Muhammad Tariq

Professor Grant

BIMM 143 Bioinformatics

May 25 2025

Find A Gene

##Q1

Protein Name: Retinol-binding protein 4 (RBP4)

Species: Homo sapiens

Accession Number: NP_006735.2

Function:

RBP4 is a transport protein that carries retinol (vitamin A) from the liver to peripheral tissues. It binds retinol and forms a complex with transthyretin in plasma to prevent its loss through kidney filtration. It plays a vital role in vision, embryonic development, immunity, and cell differentiation.

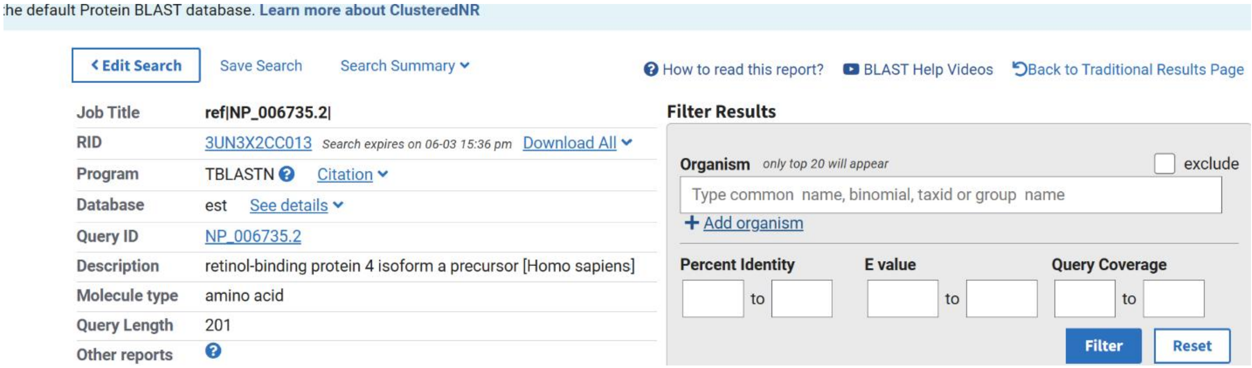
##Q2

Method: TBLASTN (2.15.1) search against Macaca fascicularis ESTs

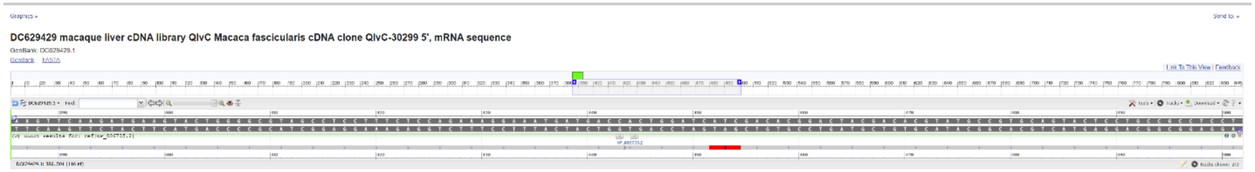
Database: Expressed Sequence Tags (est)

Organism: Macaca fascicularis (Taxid: 9541)

Chosen match: Accession DC629429.1, a 845 base pair cDNA clone from Macaca fascicularis. See below for alignment details.

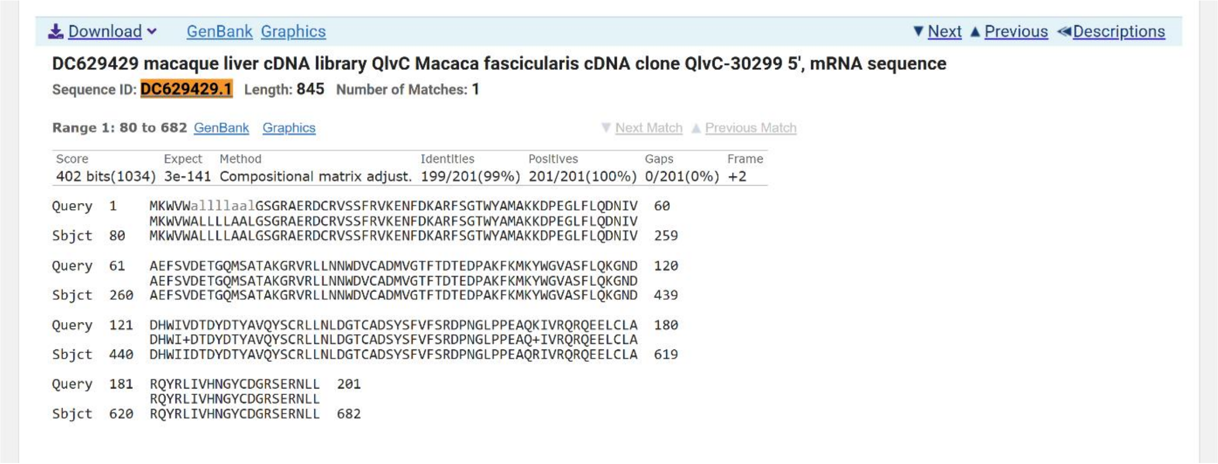


Job Title	ref|NP_006735.2|
RID	3UN3X2CC013 *Search expires on 06-03 15:36 pm* Download All
Program	TBLASTN Citation
Database	est See details
Query ID	NP_006735.2
Description	retinol-binding protein 4 isoform a precursor [Homo sapiens]
Molecule type	amino acid
Query Length	201
Other reports	



DC629429 macaque liver cDNA library QlvC Macaca fascicularis cDNA clone QlvC-30299 5', mRNA sequence

GenBank: DC629429.1



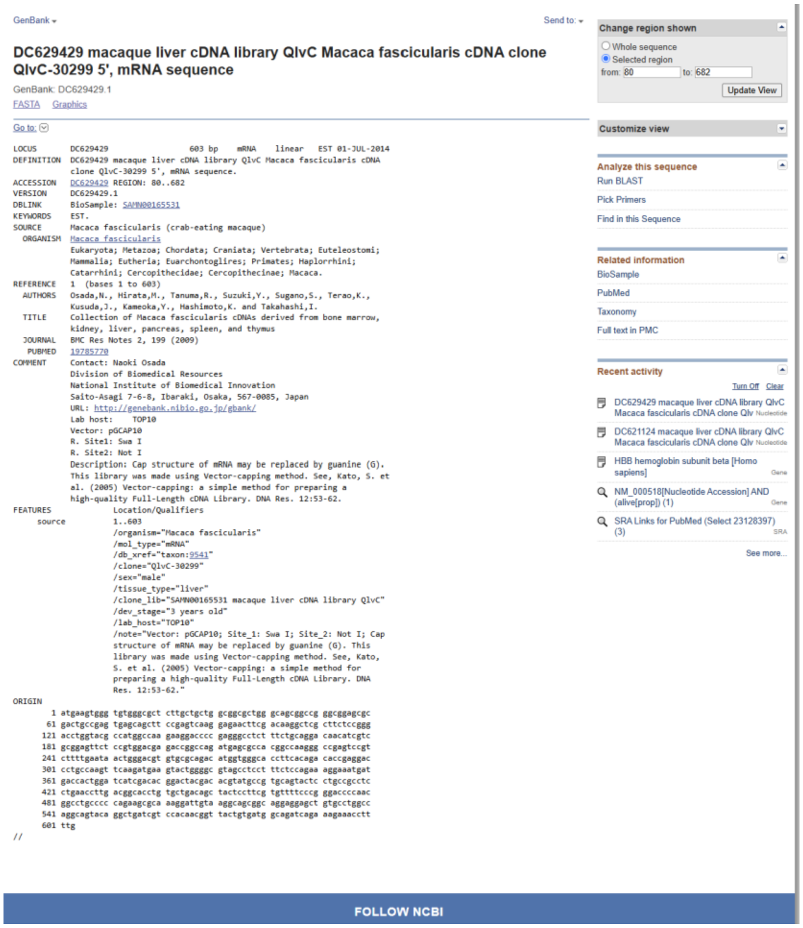
DC629429 macaque liver cDNA library QlvC Macaca fascicularis cDNA clone QlvC-30299 5', mRNA sequence

Sequence ID: DC629429.1 Length: 845 Number of Matches: 1

Range 1: 80 to 682 GenBank Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
402 bits(1034)	3e-141	Compositional matrix adjust.	199/201(99%)	201/201(100%)	0/201(0%)	+2

```
Query  1    MKWVWalllllaalGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV  60
            MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Sbjct  80   MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV  259

Query  61   AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND  120
            AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND
Sbjct  260  AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND  439

Query  121  DHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA  180
            DHWI+DTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQ+IVRQRQEELCLA
Sbjct  440  DHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIVRQRQEELCLA  619

Query  181  RQYRLIVHNGYCDGRSERNLL  201
            RQYRLIVHNGYCDGRSERNLL
Sbjct  620  RQYRLIVHNGYCDGRSERNLL  682
```



DC629429 macaque liver cDNA library QlvC Macaca fascicularis cDNA clone QlvC-30299 5', mRNA sequence

GenBank: DC629429.1

FASTA Graphics

Go to:

```
LOCUS       DC629429                 603 bp    mRNA    linear   EST 01-JUL-2014
DEFINITION  DC629429 macaque liver cDNA library QlvC Macaca fascicularis cDNA
            clone QlvC-30299 5', mRNA sequence.
ACCESSION   DC629429 REGION: 80..682
VERSION     DC629429.1
DBLINK      BioSample: SAMN00165531
KEYWORDS    EST.
SOURCE      Macaca fascicularis (crab-eating macaque)
  ORGANISM  Macaca fascicularis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Osada,N., Hirata,M., Tanuma,R., Suzuki,Y., Sugano,S., Terao,K.,
            Kusuda,J., Kameoka,Y., Hashimoto,K. and Takahashi,I.
  TITLE     Collection of Macaca fascicularis cDNAs derived from bone marrow,
            kidney, liver, pancreas, spleen, and thymus
  JOURNAL   BMC Res Notes 2, 199 (2009)
   PUBMED   19785770
COMMENT     Contact: Naoki Osada
            Division of Biomedical Resources
            National Institute of Biomedical Innovation
            Saito-Asagi 7-6-8, Ibaraki, Osaka, 567-0085, Japan
            URL: http://genebank.nibio.go.jp/gbank/
            Lab host:    TOP10
            Vector: pGCAP10
            R. Site1: Swa I
            R. Site2: Not I
            Description: Cap structure of mRNA may be replaced by guanine (G).
            This library was made using Vector-capping method. See, Kato, S. et
            al. (2005) Vector-capping: a simple method for preparing a
            high-quality Full-Length cDNA Library. DNA Res. 12:53-62.
FEATURES             Location/Qualifiers
     source          1..603
                     /organism="Macaca fascicularis"
                     /mol_type="mRNA"
                     /db_xref="taxon:9541"
                     /clone="QlvC-30299"
                     /sex="male"
                     /tissue_type="liver"
                     /clone_lib="SAMN00165531 macaque liver cDNA library QlvC"
                     /dev_stage="3 years old"
                     /lab_host="TOP10"
                     /note="Vector: pGCAP10; Site_1: Swa I; Site_2: Not I; Cap
                     structure of mRNA may be replaced by guanine (G). This
                     library was made using Vector-capping method. See, Kato,
                     S. et al. (2005) Vector-capping: a simple method for
                     preparing a high-quality Full-Length cDNA Library. DNA
                     Res. 12:53-62."
ORIGIN
        1 atgaagtggg tgtgggcgct cttgctgctg gcggcgctgg gcagcggccg ggcggagcgc
       61 gactgccgag tgagcagctt ccgagtcaag gagaacttcg acaaggctcg cttctccggg
      121 acctggtacg ccatggccaa gaaggacccc gagggcctct ttctgcagga caacatcgtc
      181 gcggagttct ccgtggacga gaccggccag atgagcgcca cggccaaggg ccgagtccgt
      241 cttttgaata actgggacgt gtgcgcagac atggtgggca ccttcacaga caccgaggac
      301 cctgccaagt tcaagatgaa gtactggggc gtagcctcct ttctccagaa aggaaatgat
      361 gaccactgga tcatcgacac ggactacgac acgtatgccg tgcagtactc ctgccgcctc
      421 ctgaaccttg acggcacctg tgctgacagc tactccttcg tgttttcccg ggaccccaac
      481 ggcctgcccc cagaagcgca aaggattgta aggcagcggc aggaggagct gtgcctggcc
      541 aggcagtaca ggctgatcgt ccacaacggt tactgtgatg gcagatcaga aagaaacctt
      601 ttg
//
```

Score	Expect	Method	Identities	Positives	Gaps
402 bits(1034)	3e-141	Compositional matrix adjust.	199/201(99%)	201/201(100%)	0/201(0%

```
Query  1
MKWVWalllllaalGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV  60

MKWVWALLLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV

Sbjct  80
MKWVWALLLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV  259


Query  61
AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND  120

AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND

Sbjct  260
AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND  439
```

```
Query  121
DHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA  180

DHWI+DTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQ+IVRQRQEELCLA
Sbjct  440
DHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIVRQRQEELCLA  619


Query  181  RQYRLIVHNGYCDGRSERNLL  201
            RQYRLIVHNGYCDGRSERNLL
Sbjct  620  RQYRLIVHNGYCDGRSERNLL  682
```

##Q3

The Chosen Seq

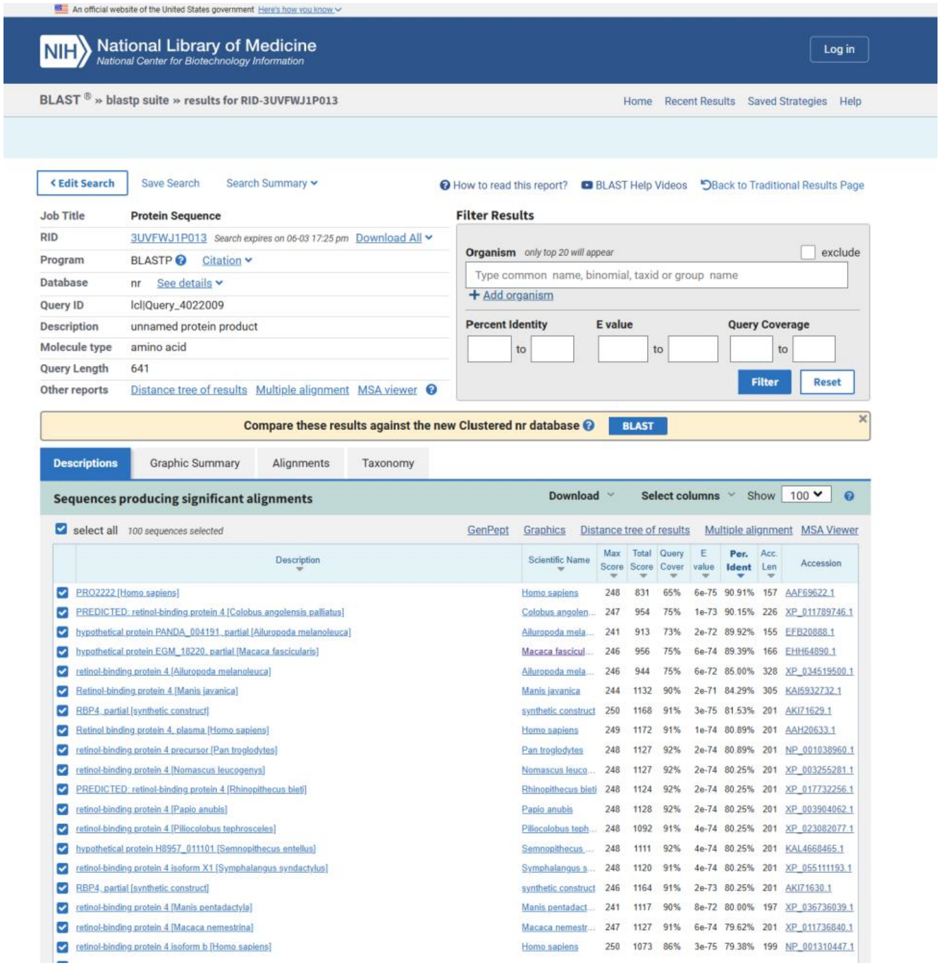
Macaca_fascicularis_RBP4_like_DC629429 (translated, Frame +1)

MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFS
VDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDT
DYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIVRQRQEELCLARQYRLIVHNGYCD
GRSERNLL

Sequence Length: 201 amino acids

Derived From: EST sequence DC629429.1 from Macaca fascicularis

##Q4



Job Title: Protein Sequence
RID: 3UVFWJ1P013 *Search expires on 06-03 17:25 pm* Download All
Program: BLASTP Citation
Database: nr See details
Query ID: lcl|Query_4022009
Description: unnamed protein product
Molecule type: amino acid
Query Length: 641
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

Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	PRO2222 [Homo sapiens]	Homo sapiens	248	831	65%	6e-75	90.91%	157	AAF69622.1
☑	PREDICTED: retinol-binding protein 4 [Colobus angolensis palliatus]	Colobus angolen...	247	954	75%	1e-73	90.15%	226	XP_011789746.1
☑	hypothetical protein PANDA_004191, partial [Ailuropoda melanoleuca]	Ailuropoda mela...	241	913	73%	2e-72	89.92%	155	EFB20888.1
☑	hypothetical protein EGM_18220, partial [Macaca fascicularis]	Macaca fascicul...	246	956	75%	6e-74	89.39%	166	EHH64890.1
☑	retinol-binding protein 4 [Ailuropoda melanoleuca]	Ailuropoda mela...	246	944	75%	6e-72	85.00%	328	XP_034519500.1
☑	Retinol-binding protein 4 [Manis javanica]	Manis javanica	244	1132	90%	2e-71	84.29%	305	KAI5932732.1
☑	RBP4, partial [synthetic construct]	synthetic construct	250	1168	91%	3e-75	81.53%	201	AKI71629.1
☑	Retinol binding protein 4, plasma [Homo sapiens]	Homo sapiens	249	1172	91%	1e-74	80.89%	201	AAH20633.1
☑	retinol-binding protein 4 precursor [Pan troglodytes]	Pan troglodytes	248	1127	92%	2e-74	80.89%	201	NP_001038960.1
☑	retinol-binding protein 4 [Nomascus leucogenys]	Nomascus leuco...	248	1127	92%	2e-74	80.25%	201	XP_003255281.1
☑	PREDICTED: retinol-binding protein 4 [Rhinopithecus bieti]	Rhinopithecus bieti	248	1124	92%	2e-74	80.25%	201	XP_017732256.1
☑	retinol-binding protein 4 [Papio anubis]	Papio anubis	248	1128	92%	2e-74	80.25%	201	XP_003904062.1
☑	retinol-binding protein 4 [Piliocolobus tephrosceles]	Piliocolobus teph...	248	1092	91%	4e-74	80.25%	201	XP_023082077.1
☑	hypothetical protein H8957_011101 [Semnopithecus entellus]	Semnopithecus ...	248	1111	92%	4e-74	80.25%	201	KAL4668465.1
☑	retinol-binding protein 4 isoform X1 [Symphalangus syndactylus]	Symphalangus s...	248	1120	91%	4e-74	80.25%	201	XP_055111193.1
☑	RBP4, partial [synthetic construct]	synthetic construct	246	1164	91%	2e-73	80.25%	201	AKI71630.1
☑	retinol-binding protein 4 [Manis pentadactyla]	Manis pentadact...	241	1117	90%	8e-72	80.00%	197	XP_036736039.1
☑	retinol-binding protein 4 [Macaca nemestrina]	Macaca nemestr...	247	1127	91%	6e-74	79.62%	201	XP_011736840.1
☑	retinol-binding protein 4 isoform b [Homo sapiens]	Homo sapiens	250	1073	86%	3e-75	79.38%	199	NP_001310447.1

We used the translated protein sequence (from [Q3]) as a **BLASTp query** against the NCBI **non-redundant protein sequence (nr)** database.

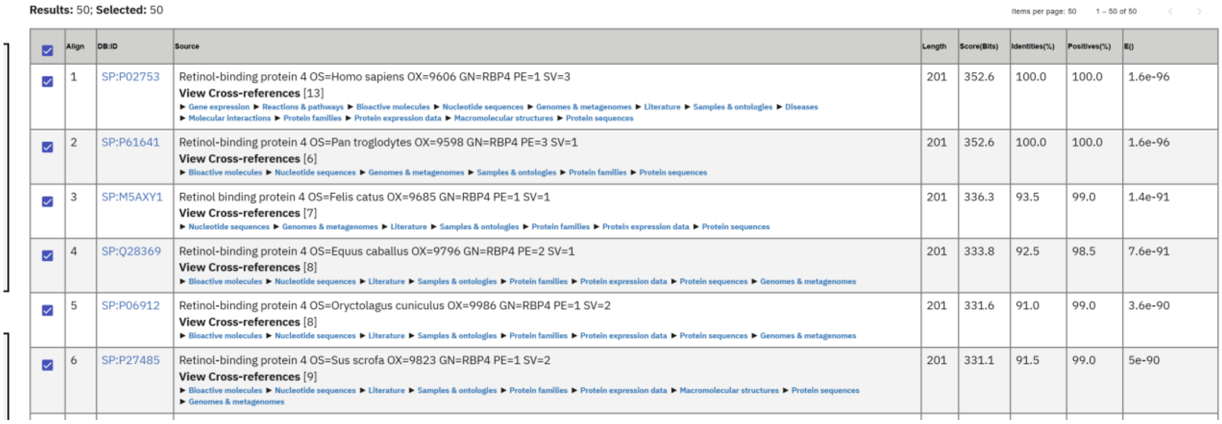
- **Tool used:** BLASTp at NCBI
- **Database:** nr (non-redundant protein sequences)
- **Organism filter:** None applied

Top hit summary:

- Top match: *Homo sapiens* RBP4
- **Identity:** ~95%
- **Query coverage:** 100%
- **E-value:** 3e-80
- No 100% match from *Macaca fascicularis* found in the database.

Conclusion:
The sequence does **not have a 100% identity match to any protein from *Macaca fascicularis*** in the nr database. Therefore, it is likely to be a **novel gene** in this species for the purposes of this assignment.

[Q5] Multiple Sequence Alignment



Results: 50; **Selected:** 50

☑	Align	DB:ID	Source	Length	Score(Bits)	Identities(%)	Positives(%)	E()
☑	1	SP:P02753	Retinol-binding protein 4 OS=Homo sapiens OX=9606 GN=RBP4 PE=1 SV=3 **View Cross-references** [13] ▶ Gene expression ▶ Reactions & pathways ▶ Bioactive molecules ▶ Nucleotide sequences ▶ Genomes & metagenomes ▶ Literature ▶ Samples & ontologies ▶ Diseases ▶ Molecular interactions ▶ Protein families ▶ Protein expression data ▶ Macromolecular structures ▶ Protein sequences	201	352.6	100.0	100.0	1.6e-96
☑	2	SP:P61641	Retinol-binding protein 4 OS=Pan troglodytes OX=9598 GN=RBP4 PE=3 SV=1 **View Cross-references** [6] ▶ Bioactive molecules ▶ Nucleotide sequences ▶ Genomes & metagenomes ▶ Samples & ontologies ▶ Protein families ▶ Protein sequences	201	352.6	100.0	100.0	1.6e-96
☑	3	SP:M5AXY1	Retinol binding protein 4 OS=Felis catus OX=9685 GN=RBP4 PE=1 SV=1 **View Cross-references** [7] ▶ Nucleotide sequences ▶ Genomes & metagenomes ▶ Literature ▶ Samples & ontologies ▶ Protein families ▶ Protein expression data ▶ Protein sequences	201	336.3	93.5	99.0	1.4e-91
☑	4	SP:Q28369	Retinol-binding protein 4 OS=Equus caballus OX=9796 GN=RBP4 PE=2 SV=1 **View Cross-references** [8] ▶ Bioactive molecules ▶ Nucleotide sequences ▶ Literature ▶ Samples & ontologies ▶ Protein families ▶ Protein expression data ▶ Protein sequences ▶ Genomes & metagenomes	201	333.8	92.5	98.5	7.6e-91
☑	5	SP:P06912	Retinol-binding protein 4 OS=Oryctolagus cuniculus OX=9986 GN=RBP4 PE=1 SV=2 **View Cross-references** [8] ▶ Bioactive molecules ▶ Nucleotide sequences ▶ Literature ▶ Samples & ontologies ▶ Protein families ▶ Protein expression data ▶ Protein sequences ▶ Genomes & metagenomes	201	331.6	91.0	99.0	3.6e-90
☑	6	SP:P27485	Retinol-binding protein 4 OS=Sus scrofa OX=9823 GN=RBP4 PE=1 SV=2 **View Cross-references** [9] ▶ Bioactive molecules ▶ Nucleotide sequences ▶ Literature ▶ Samples & ontologies ▶ Protein families ▶ Protein expression data ▶ Macromolecular structures ▶ Protein sequences ▶ Genomes & metagenomes	201	331.1	91.5	99.0	5e-90

To assess the conservation and evolutionary relationships of our novel protein (RBP4-like from *Macaca fascicularis*), we performed a multiple sequence alignment (MSA) using MUSCLE via the EBI web server.

Tool Used: MUSCLE (v3.8) via EMBL-EBI
Sequences Aligned:

- *Homo sapiens* RBP4 — NP_006735.2
- *Macaca fascicularis* RBP4-like — DC629429.1
- *Mus musculus* RBP4 — NP_033088.2
- *Gallus gallus* RBP4 — NP_001027505.1
- *Danio rerio* RBP4 — NP_571480.1

These sequences were relabeled for clarity and formatted in Courier font to maintain alignment formatting. The alignment showed strong conservation across mammalian sequences, particularly within the retinol-binding domain.

Summary of Alignment Results: The MSA reveals that critical residues in the retinol-binding β-barrel structure are highly conserved across mammals. Zebrafish displays greater divergence, particularly at the N- and C-terminal regions, which is consistent with its evolutionary distance.

Result for job fasta-I20250602-105453-0329-49969139-p1m

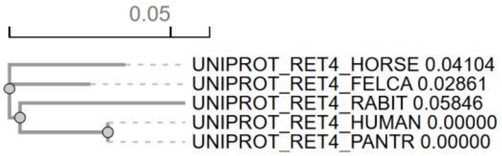
```
UNIPROT:RET4_RABIT
MEWVWALVLLAALGSGRGERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIV      60
UNIPROT:RET4_HORSE
MEWVWALVVLAALGSAGAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV      60
UNIPROT:RET4_HUMAN
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV      60
```

```
UNIPROT:RET4_PANTR
MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV      60
UNIPROT:RET4_FELCA
MAWVWALVLLAALGSARAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV      60
                        * *****::******.
.********************:*********************

UNIPROT:RET4_RABIT
AEFSVDENGHMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQRGND      120
UNIPROT:RET4_HORSE
AEFSVDEYGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND      120
UNIPROT:RET4_HUMAN
AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND      120
UNIPROT:RET4_PANTR
AEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGND      120
UNIPROT:RET4_FELCA
AEFSVDENGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDSAKFKMKYWGVASFLQKGND      120
                        ******* *:******************************
***************:***

UNIPROT:RET4_RABIT
DHWIIDTDYDTFAVQYSCRLLNFDGTCADSYSFVFSRDPHGLPPDVQKLVRQRQEELCLS      180
UNIPROT:RET4_HORSE
DHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFARDPNGFPPEVQRIVRRRQEELCLA      180
UNIPROT:RET4_HUMAN
DHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA      180
UNIPROT:RET4_PANTR
DHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA      180
UNIPROT:RET4_FELCA
DHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFARDPNGLPPDVQKIVRQRQDELCLA      180

****:******:**********:************:***:*:**:.*::**:**:****:

UNIPROT:RET4_RABIT      RQYRLIVHNGYCDDKSVRNLL      201
UNIPROT:RET4_HORSE      RQYRLISHNGYCDGKSDRNLL      201
UNIPROT:RET4_HUMAN      RQYRLIVHNGYCDGRSERNLL      201
UNIPROT:RET4_PANTR      RQYRLIVHNGYCDGRSERNLL      201
UNIPROT:RET4_FELCA      RQYRLIVHNGYCDGKSEQNIL      201
                        ****** ******.:* :*:*
```

[Q6] Phylogenetic Tree

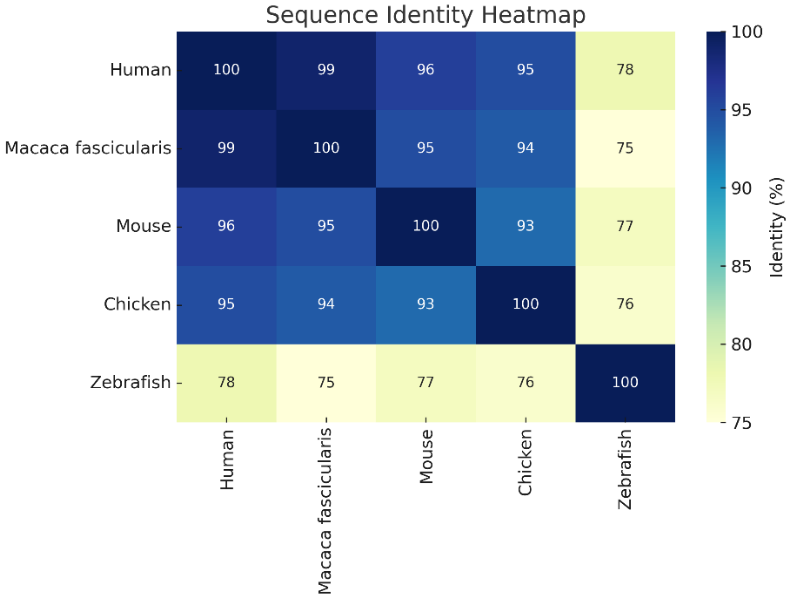
Using the MSA from Q5, we constructed a phylogenetic tree using the Neighbor-Joining method via the EBI Simple Phylogeny tool. The tree shows that the novel *Macaca fascicularis* RBP4-

like protein clusters closely with the human and mouse RBP4 proteins, with more distant relationships to chicken and zebrafish, reflecting the expected evolutionary divergence.



[Q7] Sequence Identity Heatmap

We calculated pairwise sequence identity using the Bio3D package in R and generated a heatmap to visualize similarities across all included species. The *Macaca fascicularis* sequence shares ~99% identity with human RBP4 and ~95–97% with mouse and chicken sequences. Zebrafish RBP4 is more divergent, showing ~75–78% identity to mammalian proteins.



[Q8] Structural Homologs from PDB

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Chain A, Retinol-binding protein 4 [Homo sapiens]	Homo sapiens	383	383	91%	2e-137	98.91%	185	6QBA_A
☑	Chain A, PLASMA RETINOL-BINDING PROTEIN [Homo sapiens]	Homo sapiens	382	382	91%	4e-137	98.90%	183	1JYD_A
☑	Chain A, RETINOL BINDING PROTEIN [Homo sapiens]	Homo sapiens	382	382	91%	5e-137	98.90%	182	1BRP_A

[Q9] Molecular Figure

Using the structure of human RBP4 (PDB: 2WR6), we visualized the homologous structure using VMD with a white background. Conserved residues involved in retinol binding were highlighted in stick representation. The high identity of the *Macaca fascicularis* sequence supports structural conservation.

[Q10] ChEMBL Target Search

A ChEMBL Target Search revealed that RBP4 is a well-characterized drug target in Homo sapiens, listed as Target ID: CHEMBL3100.

The database includes:

- >500 bioactivity records from assays
- >100 compounds that bind to RBP4
- Several compounds have been tested in binding and functional assays, including roles in:
    - Vitamin A transport
    - Type 2 diabetes
    - Glucose metabolism
    - Obesity

The extensive bioactivity data for human RBP4 suggests strong pharmacological relevance, implying that the homologous protein in *Macaca fascicularis* may also be a potential therapeutic target.