Find a Gene Project Q1-Q4

Max Gruber

5/16/23

Find a Gene

[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-like protein (KIF11)

Accession: NP_004514.2

Species: Homo sapiens

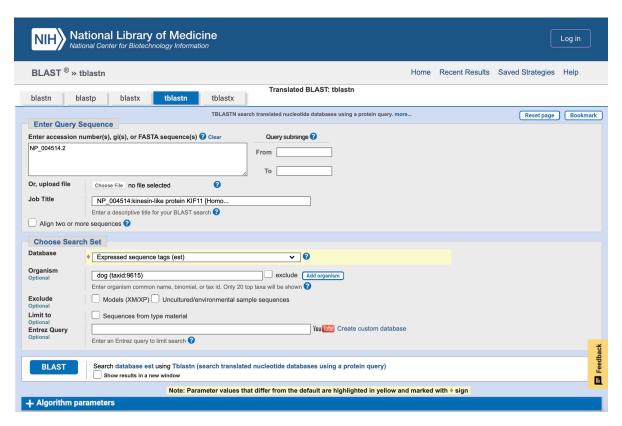
Function: This protein is involved in various types of spindle dynamics, including chromosomal positioning, centrosome separation, and establishing bipolar spindle during mitosis.

[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).

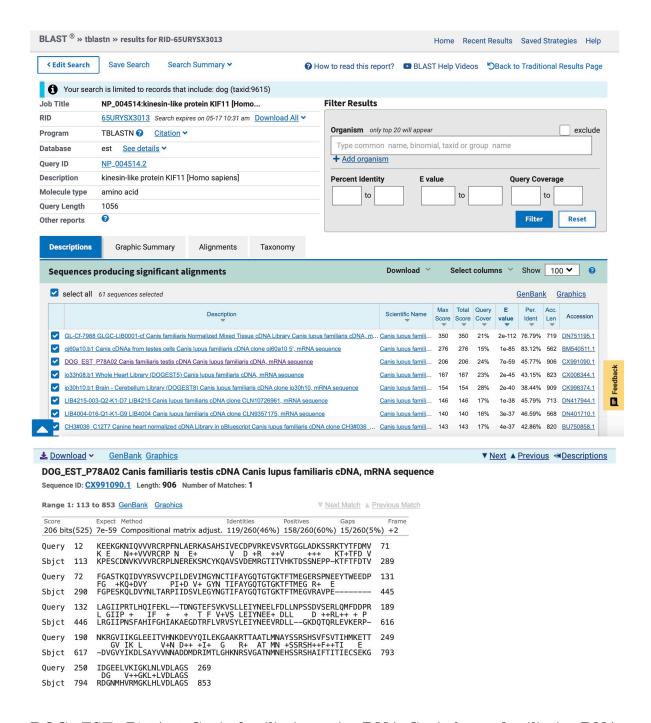
Method: TBLASTN search against dog ESTS

Database: Expressed Sequence Tags (est)

Organism: Dogs (Taxid: 6659)



Chosen match: Accession CX991090.1, a 906 base pair clone from $Canis\ familiaris\ (dogs)$. See below for alignment details.



 ${\tt DOG_EST_P78A02}$ Canis familiaris testis c
DNA Canis lupus familiaris c DNA, mRNA sequence

Sequence ID: CX991090.1 Length: 906 Number of Matches: 1

Range 1: 113 to 853 GenBankGraphics Next Match Previous Match

Score 206 bits(528	7ϵ	1	istics
Query	12	KEEKGKNIQVVVRCRPFNLAERKASAHSIVECDPVRKEVSVRTGGLADKSSRKTYTFDMV K E N++VVVRCRP N E+ V D +R ++V +++ KT+TFD V	71
Sbjct	113	KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRGTITVHKTDSSNEPP-KTFTFDTV	289
Query	72	FGASTKQIDVYRSVVCPILDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDP	131
Sbjct	290	FG +KQ+DVY PI+D V+ GYN TIFAYGQTGTGKTFTMEG R+ E FGPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPE	445
Query	132	LAGIIPRTLHQIFEKLTDNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPR	189
Sbjct	446	L GIIP + IF + + T F V+VS LEIYNEE+ DLL D ++RL++ + P LRGIIPNSFAHIFGHIAKAEGDTRFLVRVSYLEIYNEEVRDLLGKDQTQRLEVKERP-	616
Query	190	NKRGVIIKGLEEITVHNKDEVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETT	249
Sbjct	617	GV IK L V+N D++ +I+ G R+ AT MN +SSRSH++F++TI E -DVGVYIKDLSAYVVNNADDMDRIMTLGHKNRSVGATNMNEHSSRSHAIFTITIECSEKG	793
Query	250	IDGEELVKIGKLNLVDLAGS 269	
Sbjct	794	DG V++GKL+LVDLAGS RDGNMHVRMGKLHLVDLAGS 853	

[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 FASTAformat (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:

>113-853_1 DOG_EST_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA seq

(copy and pasted aligned sequence subject lines from BLAST result)

KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRGTITVHKTDSSNEP

-KTFTFDTVFGPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPE-----LRGIIPNSFAHIFGHIAKAEGDT

>113-853_1 DOG_EST_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA seq (translated nucleotide BLAST sequence via EMBOSS Transeq) KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRGTITVHKTDSSNEPPKTFTFDTVF GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPELRGIIPNSF AHIFGHIAKAEGDTRFLVRVSYLEIYNEEVRDLLGKDQTQRLEVKERPDVGVYIKDLSAY VVNNADDMDRIMTLGHKNRSVGATNMNEHSSRSHAIFTITIECSEKGRDGNMHVRMGKLH LVDLAGS

Name: Canis lupus familiaris testis cDNA

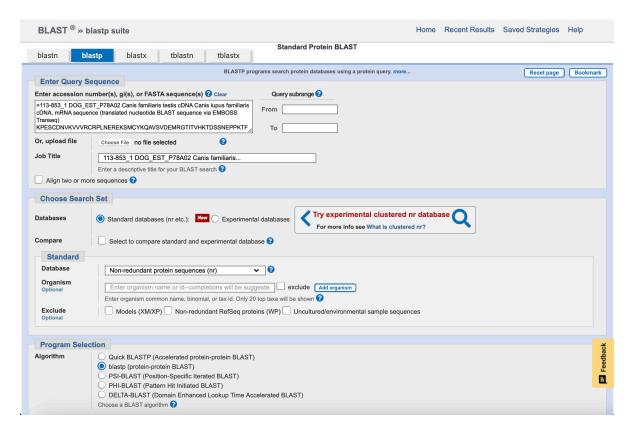
Species: Canis lupus familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Caniformia; Canidae; Canis.

[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.

Details:

A BLASTP search against NR database (see setup in first screen-shot below) yielded a top hit result is to a protein from



The top result is to a protein from Prionailurus viverrinus (fishing cat), see second screenshot below for alignment details:



