

BIMM 143 : INTRODUCTION TO BIOINFORMATICS

The find-a-gene project assignment

Q1

Name: Retinol binding protein 4 isoform a precursor (RBP4)

Accession: NP_006735.2

Species: Homo Sapiens

Q2

Method: TBLASTN 2.8.1+ search against Apicomplexans ESTs

Database: Expressed Sequence Tags (est)

Organism: apicomplexans (taxid:5794)

The screenshot shows the NCBI BLAST search interface. At the top, there are logos for NIH and NCBI, and a navigation bar with links like Home, Recent Results, Saved Strategies, and Help. The main heading is "BLAST" followed by "» tblastn". Below this, there's a sub-header "Translated BLAST: tblastn". The interface is divided into several sections: "Enter Query Sequence" with a text box containing "NP_006735.2", "Or, upload file" with a "Choose File" button, "Job Title" with a text box, and "Choose Search Set" with a dropdown menu set to "Expressed sequence tags (est)". Under "Choose Search Set", there are options for "Organism" (set to "apicomplexans (taxid:5794)"), "Exclude" (with checkboxes for "Models (XM/XP)" and "Uncultured/environmental sample sequences"), "Limit to" (with a checkbox for "Sequences from type material"), and "Entrez Query" (with a text box and a "Create custom database" link). At the bottom, there's a "BLAST" button and a "Search database Expressed sequence tags (est) using Tblastn" button. A note at the bottom states: "Note: Parameter values that differ from the default are highlighted in yellow and marked with * sign".

Chosen Match:

Accession DV107637.1, a 796 base pair clone from *Toxoplasma gondii*. See below for alignment details

NIH

U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign in to NCBI

BLAST » tblastn » RID-0BHSEDPJ015

HomeRecent ResultsSaved StrategiesHelp

BLAST Results

Your search is limited to records that include: apicomplexans (taxid:5794) » Full Entrez Query

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

Job title: ref|NP_006735.2| (201 letters)

RID0BHSEDPJ015 (Expires on 12-05 07:15 am)

Query IDNP_006735.2

Descriptionretinol-binding protein 4 isoform a precursor [Homo sapiens]

Molecule typeamino acid

Query Length201

Database Nameest

DescriptionDatabase of GenBank+EMBL+DBJ sequences from EST Divisions

ProgramTBLASTN 2.8.1+ » Citation

Other reports: » Search Summary » Taxonomy reports

Graphic Summary

Distribution of the top 6 Blast Hits on 6 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

<40

40-50

50-80

80-200

>=200

Query

14080120160200

Descriptions

Sequences producing significant alignments:

Select: AllNoneSelected 0

AlignmentsDownload » GenBankGraphics

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	EST-MOA152B11 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA152B11 5', mRNA sequence	380	380	100%	6e-135	89%	DV107637.1
<input type="checkbox"/>	EST-MOA166B10 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA166B10 5', mRNA sequence	380	380	100%	2e-134	89%	DV108899.1
<input type="checkbox"/>	EST-MOA159D12 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA159D12 5', mRNA sequence	163	163	41%	2e-49	88%	DV108308.1
<input type="checkbox"/>	EST-MOA163B07 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA163B07 5', mRNA sequence	130	130	37%	2e-36	82%	DV108631.1
<input type="checkbox"/>	EST-MOA152G08 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA152G08 5', mRNA sequence	60.5	60.5	14%	6e-11	87%	DV107693.1
<input type="checkbox"/>	EST-MOA151E09 Toxoplasma gondii Tachyzoite Library Toxoplasma gondii cDNA clone MOA151E09 5', mRNA sequence	36.2	36.2	8%	0		

Questions/comments

Alignments

Download » GenBankGraphics

EST-MOA152B11 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA152B11 5', mRNA sequence.
Sequence ID: DV107637.1 Length: 796 Number of Matches: 1

Range 1: 60 to 662 GenBankGraphics

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
380 bits(977)	6e-135	Compositional matrix adjust.	185/201(92%)	193/201(96%)	0/201(0%)	+3
Query 1	MKWVWl1111a16SGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	60				
Sbjct 60	MEWVWAL+LLAALGS +AERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	239				
Query 61	AEFSVDETGMSSATAKGRVRLNNWDVCDMVGTFDTEDPAKFKMKYWGVSFLQKGND	120				
Sbjct 240	AEFSVDENGHMSATAKGRVRLNNWDVCDMVGTFDTEDPAKFKMKYWGVSFLQKGND	419				
Query 121	DHWIIVDTDYDTYAVQYSCRLNLDGTCADSYSFVSRDPNGLPPEAQKIVRQRQEEELCLA	180				
Sbjct 420	DHWI+DTDYDTYAVQYSCRLNLDGTCADSYSFV+RDP+G PE QKIVRQRQEEELCLA	599				
Query 181	RQYRLIVHNGYCDGRSERNLL 201					
Sbjct 600	RQYRLI HNGYCDG+SERN+L 662					

Chinmay Kalluraya
PID: A13623530
ckallura@ucsd.edu

Alignment Details:

>DV107637.1 | EST-MOA152B11 Toxoplasma gondii Bradyzoite Library Toxoplasma gondii cDNA clone MOA152B11 5', mRNA sequence.

Length=796

Score = 380 bits (977), Expect = 6e-135, Method: Compositional matrix adjust.,
Identities = 185/201 (92%), Positives = 193/201 (96%), Gaps = 0/201(0%), Frame = +3

```
Query 1      MKWVWalllllaalGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV 60
              M+WVWAL+LLAALGS +AERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Sbjct 60      MEVWVALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV 239

Query 61      AEFSVDETGQMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGND 120
              AEFSVDE G MSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGND
Sbjct 240      AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGND 419

Query 121     DHWIVDTDYDTYAVQYSCRLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA 180
              DHWI+DTDYDTYAVQYSCRL NLDGTCADSYSFVF+RDP+G PE QKIVRQRQEELCLA
Sbjct 420      DHWIIDTDYDTYAVQYSCRLQNLGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA 599

Query 181     RQYRLIVHNGYCDGRSERNLL 201
              RQYRLI HNGYCDG+SERN+L
Sbjct 600     RQYRLITHNGYCDGKSERNIL 662
```

Q3

Chosen Sequence:

>T. gondii (sequence taken from BLAST result)
MEVWVALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGND
DHWIIDTDYDTYAVQYSCRLQNLGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA
RQYRLITHNGYCDGKSERNIL

Protein: Toxoplasma RBP4

Species: Toxoplasma gondii

Eukaryota; Alveolata; Apicomplexa; Conoidasida; Coccidia;
Eucoccidiorida; Eimeriorina; Sarcocystidae; Toxoplasma.

Q4

A BLASTP search against NR database (see setup in first screen-shot below) yielded a top hit result is to a protein from *Sus scrofa* (Pig). See additional screen shots below for top hits and selected alignment details:

BLAST® >> blastp suite

HomeRecent ResultsSaved StrategiesHelp

Standard Protein BLAST

blastnblastpblastxtblastntblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Reset pageBookmark

Enter Query Sequence

Enter accession number(s), g(i)s, or FASTA sequence(s)

>DV107637.1_1 EST-MOA152B11 Toxoplasma gondii Bradyzoite Library
Toxoplasma gondii cDNA clone MOA152B11 5', mRNA sequence.
NEWNAVLVLLALGSAQERDCRVSSFRVKEIPDKARFSGTNWMAKKDPEGLFLQDNIV
AEFSVDENGHSATAGRVRLINIMDVCAIMVGTFTDTEDPAKFKMKYMGVASFLQKGNID
DHMIIDTDYDTYAVQVSCRLQILQDGTCAADSYSFVFARDPHGFSPEVQKIVRQRQEEELCLA

ClearQuery subrange

From

To

Or, upload file

Choose FileNo file chosen

Job Title

DV107637.1_1 EST-MOA152B11 Toxoplasma gondii...

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism

Optional

Enter organism name or id—completions will be suggested

Exclude

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

Entrez Query

Optional

YouTubeCreate custom database

Enter an Entrez query to limit search

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)

Choose a BLAST algorithm

BLAST

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

The top result is to a protein from *Sus scrofa* (Pig), see second screen shot below for alignment details:

Sequences producing significant alignments:

Select: AllNoneSelected:0

AlignmentsDownloadGenPeptGraphicsDistance tree of resultsMultiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	retinol-binding protein 4 isoform X1 [Sus scrofa]	423	423	100%	5e-150	100%	XP_020927762.1
<input type="checkbox"/>	RecName: Full=Retinol-binding protein 4; AltName: Full=Plasma retinol-binding protein; Short=PRBP; Short=RBP; Flags=Precursor	422	422	100%	1e-149	99%	P27485.2
<input type="checkbox"/>	retinol-binding protein 4 precursor [Sus scrofa]	422	422	100%	2e-149	99%	NP_999222.1
<input type="checkbox"/>	PREDICTED: retinol-binding protein 4 [Camelus bactrianus]	412	412	100%	1e-145	97%	XP_010969515.1
<input type="checkbox"/>	PREDICTED: retinol-binding protein 4 isoform X2 [Ovis aries musimon]	411	411	100%	3e-145	97%	XP_011992224.1
<input type="checkbox"/>	PREDICTED: retinol-binding protein 4 isoform X2 [Vicugna pacos]	410	410	100%	5e-145	96%	XP_006211089.1
<input type="checkbox"/>	retinol-binding protein 4 precursor [Bos taurus]	409	409	100%	1e-144	97%	NP_001035565.2

Alignments

Download

GenPept

Graphics

retinol-binding protein 4 isoform X1 [Sus scrofa]
Sequence ID: [XP_020927762.1](#) Length: 201 Number of Matches: 1

Range 1: 1 to 201

GenPept

Graphics

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
423 bits(1088)	5e-150	Compositional matrix adjust.	201/201(100%)	201/201(100%)	0/201(0%)
Query 1	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	60			
Sbjct 1	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	60			
Query 61	AEFSVDENGHMSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND	120			
Sbjct 61	AEFSVDENGHMSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND	120			
Query 121	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA	180			
Sbjct 121	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA	180			
Query 181	RQYRLITHNGYCDGKSERNIL 201				
Sbjct 181	RQYRLITHNGYCDGKSERNIL 201				

Download

GenPept

Graphics

RecName: Full=Retinol-binding protein 4; AltName: Full=Plasma retinol-binding protein; Short=PRBP; Short=RBP; Flags: Precursor
Sequence ID: [P27485.2](#) Length: 201 Number of Matches: 1

Range 1: 1 to 201

GenPept

Graphics

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
422 bits(1085)	1e-149	Compositional matrix adjust.	200/201(99%)	201/201(100%)	0/201(0%)
Query 1	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	60			
Sbjct 1	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV	60			
Query 61	AEFSVDENGHMSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND	120			
Sbjct 61	AEFSVDENGHMSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND	120			
Query 121	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA	180			
Sbjct 121	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA	180			
Query 181	RQYRLITHNGYCDGKSERNIL 201				
Sbjct 181	RQYRIITHNGYCDGKSERNIL 201				

Q5

Re-labelled sequences for this alignment

```
>NP_006735.2 retinol-binding protein 4 isoform a precursor [Homo sapiens]
MKVWVALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA
AEFSVDETGQ
MSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND
DHWIIVDTDYDTYAVQYSCRL
LNLDTGTCADSYSFVSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL
```

```
>T. gondii (sequence taken from BLAST result)
MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
AEFSVDENGHMSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND
DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA
RQYRLITHNGYCDGKSERNIL
```

```
>XP_020927762.1 retinol-binding protein 4 isoform X1 [Sus scrofa]
MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA
AEFSVDENGH
MSATAKGRVRLNNWVDCADMVGTFDTEDPAKFKMKYWGVASFLQKGND
DHWIIDTDYDTYAVQYSCRL
QNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLARQYRLITHNGYCDGKSERNIL
```

>NP_001035565.2 retinol-binding protein 4 precursor [Bos taurus]
MEWVWALVLLAALGSARAERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVAEFSDENGH
MSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYETFAVQYSCRL
LNLDGTCADSYSFVFARDPSGFSPEVQKIVRQRQEELCLARQYRLIPHNGYCDGKSERNIL

>NP_001301158.1 retinol-binding protein 4 precursor [Capra hircus]
MEWVCALVLLAALGSARAERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVAEFSDENGH
MSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYETFAVQYSCRL
LNLDGTCADSYSFVFARDPSGFAPEVQKIVRQRQEELCLARQYRLIPHNGYCDGKSERNIL

>NP_001075420.1 retinol-binding protein 4 precursor [Equus caballus]
MEWVWALVLLAALGSAGAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSDVEYGO
MSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTYAVQYSCRL
LNLDGTCADSYSFVFARDPNGFPPEVQKIVRRRQEELCLARQYRLISHNGYCDGKSDRNLL

>NP_001277178.1 retinol binding protein 4 precursor [Felis catus]
MAWVWALVLLAALGSARAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSDENGH
MSATAKGRVRLNNWDVCADMVGTFDTEDSAKFKMKYWGVASFLQKGNDDHWIIDTDYDTYAVQYSCRL
LNLDGTCADSYSFVFARDPNGLPPDVQKIVRQRQEELCLSRQYRLIVHNGYCDGKSEQNIL

>NP_001075790.1 retinol-binding protein 4 precursor [Oryctolagus cuniculus]
MEWVWALVLLAALGSGRGERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVAEFSDENGH
MSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTFAVQYSCRL
LNFDTGTCADSYSFVFSRDPHGLPPDVQKIVRQRQEELCLSRQYRLIVHNGYCDKSVRNLL

>NP_035385.1 retinol-binding protein 4 isoform 2 precursor [Mus musculus]
MEWVWALVLLAALGGGSAERDCRVSSFRVKENFDKARFSGLWYAIKKDPEGLFLQDNIIAEFSVDEKGH
MSATAKGRVRLLSNWEVCADMVGTFDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTFALQYSCRL
QNLDGTCADSYSFVFSRDPNGLSPETRRLVRQRQEELCLERQYRWIEHNGYCQSRPSRNSL

Alignment:
Obtained using MUSCLE (version 3.8) at EBI:
CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Mus_musculus	MEWVWALVLLAALGGGSAERDCRVSSFRVKENFDKARFSGLWYAIKKDPEGLFLQDNII
Oryctolagus_cuniculus	MEWVWALVLLAALGSGRGERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIV
Homo_sapiens	MKVVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Equus_caballus	MEWVWALVLLAALGSAGAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Felis_catus	MAWVWALVLLAALGSARAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Toxoplasma_gondii	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Sus_scrofa	MEWVWALVLLAALGSAQAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV
Bos_taurus	MEWVWALVLLAALGSARAERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIV
Capra_hircus	MEWVCALVLLAALGSARAERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIV
	* * * *: : ***** .. . ***** : * * *: ***** :
Mus_musculus	AEFSVDEKGHMSATAKGRVRLLSNWEVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Oryctolagus_cuniculus	AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Homo_sapiens	AEFSVDETGQMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Equus_caballus	AEFSVDEYGOQMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Felis_catus	AEFSVDENGQMSATAKGRVRLNNWDVCADMVGTFDTEDSAKFKMKYWGVASFLQKGN
Toxoplasma_gondii	AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Sus_scrofa	AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Bos_taurus	AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
Capra_hircus	AEFSVDENGHMSATAKGRVRLNNWDVCADMVGTFDTEDPAKFKMKYWGVASFLQKGN
	***** * : ***** . * : ***** . ***** . * *

Mus_musculus	DHWIIDTDYDTFALQYSCRLQNLDGTCADSYSFVFSRDPNGLSPETRRLVLRQRQEELCLE
Oryctolagus_cuniculus	DHWIIDTDYDTFAVQYSCRLNFDGTCADSYSFVFSRDPHGLPPDVQKIVRQRQEELCLS
Homo_sapiens	DHWIVDTDYDTYAVQYSCRLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA
Equus_caballus	DHWIIDTDYDTYAVQYSCRLNLDGTCADSYSFVFARDPNGLFPPEVQRIVRRRQEELCLA
Felis_catus	DHWIIDTDYDTYAVQYSCRLNLDGTCADSYSFVFARDPNGLPPDVQKIVRQRQDELCLA
Toxoplasma_gondii	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA
Sus_scrofa	DHWIIDTDYDTYAVQYSCRLQNLDGTCADSYSFVFARDPHGFSPEVQKIVRQRQEELCLA
Bos_taurus	DHWIIDTDYETFAVQYSCRLNLDGTCADSYSFVFARDPSGFSPEVQKIVRQRQEELCLA
Capra_hircus	DHWIIDTDYETYAVQYSCRLNLDGTCADSYSFVFARDPSGFAPEVQKIVRQRQEELCLA
	****.****.*.*.***** *.*****.*** *:.*:..:*.*.****

Mus_musculus	RQYRWIEHNGYCCSRPSRNSL
Oryctolagus_cuniculus	RQYRLIVHNGYCDCKSVRNLL
Homo_sapiens	RQYRLIVHNGYCDGRSERNLL
Equus_caballus	RQYRLISHNGYCDGKSDRNLL
Felis_catus	RQYRLIVHNGYCDGKSEQNIL
Toxoplasma_gondii	RQYRLITHNGYCDGKSERNIL
Sus_scrofa	RQYRLITHNGYCDGKSERNIL
Bos_taurus	RQYRLIPHNGYCDGKSERNIL
Capra_hircus	RQYRLIPHNGYCDGKSERNIL
	**** * *****:... .* *

Q6

Import the multiple alignment sequences into Simple Phylogeny online from the EBI, and create a neighbor-joining tree:

Simple Phylogeny

[Input form](#) | [Web services](#) | [Help & Documentation](#) | [Bioinformatics Tools FAQ](#)

Tools > Phylogeny > Simple Phylogeny

Results for job simple_phylogeny-l20181204-012902-0496-31782507-p1m

Phylogenetic Tree

Result Summary

Submission Details

Phylogram

Branch length: ☒ Cladogram ☐ Real

Mus_musculus 0.10025

Oryctolagus_cuniculus 0.03905

Homo_sapiens 0.03265

Felis_catus 0.03102

Equus_caballus 0.03475

Toxoplasma_gondii 0

Sus_scrofa 0

Bos_taurus 0.00498

Capra_hircus 0.00995

Q7

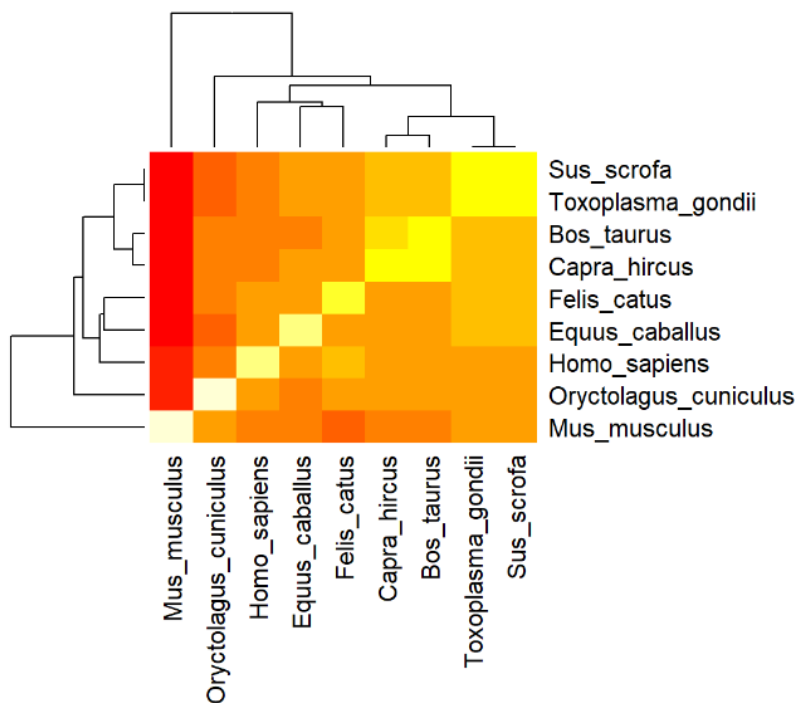
Chinmay Kalluraya
December 3, 2018

Read Fasta File

```
library("bio3d")  
file <- read.fasta("sample.fasta")  
file_seqidmat <- seqidentity(file)
```

Generate heatmap

```
heatmap(file_seqidmat, margins = c(12,8))
```



Q8

Table shown at the bottom of the attached screenshot. Also shown R code to generate the table.

Generate Consensus sequence and run blast to find the most similar atomic resolution structures to your aligned sequences

```
file.consensus <- consensus(file)
pdb_blast <- blast.pdb(file.consensus$seq)
```

```
## Searching ... please wait (updates every 5 seconds) RID = 0C240TKC014
## .....
## Reporting 76 hits
```

```
pdb_annotate <- pdb.annotate(pdb_blast$hit.tbl$pdb.id)
```

```
## Warning in pdb.annotate(pdb_blast$hit.tbl$pdb.id): ids should be standard 4
## character PDB-IDs: trying first 4 characters...
```

Merge columns from pdb_blast and pdb_annotate and find top 3 unique hits

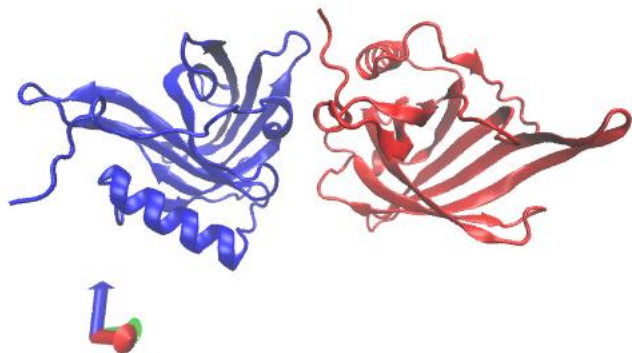
```
library(knitr)
pdb_data <- merge(pdb_annotate, pdb_blast$hit.tbl, by.y = "pdb.id", by.x = 0)
pdb_data.eval <- pdb_data[order(pdb_data$eval),]
unique_hits <- pdb_data.eval[1:3, c("structureId", "experimentalTechnique", "resolution", "source", "evalu
e", "identity")]
colnames(unique_hits) <- c("ID", "Technique", "Resolution", "Source", "Evaluate", "Identity")
kable(unique_hits, row.names = FALSE, digits = 130)
```

ID	Technique	Resolution	Source	Evalue	Identity
1AQB	X-RAY DIFFRACTION	1.65	Sus scrofa	8.10e-129	95.604
1KT3	X-RAY DIFFRACTION	1.4	Bos taurus	1.81e-128	95.055
4O9S	X-RAY DIFFRACTION	2.3	Homo sapiens	1.83e-127	93.478

Q9

Very likely to be similar in structure to Toxoplasma RBP4 given the high sequence similarity (>90%). In the figure below the RBP4 chain A is colored blue and corresponds to the Toxoplasma RBP subject of this report.

To generate high quality image: File -> Render -> Render using Tachyon(internal), replace Render Command with "C:\Program Files (x86)\University of Illinois\VMD\tachyon_WIN32.exe" tmp.dat -format BMP -o path/to/your/image.bmp



Q10

CHEMBL Details:

ChEMBL BLAST Search Results: 2

#1:

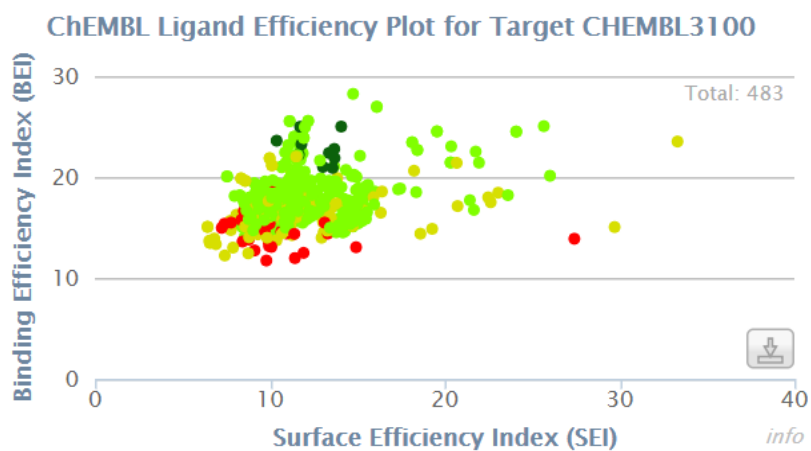
Target ID - CHEMBL3100

Target Type - Single Protein

15 Binding Assays

3 Functional Assays

Target Ligand Efficiency Data:



<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL3100>

Chinmay Kalluraya
PID: A13623530
ckallura@ucsd.edu

#2

Target ID - CHEMBL3430865

Target Type - Single Protein

1 Binding Assay ([CHEMBL3873621](#))

Binding assay linked manuscript discovered potent, selective dioxane-fused tricyclic benz[d]imidazole derivatives that act as as mPGES-1 inhibitor with results suggesting that one molecule, 17d, "proved to be suitable for further profiling *in vivo*."

Nagarajan Muthukaman, Sanjay Deshmukh, Neelam Sarode, Shital Tondlekar, Macchindra Tambe, Dnyandeo Pisal, Mahamadhanif Shaikh, Vidya G. Kattige, Srinivasa Honnegowda, Vikas Karande, Abhay Kulkarni, Satyawar B. Jadhav, Mahamad Yunnus A. Mahat, Girish S. Gudi et al. J. Med. Chem., (2003) 46:6:936 <https://doi.org/10.1016/j.bmcl.2016.10.079>

<https://www.sciencedirect.com/science/article/pii/S0960894X16311271?via%3Dihub>

0 Functional Assays

No Target Ligand Efficiency Data

<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL3430865>