

Alberto Carreno
alcarren@ucsd.edu
A10216025

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
Accession: NP_001310446
Species: Homo Sapiens

blastnblastpblastxtblastntblastx

Enter Query Sequence

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

ref|NP_001310446

ClearQuery subrange

From

To

Or, upload file

Choose FileNo file chosen

Job Title

ref|NP_001310446|

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Expressed sequence tags (est)

Organism

Drosophila (taxid:7215)

Exclude

Models (XM/XP)Uncultured/environmental sample sequences

Limit to

Sequences from type material

Entrez Query

Enter an Entrez query to limit search

BLAST

Search database est using Tblastn (search translated nucleotide databases using a protein query)

Your search is limited to records that include: Drosophila (taxid:7215)

Job Title

ref|NP_001310446|

RID

4SCJRM7016

Search expires on 02-20 01:52 am

Download All

Program

TBLASTN

Citation

Database

est

See details

Query ID

NP_001310446.1

Description

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

Molecule type

amino acid

Query Length

201

Other reports

[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)

Method: TBLASTN search against drosophila.
Database: Expressed Sequence Tag
Organism: (Drosophila Taxid:7215)

| select all 27 sequences selected | | GenBankGraphics | | | | | |
|-------------------------------------|---|-----------------|-------------|-------------|---------|------------|------------|
| | Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Accession |
| <input checked="" type="checkbox"/> | B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence | 40.8 | 40.8 | 32% | 0.003 | 32.84% | FE040818.1 |
| <input checked="" type="checkbox"/> | DK274535 Drosophila full-length cDNA library, Drosophila auraria whole body adult mixed Drosophila auraria cDNA clone daya25m14.5 | 38.1 | 38.1 | 79% | 0.022 | 20.75% | DK274535.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51871357 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-064_E21.5', mRNA sequence | 35.4 | 35.4 | 59% | 0.22 | 22.69% | EB558760.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51187444 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-034_L24.5', mRNA sequence | 35.4 | 35.4 | 59% | 0.22 | 22.69% | EB557394.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51774687 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-057_F3.5', mRNA sequence | 35.0 | 35.0 | 59% | 0.25 | 22.69% | EB570064.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51105651 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-015_J6.5', mRNA sequence | 35.0 | 35.0 | 59% | 0.29 | 22.69% | EB574233.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51356556 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-043_E4.5', mRNA sequence | 35.0 | 35.0 | 59% | 0.29 | 22.69% | EB564290.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51356657 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-043_H7.5', mRNA sequence | 35.0 | 35.0 | 59% | 0.32 | 22.69% | EB557412.1 |
| <input checked="" type="checkbox"/> | AGENCOURT_51803700 D. virilis EST Drosophila virilis cDNA clone 161524_D.VIRILIS_CDNA-052_D22.5', mRNA sequence | 34.7 | 34.7 | 59% | 0.32 | 22.69% | EB558613.1 |

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

B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
Sequence ID: [FE040818.1](#) Length: 603 Number of Matches: 1

Range 1: 263 to 448 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

| Score | Expect | Method | Identities | Positives | Gaps | Frame |
|---------------|--------|------------------------------|------------|------------|----------|-------|
| 40.8 bits(94) | 0.003 | Compositional matrix adjust. | 22/67(33%) | 40/67(59%) | 6/67(8%) | +2 |

Query 121 DHWIVDTDYDYAVQVSCRLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQKQ-EELCL 179

D ++++D+D+Y T+++QY+C + +DG +Y+ + +RD P+++ I + R+ EL

Sbjct 263 DIYVLDTDYKTFSIQYAC--IEVDGNLNLAYAIILTRDM--PDSKVIKKSRLAELSG 427

Query 180 ARQYRLI 186

R LI

Sbjct 428 VELRRLI 448

```
>FE040818.1_1 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
HNIRWKLEYEKVSGYMVSTIVVSTLSSSFARMCKV*YC*GQKW*L*NSQIGS*RQEWSFA
SANFIDNKGSG*KGWQIRFKNEKFT*RP*YICAGHRL*NLFNSICVHRSGWQSKLSVCHHI
DAGPNAGL*SD*EVTKIGRTLWS*AKKVDTHFTEELP*GCLKVVGLNMLH*VFLTKDSSL
FLFESY*NK*IRLSYQLAELV
>FE040818.1_2 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
ITSVGNLNMKKYQGIWYPQLSYPLYLHPLPECAKFNIVKGKNGNYEIHRSDDLKGKSGLLR
QRTSLITKVDKKGKGYALKMKNSHDGLDIYVLDTDYKTFSIQYACIEVDGNLNLAYAIIL
TRDRMPDSKVIKKSRLAELSGVELRRLIPISQKSCPDA*KLLGLICFIRYS*QKIQVC
FYSSLIRINKSDFLIN*LS*X
>FE040818.1_3 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
*HPLET*I*KSIRVYGIHNCRIFHIFILCPNVQSLILLRAKMVIMKFTDRILTARVVFCV
SELH**QKWIKRVANTL*K*KIHMATALIYMCWTPTIKPFQFNMRA*KWMAI*T*RMPSY*
RGTECRTLK*LRSHENWPNLSLELS*EG*YPFHRRALRMPKSCWA*YASLGILNKRKFV
SIRVLE*INQTFLSIS*ASX
>FE040818.1_4 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
N*LS*LIRKSDLFILIRLE*KQT*IFC*EYLMKHIKPNNF*ASLGQLFCEMGINLLSSTP
ESSANFRDFLITLESGIRSRVNMAYAKFRLPSTSMHAY*IEKVL*SVSSTYISRPSCEF
FIFKAYLPPFLSTFVINEVR*RKRPLLPSRSDL*IS*LPFLPLTILNFAHSGKG*R*SGY
DNCGYHIP*YFFIFKFPTDVM
>FE040818.1_5 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
LAQLIDKKV*FIYSNKTRIETNLNLLLRIPEAY*AQQLLGILRAALL*NGYQPS*LNSR
EFGQFS*LLNHFRVRHSVPRQYDGIR*V*IAIHFYARILN*KGFIVGVQHIYIKAVM*IF
HF*SVFATLFIHFICYQ*SSLTQKTLAVKIRSVNFIITIFALNNIKLCTFGQRMKIKWIR
QLWIPYTLILFHIQVSNCGYX
>FE040818.1_6 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
TSSAN**ESLIYLF**DSNRNKLESFVKNT**SILSPTTFRHP*GSSSVKWVSTFLAQLQ
RVRPIFVTS*SL*SPAFGPASI*WHTLSLDCHPLLCETHIELKRFYSRCPAHYQGRHVNF
SFLKRICHFPFYPLLSMKFADAKDHSCRQDPICEFHNYHFCP*QY*TLHIRAKDEDKVDT
TIVDTIYPDTFSYSSFQRMXLX
```

[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.

```
>FE040818.1_2 B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
ITSVGNLNMKKYQGIWYPQLSYPLYLHPLPECAKFNIVKGKNGNYEIHRSDDLKGKSGLLR
QRTSLITKVDKKGKGYALKMKNSHDGLDIYVLDTDYKTFSIQYACIEVDGNLNLAYAIIL
TRDRMPDSKVIKKSRLAELSGVELRRLIPISQKSCPDA*KLLGLICFIRYS*QKIQVC
FYSSLIRINKSDFLIN*LS*X
```

Taxonomy: Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Holometabola; Diptera; brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Mayaguana

[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.

Ran protein-protein BLAST of the sequence above. Top result was for *D. mojavensis*

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

Clear

Query subrange [?](#)

From

To

>FE040818.1_2_B_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence
ITSVGILHKKYQGLINYPQLSYPLYLHPLPECAKFNIVKNGGNGYEDHSQDGKSGLLRQTSLSITKVD
KXGGKVALKWKHGHGLODYVLDOTDYKTFSTQVACIEVDGHLNLAIAIITRDRIIPDSKVIKKSRLAEL
SGVELRLIPTISQKSCPKDA*KLGLGICFIRYS*QKIQVCFYSSILIRINKSDFLIN*LS*X

Or, upload file No file chosen [?](#)

Job Title
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 se

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 [?](#)

| | |
|---------------|--|
| Job Title | FE040818.1_2_B_E03 Accessory gland cDNA library... |
| RID | 4SCR6MD6016 <small>Search expires on 02-20 01:54 am</small> Download All ? |
| Program | BLASTP ? Citation ? |
| Database | nr See details ? |
| Query ID | lc Query_55160 |
| Description | FE040818.1_2_B_E03 Accessory gland cDNA library Drosophi... ? |
| Molecule type | amino acid |
| Query Length | 201 |
| Other reports | Distance tree of results Multiple alignment MSA viewer ? |

| <input checked="" type="checkbox"/> select all 100 sequences selected | | GenPept | Graphics | Distance tree of results | Multiple alignment | | |
|---|--|-------------------------|--------------------------|--|------------------------------------|------------|--------------------------------|
| | Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Accession |
| <input checked="" type="checkbox"/> | uncharacterized protein Dmoj_GI26513 [Drosophila mojavensis] | 216 | 216 | 79% | 3e-68 | 57.50% | XP_015020355.1 |
| <input checked="" type="checkbox"/> | PREDICTED: apolipoprotein D isoform X1 [Drosophila arizonae] | 215 | 215 | 79% | 1e-67 | 57.50% | XP_017857158.1 |
| <input checked="" type="checkbox"/> | PREDICTED: apolipoprotein D isoform X2 [Drosophila arizonae] | 215 | 215 | 79% | 1e-67 | 57.50% | XP_017857159.1 |
| <input checked="" type="checkbox"/> | hypothetical protein AWZ03_004095 [Drosophila navojoa] | 205 | 205 | 79% | 8e-64 | 56.25% | TDG49412.1 |
| <input checked="" type="checkbox"/> | uncharacterized protein LOC115562160 [Drosophila navojoa] | 205 | 205 | 79% | 9e-64 | 56.25% | XP_030238856.1 |
| <input checked="" type="checkbox"/> | lazarillo protein isoform X1 [Drosophila hydei] | 144 | 144 | 77% | 8e-40 | 43.59% | XP_023165552.1 |
| <input checked="" type="checkbox"/> | lazarillo protein isoform X2 [Drosophila hydei] | 144 | 144 | 77% | 1e-39 | 43.59% | XP_023165553.1 |
| <input checked="" type="checkbox"/> | blast:Apolipoprotein D [Drosophila guanche] | 115 | 115 | 79% | 1e-28 | 35.80% | SPP79949.1 |
| <input checked="" type="checkbox"/> | PREDICTED: lopap [Drosophila miranda] | 113 | 113 | 79% | 8e-28 | 37.65% | XP_017151350.1 |
| <input checked="" type="checkbox"/> | GL14848 [Drosophila persimilis] | 112 | 112 | 79% | 9e-28 | 37.04% | EDW39695.1 |
| <input checked="" type="checkbox"/> | lopap [Drosophila persimilis] | 113 | 113 | 79% | 1e-27 | 37.04% | XP_002020702.2 |

