

Hope Adams
A15652616

Find a gene project

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 its 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: Kinesin-like protein KIF11

Accession: P52732

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 Gallus gallus EST's

Database: Expressed Sequence Tags (est)

Organism: Gallus gallus (taxid:9031)

Also include the output of that BLAST search in your document. If appropriate, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output (e.g. alt print screen on a PC or on a MAC press ⌘-shift-4. The pointer becomes a bulls eye. Select the area you wish to capture and release. The image is saved as a file called Screen Shot [].png in your Desktop directory). It is **not** necessary to print out all of the blast results if there are many pages.

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

New Select columns

Show 100

☒ select all 100 sequences selected

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	AJ741668 riken1 Gallus gallus cDNA clone 1f15e6, mRNA sequence	Gallus gallus	473	473	23%	2e-159	91.36%	752	AJ741668.1
<input checked="" type="checkbox"/>	AJ741657 riken1 Gallus gallus cDNA clone 1f15r2, mRNA sequence	Gallus gallus	467	467	25%	9e-157	84.64%	807	AJ741657.1
<input checked="" type="checkbox"/>	GGEZEB2005E01 g embryo breast muscle - EB2 Gallus gallus cDNA clone GGEZEB2005E01, mRNA sequence	Gallus gallus	419	419	19%	3e-139	93.81%	650	CQ507391.1
<input checked="" type="checkbox"/>	603744789F1 CSEQCHN38 Gallus gallus cDNA clone ChEST647b23 5', mRNA sequence	Gallus gallus	369	369	19%	1e-118	89.37%	899	BU255828.1
<input checked="" type="checkbox"/>	603763571F1 CSEQRBN21 Gallus gallus cDNA clone ChEST685n17 5', mRNA sequence	Gallus gallus	350	350	17%	3e-113	85.87%	571	BU427444.1
<input checked="" type="checkbox"/>	AJ447422 riken1 Gallus gallus cDNA clone 16i6r1, mRNA sequence	Gallus gallus	328	328	19%	1e-103	75.12%	787	AJ447422.1
<input checked="" type="checkbox"/>	AJ448735 riken1 Gallus gallus cDNA clone 1k5r1, mRNA sequence	Gallus gallus	311	311	16%	1e-97	82.39%	761	AJ448735.1
<input checked="" type="checkbox"/>	AJ448596 riken1 Gallus gallus cDNA clone 1f15r1, mRNA sequence	Gallus gallus	285	346	17%	6e-88	83.65%	749	AJ448596.1
<input checked="" type="checkbox"/>	AJ741667 riken1 Gallus gallus cDNA clone 1f15e5, mRNA sequence	Gallus gallus	231	340	22%	3e-85	76.51%	716	AJ741667.1
<input checked="" type="checkbox"/>	AJ450267 riken1 Gallus gallus cDNA clone 2414r1, mRNA sequence	Gallus gallus	260	260	13%	4e-79	82.64%	648	AJ450267.1

On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include no functional annotation.

Chosen match: Accession: AJ741668.1, a 752 basepair mRNA from Gallus gallus

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

[Next](#)
[Previous](#)
[Descriptions](#)

AJ741668 riken1 Gallus gallus cDNA clone 1f15s6, mRNA sequence

Sequence ID: [AJ741668.1](#) Length: 752 Number of Matches: 1

Range 1: 3 to 731 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
473 bits(1217)	2e-159	Compositional matrix adjust.	222/243(91%)	232/243(95%)	0/243(0%)	-1
Query 72	FCASTKQIDVYRSVVCPIlDEVIMGYNCTIFAYGQTGTGKFTTMEGERSPNEEYTWEEDP	131	F A KQIDVYRSVVCPIlDEVIMGYNCTIFAYGQTGTGKFTTMEGERSPNEEYTWEEDP			
Sbjct 731	FRAQAQKIDVYRSVVCPIlDEVIMGYNCTVFAYGQTGTGKFTTMEGERSPNEEYTWEEDP	552	FRAQAQKIDVYRSVVCPIlDEVIMGYNCTVFAYGQTGTGKFTTMEGERSPNEEYTWEEDP			
Query 132	LAGIIPRTLHQIFEKLTdNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPNRK	191	LAGIIPRTLHQIFEKLT+NGTEFSVKVSLLEIYNEELFDLLNP+ DV ERLQMFDDPNRK			
Sbjct 551	LAGIIPRTLHQIFEKLTENGTEFSVKVSLLEIYNEELFDLLNPADPVGRLQMFDDPNRK	372	LAGIIPRTLHQIFEKLTENGTEFSVKVSLLEIYNEELFDLLNPADPVGRLQMFDDPNRK			
Query 192	RGVIIIGLEEITVHNKDEVYQILEGKAARKTTAATLMNAYSSRSHSVFSVTIHMKETTID	251	RGVIIIGLEE+TVHNK+EVYQILE+GAARKTTAAT MNAYSSRSHSVFS+TIHMKETT+D			
Sbjct 371	RGVIIIGLEEVTVHNKNEVYQILERGAARKRTTAATYMNAYSSRSHSVFSITIHMKETTVD	192	RGVIIIGLEEVTVHNKNEVYQILERGAARKRTTAATYMNAYSSRSHSVFSITIHMKETTVD			
Query 252	GEELVKIGKLNlVDLAGSENIGRSGAVDKRKAREAGNIQSLLTLGRVITALVERTHPVY	311	GEELVKIGKLNlVDLAGSENIGRSGAVDKRKAREAGNIQSLLTLGRVI+ALVER P			
Sbjct 191	GEELVKIGKLNlVDLAGSENIGRSGAVDKRKAREAGNIQSLLTLGRVISALVERAPAYSI	12	GEELVKIGKLNlVDLAGSENIGRSGAVDKRKAREAGNIQSLLTLGRVISALVERAPAYSI			
Query 312	RES 314					
	+ES					
Sbjct 11	QES 3					

In general, [Q2] is the most difficult for students because it requires you to have a “feel” for how to interpret BLAST results. You need to distinguish between a perfect match to your query (i.e. a sequence that is not “novel”), a near match (something that might be “novel”, depending on the results of [Q4]), and a non-homologous result.

If you are having trouble finding a novel gene try restricting your search to an organism that is poorly annotated.

AJ741668 riken1 Gallus gallus cDNA clone 1f15s6, mRNA sequence

GenBank: AJ741668.1

[FASTA](#) [Graphics](#)

[Go to:](#) ☒

LOCUS AJ741668 752 bp mRNA linear EST 12-FEB-2011
DEFINITION AJ741668 riken1 Gallus gallus cDNA clone 1f15s6, mRNA sequence.
ACCESSION AJ741668
VERSION AJ741668.1
DBLINK BioSample: [SAMN00170383](#)
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM [Gallus gallus](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda;
Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes;
Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 752)
AUTHORS Caldwell,R.B., Kierzek,A.M., Arakawa,H., Bezzubov,Y., Zaim,J.,
Fiedler,P., Kutter,S., Blagodatski,A., Kostovska,D., Koter,M.,
Plachy,J., Carninci,P., Hayashizaki,Y. and Buerstedde,J.M.
TITLE Full-length cDNAs from chicken bursal lymphocytes to facilitate
gene function analysis
JOURNAL Genome Biol. 6 (1), R6 (2005)
PUBMED [15642098](#)
COMMENT Contact: Caldwell RB
GSF - Forschungszentrum, Institut fuer Molekulare Strahlenbiologie
Ingolstaedter Landstr. 1, D-85764 Neuherberg, GERMANY.
FEATURES Location/Qualifiers
source 1..752
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:[9031](#)"
/catalytic="1f15s6"

[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.

➤ *Gallus gallus* protein (taken from BLAST result)

```
Query 72 FGASTKQIDVYRSVVCPIIDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDP 131
      F A KQIDVYRSVVCPIIDEVIMGYNCT+FAYGQTGTGKTFTMEGERSPNEEYTWEEDP
Sbjct 731 FRAQAKQIDVYRSVVCPIIDEVIMGYNCTVFAYGQTGTGKTFTMEGERSPNEEYTWEEDP 552

Query 132 LAGIIPRTLHQIFEKLT DNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNK 191
      LAGIIPRTLHQIFEKLT+NGTEFSVKVSLLEIYNEELFDLLNP+ DV ERLQMFDDPRNK
Sbjct 551 LAGIIPRTLHQIFEKLTENGTEFSVKVSLLEIYNEELFDLLNPAPDVGERLQMFDDPRNK 372

Query 192 RGVIKGLLEEITVHNKDEVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTID 251
      RGVIKGLLEE+TVHNK+EVYQILE+GAAKRTTAAT MNAYSSRSHSVFS+TIHMKETT+D
Sbjct 371 RGVIKGLLEEVTVHNKNEVYQILERGAAKRTTAATYMNAYSSRSHSVFSITIHMKETTVD 192

Query 252 GEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVITALVERTPHVPY 311
      GEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVI+ALVER P
Sbjct 191 GEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVISALVERAPAYSI 12

Query 312 RES 314
      +ES
Sbjct 11 QES 3
```

➤ *Gallus gallus* protein (taken from BLAST result)

```
FRAQAKQIDVYRSVVCPIIDEVIMGYNCTVFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTLHQIFEKLT
ENGTEFSVKVSLLEIYNEELFDLLNPAPDVGERLQMFDDPRNKRGVIKGLLEEVTVHNKNEVYQILERGAAKRTTAA
TYMNAYSSRSHSVFSITIHMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVISAL
VERAPAYSIQES
```

Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as *S. cerevisiae*, human or mouse, because those genomes have already been thoroughly annotated. It is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

Name: Kinesin-like protein KIF11

Species: *Gallus gallus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archelosauria; Archosauria; Dinosauria; Saurischia; Theropoda;

Coelurosauria; Aves; Neognathae; Galloanserae; Galliformes;

Phasianidae; Phasianinae; Gallus.

[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.

A BLASTP search against the nr database gave a top hit result to a protein from Gallus Gallus (Red junglefowl Birds):

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query sequence.

Enter Query Sequence

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

Query subrange [?](#)

From

To

> Gallus protein (sequence taken from BLAST result)
FRQAKQIDYRISVCPILDEVI
MGYNCTVFAYGQTGTGKFTMEGERSPNEEYTWEEPLAGIIPRTLHQIFEKLT
ENGTEFSVKVSLLEIY

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 [?](#)

Organism ☐ exclude [Add organism](#)

Optional Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude ☐ 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 [?](#)

BLAST Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

The highest percent identity was for the species Gallus Gallus but is still under 100%

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download ▼ New Select columns ▼ Show 100 ▼								
<input type="checkbox"/> select all 100 sequences selected	GenPept	Graphics	Distance tree of results Multiple alignment New MSA Viewer					
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein EK904_009114 [Melospiza melodia maxima]	Melospiza melodia maxima	484	484	100%	7e-170	95.88%	415	KAF2975241.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Gallus gallus]	Gallus gallus	499	499	100%	7e-168	97.53%	1005	XP_040558358.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Meleagris gallopavo]	Meleagris gallopavo	499	499	100%	2e-167	97.53%	1043	XP_019473221.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 [Gallus gallus]	Gallus gallus	499	499	100%	3e-167	97.53%	1067	NP_001026401.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Parus major]	Parus major	493	493	100%	4e-167	95.88%	869	XP_015488042.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 isoform X3 [Anser cygnoides domesticus]	Anser cygnoides domesticus	497	497	100%	6e-167	96.30%	996	XP_013044056.1
<input checked="" type="checkbox"/> Kinesin-like protein KIF11 [Anas platyrhynchos]	Anas platyrhynchos	497	497	100%	2e-166	96.71%	1040	EOB05693.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 [Actinotomus carolinensis]	Actinotomus carolinensis	496	496	100%	2e-166	96.71%	1039	XP_028943088.1
<input checked="" type="checkbox"/> Kinesin-like KIF11 [Actinotomus carolinensis]	Actinotomus carolinensis	496	496	100%	2e-166	96.71%	1040	KFZ61578.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 [Aythya fuligula]	Aythya fuligula	497	497	100%	3e-166	96.71%	1065	XP_032047559.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 isoform X2 [Anser cygnoides domesticus]	Anser cygnoides domesticus	496	496	100%	4e-166	96.30%	1047	XP_013043975.1
<input checked="" type="checkbox"/> Kinesin-like KIF11 [Actenodytes forsteri]	Actenodytes forsteri	496	496	100%	4e-166	96.71%	1040	KFM06497.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 [Corvus brachyrhynchos]	Corvus brachyrhynchos	494	494	100%	4e-166	96.30%	995	XP_017581548.1
<input checked="" type="checkbox"/> Kinesin-like KIF11 [Fulmarus glacialis]	Fulmarus glacialis	496	496	100%	5e-166	96.30%	1039	KFW02524.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 [Actenodytes forsteri]	Actenodytes forsteri	496	496	100%	5e-166	96.71%	1064	XP_009276301.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 [Gavia stellata]	Gavia stellata	496	496	100%	5e-166	96.30%	1039	XP_009816548.1
<input checked="" type="checkbox"/> PREDICTED: kinesin-like protein KIF11 isoform X1 [Anser cygnoides domesticus]	Anser cygnoides domesticus	496	496	100%	5e-166	96.30%	1066	XP_013043905.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 [Anas platyrhynchos]	Anas platyrhynchos	496	496	100%	6e-166	96.71%	1065	XP_027316461.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Numida meleagris]	Numida meleagris	494	494	100%	6e-166	96.30%	1005	XP_021255302.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Camachyochus papouas]	Camachyochus papouas	494	494	100%	6e-166	96.30%	1007	XP_030807536.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X1 [Meleagris gallopavo]	Meleagris gallopavo	498	498	100%	7e-166	97.53%	1130	XP_010712861.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 isoform X2 [Zonotrichia albicollis]	Zonotrichia albicollis	494	494	100%	7e-166	96.30%	994	XP_014131087.1
<input checked="" type="checkbox"/> kinesin-like protein KIF11 [Oxyura jamaicensis]	Oxyura jamaicensis	496	496	100%	7e-166	96.30%	1066	XP_035185414.1

[Q5] Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.

Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in the output alignment and in the subsequent answers below). The goal in this step is to create an interesting an alignment for building a phylogenetic tree that illustrates species divergence.

>Human_KIF11 | CX788535.1

```
VQENIQQKSKDIVNKMTHFSQKFCADSDGFSQELRNFNQEGTKLVEESVKHSDKLNGLNLEKISQETE
QRCESLNTRTVYFSEQWVSSLNEREQELHNLLEVVSQCCEASSSDITEKSDGRKAAHEKQHNIFLDQ
MTIDEDKLI AQNLELNETIKIGLTKLNCFLQDLKLDIPTGTTPQRKSYLYPSTLVRTEPREHLLDQ
LKRKQPELLMMLNCSENNKEETIPDVDVEEAVLGQYTEEPLSQEPSVDAGVDCSSIGGVPFFQHKKSH
GKDKE
```

>GallusGallus protein (taken from BLAST result)

```
FRAQAKQIDVYRSVVCPIILDEVIMGYNCTVFAYGQTGTGKFTMEGERSPNEEYTWEEEDPLAGIIPRTLHQIFEKLT
ENGTEFSVKVSLLEIYNEELFDLLNPAPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKNEVYQILERGAAKRTTAA
```

TYMNAYSSRSHSVFSITIHMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVISAL
VERAPAYSIQES

>RedGrouse_KIF11 | GW700457.1

QTFRFDYAFDETAPNEMVYRFTARPLVETIFERGMATCFAYGQTGSGKTHTMGGDFSGK-----
NQDCSKGIYALAARDVFLMLKKPNYKKLELQVYATFFEIYSGKVFDLLNRKT----KLRVLED--
GKQQVQVVGLQEREVKCEDVLKLEIGNSCRTSGQTSANAHSSRSHAVFQIILRRKGKL-----
-HGKFSLIDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQV

>ZebraFinch_KIF11 | FE719925.1

KSKSSESVRVVRCRPMNSKEQTASYEKVVNVVDVKLGQVSVKNPRGSSHELPKTFTFDVAV
YDWSNKQVELYDETFRPLVDSVLQGFNGTIFAYGQTGTGKTYTMEGVRGDPEK-----
RGVIPNSFDHIFTHISRSQNQQYLVRASYLEIYQEEIRDLL--SKDQSKRLELKERPDT-
GVFVKDLTTIVTKSVKEIEHIMNLGNQNRSVGATNMNEHSSRSHAFQITIECSELGLD

>Mallard_KIF11 | DR765078.1

RGEDEKGIPIVRVALRCRPLVPKETSEGCQXCLSFVPGEPOVVV-----
GSDKVFSYDFVFXPTVXQEEVFNTXVAPLVRGIFKGYNATVIAYGQTGSGKTYSMGGAYTASQEH--
-- DPSVGIIIPRVINAAISGEGSGGXDWGI

>SocietyFinch_KIF11 | DC282391.1

DGTEFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQVYQILER
GAAKRRTTAATYMNAYSSRSHSVFSITIHMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDKRAR
EAGNINQSLTLGRVITALVERAPHIPYRESKLTRILQDSLGGRTKTSIIATISPASINLEETLSTL
EYAHRAKNIMNKPEVNQKLTCCA

>WhiteSparrow_KIF11 | XP_005481265.1

KKEEKGNIQVVRCRPFNASSELKVSSYAVVDCDQARKEVSIIRTGGVTDKSSRKTYTFDMVFGAQAK
QIDVYRSVVCPIDEVIMGYNCTVFAYGQTGTGKTFTMEGERSPNEEYTWEEEDPLAGIIPRTLHQIF
EKLTENGTEFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQVY
QILERGAAKRRTTAATYMNAYSSRSHSVFSITIHMKETTVDGEELVKIGKLNLDLAGSENIGRSGAV
DKRAREAGNINQSLTLGRVITALVERAPHIPYRESKLTRILQDSLGGRTKTSIIATISPASVNL
TLSTLEYAHRAKNIMNKPEVNQKLTCKALIKETEEIERLKRDLAAAREKNGVYISAENYEALNGKL
TVQEEQITEYIDKISVMEEVVRVTELFVSKNELEQCKTDLQIKEKELEETQKDLQETKVQLAEEE
YVVSVLESTEQELHDTASQLLTTVEETTRDVSGLHAKLERKRAVDQHNAAVQNTFAGQMNASF
DSITENSLKQQQMLTYTNCIGDLLSTSSSTADMFASVVSASFACLKELVSTEVSHISEKITQH
ENLSLDCKAELLRLIEEHQTGLGRAVNS-
LTPMVEFVLGLNCQFQSNMCKYSAVADQMEDHKKEMDTFFADLSLTLKKIQEQTAGGFAQLQHNCDS
LKEEVEMMRLAHRKSAAELMSSLQSLDLFAQETQKSLTDVLTRNGSLKTTITAMQENVHLKTTDLV
SSTNSNHNFKAASLDNFSQELRSINAENKAMLEESNDHCQHLLTNLKNVAQHTNTWGEFTTAQMVNF
TNQHLLSFKDEKQQFQYLQKKNEENCDAIAEIAHDHIGSQKAAEEKVLNGLLDQIKVDQEILVEQKL

ALREQVQHGLTQVNGFLQEDLKVDVPTGTTTPQRKDYSYPVTLVRTEPRQLLLEQLRQKQPNLDAMLS
SVGKEMEDSAGQDLLEEGVLQEPSESLACDKYSMDTNVYCHTNGGIPFFQHKRSLKKGKENKSAAPL
E-NKMEDMTEELLQKSKHPLR

Multiple sequence alignment using MUSCLE at EBI:

```
Human_KIF11 -----VQENIQ
Mallard_KIF11 RGEDEKGIPVRVALRCRPLVPKETSEGCQXCLSFVPGEPPQVVVGSD-----KVFSYD
RedGrouse_KIF11 -----QTFRFD
ZebraFinch_KIF11 --KSKSSESVRVVVRCRPMNSKEQTASYEKVVNVDVKLGQVSVKNPRGSSHELPKTFTTFD
GallusGallus -----
SocietyFinch_KIF11 -----
WhiteSparrow_KIF11 -KKEEKGNIQVVVRCRPFNASSELKVSSYAVVDCDQARKEVSIRTGGVTDKSSRKTYTFD
```

```
Human_KIF11 QKSKDIVNKMTFHSQKFCADSD----GFSQELRNFNQEGT--KLVEESVKHSDKL-----
Mallard_KIF11 FVFXPTVXQEEVFNTXVAPLVRGIFKGYNATVIAYGQTGSGKTYSMGGAYTASQEH-----
RedGrouse_KIF11 YAFDETAPNEMVYRFTARPLVETIFERGMATCFAYGQTGSGKTHTMGGDFSGKNQ-----
ZebraFinch_KIF11 AVYDWNKQVELYDETFRPLVDSVLQGFNGTIFAYGQTGTGKTYTMEGVRGDPEK-----
GallusGallus --FRAQAKQIDVYRSVVCPIDEVIMGYNCTVFAYGQTGTGKTFTEGERSPNEEYTWEE
SocietyFinch_KIF11 -----DGT-----
WhiteSparrow_KIF11 MVFGAQAKQIDVYRSVVCPIDEVIMGYNCTVFAYGQTGTGKTFTEGERSPNEEYTWEE
*:
```

```
Human_KIF11 ---NGNLEKISQETEQRCESLNTRTVYFSEQWVSSLNEREQELHNLLEVVSQCCEASSSD
Mallard_KIF11 DPSVGIIIP-----
RedGrouse_KIF11 DCSKGIYALAARDVFLMLKKPNYKKLELQVY-ATFFEIYSGKVFDDL-----N
ZebraFinch_KIF11 ---RGVIPNSFDHIFTHISRSQNQ--QYLVR-ASYLEIYQEEIRDLL-----SKD
GallusGallus DPLAGIIPRTLHQIFEKLTENGT---EFSVK-VSLLEIYNEELFDLL-----NPAPD
SocietyFinch_KIF11 -----EFSVK-VSLLEIYNEELFDLL-----NPTPD
WhiteSparrow_KIF11 DPLAGIIPRTLHQIFEKLTENGT---EFSVK-VSLLEIYNEELFDLL-----NPTPD
```

```
Human_KIF11 ITEKSDGRKAAHEKQHNIF--LDQMTIDEDKLIAQNLELNETIK-IGLTKLNCFLQDLK
Mallard_KIF11 -----RVINAAISGEGSGG-----
RedGrouse_KIF11 RKTCLRVLLEDG--KQQVQVVLQEREVKCVEDVLKLEIGNSCRTSGQTSANAH-----
ZebraFinch_KIF11 QSKRLELKERP--DTGVFVKDLTTIVTKSVKEIEHIMNLGNQNRSVGATNMNEH-----
GallusGallus VGERLQMFDDPRNKRGVIIKGLEEVTVHNKNEVYQILERGAAKRTTAATYMNAY-----
SocietyFinch_KIF11 VGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQVYQILERGAAKRTTAATYMNAY-----
WhiteSparrow_KIF11 VGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQVYQILERGAAKRTTAATYMNAY-----
. :: .
```

```
Human_KIF11 LDIPTGTTTPQRKSYLYPSTLVRTEPREHLLDQLKRKQPELLMMLNCSENNKEETIPDVDV
Mallard_KIF11 -----XDWGI-----
RedGrouse_KIF11 -----SSRSHAVFQIIL---RRKGKLH-----GKFSLIDLAGNERGADTSSADR
ZebraFinch_KIF11 -----SSRSHAIFQITI---ECSELGLD-----
GallusGallus -----SSRSHSVFSITI---HMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDK
SocietyFinch_KIF11 -----SSRSHSVFSITI---HMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDK
WhiteSparrow_KIF11 -----SSRSHSVFSITI---HMKETTVDGEELVKIGKLNLDLAGSENIGRSGAVDK
. :
```

```
Human_KIF11 EEAVLGQYTEEPLSQEPSVDAGVDCSSIGGVFFQHKKSHGKDKE-----
Mallard_KIF11 -----
RedGrouse_KIF11 QTRLEGAEINKSLLALKECIRAL----GRNKPHTPFRASKLTQV-----
ZebraFinch_KIF11 -----
GallusGallus RAREAGN-INQSLTLGRVISAL----VERAPAYSIQES-----
SocietyFinch_KIF11 RAREAGN-INQSLTLGRVITAL----VERAPHIPYRESKLTRILQDSLGGRTKTSIIAT
WhiteSparrow_KIF11 RAREAGN-INQSLTLGRVITAL----VERAPHIPYRESKLTRILQDSLGGRTKTSIIAT
```


Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	ISPASINLEETLSTLEYAHRAKNIMNKPEVNQKLTKKA-----
WhiteSparrow_KIF11	ISPASVNLEETLSTLEYAHRAKNIMNKPEVNQKLTKKALIKKEYTEEIERLKRDLAAREK

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	NGVYISAENYEALNGKLTVQEEQITEYIDKISVMEEEVKRVTELFVRVSKNELEQCKTDLQ

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	IKEKELEETQKDLQETKVQLAEEYVVSVESTEQELHDTASQLLTTVEETTRDVSGLHA

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	KLERKRAVDQHNAAVQNTFAGQMNASFSKIQDSITENSLKQQQMLTTYTNCIGDLLSTSS

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	STADMFAVSVSASFACLKELVSTEVSHISEKITQHENLSLDCKAELLRLIEEHQTGLGRA

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	VNSLTPMVEFVLGLNCQFQSNMKKYSAVADQMEDHKKEMDTFFADLSLTLKKIQEQTAGG

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	FAQLQHNCDSLKEEVEMMRLAHRKSAAELMSSLQSQDLDLFAQETQKSLTDVLTRNGSLKT

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	TITAMQENVHLKTTDLVSSTNSNHNKFAASLDNFSQELRSINAENKAMLEESNDHCQHLL

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	TNLKNVAQHTNTWGEFTTAQMVNFTNQHLLSFKDEKQQFQYLQKKNEENCDAQIAEADIH

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	IGSQKAAEEKVLNGLLDQIKVDQEILVEQKLALREQVQHGLTQVNGFLQEDLKVDVPTGT

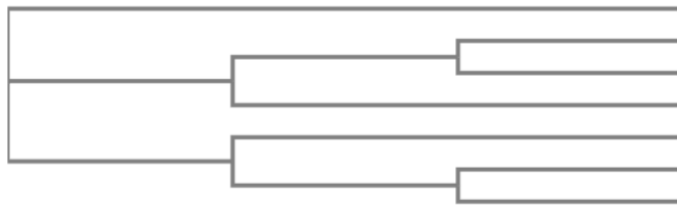
Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	TPQRKDYSYPVTLVRTEPRQLLLEQLRQKQPNLDAMLSSVGKEMEDSAGQDLLEEGVLQE

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	PSESLACDKYSMDTNVYCHTNGGIPFFQHKRSLKKGKENKSAAPLENKMEDMTEELLQKS

Human_KIF11	-----
Mallard_KIF11	-----
RedGrouse_KIF11	-----
ZebraFinch_KIF11	-----
GallusGallus	-----
SocietyFinch_KIF11	-----
WhiteSparrow_KIF11	KHPLR

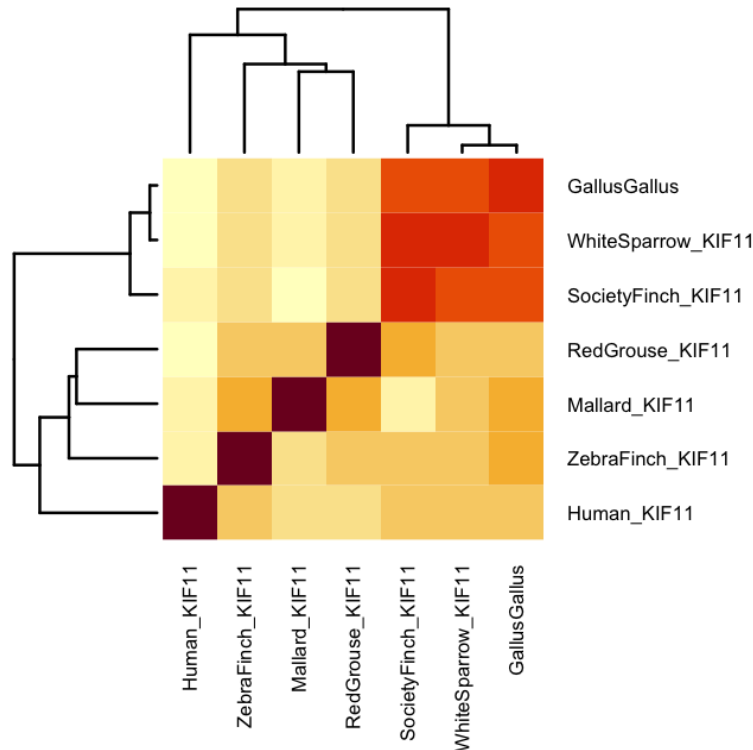
[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

Branch length: ☒ Cladogram ☐ Real



Human_KIF11 0.52099
Mallard_KIF11 0.33278
RedGrouse_KIF11 0.28085
ZebraFinch_KIF11 0.2905
GallusGallus 0.00388
SocietyFinch_KIF11 0.02702
WhiteSparrow_KIF11 -0.01809

[Q7] Generate a sequence identity based heatmap of your aligned sequences using R. If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and “Save as” FASTA format for example). Read this FASTA format alignment into R with the help of functions in the Bio3D package. Calculate a sequence identity matrix (again using a function within the Bio3D package). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.



[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

List the top 3 unique hits (i.e. not hits representing different chains from the same structure) along with their E-value and sequence identity to your query. Please also add annotation details of these structures. For example, include the annotation terms PDB identifier

(structure Id), Method used to solve the structure (experimental Technique), resolution (resolution), and source organism (source).

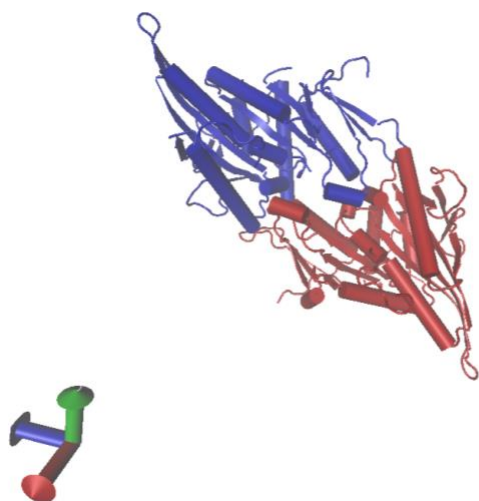
HINT: You can use a single sequence from your alignment or generate a consensus sequence from your alignment using the Bio3D function `consensus()`. The Bio3D functions `blast.pdb()`, `plot.blast()` and `pdb.annotate()` are likely to be of most relevance for completing this task.

Note that the results of `blast.pdb()` contain the hits PDB identifier (or `pdb.id`) as well as Evalue and identity. The results of `pdb.annotate()` contain the other annotation terms noted above.

Note that if your consensus sequence has lots of gap positions then it will be better to use an original sequence from the alignment for your search of the PDB. In this case you could chose the sequence with the highest identity to all others in your alignment by calculating the row-wise maximum from your sequence identity matrix.

ID	Technique	Resolution	Source	E-value	Identity
4ZCA_A	X-ray Diffraction	2.3	Homo sapiens	0.00e+00	90.529
1IA0_K	Electron microscope	15.0	Mus musculus	1.67e-66	37.500
2WBE_C	Electron microscope	9.4	Drosophila melanogaster	3.68e-140	57.713

[Q9] Generate a molecular figure of one of your identified PDB structures using VMD. You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black). Based on sequence similarity. How likely is this structure to be similar to your “novel” protein?



This structure is very likely to be similar in structure to Kinesin-like protein according to the high similarity (90.529%) and the very low e value of 0.00e+00.

[Q10] Perform a “Target” search of ChEMBL (<https://www.ebi.ac.uk/chembl/>) with your novel sequence. Are there any Target Associated Assays and ligand efficiency data reported that may be useful starting points for exploring potential inhibition of your novel protein? ChEMBL details 8 Binding Assay (CHEMBL4581) and 163 Functional Assays. There were multiple points on the ligand efficiency graph shown below.

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL4581/

