Find a Gene Project Q1-4

Katherine Quach (kpquach@ucsd.edu, A1854014)

Table of contents

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.

Name: kinesin family member 11

Accession: NP_004514

Species: Homo sapiens

Function: Enables ATP, microtubule, nucleotide, protein, and protein kinase binding. It also enables microtubule and plus-end-directed microtubule motor activities.

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 Mus musculus (House mouse) ESTs

Database: Expressed Sequence Tags (est)

Organism: Mus musculus (taxid:10090)

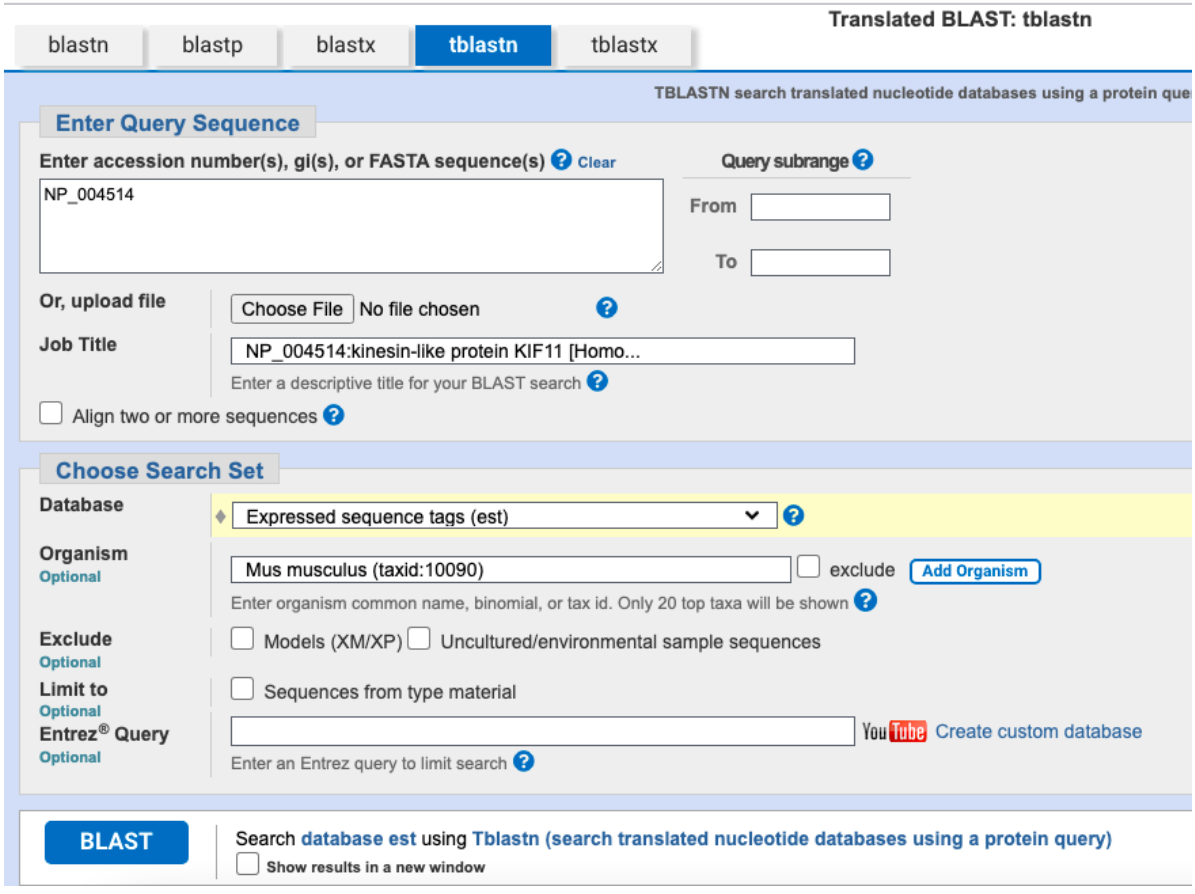

Figure 1: Blast Method Used

Chosen Match: Accession CN533430.1, a 722 base pair clone from Mus musculus.

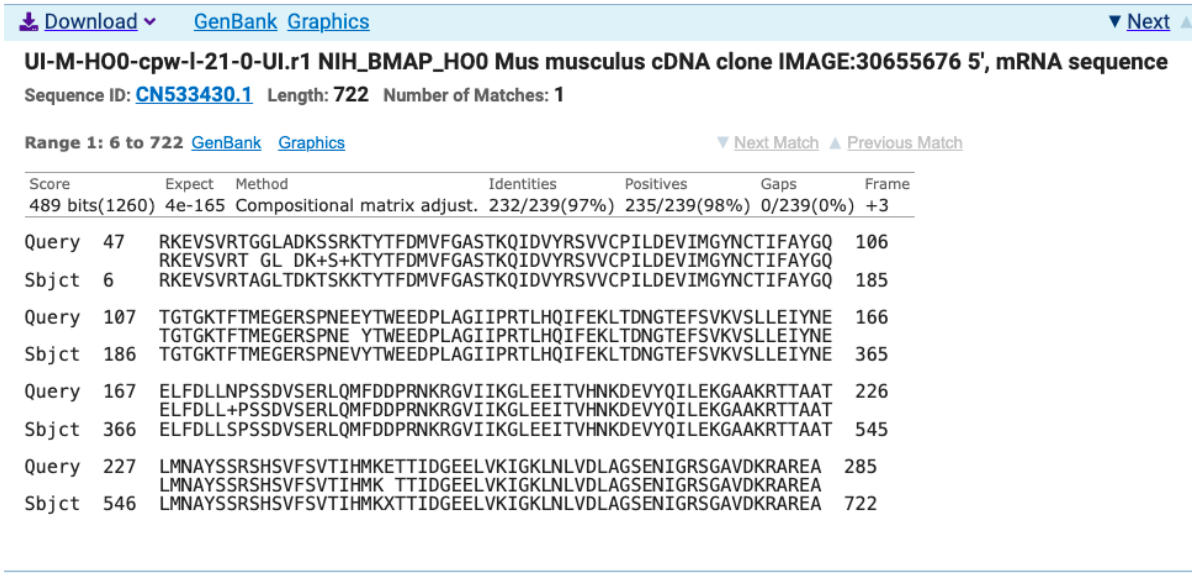

Figure 2: Chosen Match

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.

Chosen Sequence (FASTA Format): >CN533430.1 UI-M-HO0-cpw-l-21-0-UI.r1 NIH_BMAP_HO0 Mus musculus cDNA clone IMAGE:30655676 5', mRNA sequence
RKEVSVRTAGLTDKTSKKTYTFDMVFGASTKQIDVYRSVVCPILDEVIMGYNCTI-
FAYGQTGTGKTFTMEGERSPNEVYTWEEDPLAGIIPRTLHQIFEKLTDNGTEFSVKVS-
LLEIYNEELFDLLSPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEVYQILEK-
GAAKRTTAATLMNAYSSRSHSVFSVTIHMKXTTIDGEELVKIGKLNLVDLAGSENIGRS-
GAVDKRAREA

Protein Name: UI-M-HO0-cpw-l-21-0-UI.r1 NIH_BMAP_HO0 Mus musculus cDNA clone IMAGE:30655676 5', mRNA sequence

Species: Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus.

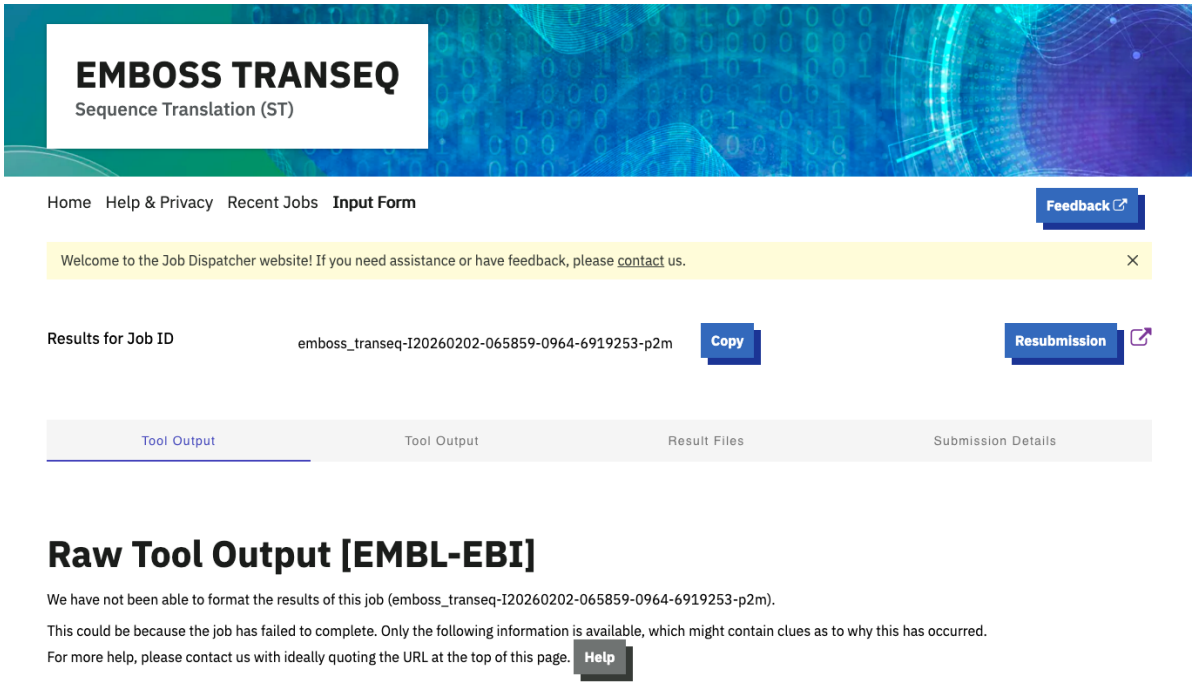

Figure 3: EMBOSS Transeq Search

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.

Description: A BLASTP search against NR database yielded a top hit result to a protein from Rattus norvegicus (NP_001162583.1).

Species: Rattus norvegicus

Name: kinesin-like protein KIF11 [Rattus norvegicus]

Percent Identity: 99.16%. Since the top match has less than 100% identity, this must mean the protein is novel because there is no identical match that is 100% to this specific species in the protein database.

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
kinesin family member 11, isoform CRA_b, partial [Mus musculus]	Mus musculus	517	517	100%	1e-153	99.58%	1064	EDL41788.1
kinesin-like protein KIF11 [Mus musculus]	Mus musculus	516	516	100%	2e-153	99.58%	1052	NP_034745.1
kinesin-like protein KIF11 [Rattus norvegicus]	Rattus norvegicus	515	515	100%	5e-153	99.16%	1056	NP_001162583.1
kinesin-like protein KIF11 [Rattus rattus]	Rattus rattus	515	515	100%	5e-153	99.16%	1056	XP_032747743.1
rCG48024 [Rattus norvegicus]	Rattus norvegicus	515	515	100%	5e-153	99.16%	850	EDM13198.1
kinesin-like protein KIF11 [Peromyscus maniculatus bairdii]	Peromyscus maniculatus bairdii	514	514	100%	1e-152	98.74%	1055	XP_042118941.2
kinesin-like protein KIF11 [Peromyscus leucopus]	Peromyscus leucopus	514	514	100%	1e-152	98.74%	1055	XP_028738726.1
kinesin-like protein KIF11 isoform X1 [Peromyscus maniculatus bairdii]	Peromyscus maniculatus bairdii	514	514	100%	2e-152	98.74%	1055	XP_042118941.1
kinesin-like protein KIF11 [Arvicanthis niloticus]	Arvicanthis niloticus	514	514	100%	2e-152	99.16%	1054	XP_034350713.1
kinesin-like protein KIF11 [Mastomys coucha]	Mastomys coucha	513	513	100%	2e-152	98.74%	1055	XP_031246053.1
Chain A, KINESIN-LIKE PROTEIN KIF11 [Homo sapiens]	Homo sapiens	500	500	100%	5e-152	97.07%	348	3ZCW_A
Chain A, Kinesin-like protein KIF11 [Homo sapiens]	Homo sapiens	503	503	100%	5e-152	97.07%	382	3WPN_A
Chain A, Kinesin-like protein KIF11 [Homo sapiens]	Homo sapiens	501	501	100%	5e-152	97.07%	367	1Q0B_A
kinesin-like protein KIF11 isoform X1 [Phodopus roborovskii]	Phodopus roborovskii	512	512	100%	5e-152	98.33%	1055	XP_051055532.1
Chain A, Kinesin-like protein KIF11 [Homo sapiens]	Homo sapiens	501	501	100%	5e-152	97.07%	367	5ZO7_A
Chain A, Kinesin-like protein KIF11 [Homo sapiens]	Homo sapiens	501	501	100%	5e-152	97.07%	367	6TLE_A
Chain A, KINESIN-RELATED MOTOR PROTEIN Eg5 [Homo sapiens]	Homo sapiens	501	501	100%	5e-152	97.07%	368	1II6_A

Figure 4: BLASTP Search Against NR Dataset