

Graphs for Find a Gene

Morgan Farrell

2/22/2022

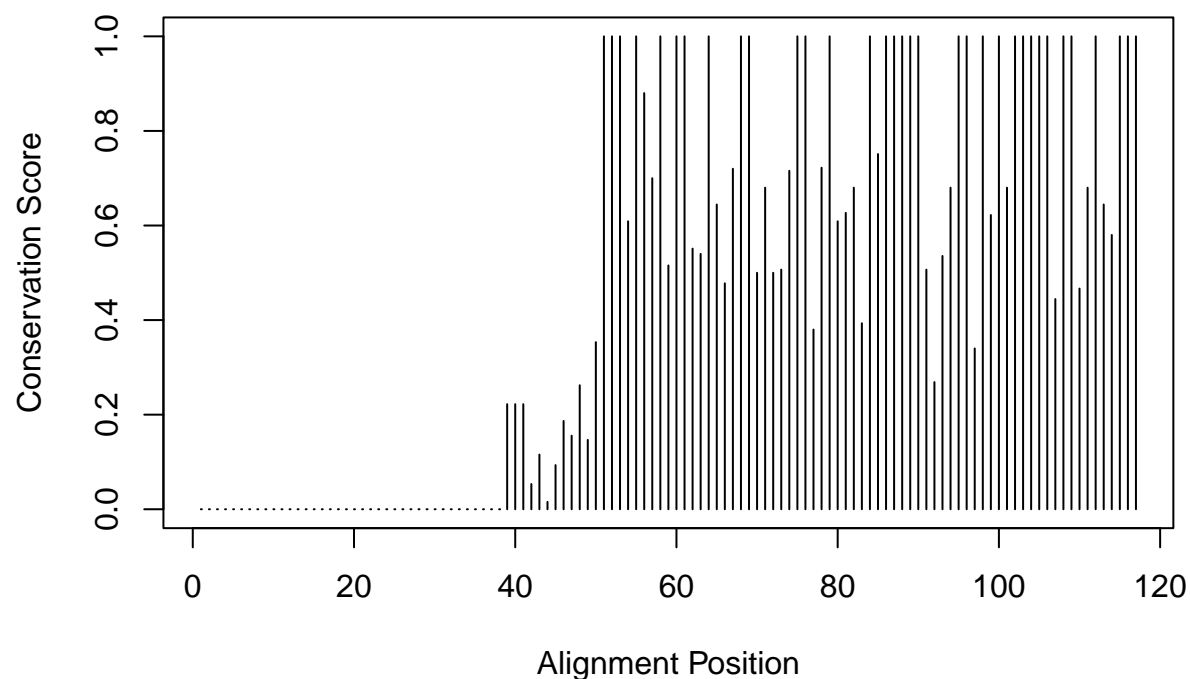
Install proper packages

```
library(bio3d)
#install.packages("pheatmap")
library(pheatmap)
```

Read in the alignment that was converted in SEAVIEW from the original CLUSTALW format

```
romo1 <- read.fasta("alignment-converted fasta.fst")
romo1
```

```
##                               1           .           .           .           40
## [Truncated_Name:1]Human      -----MP
## [Truncated_Name:2]Mouse      -----MP
## [Truncated_Name:3]Brown_Sea_ -----
## [Truncated_Name:4]Starlet_Se -----
## [Truncated_Name:5]Cauliflowe -----
## [Truncated_Name:6]Elkhorn_Co -----
## [Truncated_Name:7]Finger_Cor -----
## [Truncated_Name:8]Great_Scal MYQYTWQNITFSNRKCVREKTLPVWTCLSARLISQTAEMP
## [Truncated_Name:9]Brown_Dog_ -----MP
## [Truncated_Name:10]Scarce_Cha -----MP
##
##                               1           .           .           .           40
##
##                               41           .           .           .           80
## [Truncated_Name:1]Human      VAVGPYGGSQSPSCFDRVKMGFVMGCAVGMAAGALFGTFSC
## [Truncated_Name:2]Mouse      VAVGPYGGSQSPSCFDRVKMGFVMGCAVGMAAGALFGTFSC
## [Truncated_Name:3]Brown_Sea_ ----MQRPQPPSCFDRVKVGFMMGFVAVGMSSAGLFGMYSA
## [Truncated_Name:4]Starlet_Se ----MQRAPQPSCFDRVKVGFMMGFVAVGMSSAGLFGTYSA
## [Truncated_Name:5]Cauliflowe ----MQRAPQPSCFDRVKVGFMMGFVAVGMSSAALFGTYSA
## [Truncated_Name:6]Elkhorn_Co -----QPSCFDRVKVGFMMGFVAVGMSSAALFGTYSA
## [Truncated_Name:7]Finger_Cor ---MMRAPQPSCFDRVKVGFMMGFVAVGMSSAALFGTYSA
## [Truncated_Name:8]Great_Scal VGNIGYGGSQSPSCFDKMKIGFMMGFSVGMASGIIFFGGFSA
## [Truncated_Name:9]Brown_Dog_ VPVQGGYQRGPSCWDRVKMGFTIGFCVGMASGALFGGFSA
```

Begin building heatmap based on conservation scores

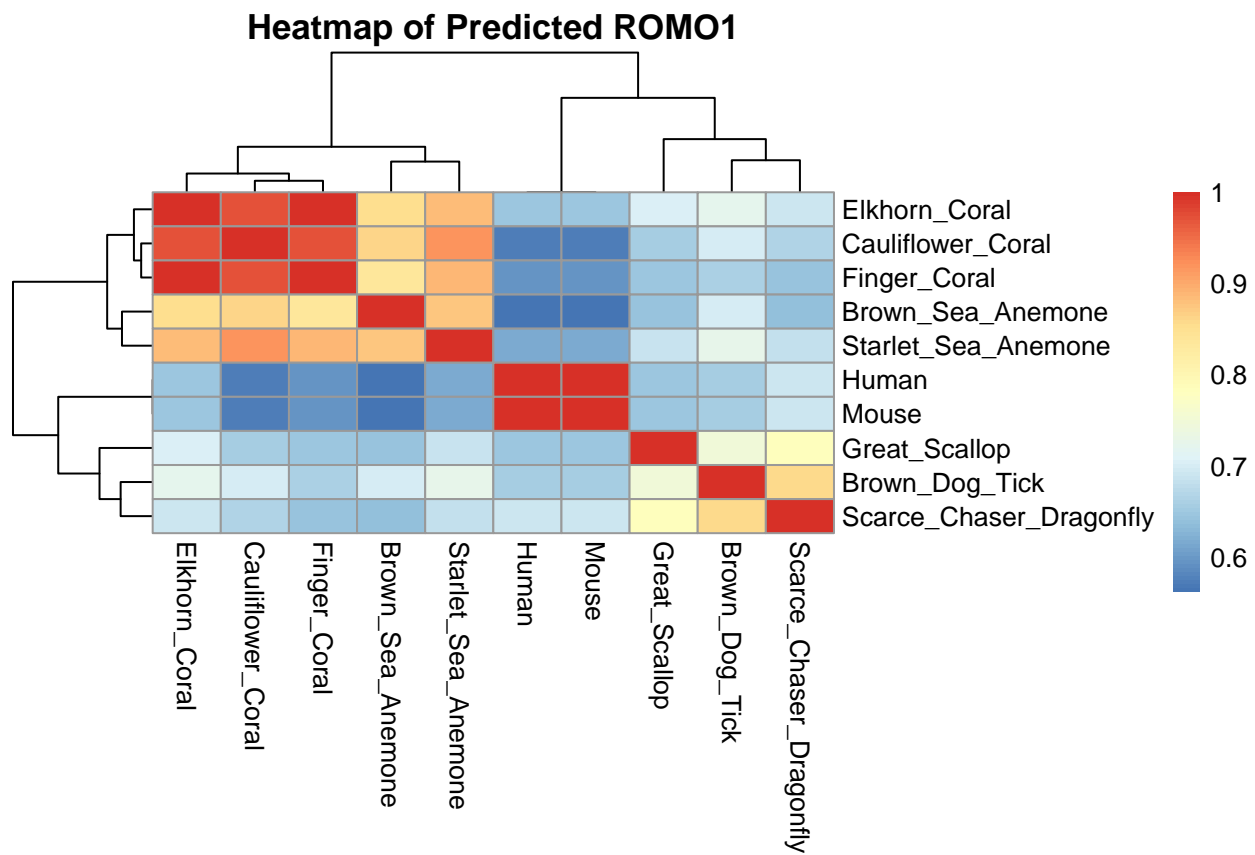
```
identity_romo1 <- seqidentity(romo1)
identity_romo1
```

```
##           Human Mouse Brown_Sea_Anemone Starlet_Sea_Anemone
## Human      1.000 1.000                0.562                0.616
## Mouse      1.000 1.000                0.562                0.616
## Brown_Sea_Anemone 0.562 0.562            1.000                0.877
## Starlet_Sea_Anemone 0.616 0.616            0.877                1.000
## Cauliflower_Coral 0.575 0.575            0.863                0.918
## Elkhorn_Coral 0.647 0.647            0.853                0.882
## Finger_Coral 0.595 0.595            0.836                0.890
## Great_Scallop 0.646 0.646            0.644                0.685
## Brown_Dog_Tick 0.658 0.658            0.699                0.726
## Scarce_Chaser_Dragonfly 0.692 0.692            0.639                0.681
##           Cauliflower_Coral Elkhorn_Coral Finger_Coral
## Human              0.575            0.647            0.595
## Mouse              0.575            0.647            0.595
## Brown_Sea_Anemone 0.863            0.853            0.836
## Starlet_Sea_Anemone 0.918            0.882            0.890
## Cauliflower_Coral 1.000            0.971            0.973
## Elkhorn_Coral     0.971            1.000            1.000
```

```
## Finger_Coral          0.973          1.000          1.000
## Great_Scallop         0.658          0.706          0.649
## Brown_Dog_Tick       0.699          0.721          0.662
## Scarce_Chaser_Dragonfly 0.667          0.691          0.644
##
##           Great_Scallop Brown_Dog_Tick Scarce_Chaser_Dragonfly
## Human          0.646          0.658          0.692
## Mouse          0.646          0.658          0.692
## Brown_Sea_Anemone 0.644          0.699          0.639
## Starlet_Sea_Anemone 0.685          0.726          0.681
## Cauliflower_Coral  0.658          0.699          0.667
## Elkhorn_Coral     0.706          0.721          0.691
## Finger_Coral     0.649          0.662          0.644
## Great_Scallop     1.000          0.747          0.782
## Brown_Dog_Tick    0.747          1.000          0.859
## Scarce_Chaser_Dragonfly 0.782          0.859          1.000
```

Assemble heat map

```
heatmap(identity_romo1, main="Heatmap of Predicted ROM01", scale="none")
```



```
#Make consensus sequence from alignment
```

```
#cons_romo1 <- consensus(romo1)
#cons_romo1
```

My consensus sequence has too many gaps to use. I'll need to find the sequence with the highest identity to the others and use that

```
#Find highest identity sequence
```

```
#install.packages("matrixStats")
library(matrixStats)
```

```
romo1_wo_self <- 1-identity_romo1
romo1_noself_matrix <- as.matrix(romo1_wo_self)
romo1_noself_matrix
```

```
##           Human Mouse Brown_Sea_Anemone Starlet_Sea_Anemone
## Human      0.000 0.000             0.438             0.384
## Mouse      0.000 0.000             0.438             0.384
## Brown_Sea_Anemone 0.438 0.438             0.000             0.123
## Starlet_Sea_Anemone 0.384 0.384             0.123             0.000
## Cauliflower_Coral 0.425 0.425             0.137             0.082
## Elkhorn_Coral 0.353 0.353             0.147             0.118
## Finger_Coral 0.405 0.405             0.164             0.110
## Great_Scallop 0.354 0.354             0.356             0.315
## Brown_Dog_Tick 0.342 0.342             0.301             0.274
## Scarce_Chaser_Dragonfly 0.308 0.308             0.361             0.319
##           Cauliflower_Coral Elkhorn_Coral Finger_Coral
## Human              0.425             0.353             0.405
## Mouse              0.425             0.353             0.405
## Brown_Sea_Anemone 0.137             0.147             0.164
## Starlet_Sea_Anemone 0.082             0.118             0.110
## Cauliflower_Coral 0.000             0.029             0.027
## Elkhorn_Coral 0.029             0.000             0.000
## Finger_Coral 0.027             0.000             0.000
## Great_Scallop 0.342             0.294             0.351
## Brown_Dog_Tick 0.301             0.279             0.338
## Scarce_Chaser_Dragonfly 0.333             0.309             0.356
##           Great_Scallop Brown_Dog_Tick Scarce_Chaser_Dragonfly
## Human              0.354             0.342             0.308
## Mouse              0.354             0.342             0.308
## Brown_Sea_Anemone 0.356             0.301             0.361
## Starlet_Sea_Anemone 0.315             0.274             0.319
## Cauliflower_Coral 0.342             0.301             0.333
## Elkhorn_Coral 0.294             0.279             0.309
## Finger_Coral 0.351             0.338             0.356
## Great_Scallop 0.000             0.253             0.218
## Brown_Dog_Tick 0.253             0.000             0.141
## Scarce_Chaser_Dragonfly 0.218             0.141             0.000
```

```
#install.packages("dplyr")
#library(dplyr)
```

```
rom01_noself_matrix$row_maximum = rowMaxs(rom01_noself_matrix)
```

```
## Warning in rom01_noself_matrix$row_maximum = rowMaxs(rom01_noself_matrix):  
## Coercing LHS to a list
```

```
rom01_noself_matrix
```

```
## [[1]]  
## [1] 0  
##  
## [[2]]  
## [1] 0  
##  
## [[3]]  
## [1] 0.438  
##  
## [[4]]  
## [1] 0.384  
##  
## [[5]]  
## [1] 0.425  
##  
## [[6]]  
## [1] 0.353  
##  
## [[7]]  
## [1] 0.405  
##  
## [[8]]  
## [1] 0.354  
##  
## [[9]]  
## [1] 0.342  
##  
## [[10]]  
## [1] 0.308  
##  
## [[11]]  
## [1] 0  
##  
## [[12]]  
## [1] 0  
##  
## [[13]]  
## [1] 0.438  
##  
## [[14]]  
## [1] 0.384  
##  
## [[15]]  
## [1] 0.425  
##  
## [[16]]
```

```

## [1] 0.353
##
## [[17]]
## [1] 0.405
##
## [[18]]
## [1] 0.354
##
## [[19]]
## [1] 0.342
##
## [[20]]
## [1] 0.308
##
## [[21]]
## [1] 0.438
##
## [[22]]
## [1] 0.438
##
## [[23]]
## [1] 0
##
## [[24]]
## [1] 0.123
##
## [[25]]
## [1] 0.137
##
## [[26]]
## [1] 0.147
##
## [[27]]
## [1] 0.164
##
## [[28]]
## [1] 0.356
##
## [[29]]
## [1] 0.301
##
## [[30]]
## [1] 0.361
##
## [[31]]
## [1] 0.384
##
## [[32]]
## [1] 0.384
##
## [[33]]
## [1] 0.123
##
## [[34]]

```

```

## [1] 0
##
## [[35]]
## [1] 0.082
##
## [[36]]
## [1] 0.118
##
## [[37]]
## [1] 0.11
##
## [[38]]
## [1] 0.315
##
## [[39]]
## [1] 0.274
##
## [[40]]
## [1] 0.319
##
## [[41]]
## [1] 0.425
##
## [[42]]
## [1] 0.425
##
## [[43]]
## [1] 0.137
##
## [[44]]
## [1] 0.082
##
## [[45]]
## [1] 0
##
## [[46]]
## [1] 0.029
##
## [[47]]
## [1] 0.027
##
## [[48]]
## [1] 0.342
##
## [[49]]
## [1] 0.301
##
## [[50]]
## [1] 0.333
##
## [[51]]
## [1] 0.353
##
## [[52]]

```



```
## [1] 0.353
##
## [[53]]
## [1] 0.147
##
## [[54]]
## [1] 0.118
##
## [[55]]
## [1] 0.029
##
## [[56]]
## [1] 0
##
## [[57]]
## [1] 0
##
## [[58]]
## [1] 0.294
##
## [[59]]
## [1] 0.279
##
## [[60]]
## [1] 0.309
##
## [[61]]
## [1] 0.405
##
## [[62]]
## [1] 0.405
##
## [[63]]
## [1] 0.164
##
## [[64]]
## [1] 0.11
##
## [[65]]
## [1] 0.027
##
## [[66]]
## [1] 0
##
## [[67]]
## [1] 0
##
## [[68]]
## [1] 0.351
##
## [[69]]
## [1] 0.338
##
## [[70]]
```

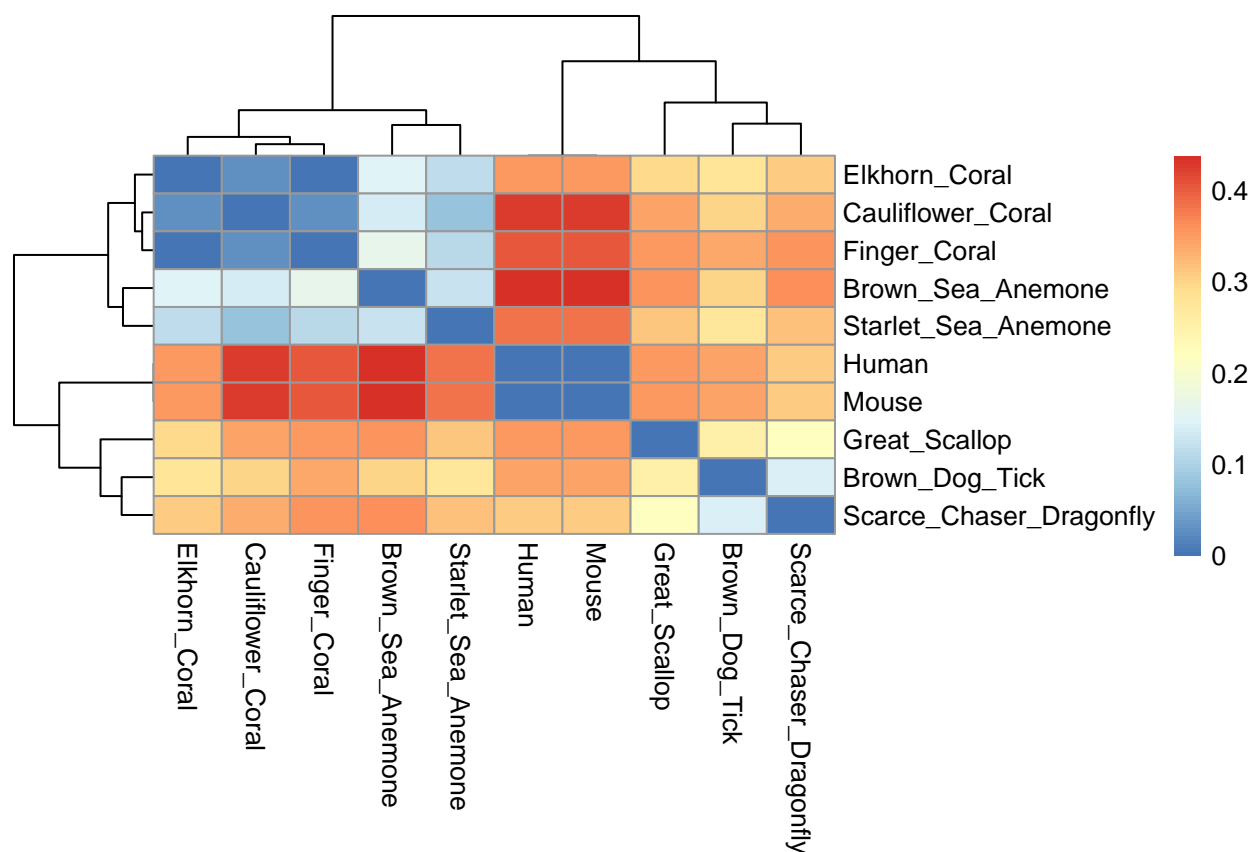
```
## [1] 0.356
##
## [[71]]
## [1] 0.354
##
## [[72]]
## [1] 0.354
##
## [[73]]
## [1] 0.356
##
## [[74]]
## [1] 0.315
##
## [[75]]
## [1] 0.342
##
## [[76]]
## [1] 0.294
##
## [[77]]
## [1] 0.351
##
## [[78]]
## [1] 0
##
## [[79]]
## [1] 0.253
##
## [[80]]
## [1] 0.218
##
## [[81]]
## [1] 0.342
##
## [[82]]
## [1] 0.342
##
## [[83]]
## [1] 0.301
##
## [[84]]
## [1] 0.274
##
## [[85]]
## [1] 0.301
##
## [[86]]
## [1] 0.279
##
## [[87]]
## [1] 0.338
##
## [[88]]
```

```

## [1] 0.253
##
## [[89]]
## [1] 0
##
## [[90]]
## [1] 0.141
##
## [[91]]
## [1] 0.308
##
## [[92]]
## [1] 0.308
##
## [[93]]
## [1] 0.361
##
## [[94]]
## [1] 0.319
##
## [[95]]
## [1] 0.333
##
## [[96]]
## [1] 0.309
##
## [[97]]
## [1] 0.356
##
## [[98]]
## [1] 0.218
##
## [[99]]
## [1] 0.141
##
## [[100]]
## [1] 0
##
## $row_maximum
## [1] 0.438 0.438 0.438 0.384 0.425 0.353 0.405 0.356 0.342 0.361

```

```
pheatmap(1-identity_roma1)
```



I think the brown sea anemone has the highest sequence identity

#R/Bio3D to blast online, find the top 3 unique hi

```
romo1_bsa_fasta <- read.fasta("brown sea anemone.fst")
romo1_bsa_blast <- blast.pdb(romo1_bsa_fasta)
```

```
## Searching ... please wait (updates every 5 seconds) RID = 1DHU5U6Z016
## .
## Reporting 4 hits
```

```
romo1_bsa_blast
```

```
## $hit.tbl
##      queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_23705 1ZZH_A   33.929           56          32         2         1
## 2 Query_23705 6BUM_A   39.535           43          23         2        21
## 3 Query_23705 3JOG_M   28.125           32          23         0         5
## 4 Query_23705 1IK6_A   28.814           59          39         1        16
##      q.end s.start s.end evalue bitscore positives mlog.evalue pdb.id   acc
## 1      51      18    73    6.8    25.8    51.79    -1.916923 1ZZH_A 1ZZH_A
## 2      63      51    90    7.3    25.8    53.49    -1.987874 6BUM_A 6BUM_A
## 3      36      18    49    8.4    24.6    46.88    -2.128232 3JOG_M 3JOG_M
## 4      71     108   166    9.6    25.4    52.54    -2.261763 1IK6_A 1IK6_A
##
## $raw
```

```
##          queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_23705      1ZZH_A   33.929              56          32          2          1
## 2 Query_23705      6BUM_A   39.535              43          23          2         21
## 3 Query_23705      3JOG_M   28.125              32          23          0          5
## 4 Query_23705      1IK6_A   28.814              59          39          1         16
##    q.end s.start s.end evalue bitscore positives
## 1    51      18    73    6.8     25.8     51.79
## 2    63      51    90    7.3     25.8     53.49
## 3    36      18    49    8.4     24.6     46.88
## 4    71     108   166    9.6     25.4     52.54
##
## $url
##
## "https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tabular&RES
##
## attr(,"class")
## [1] "blast"
```

#Try again with different sequence to see if I can get higher e values

```
romo1_cc_fasta <- read.fasta("cauliflower coral.fst")
romo1_cc_blast <- blast.pdb(romo1_cc_fasta)
```

```
## Searching ... please wait (updates every 5 seconds) RID = 1DHUDP73016
## .....
## Reporting 19 hits
```

```
romo1_cc_blast
```

```
## $hit.tbl
##          queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_30073      1M4A_A   37.037              27          17          0         39
## 2 Query_30073      1M4B_A   35.714              28          18          0         38
## 3 Query_30073      1IRL_A   37.037              27          17          0         39
## 4 Query_30073      1M47_A   37.037              27          17          0         39
## 5 Query_30073      2ERJ_D   37.037              27          17          0         39
## 6 Query_30073      7M2G_A   37.037              27          17          0         39
## 7 Query_30073      6YE3_C   37.037              27          17          0         39
## 8 Query_30073      1PY2_A   37.037              27          17          0         39
## 9 Query_30073      1QVN_A   37.037              27          17          0         39
## 10 Query_30073     5LQB_A   37.037              27          17          0         39
## 11 Query_30073     5A8F_A   30.189              53          34          2         14
## 12 Query_30073     4NEJ_A   37.037              27          17          0         39
## 13 Query_30073     5M5E_D   35.714              28          18          0         38
## 14 Query_30073     4AOG_A   28.125              32          23          0          5
## 15 Query_30073     1HL8_A   40.476              42          22          2         30
## 16 Query_30073     2WSP_A   40.476              42          22          2         30
## 17 Query_30073     1ODU_A   40.476              42          22          2         30
## 18 Query_30073     2ZWY_A   40.476              42          22          2         30
## 19 Query_30073     5NUG_A   38.462              52          22          3         10
##    q.end s.start s.end evalue bitscore positives mlog.evalue pdb.id    acc
## 1    65      78   104    1.8     27.3     66.67 -0.5877867 1M4A_A 1M4A_A
## 2    65      77   104    1.8     27.3     64.29 -0.5877867 1M4B_A 1M4B_A
```

```

## 3      65      78     104     1.9     27.3     66.67 -0.6418539 1IRL_A 1IRL_A
## 4      65      78     104     1.9     27.3     66.67 -0.6418539 1M47_A 1M47_A
## 5      65      78     104     2.0     27.3     66.67 -0.6931472 2ERJ_D 2ERJ_D
## 6      65      78     104     2.0     27.3     66.67 -0.6931472 7M2G_A 7M2G_A
## 7      65      79     105     2.0     27.3     66.67 -0.6931472 6YE3_C 6YE3_C
## 8      65      78     104     2.0     27.3     66.67 -0.6931472 1PY2_A 1PY2_A
## 9      65      78     104     2.0     27.3     66.67 -0.6931472 1QVN_A 1QVN_A
## 10     65      77     103     2.1     26.9     66.67 -0.7419373 5LQB_A 5LQB_A
## 11     63     146     198     2.2     27.3     50.94 -0.7884574 5A8F_A 5A8F_A
## 12     65      75     101     2.3     26.9     66.67 -0.8329091 4NEJ_A 4NEJ_A
## 13     65      96     123     2.7     26.9     64.29 -0.9932518 5M5E_D 5M5E_D
## 14     36     657     688     4.0     26.6     62.50 -1.3862944 4AOG_A 4AOG_A
## 15     68     134     175     6.2     26.2     57.14 -1.8245493 1HL8_A 1HL8_A
## 16     68     134     175     6.2     26.2     57.14 -1.8245493 2WSP_A 2WSP_A
## 17     68     134     175     6.2     26.2     57.14 -1.8245493 1ODU_A 1ODU_A
## 18     68     134     175     6.2     26.2     57.14 -1.8245493 2ZWY_A 2ZWY_A
## 19     51    3868    3919     9.0     25.8     53.85 -2.1972246 5NUG_A 5NUG_A
##
## $raw
##      queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_30073    1M4A_A   37.037                27          17         0        39
## 2 Query_30073    1M4B_A   35.714                28          18         0        38
## 3 Query_30073    1IRL_A   37.037                27          17         0        39
## 4 Query_30073    1M47_A   37.037                27          17         0        39
## 5 Query_30073    2ERJ_D   37.037                27          17         0        39
## 6 Query_30073    7M2G_A   37.037                27          17         0        39
## 7 Query_30073    6YE3_C   37.037                27          17         0        39
## 8 Query_30073    1PY2_A   37.037                27          17         0        39
## 9 Query_30073    1QVN_A   37.037                27          17         0        39
## 10 Query_30073   5LQB_A   37.037                27          17         0        39
## 11 Query_30073   5A8F_A   30.189                53          34         2        14
## 12 Query_30073   4NEJ_A   37.037                27          17         0        39
## 13 Query_30073   5M5E_D   35.714                28          18         0        38
## 14 Query_30073   4AOG_A   28.125                32          23         0         5
## 15 Query_30073   1HL8_A   40.476                42          22         2        30
## 16 Query_30073   2WSP_A   40.476                42          22         2        30
## 17 Query_30073   1ODU_A   40.476                42          22         2        30
## 18 Query_30073   2ZWY_A   40.476                42          22         2        30
## 19 Query_30073   5NUG_A   38.462                52          22         3        10
##      q.end s.start s.end evalule bitscore positives
## 1      65      78     104     1.8     27.3     66.67
## 2      65      77     104     1.8     27.3     64.29
## 3      65      78     104     1.9     27.3     66.67
## 4      65      78     104     1.9     27.3     66.67
## 5      65      78     104     2.0     27.3     66.67
## 6      65      78     104     2.0     27.3     66.67
## 7      65      79     105     2.0     27.3     66.67
## 8      65      78     104     2.0     27.3     66.67
## 9      65      78     104     2.0     27.3     66.67
## 10     65      77     103     2.1     26.9     66.67
## 11     63     146     198     2.2     27.3     50.94
## 12     65      75     101     2.3     26.9     66.67
## 13     65      96     123     2.7     26.9     64.29
## 14     36     657     688     4.0     26.6     62.50

```

```
## 15      68      134      175      6.2      26.2      57.14
## 16      68      134      175      6.2      26.2      57.14
## 17      68      134      175      6.2      26.2      57.14
## 18      68      134      175      6.2      26.2      57.14
## 19      51     3868     3919      9.0      25.8      53.85
##
## $url
##
## "https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tabular&RES
##
## attr(,"class")
## [1] "blast"
```

```
romol_m_fasta <- read.fasta("mouse.fst")
romol_m_blast <- blast.pdb(romol_m_fasta)
```

```
## Searching ... please wait (updates every 5 seconds) RID = 1DHWPMW5013
## .
## Reporting 4 hits
```

```
romol_m_blast
```

```
## $hit.tbl
##      queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_25757 1PWA_A 43.750          32          16          1          6
## 2 Query_25757 6KTR_C 43.750          32          16          1          6
## 3 Query_25757 2P23_A 43.750          32          16          1          6
## 4 Query_25757 6YII_A 28.571          35          25          0         29
##      q.end s.start s.end evalue bitscore positives mlog.evalue pdb.id  acc
## 1      37      17      46      1.9      27.7      62.5 -0.6418539 1PWA_A 1PWA_A
## 2      37      29      58      2.4      27.3      62.5 -0.8754687 6KTR_C 6KTR_C
## 3      37      29      58      2.4      27.3      62.5 -0.8754687 2P23_A 2P23_A
## 4      63     390     424      7.6      26.2      60.0 -2.0281482 6YII_A 6YII_A
##
## $raw
##      queryid subjectids identity alignmentlength mismatches gapopens q.start
## 1 Query_25757 1PWA_A 43.750          32          16          1          6
## 2 Query_25757 6KTR_C 43.750          32          16          1          6
## 3 Query_25757 2P23_A 43.750          32          16          1          6
## 4 Query_25757 6YII_A 28.571          35          25          0         29
##      q.end s.start s.end evalue bitscore positives
## 1      37      17      46      1.9      27.7      62.5
## 2      37      29      58      2.4      27.3      62.5
## 3      37      29      58      2.4      27.3      62.5
## 4      63     390     424      7.6      26.2      60.0
##
## $url
##
## "https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tabular&RES
##
## attr(,"class")
## [1] "blast"
```