

# 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)

National Library of Medicine  
National Center for Biotechnology Information

Log in

BLAST® » tblastn

HomeRecent ResultsSaved StrategiesHelp

blastnblasttblastntblastx

Translated BLAST: tblastn

TBLASTN search translated nucleotide databases using a protein query. more...

Reset pageBookmark

Enter Query Sequence

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

NP\_004514.2

Query subrange ?

From

To

Or, upload file

Choose File no file selected ?

Job Title

NP\_004514:kinesin-like protein KIF11 [Homo...

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

Expressed sequence tags (est) ?

Organism

Optional

dog (taxid:9615) ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude

Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to

Optional

☐ Sequences from type material

Entrez Query

Optional

[YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search ?

BLAST

☐ Search database est using Tblastn (search translated nucleotide databases using a protein query)  
☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

+ Algorithm parameters

Feedback

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

BLAST® » tblastn » results for RID-65URYSX3013

[Home](#)
[Recent Results](#)
[Saved Strategies](#)
[Help](#)

[< Edit Search](#)
[Save Search](#)
[Search Summary ▾](#)

[? How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

**i**

Your search is limited to records that include: dog (taxid:9615)

Job Title

NP\_004514:kinesin-like protein KIF11 [Homo...

RID

65URYSX3013 [Search expires on 05-17 10:31 am](#) [Download All ▾](#)

Program

TBLASTN [Citation ▾](#)

Database

est [See details ▾](#)

Query ID

NP\_004514.2

Description

kinesin-like protein KIF11 [Homo sapiens]

Molecule type

amino acid

Query Length

1056

Other reports

[?](#)

Filter Results

Organism

only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

to

E value

to

Query Coverage

to

[Filter](#) [Reset](#)

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▾

Select columns ▾

Show 100 ▾

[?](#)

☒ select all 61 sequences selected

[GenBank](#) [Graphics](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	GL-Cf-7988 GLGC-LIB0001-cl Canis familiaris Normalized Mixed Tissue cDNA Library Canis lupus familiaris cDNA, m...	Canis lupus famili...	350	350	21%	2e-112	76.79%	719	DN751195.1
<input checked="" type="checkbox"/>	gi60a10.b1 Canis cDNAs from testes cells Canis lupus familiaris cDNA clone qj60a10.5', mRNA sequence	Canis lupus famili...	276	276	15%	1e-85	83.12%	562	BM540511.1
<input checked="" type="checkbox"/>	DOG_EST_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA sequence	Canis lupus famili...	206	206	24%	7e-59	45.77%	906	CX991090.1
<input checked="" type="checkbox"/>	io33h08.b1 Whole Heart Library (DOGEST5) Canis lupus familiaris cDNA, mRNA sequence	Canis lupus famili...	167	167	23%	2e-45	43.15%	823	CX008344.1
<input checked="" type="checkbox"/>	io30h10.b1 Brain - Cerebellum Library (DOGEST8) Canis lupus familiaris cDNA clone io30h10, mRNA sequence	Canis lupus famili...	154	154	28%	2e-40	38.44%	909	CK998374.1
<input checked="" type="checkbox"/>	LIB4215-003-Q2-K1-D7 LIB4215 Canis lupus familiaris cDNA clone CLN10726961, mRNA sequence	Canis lupus famili...	146	146	17%	1e-38	45.79%	713	DN417944.1
<input checked="" type="checkbox"/>	LIB4004-016-Q1-K1-G8 LIB4004 Canis lupus familiaris cDNA clone CLN9357175, mRNA sequence	Canis lupus famili...	140	140	16%	3e-37	46.59%	568	DN401710.1
<input checked="" type="checkbox"/>	CH3#036_C1217 Canine heart normalized cDNA Library in pBluescript Canis lupus familiaris cDNA clone CH3#036...	Canis lupus famili...	143	143	17%	4e-37	42.86%	820	BU750858.1

[Download ▾](#)
[GenBank](#)
[Graphics](#)

[▼ Next](#)
[▲ Previous](#)
[◀ Descriptions](#)

DOG\_EST\_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA sequence

Sequence ID: [CX991090.1](#) Length: 906 Number of Matches: 1

Range 1: 113 to 853 [GenBank](#) [Graphics](#)

[▼ Next Match](#)
[▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
206 bits(525)	7e-59	Compositional matrix adjust.	119/260(46%)	158/260(60%)	15/260(5%)	+2
Query 12	KEEKGNKIQVVRCPFNLAERKASAHSIVEDPVRKEVSVRTGGLADKSSRKYTFDMV	71				
Sbjct 113	K E N++VVVRCRP N E+ V D +R ++V +++ KT+TFD V	289				
Query 72	FGASTKQIDVYRSVCPILDEVIMGYNCTIFAYGQTGTGKFTMEGERSPNEEYTWEEDP	131				
Sbjct 290	FG +KQ+DVY PI+D V+ GYN TIFAYGQTGTGKFTMEG R+ E	445				
Query 132	LAGIIPRTLHQIFEKL---TDNGTEFSVKVSLLEIYNEELFDLLNPSSDVSRERLQMFDDPR	189				
Sbjct 446	L GIIP + IF + + T F V+VS LEIYNEE+ DLL D ++RL++ + P	616				
Query 190	NKRGVVIKGLEEITVHNKDEVYQILEKGAARKTTAATLMNAYSSRSHSVFSVTIHKETT	249				
Sbjct 617	GV IK L V+N D++ +I+ G R+ AT MN +SSRSH++F++TI E	793				
Query 250	IDGEELVKIGKLNLDVLAGS 269					
Sbjct 794	DG V++GKL+LVDLA GS 853					

DOG\_EST\_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA sequence

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

3

Range 1: 113 to 853 [GenBankGraphics](#) [Next Match](#) [Previous Match](#)

						Alignment statistics for match #1	
Score	Expect	Method	Identities	Positives	Gaps	Frame	
206 bits(525)	7e- 59	Compositional matrix adjust.	119/260(46%)	158/260(60%)	260(5%)	2	
Query	12	KEEKGKNIQVVVRCRPFNLAERKASAHSIVECDPVRKEVSVRTGGLADKSSRKTYTFDMV					71
		K E N++VVVRCRP N E+ V D +R ++V +++ KT+TFD V					
Sbjct	113	KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRTITVHKTDSSNEPP-KTFTFDTV					289
Query	72	FGASTKQIDVYRSVVCPIILDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDEP					131
		FG +KQ+DVY PI+D V+ GYN TIFAYGQTGTGKTFTMEG R+ E					
Sbjct	290	FGPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPE-----					445
Query	132	LAGIIPRTLHQIFEKL--TDNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPR					189
		L GIIP + IF + + T F V+VS LEIYNEE+ DLL D ++RL++ + P					
Sbjct	446	LRGIIPNSFAHIFGHIAKAEGDTRFLVRVSYLEIYNEEVRDLL--GKDQTRLEVKERP-					616
Query	190	NKRGVIIKGLEEITVHNKDEVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHKETT					249
		GV IK L V+N D++ +I+ G R+ AT MN +SSRSH++F++TI E					
Sbjct	617	-DVGVYIKDLSAYVVNNADDMDRIMTLGHKNRSVGATNMNEHSSRSHAIFTITIECSEKG					793
Query	250	IDGEELVKIGKLNLDLAGS					269
		DG V++GKL+LVDLAGS					
Sbjct	794	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)

```
KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRTITVHKTDSSNEP  
-KTFTFDTVFGPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPE-----LRGIIPNSFAHIFGHIKAEGDT
```

```
>113-853_1 DOG_EST_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA seq  
(translated nucleotide BLAST sequence via EMBOSS Transeq)  
KPESCDNVKVVVRCRPLNEREKSMCYKQAVSVDEMRTITVHKTDSSNEPPKTFTFDTVF  
GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKTFTMEGVRAVPELRGIIPNSF  
AHIFGHIKAEGDTRFLVRVSYLEIYNNEVRDLLGKDQTRLEVKERPDVGVYIKDLSAY  
VVNNADDMDRIMTLGHKNRSGATNMNEHSSRSHAIFTITIECSEKGRDGNMHVRMGKLH  
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

BLAST® » blastp suite Home Recent Results Saved Strategies Help

---

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

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

**Enter Query Sequence**

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

>113-853\_1 DOG\_EST\_P78A02 Canis familiaris testis cDNA Canis lupus familiaris cDNA, mRNA sequence (translated nucleotide BLAST sequence via EMBOSS Transeq)  
 KPESCDNVKVVRCRPLNEREKSMCYKQAVSVDEMRTITVHKTDSSNEPPKTF

Query subrange [?](#)

From

To

Or, upload file  no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

**Choose Search Set**

Databases ☒ Standard databases (nr etc.): New ☐ Experimental databases

Compare ☐ Select to compare standard and experimental database [?](#)

**Standard**

Database  [?](#)

Organism Optional  ☐ exclude [Add organism](#)

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

**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 [?](#)

Feedback

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

Sequences producing significant alignments

DownloadSelect columnsShow100

☒ 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
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X4 [Prionailurus viverrinus]	<a href="#">Prionailurus viverrinus</a>	525	525	100%	0.0	99.60%	611	<a href="#">XP_047712621.1</a>
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X3 [Mirounga leonina]	<a href="#">Mirounga leonina</a>	525	525	100%	0.0	99.60%	679	<a href="#">XP_034869996.1</a>
<input checked="" type="checkbox"/>	LOW QUALITY PROTEIN: kinesin-like protein KIF3A [Phacochoerus africanus]	<a href="#">Phacochoerus africanus</a>	525	525	100%	0.0	99.60%	702	<a href="#">XP_047634557.1</a>
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X5 [Talpa occidentalis]	<a href="#">Talpa occidentalis</a>	525	525	100%	0.0	99.60%	713	<a href="#">XP_054550385.1</a>
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X2 [Mirounga angustirostris]	<a href="#">Mirounga angustirostris</a>	525	525	100%	0.0	99.60%	676	<a href="#">XP_054367410.1</a>
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X2 [Talpa occidentalis]	<a href="#">Talpa occidentalis</a>	525	525	100%	0.0	99.60%	722	<a href="#">XP_037365254.1</a>
<input checked="" type="checkbox"/>	kinesin-like protein KIF3A isoform X5 [Ovis aries]	<a href="#">Ovis aries</a>	525	525	100%	0.0	99.60%	699	<a href="#">XP_042106166.1</a>

**kinesin-like protein KIF3A isoform X4 [Prionailurus viverrinus]**  
Sequence ID: [XP\\_047712621.1](#) Length: **611** Number of Matches: **1**

Range 1: 8 to 254 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
525 bits(1352)	0.0	Compositional matrix adjust.	246/247(99%)	246/247(99%)	0/247(0%)
Query 1	KPESCDNVKVVVRCRPLNEREKSMCYQAVSVDEM	RGTTIVHKTDSSNEPPKTTFTFDTVF	60		
Sbjct 8	KPESCDNVKVVVRCRPLNEREKSMCYQAVSVDEM	RGTTIVHKTDSSNEPPKTTFTFDTVF	67		
Query 61	GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKFTMEGVRVPEL	RGIIPN	120		
Sbjct 68	GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKFTMEGVRVPEL	RGIIPN	127		
Query 121	AHIFGHIKAEGDTRFLVRVSYLEIYNEEVRDLGKDQ	TQRLEVKERPDVGVIK	180		
Sbjct 128	AHIFGHIKAEGDTRFLVRVSYLEIYNEEVRDLGKDQ	TQRLEVKERPDVGVIK	187		
Query 181	VVNNADDMDRIMTLGHKNRSVGATNMNEHSSRS	HAIFTITIECSEKGRDGNMHV	240		
Sbjct 188	VVNNADDMDRIMTLGHKNRSVGATNMNEHSSRS	HAIFTITIECSEKGRDGNMHV	247		
Query 241	LVDLAGS	247			
Sbjct 248	LVDLAGS	254			

**Related Information**  
[Genome Data Viewer](#) - aligned genomic context

**kinesin-like protein KIF3A isoform X3 [Mirounga leonina]**  
Sequence ID: [XP\\_034869996.1](#) Length: **679** Number of Matches: **1**

Range 1: 8 to 254 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
525 bits(1352)	0.0	Compositional matrix adjust.	246/247(99%)	246/247(99%)	0/247(0%)
Query 1	KPESCDNVKVVVRCRPLNEREKSMCYQAVSVDEM	RGTTIVHKTDSSNEPPKTTFTFDTVF	60		
Sbjct 8	KPESCDNVKVVVRCRPLNEREKSMCYQAVSVDEM	RGTTIVHKTDSSNEPPKTTFTFDTVF	67		
Query 61	GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKFTMEGVRVPEL	RGIIPN	120		
Sbjct 68	GPESKQLDVYNLTARPIIDSVLEGYNGTIFAYGQTGTGKFTMEGVRVPEL	RGIIPN	127		
Query 121	AHIFGHIKAEGDTRFLVRVSYLEIYNEEVRDLGKDQ	TQRLEVKERPDVGVIK	180		
Sbjct 128	AHIFGHIKAEGDTRFLVRVSYLEIYNEEVRDLGKDQ	TQRLEVKERPDVGVIK	187		
Query 181	VVNNADDMDRIMTLGHKNRSVGATNMNEHSSRS	HAIFTITIECSEKGRDGNMHV	240		
Sbjct 188	VVNNADDMDRIMTLGHKNRSVGATNMNEHSSRS	HAIFTITIECSEKGRDGNMHV	247		
Query 241	LVDLAGS	247			
Sbjct 248	LVDLAGS	254			

**Related Information**  
[Gene](#) - associated gene details  
[Genome Data Viewer](#) - aligned genomic context