

Jonathan Cifuentes
A17238145
jcifuentes@ucsd.edu

BIMM-143: INTRODUCTION TO BIOINFORMATICS

The find-a-gene project assignment
https://bioboot.github.io/bimm143_S20/

Dr. Barry Grant

Overview:

The find-a-gene project is a required assignment for BIMM-143. You should prepare a written report in **PDF** format that has responses to each question labeled **[Q1] - [Q10]** below. You may wish to consult the scoring rubric at the end of this document and the example report provided online.

The objective with this assignment is for you to demonstrate your grasp of database searching, sequence analysis, structure analysis and the R environment that we have covered in class.

Due Date:

Your responses to questions Q1-Q4 are due at the beginning of class **Tuesday May 5th** (05/05/20) at 12pm San Diego time. Note that these answers can be obtained very quickly (at best within 10 or 15 minutes), so if you don't succeed at first, just keep trying.

The complete assignment, including responses to all questions, is due **Friday June 5th** (06/05/20) at 12pm San Diego time.

Submission instructions:

Your report formatted as a **PDF document** should be uploaded to **GradeScope**. Please make sure to include your UCSD email and PID number on the first page.

Be sure to include your UCSD email and PID number on the first page of your report.

Submit your preliminary report with answers to Q1-Q4 as soon as you can so we can determine if you have found a novel gene. Submit this preliminary report as one document with screen shots of the results inserted appropriately.

See the demonstration report linked to on the course website for an example of format. I will email you my decision; proceed with subsequent questions only after we are sure you have found a novel gene.

For the final report add your results for Q5-Q10 to the preliminary report and submit the final document containing your results for all questions - **Please do not send only Q5-Q10**

answers as the final report.

Questions:

[Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as it's function is known.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

Name: RBP4

Function: It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein posttranslationally and results in defective delivery and supply to the epidermal cells

Accession: NP_004796

Species: homo sapiens

[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

Also include the output of that BLAST search in your document. If appropriate, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output (e.g. alt print screen on a PC or on a MAC press ⌘-shift-4. The pointer becomes a bulls eye. Select the area you wish to capture and release. The image is saved as a file called Screen Shot [].png in your Desktop directory). It is not necessary to print out all of the blast results if there are many pages.

On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include no functional annotation.

In general, [Q2] is the most difficult for students because it requires you to have a "feel" for how to interpret BLAST results. You need to distinguish between a perfect match to your query (i.e. a sequence that is not "novel"), a near match (something that might be "novel", depending on the results of [Q4]), and a non-homologous result.

If you are having trouble finding a novel gene try restricting your search to an organism that is poorly annotated.

Method: TBLASTN (2.14.0+) search against chicken ESTs

Database: Expressed Sequence Tags (est)

Organism: chicken (taxid:9031)

Chosen match: Accession [CD731562.1](#), a 552 base pair clone from Chicken Intestinal Lymphocyte Gallus gallus cDNA clone 1GAL_71G14 5', mRNA sequence

523 Das...

Bhttps://bioboot.gith...

█Adding color to a sc...

U SINGLE SIGN-ON

≡Find_A_Gene_Proje...

🔍NCBI Blast:ref|NP_0...

🌐https://ftp.ncbi.nlm...

Bhttps:...

BLAST® » tblastn » results for RID-4YX9UK41016

HomeRecent ResultsSaved StrategiesHelp

Edit Search

Save Search

Search Summary ▼

🔍How to read this report?

📺BLAST Help Videos

↩️Back to Traditional Results Page

📌Your search is limited to records that include: chicken (taxid:9031)

Job Title

ref|NP_004796|

RID

4YX9UK41016

Search expires on 05-02 16:15 pm

Download All ▼

Program

TBLASTN

?

Citation ▼

Database

est

See details ▼

Query ID

NP_004796.1

Description

DNA-directed RNA polymerase II subunit RPB4 [Homo sapiens]

Molecule type

amino acid

Query Length

142

Other reports

?

Filter Results

Organism

only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▼Select columns ▼Show100 ▼🔍

☐ select all

9 sequences selected

GenBankGraphics

	Description ▼	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover ▼	E value ▼	Per. Ident ▼	Acc. Len ▼	Accession
<input checked="" type="checkbox"/>	4041478 1GAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA clone 1GAL_71G14 5', mRNA sequence	Gallus gallus	283	283	100%	1e-98	97.18%	552	CD731562.1

Download ▼GenBankGraphics

▼Next ▲Previous ◀Descriptions

4041478 1GAL - Chicken Intestinal Lymphocyte Gallus gallus cDNA clone 1GAL_71G14 5', mRNA sequence

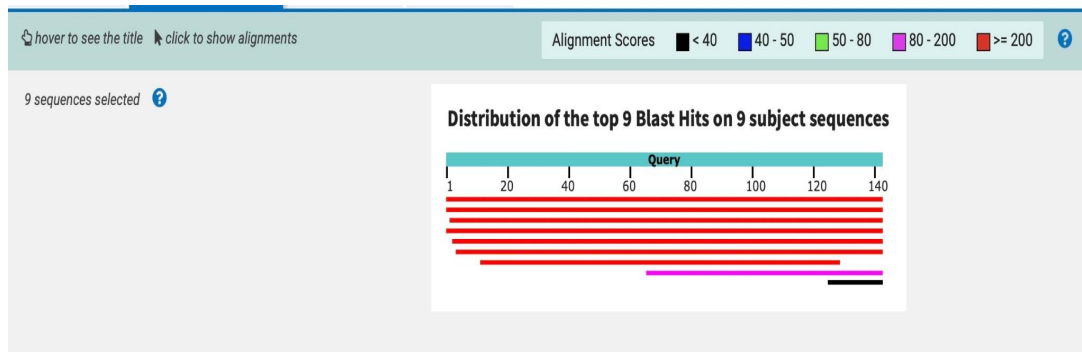
Sequence ID: CD731562.1Length: 552Number of Matches: 1

Range 1: 5 to 430

GenBankGraphics

▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
283 bits(725)	1e-98	Compositional matrix adjust.	138/142(97%)	140/142(98%)	0/142(0%)	+2
Query 1	MAAGGSDPRAGDVEEDASQLIFPKEFETAETLLNSEVHMLLEHRKQONESAEDEQELSEV					60
Sbjct 5	MAAGGSDPR DVEEDASQL+FPKEFETAETLLNSEVHMLLEHRKQONESAEDEQELSEV					184
Query 61	FMKTLNNTARFSRFKNRETIASVRSLLQKKLHKFELACLANLCPETAESKALIPSLEG					120
Sbjct 185	FMKTLNNTARFSRFKNRETIASVRSLLQKKLHKFELACLANLCPETAEEAKALIPSLEG					364
Query 121	RFEEDELQIILDDIQTKRFSQY		142			
Sbjct 365	RFEEDELQIILDDIQTKRFSQY		430			



[Q3] Gather information about this “novel” **protein**. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI. Don’t forget to translate all six reading frames; the ORF (open reading frame) is likely to be the longest sequence without a stop codon. It may not start with a methionine if you don’t have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format. Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as *S. cerevisiae*, human or mouse, because those genomes have already been thoroughly annotated. It is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

>

```
Chicken Intestinal Lymphocyte Gallus gallus cDNA
clone 1GAL_71G14 5'
```

```
MAAGGSDPRTADVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEV
FMKTLNLTARFSRFKNRETIASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEG
RFEDEELQQILDDIQTKRSFY
```

```
Name: Chicken Intestinal Lymphocyte Gallus gallus cDNA
clone
```

```
Species: Gallus gallus
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda;
Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes;
Phasianidae; Phasianinae; Gallus.
```

[Q4] Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

Novel Gene Found: top match is 100% identity but to a different species than the one started with.

A BLASTP search against NR database (see setup in first screen-shot below) yielded a top hit result to a protein from *Grus Americana* (Whooping Crane).

blastn

blastp

blastx

tblastn

tblastx

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

Query subrange ?

>unnamed protein product
MAAGGSDPRTADVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQNESAEDE
QELSEVFMKTLNYTAR
FSRFKNRETIASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEGRFEDEEL

From

To

Or, upload file

Choose File no file selected ?

Job Title

Protein Sequence

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Databases

☒ Standard databases (nr etc.): **New** ☐ Experimental databases

Compare

☐ Select to compare standard and experimental database ?

Try experimental clustered nr database ?

For more info see What is clustered nr?

Standard

Database

Non-redundant protein sequences (nr) ?

Organism

Optional

Enter organism name or id—completions will be suggested ☐ exclude

Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude

Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

[Edit Search](#)
[Save Search](#)
[Search Summary](#)

[How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Query Title
Protein Sequence

Accession
[4Z1XZ49R016](#)
[Search expires on 05-02 17:34 pm](#)
[Download All](#)

Program
BLASTP
[Citation](#)

Database
nr
[See details](#)

Query ID
lcl|Query_104120

Description
unnamed protein product

Molecule type
amino acid

Query Length
142

Other reports
[Distance tree of results](#)
[Multiple alignment](#)
[MSA viewer](#)

Filter Results

Organism
only top 20 will appear
☐ exclude

Type common name, binomial, taxid or group name

[Add organism](#)

Percent Identity
 to
E value
 to
Query Coverage
 to

[Filter](#)
[Reset](#)

Compare these results against the new Clustered nr database
[BLAST](#)

[Descriptions](#)
[Graphic Summary](#)
[Alignments](#)
[Taxonomy](#)

Sequences producing significant alignments
[Download](#)
[Select columns](#)
Show 100

☒ select all
100 sequences selected
[GenPept](#)
[Graphics](#)
[Distance tree of results](#)
[Multiple alignment](#)
[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	RPB4 polymerase [Grus americana]	Grus americana	289	289	100%	3e-98	100.00%	143	NWH25084.1
<input checked="" type="checkbox"/>	RPB4 polymerase [Picathartes gymnocephalus]	Picathartes gymnocephalus	289	289	100%	3e-98	100.00%	143	NWI39806.1

[Download](#)
[GenPept](#)
[Graphics](#)
[Next](#)
[Previous](#)
[Descriptions](#)

RPB4 polymerase [Grus americana]

Sequence ID: [NWH25084.1](#) Length: 143 Number of Matches: 1

[See 8 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 2 to 143
[GenPept](#)
[Graphics](#)
[Next Match](#)
[Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
289 bits(739)	3e-98	Compositional matrix adjust.	142/142(100%)	142/142(100%)	0/142(0%)
Query 1	MAAGGSDPRTADVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQONESAEDEQELSEV				60
Sbjct 2	MAAGGSDPRTADVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQONESAEDEQELSEV				61
Query 61	FMKTLNLTARFSRFRKNRETIASVRSLLLQKKLHKFELACLANLCPETAEAAKALIPSLEG				120
Sbjct 62	FMKTLNLTARFSRFRKNRETIASVRSLLLQKKLHKFELACLANLCPETAEAAKALIPSLEG				121
Query 121	RFEEDEELQIILDDIQTGRSFQY		142		
Sbjct 122	RFEEDEELQIILDDIQTGRSFQY		143		

Related Information
[AlphaFold Structure](#) - 3D structure displays
[Genome Data Viewer](#) - aligned genomic context
[Identical Proteins](#) - Identical proteins to NWH25084.1

[Q5] Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.

Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in

the output alignment and in the subsequent answers below). The goal in this step is to create an interesting alignment for building a phylogenetic tree that illustrates species divergence.

```
>[Homo sapiens]|DNA-directed RNA polymerase II subunit RPB4
|NP_004796.1
MAAGGSDPRAGDVEEDASQLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTA
RFSRFKNRETIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
SFQY
```

```
>Chicken |Intestinal Lymphocyte Gallus gallus cDNA clone 1GAL_71G14
5'|[sequence from blast]
MAAGGSDPRTADVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEV
FMKTLNYTARFSRFKNRETIASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEG
RFEDEELQQILDDIQTKRSFQY
```

```
>[Rattus norvegicus]|DNA-directed RNA polymerase II subunit
RPB4 |nr|NP_001102356.1
MAAGGSDPRAGDVEEDASQLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMK
TLNYTARFSRFKNRETI
ASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKRSFQY
```

```
> [Otolemur garnettii]|DNA-directed RNA polymerase II subunit
RPB4|nr|XP_003800379.1
MAAGGSDPRAGDVEEDASQLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMK
TLNYTSRFSRFKNRETI
ASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKRSFQY
```

```
>[Sus scrofa]| Chain D, RNA polymerase II subunit D|nr|70KX_D
MWGPAQPYSDSALSPKPRPFRAVFRGSALPFPVRVEVRGRSMAAGGSDPRAGDVEEDASQLI
FPKEFETAETLLNSEVH
MLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRETIASVRSLLLQKKLHKFELACLA
NLCPETAESKALIPSL
EGRFEDEELQQILDDIQTKRSFQY
```

```
>[Vombatus ursinus]|DNA-directed RNA polymerase II subunit
RPB4|nr|XP_027710460.1|
MAAGGGDPAGDVEEDASQLVFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMK
TLNYTARFSRFKNRETI
ASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEGRFEDEELQQILDDIQTKRSFQY
```

```
> [Pteropus alecto]|DNA-directed RNA polymerase II subunit RPB4
|nr|XP_006913507.1
MAVRLPAMAAGGSDLRAGDVEEDASQLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQE
LSEVFMKTLNYTARFSR
FKNRETIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQT
KRSFQY
```

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```
[Pteropus      -----MAVRLPAMAAGGSDLRAGDVEEDAS
[Sus           MWGPAQPYSDSALSPKPRPFRAVFRGSALPFPVRVEVRGRSMAAGGSDPRAGDVEEDAS
[Homo          -----MAAGGSDPRAGDVEEDAS
[Rattus        -----MAAGGSDPRAGDVEEDAS
[Otolemur      -----MAGGS PRAGD VEDAS
Chicken        -----MAAGGSDPRTADVEEDAS
[Vombatus      -----MAAGGGDPRAGDVEEDAS

                        *****.* *:*****
```

```
[Pteropus      QLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE
[Sus           QLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE
[Homo          QLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE
[Rattus        QLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE
[Otolemur      QLIFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYSRFSRFKNRE
Chicken        QLVFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE
[Vombatus      QLVFPKEFETAETLLNSEVHMLLEHRKQQNESAEDEQELSEVFMKTLNYTARFSRFKNRE

**.:*****:*****
```

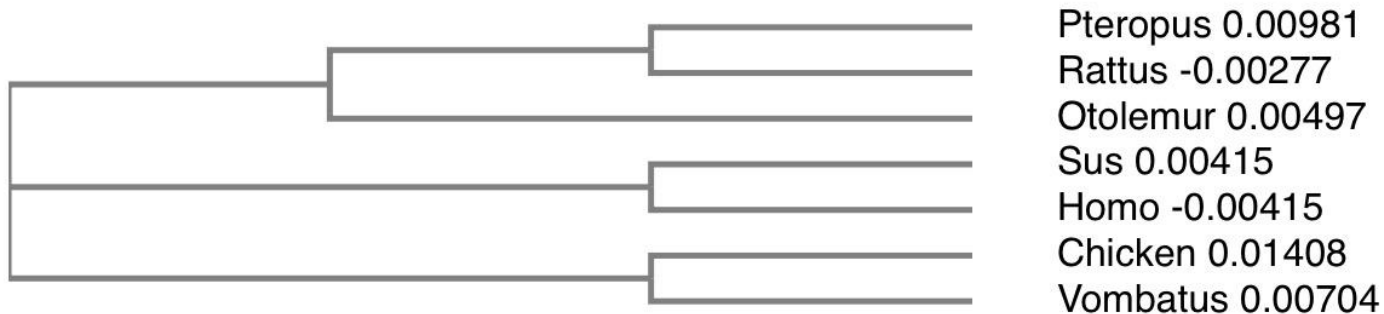
```
[Pteropus      TIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
[Sus           TIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
[Homo          TIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
[Rattus        TIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
[Otolemur      TIASVRSLLLQKKLHKFELACLANLCPETAESKALIPSLEGRFEDEELQQILDDIQTKR
Chicken        TIASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEGRFEDEELQQILDDIQTKR
[Vombatus      TIASVRSLLLQKKLHKFELACLANLCPETAEEAKALIPSLEGRFEDEELQQILDDIQTKR

*****.:*****
```

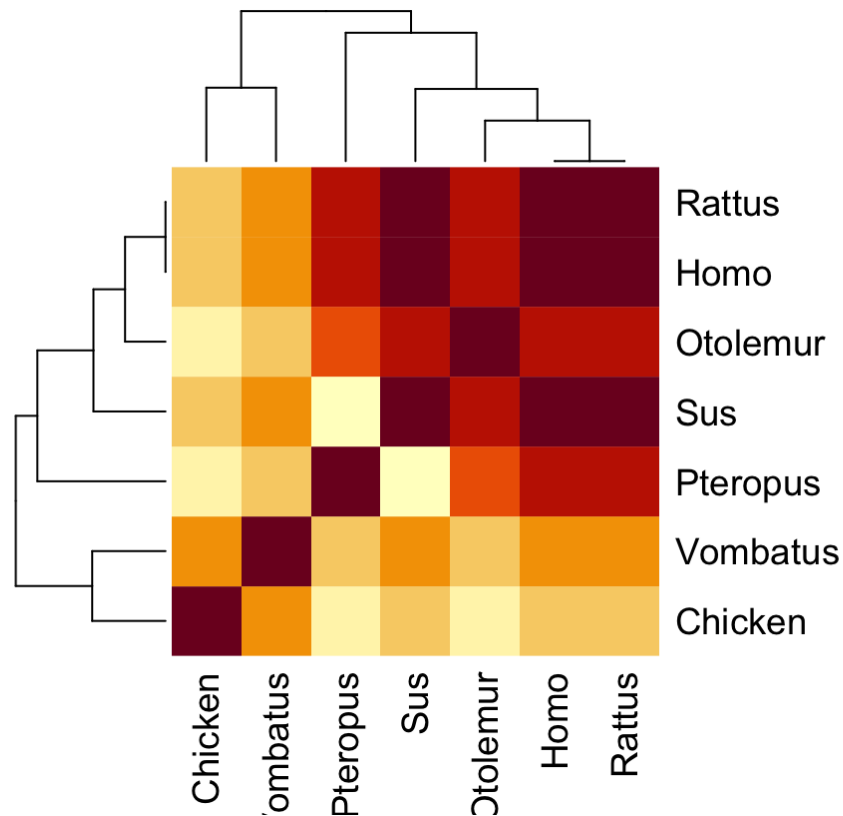
```
[Pteropus      SFQY
[Sus           SFQY
[Homo          SFQY
[Rattus        SFQY
[Otolemur      SFQY
Chicken        SFQY
[Vombatus      SFQY
*****
```


[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

Cladogram produced through EBI's simple phylogeny



[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.



[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

List the top 3 *unique* hits (i.e. not hits representing different chains from the same structure) along with their Evalue and sequence identity to your query. Please also add annotation details of these structures. For example include the annotation terms PDB identifier (structureId), Method used to solve the structure (experimentalTechnique), resolution (resolution), and source organism (source).

HINT: You can use a single sequence from your alignment or generate a consensus sequence from your alignment using the Bio3D function `consensus()`. The Bio3D functions `blast.pdb()`, `plot.blast()` and `pdb.annotate()` are likely to be of most relevance for completing this task. Note that the results of `blast.pdb()` contain the hits PDB identifier (or `pdb.id`) as well as Evalue and identity. The results of `pdb.annotate()` contain the other annotation terms noted above.

Note that if your consensus sequence has lots of gap positions then it will be better to use an original sequence from the alignment for your search of the PDB. In this case you could choose the sequence with the highest identity to all others in your alignment by calculating the row-wise maximum from your sequence identity matrix.

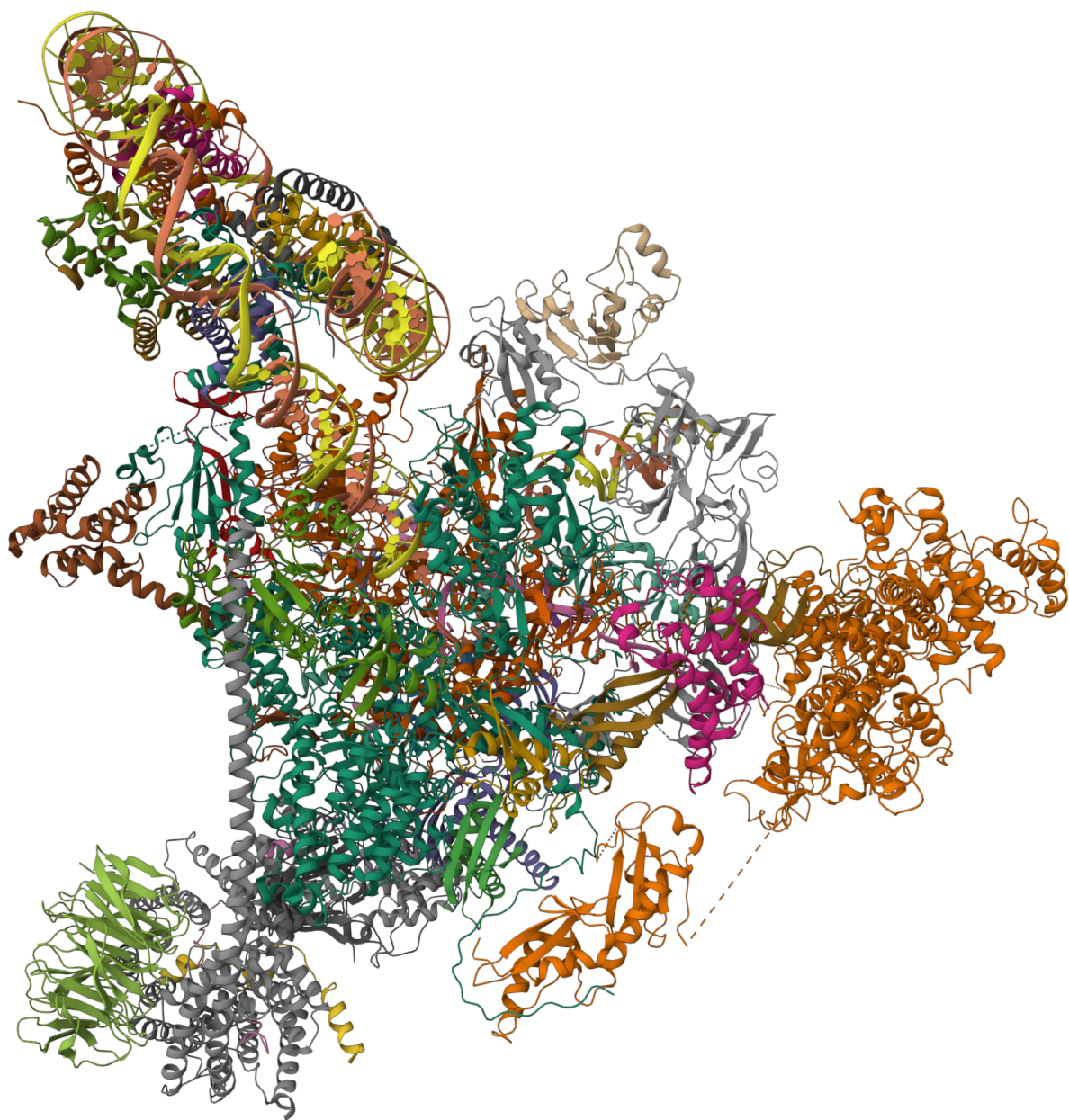
Used Blast w/ original query sequence homo sapiens

ID	Technique	Resolution Å	Source	Evalue	Identity
5IY6	Electron Microscopy	7.2	Homo sapiens (human)	5e-102	100
5OIK	Electron Microscopy	3.7	Bos taurus (cattle)	2e-101	99.3
7UND	Electron Microscopy	3	Sus scrofa (pig)	3e-101	100

[Q9] Generate a molecular figure of one of your identified PDB structures using the **NGL viewer** online (or **VMD/PyMol**). You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black).

Based on sequence similarity. How likely is this structure to be similar to your “novel” protein?

Sequence (7UND) is very likely to be similar in structure to the original homo sapiens RBP4 sequence given that the sequence similarity is >80%.



[Q10] Perform a “Target” search of ChEMBEL (<https://www.ebi.ac.uk/chembl/>) with your novel sequence. Are there any **Target Associated Assays** and **ligand efficiency data** reported that may be useful starting points for exploring potential inhibition of your novel protein?

According to what chatGPT says, a lower k_i value denotes higher potency and stronger inhibition of the target. Unfortunately, not all the chembl results returned a k_i value. However, a common value that was returned was the IC50. ChatGPT states, “the compound with the lowest IC50 value is considered the most potent or strongest inhibitor among the set.” Of the Chembl results that were returned (Chembl5701, CHEMBL1075195, CHEMBL5261) 5701 and 5261 showed the most promise for inhibitory potential.

Chembl5701

IC50 count 113, 11.38%. Associated assays: 104 bound, 0 functional. Ligand efficiency data available.

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL5701/

Chembl 5261

IC50 count 205, 10.22%. Associated assays: 124 bound, 1 functional. Ligand efficiency data available.

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL5261/

Scoring Rubric:

[45 total points available]