS	GGGGGGGGAG	QKREEEEVA <i>I</i>	<i>1</i>

## ##	Chimpanzee Mouse Chicken Horse	FPGNG YPGNG	DEQPLTEN HLQPLSES	IPRKYVMGHFR SIRKYVMSHFR	WDRFGRRNSSS- WDRFGPRNSSS- WNKFGRRNSSS- WDRFGRRNSSS-	AGSA <i>A</i> G	AQRRAEEEAV SHKREEVA	W .GLA
##		61						120
##		121						180
	Deer		· 	· 	MEHFI	RWGKPVGKKRF	· RPVKVYPNGA	
	Human	PLPEG	GPEPR	-SDGA-KPGPR	EGKRSYSMEHFI			•
	Frog				ENKRAYSMEHFI			
	Sheep		-		EDKRSYSMEHFI			
##	Cow	GE	GPGPR	-GDDA-ETGPR	EDKRSYSMEHFI	PWGKPVGKKRF	RPVKVYPNGA	EDESAQ
##	Goat	GE	GPGPR	-GDGA-ETGPR	EDKRSYSMEHF	RWGKPVGKKRF	RPVKVYPNGA	EDESAQ
	Macaque	LLPEG	GPEPR	-GDGA-EPGPR	EGKRSYSMEHFI	RWGKPVGKKRF	RPVKVYPNGA	EDESAE
	Pig				QDKRSYSMEHFI			
	${\tt Chimpanzee}$				EGKRSYSMEHFI			
	Mouse				EGKRSYSMEHFI			
	Chicken				EGKRSYSMEHFI			
	Horse	LG	GPGPRGDC	iGDGG-EAGPR	EGKRSYSMEHFI			
##		121			****	******	** ****	^ * * 180
##		121	•	•	•	•	•	100
##		181						240
	Deer		FKRELTGE	:RI.DQARGPE-	AQAESA	AAARAEI.EYGI	.VAEAEAA	
	Human				GPADDG			
	Frog			•				
	Sheep	AFPLE	FKRELTGE	ERLEQARXPE-	AQAESA	AAARAELEYGI	LVAEAEAA	EKKDSG
##	Cow	AFPLE	FKRELTGE	ERLEQARGPE-	AQAESA	AARPELEYGI	LVAEAEAEAA	EKKDSG
##	Goat			-	QAESAAAQAESA			
	Macaque			-	GPADDG			
	Pig				APAEGA			
	Chimpanzee				GPADDG			
	Mouse							
	Chicken				GDPFGI			
	Horse		FKKELAGI : ^**^	:кРĿ- *	G <i>I</i>	AAAKAELGYSI	LVALALAA	.EKKDEG * *
##		181	* **	*				240
##		101	•	•	•	•	•	240
##		241		_				300
	Deer		HFRWGSPI	PKDKRYGGFMT	SEKSLTPLVTLI	- FKTPSSRTPTF	RRASEGAAGR	
##	Human				SEKSQTPLVTLI			
##	Frog	NYRMH	HFRWGSPI	KDKRYGGFMT	PERSQTPLMTLI	FKNAIIKNSH	KKGQ	
##	Sheep	PYKME	HFRWGSPI	PKDKRYGGFMT	SEKSQTPLVTLI	FKNAIIKNAH	KKGQ	
##	Cow	PYKME	HFRWGSPI	KDKRYGGFMT	SEKSQTPLVTLI	FKNAIIKNAH	KKGQ	
##	Goat				SEKSQTPLVTLI		•	
##	Macaque				SEKSQTPLVTLI		-	
	Pig				SEKSQTPLVTLI		-	
	Chimpanzee				SEKSQTPLVTL			
	Mouse				SEKSQTPLVTL		-	
##	Chicken				LEHSQTPLMTLI		-	
##	Horse				SEKSQTPLVTLI *^* ***^**		\r\G \ U=====	
##		241	-1·ጥጥጥ ↑	⊕ጥጥ ጥጥጥጥ	⊸ ጥ ጥጥጥ ላች?	17-17-		300
11.11			•	•	•	•	•	500

```
##
##
          301
                          . 323
           VDPEGLSSALLPPRSPGEDSPRQ
## Deer
## 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



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
```