

Extra Credit

2022-10-06

```
#Install package
renv::install('bio3d')
```

```
## Installing bio3d [2.4-3] ...
## OK [copied cache]
```

```
library('bio3d')
```

```
## Warning: package 'bio3d' was built under R version 4.1.3
```

```
#Load in fasta file from PSI-Blast
aln <- read.fasta(file = "KYE1UJR421S-alignment.fa")
aln
```

```
##          1          .          .          .          .          50
## Query_30566 -----MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLL
## ref|NP_000510.1| -----MVHLTPEEKTAVNALWGKV--NVDAVGGEALGRLL
## ref|NP_000175.1| -----MGHFTTEEDKATITSLWGKV--NVEDAGGETLGRLL
## ref|NP_000509.1| -----MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLL
## ref|NP_000550.2| -----MGHFTTEEDKATITSLWGKV--NVEDAGGETLGRLL
## ref|NP_005321.1| -----MVHFTAEEKAAVTSLSWKM--NVEEAGGEALGRLL
## ref|NP_001349775.1| -----MGLSDGEWQLVLNVWGKVEADIPGHGQEVLRIRLF
## ref|XP_005257062.1| MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYANCEDVGVAAILVRFF
## ref|NP_005323.1| -----MSLTKTERTIIVSMWAKISTQADTIGTETLERLF
## ref|NP_599030.1| MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYANCEDVGVAAILVRFF
## ref|NP_000508.1| -----MVLSPADKTNVKAAWGKVGGAHAGEYGAEALERMF
## ref|NP_005322.1| -----MALSAEDRALVRALWKKLGSNVGVYTTEALERTF
## ref|NP_001003938.1| -----MLSAQERAQIAQVWDLIAGHEAQFGAELLLRLF
## ref|NP_067080.1| -----MERPEPELIRQSWRAVSRSPLEHGTVLFRRLF
## ref|XP_016879605.1| -----
## ref|NP_001369741.1| -----
##
##          1          .          .          .          .          50
##
##          51          .          .          .          .          100
## Query_30566 VVYPWTQRFFE-SFGDLSTPDAVM-GNPKVKAHGKKVLGAFSDGLAHLDN
## ref|NP_000510.1| VVYPWTQRFFE-SFGDLSSPDAVM-GNPKVKAHGKKVLGAFSDGLAHLDN
## ref|NP_000175.1| VVYPWTQRFFD-SFGNLSASAIM-GNPKVKAHGKKVLTSLGDAIKHLDD
## ref|NP_000509.1| VVYPWTQRFFE-SFGDLSTPDAVM-GNPKVKAHGKKVLGAFSDGLAHLDN
## ref|NP_000550.2| VVYPWTQRFFD-SFGNLSASAIM-GNPKVKAHGKKVLTSLGDATKHLDD
## ref|NP_005321.1| VVYPWTQRFFD-SFGNLSSPSAIL-GNPKVKAHGKKVLTSLGDAIKNMDN
## ref|NP_001349775.1| KGHPTLEKFD-KFKHLKSEDEMK-ASEDLKKHGATVLTALGGILKKKGH
## ref|XP_005257062.1| VNFPSAKQYFS-QFKHMEDEPLEME-RSPQLRKHACRVMGALNTVVENLHD
```

```

## ref|NP_005323.1|      LSHPQTKTYFP-HF-----DLHPGSAQLRAHGSKVVAAVGDAVKSIDD
## ref|NP_599030.1|      VNFPSAKQYFS-QFKHMEPLEME-RSPQLRKHACRVMGALNTVVENLHD
## ref|NP_000508.1|      LSFPTTKTYFP-HF-----DLSHGSAQVKGHGKKVADALTNAVAHVDD
## ref|NP_005322.1|      LAFPATKTYFS-H-----LDLSPGSSQVRAHGQKVADALSLAVERLDD
## ref|NP_001003938.1|   TVYPSTKVYFP-HL-----SACQ-DATQLLSHGQRMMLAAVGAAVQHVDN
## ref|NP_067080.1|      ALEPDLLPLFYQNCRFSSPEDCL-SSPEFLDHIRKVMLVIDAAVTNVED
## ref|XP_016879605.1|   -----MEDPLEME-RSPQLRKHACRVMGALNTVVENLHD
## ref|NP_001369741.1|   -----MK-ASEDLKKHGATVLTALGGILKKKGH
##
##                                     *   ^
##
##          51          .          .          .          .          100
##
##
##          101         .          .          .          .          150
##
## Query_30566          LKGT---FATLSELHCDKLHVDPENFRLGNVLCVLAHHFGKEFTPPVQ
## ref|NP_000510.1|      LKGT---FSQSELHCDKLHVDPENFRLGNVLCVLAARNGKEFTPPMQ
## ref|NP_000175.1|      LKGT---FAQLSELHCDKLHVDPENFKLLGNVLTVLAIHFGKEFTPEVQ
## ref|NP_000509.1|      LKGT---FATLSELHCDKLHVDPENFRLGNVLCVLAHHFGKEFTPPVQ
## ref|NP_000550.2|      LKGT---FAQLSELHCDKLHVDPENFKLLGNVLTVLAIHFGKEFTPEVQ
## ref|NP_005321.1|      LKPA---FAKLSELHCDKLHVDPENFKLLGNVMVILATHFGKEFTPEVQ
## ref|NP_001349775.1|   HEAE---IKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQ
## ref|XP_005257062.1|   PDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASDFPPETQ
## ref|NP_005323.1|      IGGA---LSKLSELHAYILRVDPVNFKLLSHCLLVTLAARFPADFTAEAH
## ref|NP_599030.1|      PDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASDFPPETQ
## ref|NP_000508.1|      MPNA---LSALSDLHAHKL RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVH
## ref|NP_005322.1|      LPHA---LSALSHLHACQLRVDPASFQLLGHCLLVTLARHYPGDFSPALQ
## ref|NP_001003938.1|   LRAA---LSPLADLHALVLRVDPANFPLLIQCFHVVLASHLQDEFTVQMQ
## ref|NP_067080.1|      LSSLEEYLASLGRKHRA-VGVKLSSFSTVGESLLYMLEKCLGPAFTPATR
## ref|XP_016879605.1|   PDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFASDFPPETQ
## ref|NP_001369741.1|   HEAE---IKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQ
##
##          ^      *      ^      ^      ^      ^      *
##
##          101         .          .          .          .          150
##
##
##          151         .          .          .          .          200
##
## Query_30566          AAYQKVAVAGVANALAHKYH-----
## ref|NP_000510.1|      AAYQKVAVAGVANALAHKYH-----
## ref|NP_000175.1|      ASWQKMVTGVASALSSRYH-----
## ref|NP_000509.1|      AAYQKVAVAGVANALAHKYH-----
## ref|NP_000550.2|      ASWQKMVTAVASALSSRYH-----
## ref|NP_005321.1|      AAWQKLVSATAIALAHKYH-----
## ref|NP_001349775.1|   GAMNKALELFRKDMASNYKELGFQG-----
## ref|XP_005257062.1|   RAWAKLRGLIYSHVTAAYKEVGWVQVQPNATTHSSWRRSPEGSWGRQASC
## ref|NP_005323.1|      AAWDKFLSVVSSVLTEKYR-----
## ref|NP_599030.1|      RAWAKLRGLIYSHVTAAYKEVGWVQVQPNATTPP-----ATL
## ref|NP_000508.1|      ASLDKFLASVSTVLTSEYR-----
## ref|NP_005322.1|      ASLDKFLSHVISALVSEYR-----
## ref|NP_001003938.1|   AAWDKFLTGVAVVLTEKYR-----
## ref|NP_067080.1|      AAWSQLYGAVVQAMSRGWDGE-----
## ref|XP_016879605.1|   RAWAKLRGLIYSHVTAAYKEVGWVQVQPNATTHSSWRRSPEGSWGRQASC
## ref|NP_001369741.1|   GAMNKALELFRKDMASNYKELGFQG-----
##
##          ^      ^
##
##          151         .          .          .          .          200
##
##
##          201      205
##
## Query_30566          -----
## ref|NP_000510.1|      -----

```

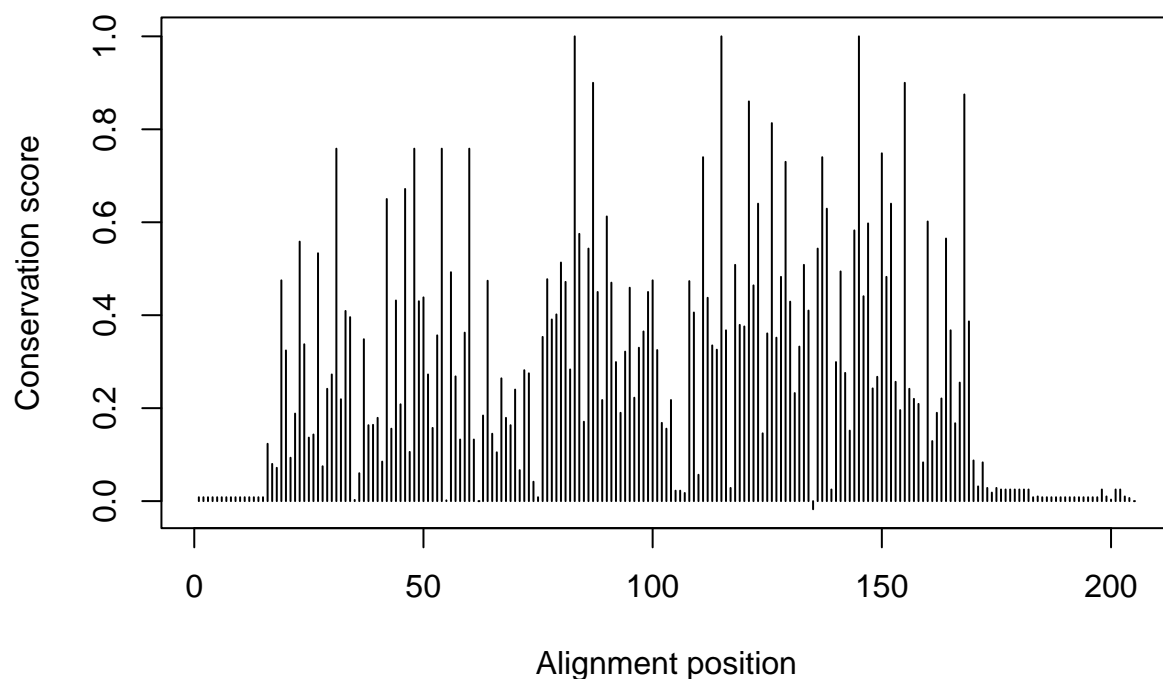
```

## ref|NP_000175.1|      -----
## ref|NP_000509.1|      -----
## ref|NP_000550.2|      -----
## ref|NP_005321.1|      -----
## ref|NP_001349775.1|    -----
## ref|XP_005257062.1|    PSSC-
## ref|NP_005323.1|      -----
## ref|NP_599030.1|      PSSGP
## ref|NP_000508.1|      -----
## ref|NP_005322.1|      -----
## ref|NP_001003938.1|    -----
## ref|NP_067080.1|      -----
## ref|XP_016879605.1|    PSSC-
## ref|NP_001369741.1|    -----
##
##                      201    205
##
## Call:
##   read.fasta(file = "KYE1UJR421S-alignment.fa")
##
## Class:
##   fasta
##
## Alignment dimensions:
##   16 sequence rows; 205 position columns (92 non-gap, 113 gap)
##
## + attr: id, ali, call

#Step 2: Identifying conserved positions
sim <- conserv(aln)

#Alignment plot
plot(sim, typ="h", xlab="Alignment position", ylab="Conservation score")

```



```
#Ordering of most conserved AA's
inds <- order(sim, decreasing=TRUE)
head(sim[inds])
```

```
## [1] 1.000 1.000 1.000 0.900 0.900 0.875
```

```
#Make a data frame
positions <- data.frame(pos=1:length(sim),
aa=aln$ali[1,],
score=sim)
head(positions)
```

```
##   pos aa      score
## 1   1 - 0.008333333
## 2   2 - 0.008333333
## 3   3 - 0.008333333
## 4   4 - 0.008333333
## 5   5 - 0.008333333
## 6   6 - 0.008333333
```

```
#Most conserved positions
head( positions[inds,] )
```

```
##      pos aa score
```

```
## 83 83 H 1.000
## 115 115 H 1.000
## 145 145 F 1.000
## 87 87 V 0.900
## 155 155 K 0.900
## 168 168 Y 0.875
```

```
aa123(positions[inds,]$aa)[1:3]
```

```
## [1] "HIS" "HIS" "PHE"
```

```
#Step 3: Relationship between sequences
```

```
install.packages("pheatmap")
```

```
## Installing package into 'C:/Users/laryb/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'pheatmap' successfully unpacked and MD5 sums checked
```

```
##
```

```
## The downloaded binary packages are in
```

```
## C:\Users\laryb\AppData\Local\Temp\Rtmp2LHiN6\downloaded_packages
```

```
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.1.3
```

```
ide <- seqidentity(aln)
pheatmap((1-ide))
```

