Alberto Carreno

[alcarren@ucsd.edu](mailto:alcarren@ucsd.edu)

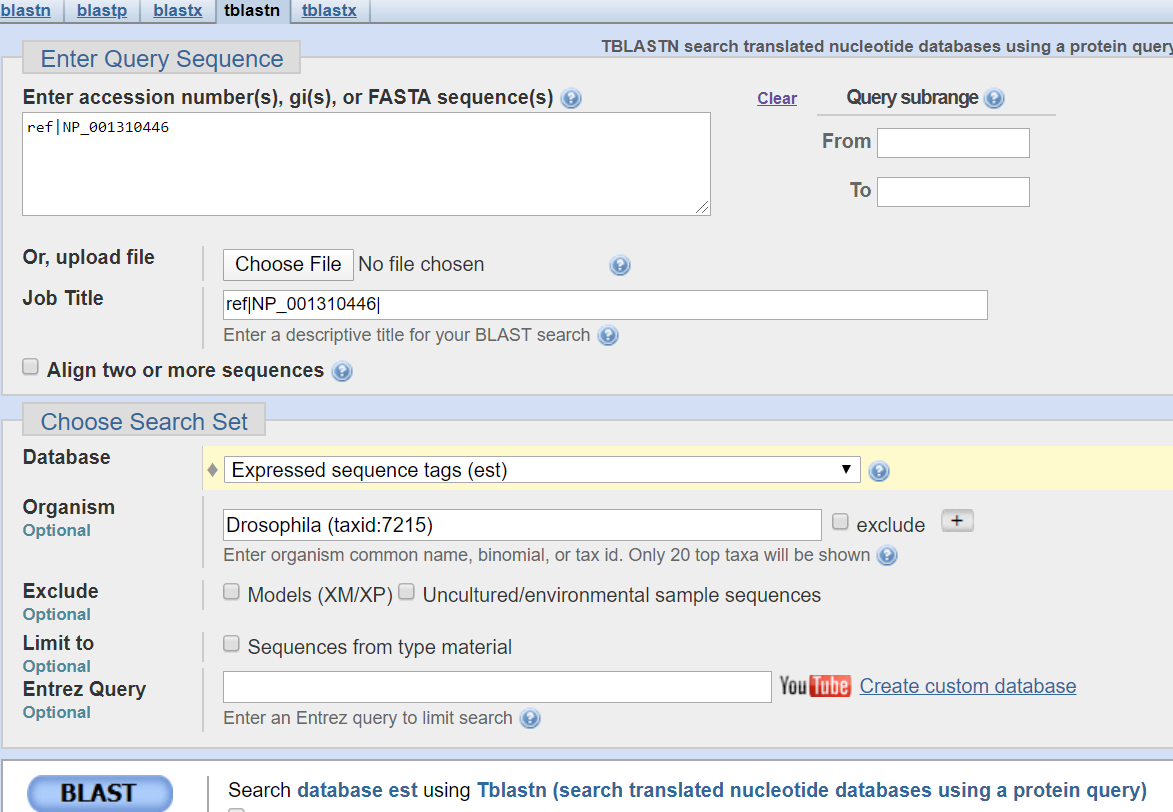
A10216025

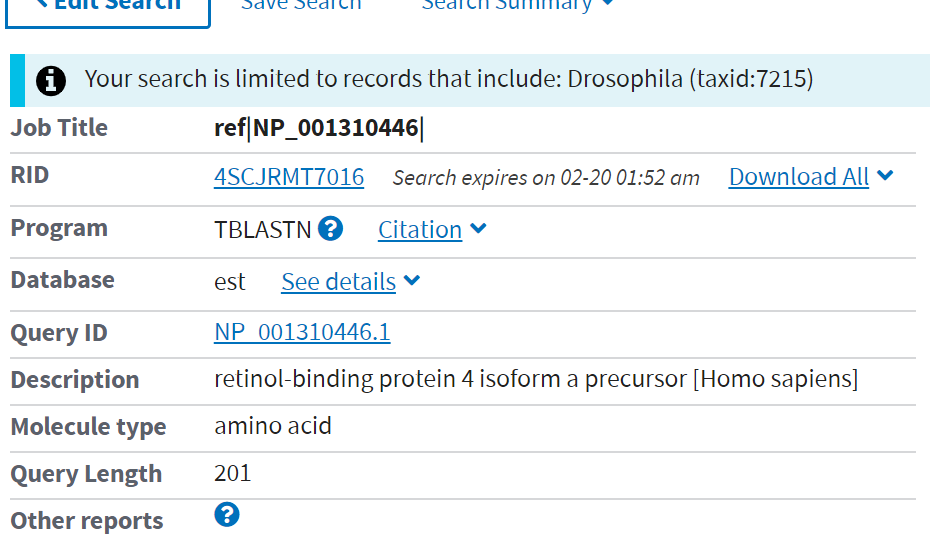
Questions: [Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as it's function is known. If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online

Name: RBP4

Accession: NP\_001310446

Species: Homo Sapiens



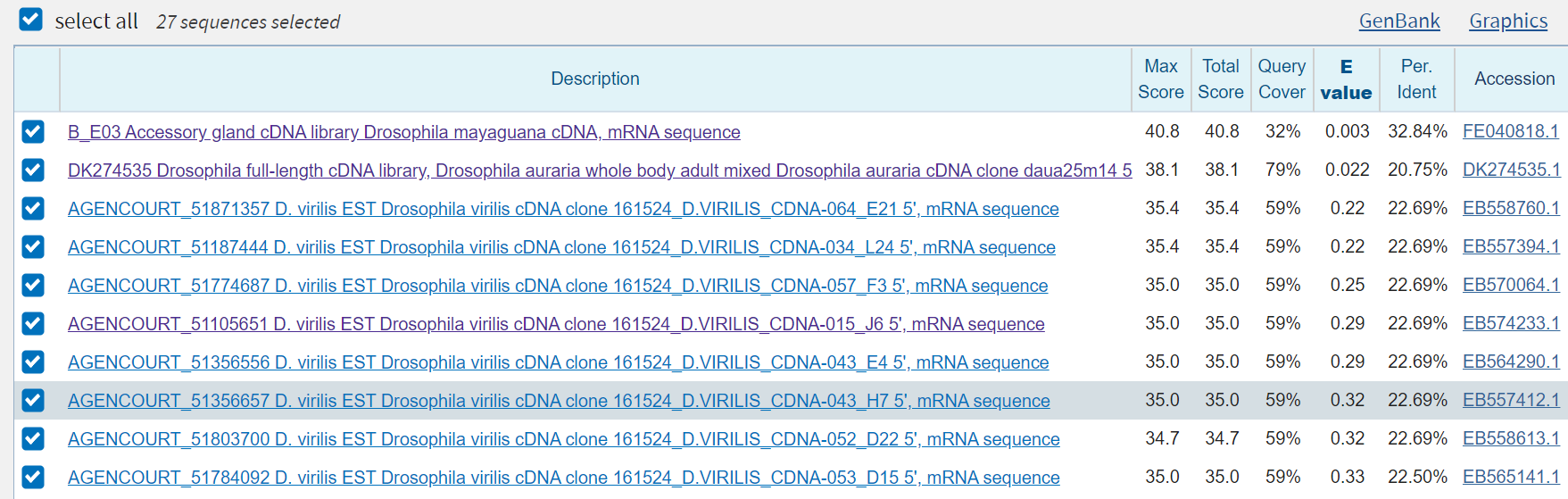


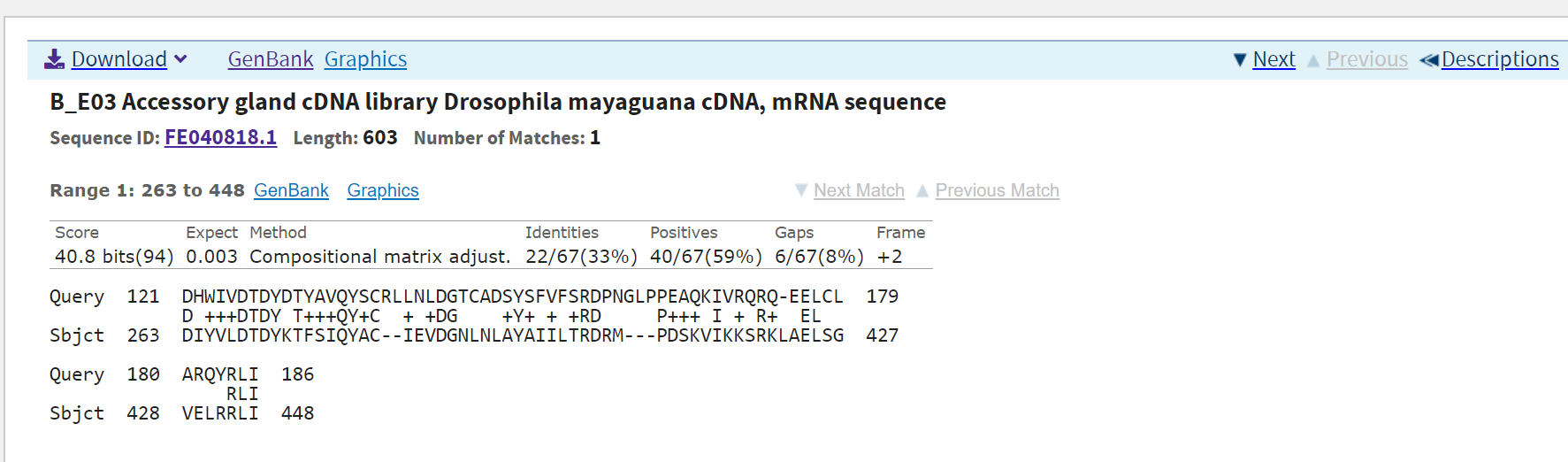
[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism)

Method: TBLASTN search against drosophila.

Database: Expressed Sequence Tag

Organism: (Drosophila Taxid:7215)





>FE040818.1\_1 B\_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence

HNIRWKLEYEKVSGYMVSTIVVSTLSSSFARMCKV\*YC\*GQKW\*L\*NSQIGS\*RQEWSFA

SANFIDNKSG\*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**

**ITSVGNLNMKKYQGIWYPQLSYPLYLHPLPECAKFNIVKGKNGNYEIHRSDLDGKSGLLR**

**QRTSLITKVDKKGGKYALKMKNSHDGLDIYVLDTDYKTFSIQYACIEVDGNLNLAYAIIL**

**TRDRMPDSKVIKKSRKLAELSGVELRRLIPISQKSCPKDA\*KLLGLICFIRYS\*QKIQVC**

**FYSSLIRINKSDFLIN\*LS\*X**

>FE040818.1\_3 B\_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence

\*HPLET\*I\*KSIRVYGIHNCRIHFIFILCPNVQSLILLRAKMVIMKFTDRILTARVVFCV

SELH\*\*QKWIKRVANTL\*K\*KIHMTALIYMCWTPTIKPFQFNMRA\*KWMAI\*T\*RMPSY\*

RGTECRTLK\*LRSHENWPNSLELS\*EG\*YPFHRRAALRMPKSCWA\*YASLGILNKRFKFV

SIRVLLE\*INQTFLSIS\*ASX

>FE040818.1\_4 B\_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence

N\*LS\*LIRKSDLFILIRLE\*KQT\*IFC\*EYLMKHIKPNNF\*ASLGQLFCEMGINLLSSTP

ESSANFRDFLITLESGIRSRVNMMAYAKFRLPSTSMHAY\*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\*FIYSNKTRIETNLNLLLRIPNEAY\*AQQLLGILRAALL\*NGYQPS\*LNSR

EFGQFS\*LLNHFRVRHSVPRQYDGIR\*V\*IAIHFYARILN\*KGFIVGVQHIYIKAVM\*IF

HF\*SVFATLFIHFCYQ\*SSLTQKTTLAVKIRSVNFIITIFALNNIKLCTFGQRMKIKWIR

QLWIPYTLILFHIQVSNGCYX

>FE040818.1\_6 B\_E03 Accessory gland cDNA library Drosophila mayaguana cDNA, mRNA sequence

TSSAN\*\*ESLIYLF\*\*DSNRNKLESFVKNT\*\*SILSPTTFRHP\*GSSSVKWVSTFLAQLQ

RVRPIFVTS\*SL\*SPAFGPASI\*WHTLSLDCHPLLCTHIELKRFYSRCPAHIYQGRHVNF

SFLKRICHPFYPLLLSMKFADAKDHSCRQDPICEFHNYHFCP\*QY\*TLHIRAKDEDKVDT

TIVDTIYPDTFSYSSFQRMLX

[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**

**ITSVGNLNMKKYQGIWYPQLSYPLYLHPLPECAKFNIVKGKNGNYEIHRSDLDGKSGLLR**

**QRTSLITKVDKKGGKYALKMKNSHDGLDIYVLDTDYKTFSIQYACIEVDGNLNLAYAIIL**

**TRDRMPDSKVIKKSRKLAELSGVELRRLIPISQKSCPKDA\*KLLGLICFIRYS\*QKIQVC**

**FYSSLIRINKSDFLIN\*LS\*X**

Taxonomy:Eukaryota;Metazoa,Ecdysozoa;Arthropoda;Hexapoda;Insecta;Pterygote;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*

