BIMM 143 Find-A-Gene Project

Kira Jung

Date: 2023-06-15
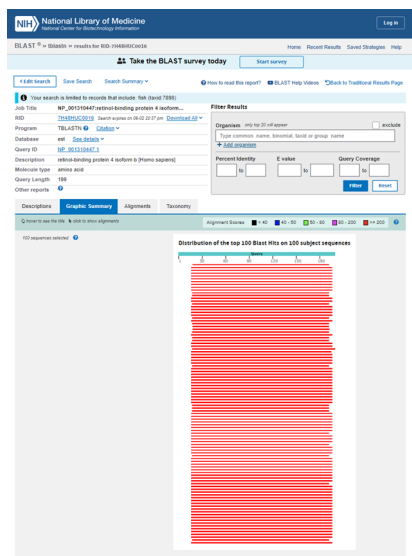
Q1.

Name: Retinol binding protein 4 (RBP4) isoform b Accession: NP_001310447 Species: Homo Sapiens

Q2.

Method: TBLASTN (2.7.1) search against fish ESTs Database: Expressed Sequence Tags (est) Organism: fish (Taxis: 7898)

```
knitr::include_graphics("Q2_1.png")
```

```r
knitr::include_graphics("Q2_2.png")
```



```r
knitr::include_graphics("Q2_3.png")
```

Q3.

Chosen sequence: >CN976133.1 26058_125_035_F12 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to Plasma retinol-binding protein I (PRBP-I), mRNA sequence TGCAGTGGGCTTCTGGCGCTGTCCTGGGCACAGGATTGCAAGGTAGAGAA-CATCCAGGTCAAGCAGGATT TTGACAAAACCAGGTATGCAGGGACCTGGTAT-GCTGTGGGAAAGAAGGACCCAGAGGGATTGTTTTTACT TGACAACGTCGTAGC-CCAGTTTAACATCGAGGATGATGGCAGAATGACCGCGACCGCAAAGGGCAGAGTC ATCATTCTGAACAACTGGGAAATGTGTGCCCACATGTTTGCTACCTTCGAGGGATC-TACCGACCCTGCCA AGTTCAGGATGAAGTACTGGGGAGCGGCTTCATACCTGCA-GACAGGGAATGACGAACACTGGGTTATCGA CACCGACTACGACAACTACGC-CATCCACTACTCCTGCAGACTGCTCGACGCAGACGGCACGTGCTTGGAC AGC-TACTCCTTCGTCTTCTCCCGTCACCCGACGGGCCTGAGGCCAGAGGACCAGGC-CATCATCACGCAGA AGAAGATGGACCTCTGCCTTTTGGGCAAATACAGACGCG-TAGCGCACAACGGTTTCTGCGAGAACAGTGC GACGCAGAGCCGCCTCAGT-GACACCCTTCCTGATGCTCATATTTTCGTTCTGCTGCTTCAGATTGATCCA TCTCTCTCATCTACGTCTGTTCTCTCGAGAGAGAGTAGTCTTTCTTCGCTCTGC-TATAAGTACACTACAC TCAAAAAAACATACACAAAACCTCGCAGGTATTCGCTTGA-CATAATCTACATCCTGAGTCA

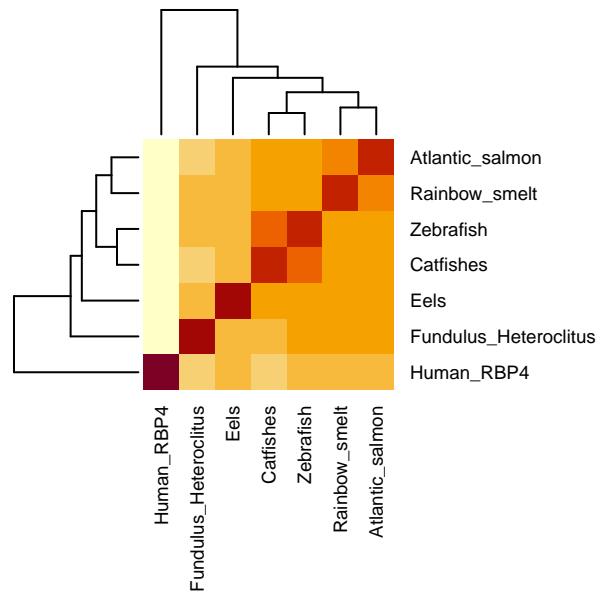## Generating a heatplot for Question 7:

After reading the muscle alignment file into Seaview and saving the file as FASTA:

```
library(bio3d)

muscle_al <- read.fasta("bioinfproj")

identity <- seqidentity(muscle_al)

heatmap(identity, margins = c(9, 9), cexRow = 0.7, cexCol = 0.7)
```

3

Q8.

First calculating the sum of each row in the identity matrix to determine which sequence has the highest identity in the alignment:

```
sums <- rowSums(identity)
which.max(sums)
```

```
Zebrafish
        4
```

In the main protein structure database, using the original sequence to search for the most similar atomic resolution structure to the aligned sequences.

```
human_rbp <- read.fasta("sequence.fasta")
hits <- blast.pdb(human_rbp, database = "pdb")
```

```
 Searching ... please wait (updates every 5 seconds) RID = 8SACAKTZ016
 .................................................................................
 Reporting 43 hits
```

```
head(hits, 3)
```

$hit.tbl
```
      queryid subjectids identity alignmentlength mismatches gapopens q.start
1  Query_91489     4O9S_A   95.408             196          7        1       6
2  Query_91489     3FMZ_A   95.408             196          7        1       6
3  Query_91489     6QBA_A  100.000             183          0        0      17
4  Query_91489     1JYD_A  100.000             182          0        0      17
5  Query_91489     1BRP_A  100.000             182          0        0      17
6  Query_91489     1JYJ_A   98.901             182          2        0      17
7  Query_91489     1QAB_E   98.333             180          3        0      20
8  Query_91489     2WQA_E  100.000             176          0        0      17
9  Query_91489     3BSZ_E  100.000             176          0        0      17
10 Query_91489     2WQ9_A  100.000             174          0        0      17
11 Query_91489     2WR6_A  100.000             174          0        0      17
12 Query_91489     1AQB_A   92.896             183         13        0      17
13 Query_91489     1HBQ_A   92.350             183         14        0      17
14 Query_91489     1ERB_A   92.350             183         14        0      17
15 Query_91489     1KT5_A   93.143             175         12        0      17
16 Query_91489     1RLB_E   92.529             174         13        0      17
17 Query_91489     1IIU_A   87.209             172         22        0      19
18 Query_91489     7VNS_A   26.744             172        111        7      27
19 Query_91489   7O2Y_AAA   26.744             172        111        7      27
20 Query_91489     5EZ2_A   26.744             172        111        7      27
21 Query_91489     5F6Z_A   30.714             140         84        6      27
22 Query_91489     7VNL_A   26.744             172        111        7      27
23 Query_91489     7YX1_A   27.326             172        110        7      27
24 Query_91489     7O3K_A   26.163             172        112        7      27
25 Query_91489     2HZQ_A   27.273             143         83        7      27
26 Query_91489     2NND_A   23.684             152         93        6      30
27 Query_91489     1IW2_A   26.829             123         80        4      22
28 Query_91489     2OVD_A   25.410             122         83        3      22
29 Query_91489     2QOS_C   25.439             114         77        3      22
30 Query_91489     2OVA_A   25.410             122         83        3      22
31 Query_91489     4ES7_A   23.140             121         65        5      24
32 Query_91489     3OJY_C   24.561             114         78        3      22
33 Query_91489     3QKG_A   23.967             121         64        5      24
34 Query_91489     8B0F_F   24.561             114         78        3      22
35 Query_91489     2RD7_C   24.561             114         78        3      22
36 Query_91489     1GKA_B   23.034             178        119        7      21
37 Query_91489     5NGH_A   23.226             155         93        6      27
38 Query_91489     2K23_A   21.368             117         77        3      24
```

39	Query_91489	1JZU_A	22.302	139	96	4	32
40	Query_91489	1Z24_A	28.440	109	70	4	30
41	Query_91489	5X7Y_A	22.078	154	87	6	27
42	Query_91489	2XST_A	23.423	111	71	3	27
43	Query_91489	6NRE_A	22.078	154	87	6	27

	q.end	s.start	s.end	evalue	bitscore	positives	mlog.evalue	pdb.id	acc
1	199	20	215	1.31e-139	389.0	95.92	319.7893008	4O9S_A	4O9S_A
2	199	17	212	1.90e-139	389.0	95.92	319.4174740	3FMZ_A	3FMZ_A
3	199	3	185	1.60e-138	385.0	100.00	317.2867392	6QBA_A	6QBA_A
4	198	2	183	4.72e-138	384.0	100.00	316.2049340	1JYD_A	1JYD_A
5	198	1	182	5.56e-138	384.0	100.00	316.0411447	1BRP_A	1BRP_A
6	198	2	183	1.83e-135	377.0	98.90	310.2446716	1JYJ_A	1JYJ_A
7	199	1	180	6.78e-134	373.0	98.33	306.6324254	1QAB_E	1QAB_E
8	192	2	177	1.75e-133	372.0	100.00	305.6842016	2WQA_E	2WQA_E
9	192	1	176	2.45e-133	372.0	100.00	305.3477293	3BSZ_E	3BSZ_E
10	190	1	174	7.31e-132	368.0	100.00	301.9519890	2WQ9_A	2WQ9_A
11	190	2	175	8.37e-132	368.0	100.00	301.8165784	2WR6_A	2WR6_A
12	199	1	183	5.70e-131	366.0	96.17	299.8981810	1AQB_A	1AQB_A
13	199	1	183	2.48e-130	364.0	97.27	298.4278035	1HBQ_A	1HBQ_A
14	199	1	183	2.89e-130	364.0	96.72	298.2748056	1ERB_A	1ERB_A
15	191	1	175	6.08e-125	350.0	96.57	286.0181319	1KT5_A	1KT5_A
16	190	1	174	1.20e-123	347.0	96.55	283.0356449	1RLB_E	1RLB_E
17	190	2	173	1.77e-116	329.0	94.77	266.5288912	1IIU_A	1IIU_A
18	196	13	171	2.90e-08	52.8	42.44	17.3559700	7VNS_A	7VNS_A
19	196	13	171	6.25e-08	52.0	42.44	16.5880993	7O2Y_a	7O2Y_AAA
20	196	13	171	7.77e-08	51.6	42.44	16.3704106	5EZ2_A	5EZ2_A
21	166	13	139	1.07e-07	50.8	45.00	16.0504370	5F6Z_A	5F6Z_A
22	196	13	171	1.28e-07	50.8	42.44	15.8712356	7VNL_A	7VNL_A
23	196	13	171	1.64e-07	50.8	42.44	15.6233994	7YX1_A	7YX1_A
24	196	13	171	1.88e-07	50.4	42.44	15.4868239	7O3K_A	7O3K_A
25	164	11	137	1.22e-05	45.4	47.55	11.3140746	2HZQ_A	2HZQ_A
26	176	21	154	1.69e-04	42.0	42.76	8.6856118	2NND_A	2NND_A
27	142	13	127	1.00e-03	39.7	45.53	6.9077553	1IW2_A	1IW2_A
28	142	13	127	4.00e-03	38.1	42.62	5.5214609	2OVD_A	2OVD_A
29	134	4	110	5.00e-03	37.7	43.86	5.2983174	2QOS_C	2QOS_C
30	142	13	127	9.00e-03	37.0	42.62	4.7105307	2OVA_A	2OVA_A
31	132	36	140	7.20e-02	34.7	43.80	2.6310892	4ES7_A	4ES7_A
32	134	13	119	1.40e-01	33.5	42.98	1.9661129	3OJY_C	3OJY_C
33	132	9	113	1.50e-01	33.5	43.80	1.8971200	3QKG_A	3QKG_A
34	134	33	139	1.50e-01	33.5	42.98	1.8971200	8B0F_F	8B0F_F
35	134	15	121	1.60e-01	33.5	42.98	1.8325815	2RD7_C	2RD7_C
36	194	11	174	2.20e-01	32.7	39.89	1.5141277	1GKA_B	1GKA_B
37	176	7	140	4.50e-01	32.0	41.94	0.7985077	5NGH_A	5NGH_A

38	134	15	122	5.00e-01	32.0	42.74	0.6931472	2K23_A	2K23_A
39	170	5	131	5.80e-01	31.6	39.57	0.5447272	1JZU_A	1JZU_A
40	136	17	119	2.40e+00	30.0	42.20	-0.8754687	1Z24_A	1Z24_A
41	169	10	141	4.20e+00	29.3	40.26	-1.4350845	5X7Y_A	5X7Y_A
42	134	8	107	4.40e+00	28.9	39.64	-1.4816045	2XST_A	2XST_A
43	169	8	139	4.50e+00	28.9	40.26	-1.5040774	6NRE_A	6NRE_A

$raw

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_91489	4O9S_A	95.408	196	7	1	6
2	Query_91489	3FMZ_A	95.408	196	7	1	6
3	Query_91489	6QBA_A	100.000	183	0	0	17
4	Query_91489	1JYD_A	100.000	182	0	0	17
5	Query_91489	1BRP_A	100.000	182	0	0	17
6	Query_91489	1JYJ_A	98.901	182	2	0	17
7	Query_91489	1QAB_E	98.333	180	3	0	20
8	Query_91489	2WQA_E	100.000	176	0	0	17
9	Query_91489	3BSZ_E	100.000	176	0	0	17
10	Query_91489	2WQ9_A	100.000	174	0	0	17
11	Query_91489	2WR6_A	100.000	174	0	0	17
12	Query_91489	1AQB_A	92.896	183	13	0	17
13	Query_91489	1HBQ_A	92.350	183	14	0	17
14	Query_91489	1ERB_A	92.350	183	14	0	17
15	Query_91489	1KT5_A	93.143	175	12	0	17
16	Query_91489	1RLB_E	92.529	174	13	0	17
17	Query_91489	1IIU_A	87.209	172	22	0	19
18	Query_91489	7VNS_A	26.744	172	111	7	27
19	Query_91489	7O2Y_AAA	26.744	172	111	7	27
20	Query_91489	5EZ2_A	26.744	172	111	7	27
21	Query_91489	5F6Z_A	30.714	140	84	6	27
22	Query_91489	7VNL_A	26.744	172	111	7	27
23	Query_91489	7YX1_A	27.326	172	110	7	27
24	Query_91489	7O3K_A	26.163	172	112	7	27
25	Query_91489	2HZQ_A	27.273	143	83	7	27
26	Query_91489	2NND_A	23.684	152	93	6	30
27	Query_91489	1IW2_A	26.829	123	80	4	22
28	Query_91489	2OVD_A	25.410	122	83	3	22
29	Query_91489	2QOS_C	25.439	114	77	3	22
30	Query_91489	2OVA_A	25.410	122	83	3	22
31	Query_91489	4ES7_A	23.140	121	65	5	24
32	Query_91489	3OJY_C	24.561	114	78	3	22
33	Query_91489	3QKG_A	23.967	121	64	5	24
34	Query_91489	8B0F_F	24.561	114	78	3	22

35	Query_91489	2RD7_C	24.561	114	78	3	22
36	Query_91489	1GKA_B	23.034	178	119	7	21
37	Query_91489	5NGH_A	23.226	155	93	6	27
38	Query_91489	2K23_A	21.368	117	77	3	24
39	Query_91489	1JZU_A	22.302	139	96	4	32
40	Query_91489	1Z24_A	28.440	109	70	4	30
41	Query_91489	5X7Y_A	22.078	154	87	6	27
42	Query_91489	2XST_A	23.423	111	71	3	27
43	Query_91489	6NRE_A	22.078	154	87	6	27

	q.end	s.start	s.end	evalue	bitscore	positives
1	199	20	215	1.31e-139	389.0	95.92
2	199	17	212	1.90e-139	389.0	95.92
3	199	3	185	1.60e-138	385.0	100.00
4	198	2	183	4.72e-138	384.0	100.00
5	198	1	182	5.56e-138	384.0	100.00
6	198	2	183	1.83e-135	377.0	98.90
7	199	1	180	6.78e-134	373.0	98.33
8	192	2	177	1.75e-133	372.0	100.00
9	192	1	176	2.45e-133	372.0	100.00
10	190	1	174	7.31e-132	368.0	100.00
11	190	2	175	8.37e-132	368.0	100.00
12	199	1	183	5.70e-131	366.0	96.17
13	199	1	183	2.48e-130	364.0	97.27
14	199	1	183	2.89e-130	364.0	96.72
15	191	1	175	6.08e-125	350.0	96.57
16	190	1	174	1.20e-123	347.0	96.55
17	190	2	173	1.77e-116	329.0	94.77
18	196	13	171	2.90e-08	52.8	42.44
19	196	13	171	6.25e-08	52.0	42.44
20	196	13	171	7.77e-08	51.6	42.44
21	166	13	139	1.07e-07	50.8	45.00
22	196	13	171	1.28e-07	50.8	42.44
23	196	13	171	1.64e-07	50.8	42.44
24	196	13	171	1.88e-07	50.4	42.44
25	164	11	137	1.22e-05	45.4	47.55
26	176	21	154	1.69e-04	42.0	42.76
27	142	13	127	1.00e-03	39.7	45.53
28	142	13	127	4.00e-03	38.1	42.62
29	134	4	110	5.00e-03	37.7	43.86
30	142	13	127	9.00e-03	37.0	42.62
31	132	36	140	7.20e-02	34.7	43.80
32	134	13	119	1.40e-01	33.5	42.98
33	132	9	113	1.50e-01	33.5	43.80

```
34   134      33   139  1.50e-01      33.5      42.98
35   134      15   121  1.60e-01      33.5      42.98
36   194      11   174  2.20e-01      32.7      39.89
37   176       7   140  4.50e-01      32.0      41.94
38   134      15   122  5.00e-01      32.0      42.74
39   170       5   131  5.80e-01      31.6      39.57
40   136      17   119  2.40e+00      30.0      42.20
41   169      10   141  4.20e+00      29.3      40.26
42   134       8   107  4.40e+00      28.9      39.64
43   169       8   139  4.50e+00      28.9      40.26
```

$url

"https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tabu

Using the