

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.

KIF11 (kinesin family member 11)

Accession number: NP_004514

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 Fungi ESTs

Database: Expressed sequence tag (est)

Organism: Fungi (taxid:4751)

Translated BLAST: tblastn

blastn blastp blastx **tblastn** tblastx

TBLASTN search translated nucleotide databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

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

NP_004514.2

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

Database

Organism ☐ exclude

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

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

Limit to ☐ Sequences from type material

Entrez Query [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

BLAST results will be displayed in a new format by default

You can always switch back to the Traditional Results page.

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.

Chosen match: Accession DV760029.1, 849 base pairs cDNA clone from *Phanerochaete chrysosporium*.

<input checked="" type="checkbox"/> select all 100 sequences selected		GenBank					Graphics
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	EST786001 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBBJ47 5' end, mRNA sequ	248	248	22%	2e-73	52.94%	CO015619.1
<input checked="" type="checkbox"/>	PchrSEQ9845 Phanerochaete chrysosporium pBluescript (EcoRI-XhoI) Phanerochaete chrysosporium cDNA clone PchrSEQ9845, mRNA sequence	246	246	23%	2e-72	53.78%	DV760029.1
<input checked="" type="checkbox"/>	G873P58FD23.T0 C. posadasii Silveira, 24HR SPHERULE NORMALIZED Coccidioides posadasii cDNA, mRNA sequence	237	237	23%	1e-69	51.01%	GH414355.1
<input checked="" type="checkbox"/>	CCHU13341.b1 CCHU Glomus intraradices germinated spores on water agar 3 days Rhizophagus intraradices cDNA clone CCHU13341 5', mRNA	228	228	24%	3e-66	49.42%	GW112992.1
<input checked="" type="checkbox"/>	CBWQ15429.b1 CBWO Melampsora larici-populina germinated and ungerminated urediniospores (H) Melampsora larici-populina cDNA clone CBW	215	215	24%	2e-61	45.66%	GR790348.1
<input checked="" type="checkbox"/>	EST786841 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBBO56 3' end, mRNA seq	205	205	28%	1e-57	40.89%	CO016459.1
<input checked="" type="checkbox"/>	HMA050014.T3_E07_055 GLEST05 Ganoderma lucidum cDNA 3', mRNA sequence	198	198	26%	2e-55	42.60%	HO746710.1
<input checked="" type="checkbox"/>	TlanSEQ14201 Thermomyces lanuginosa pBluescript (EcoRI-XhoI) Thermomyces lanuginosus cDNA clone TlanSEQ14201, mRNA sequence	197	197	24%	3e-55	44.06%	EB066279.1
<input checked="" type="checkbox"/>	FM240253 Library 8 Zymoseptoria tritici cDNA clone 25206, mRNA sequence	192	192	14%	1e-53	62.09%	FM240253.1
<input checked="" type="checkbox"/>	CeleSEQ13238 Cunninghamella elegans pBluescript (EcoRI-XhoI) Cunninghamella elegans cDNA clone CeleSEQ13238, mRNA sequence	187	187	23%	6e-52	44.84%	DY893802.1
<input checked="" type="checkbox"/>	CCAHK46TR C.neoformans strain JEC21 Cryptococcus neoformans var. neoformans cDNA clone CCAHK46, mRNA sequence	184	184	24%	4e-50	40.15%	CF694261.1
<input checked="" type="checkbox"/>	CCADB58TR C.neoformans strain JEC21 Cryptococcus neoformans var. neoformans cDNA clone CCADB58, mRNA sequence	184	184	24%	4e-50	40.15%	CF711531.1
<input checked="" type="checkbox"/>	G873P58RD23.T0 C. posadasii Silveira, 24HR SPHERULE NORMALIZED Coccidioides posadasii cDNA, mRNA sequence	182	182	24%	7e-50	39.34%	GH414356.1
<input checked="" type="checkbox"/>	EST793475 Coccidioides posadasii spherule cDNA library, 0.4 to 2.3 kb Coccidioides posadasii cDNA clone CIEAN39 3' end, mRNA sequence	182	182	25%	1e-49	39.86%	CO005140.1
<input checked="" type="checkbox"/>	EST694586 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBAX21 3' end, mRNA seq	182	182	24%	2e-49	39.85%	CF817204.1
<input checked="" type="checkbox"/>	ApulSEQ15914 Aureobasidium pullulans pBluescript (EcoRI-XhoI) Aureobasidium pullulans cDNA clone ApulSEQ15914, mRNA sequence	177	177	20%	1e-48	46.22%	DY862005.1
<input checked="" type="checkbox"/>	FP399961 Tuber melanosporum mycelium Tuber melanosporum cDNA clone SY0AAA27YG16 5', mRNA sequence	177	177	19%	2e-48	47.37%	FP399961.1

PchrSEQ9845 Phanerochaete chrysosporium pBluescript (EcoRI-XhoI) Phanerochaete chrysosporium cDNA clone PchrSEQ9845, mRNA sequence

Sequence ID: DV760029.1 Length: 849 Number of Matches: 1

Range 1: 8 to 760 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
246 bits(628)	2e-72	Compositional matrix adjust.	143/251(57%)	181/251(72%)	7/251(2%)	+2
Query 256	VKIGKLNLDLAGSENIGRSGAVDKRAREAGNIQSLLTLGRVITALVERTPHVPYRESK					315
Sbjct 8	+++GKLNLDLAGSENIGRSGA DKRAREAG INQSLLTLGRVI ALV+R+ HVPYRESK					
	LRVGKLNLDLAGSENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESK					187
Query 316	LTRLQDSLGGRTRTSIIATISPA ^{slnleettlstle} YAHRAKNILNKPEVNO ^{++T+} AL+					375
Sbjct 188	LTRLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNO ^{++T+} AL+					367
Query 376	KEYTEEIERLKRDAAAREKNGVYISEENFRVM ^{SGKL} TVQEEQIVELIEKIGAVEEELNR					435
Sbjct 368	KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDT ^{EMQ} EAKKQVEIVESQLRN					547
Query 436	VTE-----LFMDNKNELDQCKSDLNK ^{TQ} LETTQKHLQETKLQLVKEEYITSALEST					488
Sbjct 548	VREEFEQSMALLMRD ^{GEL} KETKERLQKKETELKAT ^{EGK} LEVVGAL ^{EEVVVR} QAYQEN					727
Query 489	E EKLHDAASKL 499					
Sbjct 728	E L A+ L 760					
	ETVLDGVATGL					

PchrSEQ9845 Phanerochaete chrysosporium pBluescript (EcoRI-XhoI) Phanerochaete chrysosporium cDNA clone PchrSEQ9845, mRNA sequence

Sequence ID: DV760029.1 Length: 849 Number of Matches: 1

Range 1: 8 to 760GenBankGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
246 bits(628)	2e-72	Compositional matrix adjust.	143/251(57%)	181/251(72%)	7/251(2%)	+2
Query	256	VKIGKLNLVDLAGSENIGRSGAVDKRAREAGNINQSLTLGRVITALVERTPHVPYRESK	315			
		+++GKLNLVDLAGSENIGRSGA DKRAREAG INQSLTLGRVI ALV+R+ HVPYRESK				
Sbjct	8	LRVGKLNLVDLAGSENIGRSGAQDKRAREAGMINQSLTLGRVINALVDRSSHVPYRESK	187			
Query	316	LTRILQDSLGGRTRTSIIATISPAslnleetlstleYAHRAKNILNKPEVNQKLTKKALI	375			
		LTR+LQDSLGGRT+T IIATISPA N+EETLSTL+YA RAK+I NKPEVNQ++T+ AL+				
Sbjct	188	LTRLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL	367			
Query	376	KEYTEEIERLKRDLAAAREKNGVYI SEENFRVMSGKLTVQEEQIVELIEKIGAVEEEELNR	435			
		KEY EIERLK D+ AAREKNG++ SEE ++ M+ + +++ ++ E +++ VE +L				
Sbjct	368	KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDTMQEAKKQVEIVESQLRN	547			
Query	436	VTE-----LFMDNKNELDQCKSDLQNKQTQELETTQKHLQETKLQLVKEEYITSALEST	488			
		V E L M EL + K LQ K EL+ T+ L+ K L +E + A +				
Sbjct	548	VREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALEEEVVVRQAYQEN	727			
Query	489	EEKLHDAASKL 499				
		E L A+ L				
Sbjct	728	ETVLDGVATGL 760				

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.

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.

[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.

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.

```
>8-760_1 PchrSEQ9845 Phanerochaete chrysosporium pBluescript (EcoRI-XhoI) Phanerochaete  
chrysosporium cDNA clone PchrSEQ9845, mRNA sequence  
LRVGKLNLDLAGSENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESK  
LTRLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL  
KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDEMQEAKKQVEIVESQLRN  
VREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALEEEVVVRQAYQEN  
ETVLDGVATGL
```

Name: *Phanerochaete chrysosporium*, hypothetical protein

Species: *Phanerochaete chrysosporium*

*Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;
Agaricomycetes; Polyporales; Phanerochaetaceae; Phanerochaete*

[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.

Blastp search against nr database:

blastn **blastp** blastx tblastn tblastx

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

Enter Query Sequence

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

Query subrange [From](#) [To](#)

LRVGKLNLDLAGSENI...
LRLQLQSLGGRTKTCIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNRQMRNALL
KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKQTEMQEAQKQVEIVSQRLN
VREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALEEVVVRQAYQEN
ETVLDGVATGL

Or, upload file [Choose File](#) no file selected

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

☐ Align two or more sequences

Choose Search Set

Database [Non-redundant protein sequences \(nr\)](#)

Organism [Optional](#) [exclude](#) [+](#)

[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](#)

BLAST [Search database nr using Blastp \(protein-protein BLAST\)](#)

☒ Show results in a new window

The top result is a hypothetical protein from *Phanerochaete camosa* with identity 94.42%, some results indicate that this is a kinesin-domain-containing protein:

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
✓	hypothetical protein PHACADRAFT_173717 [Phanerochaete carnosae HHB-10118-sp]	483	483	100%	5e-161	94.42%	XP_007395896.1
✓	hypothetical protein PHLGIDRAFT_22804 [Phlebiopsis gigantea 11061_1 CR5-6]	467	467	100%	4e-155	91.24%	KIP09572.1
✓	kinesin motor protein cin8 [Steccherinum ochraceum]	447	447	100%	1e-145	87.65%	TCD71883.1
✓	hypothetical protein EUX98_g250 [Antrodiaella citrinella]	446	446	100%	3e-145	86.45%	THH33841.1
✓	hypothetical protein EVJ58_g2903 [Fomitopsis rosea]	442	442	100%	9e-145	86.06%	TFY64019.1
✓	hypothetical protein SERLADRAFT_451502 [Serpula lacrymans var. lacrymans S7.9]	441	441	100%	1e-144	85.66%	XP_007321258.1
✓	hypothetical protein PHLZEN_2v10106 [Phlebia centrifuga]	444	444	100%	3e-144	85.66%	PSR74150.1
✓	hypothetical protein PAXINDRAFT_112407 [Paxillus involutus ATCC 200175]	442	442	100%	3e-144	85.26%	KIJ17421.1
✓	hypothetical protein SERLA73DRAFT_170393 [Serpula lacrymans var. lacrymans S7.3]	440	440	100%	3e-144	85.66%	EGN95950.1
✓	hypothetical protein HYDPIDRAFT_145382 [Hydnomerulius pinastris MD-312]	441	441	100%	1e-143	86.06%	KIJ69381.1
✓	kinesin-domain-containing protein [Daedalea quercina L-15889]	440	440	100%	1e-143	86.06%	KZT67731.1
✓	hypothetical protein PAXRUDRAFT_825572 [Paxillus rubicundulus Ve08.2h10]	440	440	100%	2e-143	85.26%	KIK96808.1
✓	hypothetical protein PISMIDRAFT_670188 [Pisolithus microcarpus 441]	438	438	100%	8e-143	85.26%	KIK31191.1
✓	hypothetical protein FOMPIDRAFT_1109816 [Fomitopsis pinicola FP-58527 SS1]	436	436	100%	2e-142	85.26%	EPT06162.1
✓	hypothetical protein M404DRAFT_951987 [Pisolithus tinctorius Marx 270]	436	436	100%	5e-142	85.66%	KIO13377.1
✓	predicted protein [Postia placenta Mad-698-R]	434	434	100%	5e-142	85.66%	EED84161.1
✓	hypothetical protein PLICRDRAFT_161615 [Plicaturopsis crispa FD-325 SS-3]	436	436	100%	7e-142	84.46%	KII88479.1
✓	predicted protein [Fibroporia radiculosa]	438	438	100%	1e-141	85.26%	XP_012181653.1
✓	hypothetical protein SCLCIDRAFT_113134 [Scleroderma citrinum Foug A]	434	434	100%	2e-141	84.06%	KIM65286.1
✓	kinesin-domain-containing protein [Rhizopogon vinicolor AM-OR11-026]	434	434	100%	2e-141	84.46%	OAX44667.1

[Download](#) [GenPept](#) [Graphics](#)

hypothetical protein PHACADRAFT_173717 [Phanerochaete carnosae HHB-10118-sp]

Sequence ID: [XP_007395896.1](#) Length: 1058 Number of Matches: 1

[See 1 more title\(s\)](#)

Range 1: 320 to 570 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
483 bits(1243)	5e-161	Compositional matrix adjust.	237/251(94%)	245/251(97%)	0/251(0%)
Query 1	LRVGKLNLDLAGSENIGRSGAQDKRAREAGMINQSLTLGRVINALVDRSSHVPYRESK				60
Sbjct 320	LRVGKLNLDLAGSENIGRSGAQDKRAREAGMINQSLTLGRINALVDRS+HVPYRESK				379
Query 61	LTRLLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL				120
Sbjct 380	LTRLLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL				439
Query 121	KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDTMQEAKKQVEIVESQLRN				180
Sbjct 440	KEYVAEIERLK+DVLAAAREKNGIFFSEERWEMTAEQEL+DTMQEA+KQVEIVESQLRN				499
Query 181	VREEFEQSMALLMRRDGELKETKERLQKETELKATEGKLEVVKGALEEEVVVRQAYQEN				240
Sbjct 500	VREEFEQSMALLMRRDGELKETKE+LQK+ET+LKATEGKL VKGALEEEVVVRQAY+EN				559
Query 241	ETVLDGVATGL	251			
Sbjct 560	E LDGVATGL				
	EALDGVATGL	570			

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.

Sequence selected for comparing (re-labeled):

```
>Cricetulus_griseus RLQ59498.1 KIF11 [Cricetulus griseus]
MNGKLTVQEEQIVESAEKIGALEEEELSKVTGLFVDSKNELDQCKSDLQTKTQELETTQKHLQETKLQLVK
EEYVSSALERTEEKLHDAASKLLSTVKETTRDVSGLSKLDKRAIDENAEAQDIFGNLMSSSVSALDT
VTTTALGTLMSIPQNV SARVTQISDMILKEQSLAAQSKTVLRGLINGLETDLLGSLKTI LAPGVVSILNI
NRQLQHIFRTSLTMAEKELS QLVSSWTERFCALEKKCENIQKSLNKIQENTEH L FSEQWASCLSKSKEEL
QNLMEFINECCEATGSEITAKQSVHAAAIENQHTSFIVQITSDDKELKAGNLKLD ETIKTGLTKLNCFLQ
EDLKLDIPTGMTPERKNYLYPSTLVRTQPREQLLDQLQKKQHNQLNCSENDKEAFQHKKPHGKDKENRGI
NPVEKSKVEETSEYSVTKSRLPLRASNL
```

```
>Leptonychotes_weddellii XP_006733941.1 kinesin-like protein KIF11
[Leptonychotes weddellii]
MSGRPFNLAERKANAHSVVECDHVRKEVSVRTGGLADKSSRKYTFDMVFGASTKQIDVYRSVVC PILDE
VIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEE DPLAGIIPRTLHQIFEKLT DNGTEFSVKVSLLE
IYNEELFDLLNPSSDV SERLQMFDDPRNKRGVIIKGLEEITVHNKDEVYQILEKGAAKR TTAATLMNAYS
SRSHSVFSVTIHMKET TIDGEELVKIGKLNLDLAGSENIGRSGAVDKRAREAGNINQSLLTGRVITAL
VERTPHVPYRESKLTRILQDSLGGRT RTSIIATISPASLNLEETLSTLEYAHRKNILNKPEVNQKLTKR
ALIKEYTEEIERLKRDLAAAREKNGVY ISEENFRAMSGKLTVQEEQIVELIEKIGAIEEELNRVTELFMD
NKNELDQCKSDLQNK TQELETTQKHLQETKLQLV EEEYITSALESTEEKLHDAASRLTTVEETTKDVSG
LH SKLDRKKAIDQHNAEAQDIFGKNLNSLFSNMEELIKDSSSKQKAMLETHKTLFGNLLSSSVSALDTIT
TTALGSLTSVPENVSIRVSQISNRILKEQSLAAESKTVLQTLIVSKSLKVFFKCYFAKPSKTM TSKDIIS
KTASHSTQFCADCDGLSQELRCCNQEGTKLVEESVKHCDKLSSNLEVVSQETEQRCEALDASTLAFSEQW
VSCLNKREEELQNLLEVINQGCKAASSEITEKLN GHKAAANENQHNTFFGQIT TDEEKLMAQSLELNETIK
MGLTKLNCFLQ QDLKLDIPTGTT PQRKNYLYPSTLV RTEPREQLLDQLKRKQPELLMMLDCSENNKEEIS
QDLDEEKSVLGH SVEELPGQEPSIDASVDCSSSGGIPPFQHKKSHGKDKENRGINSVERSKVEETTEHSV
TRSRLPLRAQINL
```

```
>Symbiodinium_microadriaticum OLP75816.1 Kinesin-like protein KIF11
[Symbiodinium microadriaticum]
MHEPDGCPDNRGTACRRQARSRPATPTGASNAEIPVTNVRVVC RIRPMNDREKKADPQCIEYAGATPAAT
ASTERREVAVVRSTFHFDDVLTSFSSQEDVFRATLQPLVGQVHVARAMLV LAVVSLPGWRANWTQRKAHP
EFTESVTVSYLEIYNEELS DLLATAERHPKLDLKDIGSGRGVCCQGLSEALPHKDASKAVHGETLTPLE
RPRSLCAMPDGASPAPTVELRLLSTAAAPSPEAEHHGFLEVPSARRGHRRSSLARLTELEHQ LHRDKGHL
AEKKLAKALALVLYTLARISANLVSIDYENHIRIVNALVRPLVTGLGVGMFLIREERFSTKSRLLIWSRR
TVFVLAIFVFD ESMSTLRAALPIFSVLPWEITCNVMQVHMTNLR LQALGWRLTAKFCHIMAAIGCSFLLG
```


GLVILTSRISISEFGKEGNLTVERVGAPASFLFFSFFCCWTTLFTQCGALCQAGSKAMVAGRADLDDL
KLTAFLYLNALLVLLGPAFGWSWAQFVRRIMTWLTIDICVQIFTALALSGMMGPRNPMDAFRKLADL
SGYGFASKRITFPGHINERAAKCIVSFPGVYNKLWDKAVKAVRDDFVTEEDLCSLACVFLTDSASGLGQH
SMNPDPGKWCCHAIYGQIRASAYLRVVEVDPCEADSIESREKL RHEVADADAMNQVLLIKSEQNDLEWQ
KQYARALRTARDLGQENGGRAPWGCQWFEVWKKNVDRAVELKQELHVFYFQGRKGQGLSWEDLSSDAK
NRVRPESGLGASQTAEVAYLLKMGIPFKEHDVQDFLSFLSSESS

>Homo_sapiens NP_004514.2 kinesin-like protein KIF11 [Homo sapiens]
MASQPNSSAKKKEEKGNIQVVVRCRPFNLAERKASAHSIVECDPVRKEVSVRTGGLADKSSRKTYTFDM
VFGASTKQIDVYRSVVCPIDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEEDPLAGIIPRTL
HQIFEKLTDNNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEV
YQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTTIDGEELVKIGKLNLDLAGSENIGRSGAVDK
RAREAGNINQSLTLGRVITALVERTPHVPYRESKLTRILQDSLGGRTTSIIATISPASLNLEETLSTL
EYAHRAKNILNKPEVNQKLTKKALIKETTEEIERLKRDLAAAREKNGVYISEENFRVMSGKLTQVEEQIV
ELIEKIGAVEEELNRVTELFMDNKNELDQCKSDLQNKQTQELETTQKHLQETKLQLVKEEYITSALESTEE
KLHDAASKLLNTVEETTKDVSGLSKLDKKAQVQHNAAEQDIFGKNLNSLFNNMEELIKDGSSKQKAML
EVHKTFLGNLLSSSVSALDTITTVAGSLTSIPENVSTHVSQIFNMILKEQSLAAESKTVLQELINVLKT
DLLSSLEMILSPTVVSILKINSQKHFKTSLTVADKIEDQKKELDGFLSILCNNLHELQENTICSLSVES
QKQCGNLTEDLKTIKQTHSQELCKLMNLWTERFCALEEKECENIQKPLSSVQENIQQKSKDIVNKMFTFSQ
KFCADSDGFSQELRNFNQEGTKLVEESVKHSDKLNGLNLEKISQETEQRCESLNTRTVYFSEQWVSSLNER
EQELHNLLEVVSQCCASSSDITEKSDGRKAAHEKQHNIFLDQMTIDEDKLIAQNLELNETIKIGLTKLN
CFLEQDLKLDIPTGTTPQRKSYLYPSTLVRTEPREHLLDQLKRKQPELLMMLNCSENNKEETIPDVEDVEE
AVLGQYTEEPLSQEPSVDAGVDCSSIGGVPFFQHKKSHGKDKENRGINTLERSKVEETTEHLVTKSRLPL
RAQINL

>Phanerochaete_chrysosporium 8-760_1 PchrSEQ9845 Phanerochaete chrysosporium
pBluescript (EcoRI-XhoI) Phanerochaete chrysosporium cDNA clone PchrSEQ9845,
mRNA sequence
LRVGKLNLDLAGSENIGRSGAQDKRAREAGMINQSLTLGRVINALVDRSSHVPYRESKLTRLLQDSLGGRTKTC
IIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALLKEYVAEIERLKADVLAAREKNGIFFSEERWQE
MTAEQELKDTMQEAKKQVEIVESQLRNVREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALE
EEVVVRQAYQENETVLDGVATGL

>Nothobranchius_rachovii SBS17602.1 kinesin family member 11
[Nothobranchius rachovii]
MAGASVKVAVRVRPFNSRETEKESKCIQMSGNTTTIINPKQAKDNKSFNFDFSYSWHTSPEDVNYASQM
RVYKDIGEEMLLHAFEGYNVCIFAYGQTGAGKSYTMMGKQDVKDQGGIIPLLCEDLFTKINDNTNNSMSY
SVEVSYMEIYCERVRDLLNPKNKGNLVRHPLMGPYVEDLSKLAVTSYNDIQDLMDSGNKARTVAATNM
NETSSRSHAVFNIIFTQKRYDAETDNTSEKVSISLVDLAGSENIGRSGAVDKRAREAGNINQSLTLGR
VITALVEKRPHIPYRESKLTRILQDSLGGRTKTSIIATVSPSSSNMEETMSTLEYACRAKNIMNKPEVNQ
KLTKRTLIKEYTEEIERLKRDLIAVREKNGVYLSSENYESMMTQIMAHEEQIAEYVDKIIIMEEELRKVT
ELFEDSQTKLEQCTADLDEKHKRLEETSKDLQQTKEKLVEEEFICSEVTSVHESLYNTAGQLLSTADAST
RDVSGLHDKLDRKKKVESHNKQIQQSFAERMDGALSSMQRCVHRHGAKHNELLSNCSQAVDGLMMNEVA
LKGTVTTMESFVGGVRSVVAEGVARCREKLQQHEVCLQDKESLLQLLEEHQQDMEDVLAARTLMGLSAV
NELSDVLRSSVEKQRALADKVEESVGLFFNSLCRDLDLREEAVHRLSSLQDEQDKLEDRI SQAQERQRL
GMQQTIQCLQDQLNLLSMQSQQDYADLRASEALKAPALSLQESICSGCSSMKDQATTQADLHLTTSSNL
TSSLHQTQQHVEEMNGYCADLHSSISGLVDRDLRWNLGVRDHTRQTREHLSVMEKLSAEASLDQDVDT
CCTNQLQTAEKELSGHREEVKQALLAIQNQTSMDQTILEQQQAELEDHVEVNQQLVHKFLQSELQQDIPT
GTTPQQETSKSNESLATELSFDENLVFNESKRVPFFKQKKGKMTKIPTRSKVSENNDSLQTTPOKTRAPL
HCVN

>Hydra_vulgaris CDG70760.1 Hydra vulgaris Kinesin-like protein KIF11,
partial [Hydra vulgaris]

MGKPILKLDKNQNIQVAVRCRPRNSQEIKAGSADVVDINNQTKEITVRQDMNYMDKGNNKTF SFDKVF GP
KSKQIDIIYQSMVCPVIDEVLQGYNCTIFAYGQTGTGKTFTMEGDRLDSEEF SWQDDPLAGIIPRAMHQLF
ELLNIMEECAEFSVRVSFLEIYNEELFDLLGTS LDSQKLRLFEDTTKKGSVVIQGLEEVIVHSRNEVYHI
L ERGAARRQTAATLLNAQSSRSLSL FMVTIHKENNINGEEFLKTGKLNLDLAGSENIGRSGAVEKRAR
EAGTINQSLLT LGRVITALVENAPHV PYRESKLTRLLKDSLGGRTKTSIIATISPAACNLEETLSTLDYA
QRAKNITNKPEINQKLTKKALIK EYTEEIEKLKKDLFAAREKNGIFIAEDNYILMQNQLSSQKSSLREYV
EKIQFMEEMKKTEQLFASMQQKLEKTTQHLQWIKQERDETKVLVSKHVETEQQLFNQASELLDTVEETI
NHKNILHSSLDNRKNLHVVNLLSSQKFKENAHAKVSSLNESI INGRHSSFAYHQDIVTELNNFKSEKASS
YNQLSKIINISVEEIIYKCLEIARTKQNEIKSECTKKDLLFSHNVETCKNETLQAINNYHKEIFTNALKEF
HECILRMQECDERLVNSV NAMTEHFCETT KGFIEKQTVELRDLGQTVDEHICCCQKEAIEKVNSYVEKTVQ
NQDKMQKELKDGFIKSLQASILDVVEKNFSNLFQTTQQTQLQEDTQMICSELSCQTENESNFKNSISSKLK
DMNLTAISTVDGITT LAYQNLQYATENIDALKTCANSLKTKLDDLEKNETSFENVLIDNSNKYHDMTLKS
YKELVTLSDNLISENHKNINELQTLHNSLCDQCLSNINTQVLSCDSVIDGLTEKMKQNQADSMNHSLSLE
NGLHLLENDIIIEFVEEELKQDSPTGLTPSQKEYNYPRSLQRTQEHEQIINEYRQRHDIPPLPSLLEYESD
ENHDGSFLQQSEDKQKESEFLIPRSPDETGNVLSIDTASDKENRNALLNNYYSSASKIGNNRPRIPLRS
SNTLS

>Zonotrichia_albicollis XP_014131087.1 kinesin-like protein KIF11
isoform X2 [Zonotrichia albicollis]
MVFGAQAKQIDVYRSVVCPI LDEVIMGYNCTVFAYGQTGTGKTFTMEGERSPN EEEYTWEEDPLAGIIPRT
LHQIF EKL TENGTEFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQ
VYQIL ERGA AKRTTAATYMNAYSSRSHSVFSIT IHKETTVDGEELVKIGKLNLDLAGSENIGRSGAVD
KRAREAGNINQSLLT LGRVITALVERAPHI PYRESKLTRILQDSLGGRTKTSIIATISPA SVNLEETLST
LEYAHRAKNIMNKPEVNQKLTKKALIK EYTEEIERLKRDLAAAREKNGVYISAENYEALNGKLTVQEEQI
TEYIDKISVMEEEVKRVTE LFRVSKNELEQCKTDLQIKEKELEETQKDLQETKVQLAE EEEYVVSVLESTE
QELHDTASQLLT TVEETTRDVSGLHAKLERKRAVDQHNAAVQNTFAGQMNASF SKIQDSITENSLKQQQM
LTYTNCIGDLLSTSSSTADMFASVVSASFACLKELVSTEVSHISEKITQHENLSLDCKAELLRLIEEHQ
TGLGRAVNSLTPMVEFVLGLNCQFQSNMKKYSAVADQMEDHKKEMDTFFADLSLTLKKIQEQTAGGFAQL
QHNCDSLKEEVEMMRLAHRKSAAELMSSLSQSLDLFAQETQKSLTDVLTNRNGLKTTITAMQENVHLKTT
DLVSSSTNSNHNKFAASLDNFSQELRSINAENKAMLEESNDHCQHLLTNLKNVAQHTNTWGEFTTAQMVNF
TNQHLLSFKDEKQQFQYLQKKNEENCDAIAE IADHIGSQKAAEEKVLNGLLDQIKVDQEILVEQKLALR
EQVQHGLTQVNGFLQEDLKVDVPTGTT PQRKDYSPVTLVRTEPRQLLLEQLRQKQPNLDAMLSSVGKEM
EDSAGQDLLEEGVLQEPSESLACDKYSMDTNVYCHTNGGIPFFQHKRSLKKGKENKSAAPLENKMEDMTE
ELLQKSKHPLRLLN

Alignment using MUSCLE at EBI:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Symbiodinium_microadriaticum	MHEPDGCPDNRGTACRRQARSRPATPTGASNAEIPVTNVRVVCRIRPMNDREKKADPQCI
Nothobranchius_rachovii	-----MAGASVKVAVRVVPFNSRETEKESKCI
Hydra_vulgaris	-----MGKPILKLDKNQNIQVAVRCRPRNSQEIKAGSADV
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	-----
Cricetulus_griseus	-----
Leptonychotes_weddellii	-----MSGRPFNLAERKANAHSV
Homo_sapiens	-----MASQPNSSAKKKEEGKNIQVVVRCRPFNLAERKASAHSI
Symbiodinium_microadriaticum	-EYAGATPAATASTE--RREVAVVRSTFHFDDVLTSFSSQED-----VFRATLQPLVG
Nothobranchius_rachovii	IQMSGNTTTTII-----NPKQAKDNKSFNFDFSYSWHTSPEDVNYASQMRVYKDIGEEMLL

Hydra_vulgaris	VDINNQTKEITVRQDMNYMDKGNKTFSTFDKVFGPSKQID-----IYQSMVCPVID
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	-----MVFGAQAKQID-----VYRSVVCPILD
Cricetulus_griseus	-----
Leptonychotes_weddellii	VECDHVRKEVSVRTG-GLADKSSRKTYTFDMVFGASTKQID-----VYRSVVCPILD
Homo_sapiens	VECDPVRKEVSVRTG-GLADKSSRKTYTFDMVFGASTKQID-----VYRSVVCPILD
Symbiodinium_microadriaticum	QVHVARAMLVLA-----VVSLPGWR-----ANW-----TQRKAHPEFT-----
Nothobranchius_rachovii	HAFEGYNVCIFAYGQTGAGKSYTMMGKDQVKDQ---Q-----GIIPLLCEDLFTKINDNT
Hydra_vulgaris	EVLQGYNCTIFAYGQTGTGKTFTEGDRLDSEEFWSQDDPLAGIIPRMLHQIFELLNIME
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	EVIMGYNCTVTFAYGQTGTGKTFTEGERSPNEEYTWEEEDPLAGIIPRTLHQIFEKLTENG
Cricetulus_griseus	-----
Leptonychotes_weddellii	EVIMGYNCTIFAYGQTGTGKTFTEGERSPNEEYTWEEEDPLAGIIPRTLHQIFEKLTNG
Homo_sapiens	EVIMGYNCTIFAYGQTGTGKTFTEGERSPNEEYTWEEEDPLAGIIPRTLHQIFEKLTNG
Symbiodinium_microadriaticum	----ESTVTVSYLEIYNEELSDLLATAER-HPKLDL-KDIGSGRGVCCQGLSEALPH---
Nothobranchius_rachovii	NNSMSYSVEVSYMEIYCERVRDLLNPKNKGNLVR---EHPLMGYPVEDLSKLAVTSYN
Hydra_vulgaris	ECA-EFSVRVSFLEIYNEELFDLLGTSLD-SQKLRLFEDTTKKGSVVIQGLEEVIVHSRN
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	T---EFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEIVTHNKN
Cricetulus_griseus	-----
Leptonychotes_weddellii	T---EFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEIVTHNKD
Homo_sapiens	T---EFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEIVTHNKD
Symbiodinium_microadriaticum	-----KDASKAVHGETLTPLERPRSLCAMP-----DGASPA-PTVELRLLSTA
Nothobranchius_rachovii	DIQDLMDSGNKARTVAATNMNETSSRSHAVFNIIFTQKRYDAETDNTSEKVSISLVDLA
Hydra_vulgaris	EVYHILERGAARRQTAATLLNAQSSRSLSFMVTIHMKENNINGEFL-KTGKLNLDLA
Phanerochaete_chrysosporium	-----L-RVGKLNLDLA
Zonotrichia_albicollis	QVYQILERGAAKRTAATYMNAYSSRSHSVFSITIMKETTVDGEELV-KIGKLNLDLA
Cricetulus_griseus	-----
Leptonychotes_weddellii	EVYQILEKGAAKRTAATLMNAYSSRSHSVFSITIMKETTIDGEELV-KIGKLNLDLA
Homo_sapiens	EVYQILEKGAAKRTAATLMNAYSSRSHSVFSITIMKETTIDGEELV-KIGKLNLDLA
Symbiodinium_microadriaticum	AAPSPEAEHHGFLEVPSSARRGHRSSSLARLTELEHQLHRDKGHLA--EKKLAKAL----
Nothobranchius_rachovii	G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHIPYRESKLTRILQDSL
Hydra_vulgaris	G--SENIGRSGAVEKRAREAGTINQSLLTLGRVITALVENAPHVPYRESKLTRILLKDSL
Phanerochaete_chrysosporium	G--SENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESKLTRILLQDSL
Zonotrichia_albicollis	G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERAPHIPYRESKLTRILQDSL
Cricetulus_griseus	-----
Leptonychotes_weddellii	G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSL
Homo_sapiens	G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSL
Symbiodinium_microadriaticum	----ALVLYTLARISAN----LVSIDYENHIRIV-----NALVRPLVTGLGV-
Nothobranchius_rachovii	GRTKTSIIATVSPSSNMEETMSTLEYACRAKNIMNKPEVNQKLTKRILKEYTEEIERL
Hydra_vulgaris	GRTKTSIIATISPAACNLEETLSTLDYARAKNITNKPEINQKLTKKALIKKEYTEEIERL
Phanerochaete_chrysosporium	GRTKTSIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALLKEYVAEIERL
Zonotrichia_albicollis	GRTKTSIIATISPAVNLEETLSTLEYAHRAKNIMNKPEVNQKLTKKALIKKEYTEEIERL
Cricetulus_griseus	-----
Leptonychotes_weddellii	GRTRTSIIATISPASLNLEETLSTLEYAHRAKNINLNKPEVNQKLTKRILKEYTEEIERL
Homo_sapiens	GRTRTSIIATISPASLNLEETLSTLEYAHRAKNINLNKPEVNQKLTKKALIKKEYTEEIERL
Symbiodinium_microadriaticum	-----GMFLIREERFSTKSRLLIWSRRTV-FVLAIFVFDESM-----S
Nothobranchius_rachovii	KRDLIAVREKNGVYLSSENYESSMTQIMAHEEQIAEYVDKIIIMEEELRKVTELFEDSQT
Hydra_vulgaris	KKDLFAAREKNGIFIAEDNYILMQNLSSQKSSSLREYVEKIQFMEEEMKKTEQFLASMQQ
Phanerochaete_chrysosporium	KADVLAAREKNGIFFSEERWQEMTAEQELKDTEMQEAQKQVEIVESQLRNV-----RE
Zonotrichia_albicollis	KRDLAAREKNGVYISAENYEALNGKLTVQEEQITEYIDKISVMEEVEKRVTELFVRVSKN
Cricetulus_griseus	-----MNGKLTVQEEQIVESAEKIGALEEELSKVTGLFVDSKN
Leptonychotes_weddellii	KRDLAAREKNGVYISEENFRAMSGKLTVQEEQIVELIEKIGAEELNRVTELFMDNKN
Homo_sapiens	KRDLAAREKNGVYISEENFRVMSGKLTVQEEQIVELIEKIGAVEELNRVTELFMDNKN
	.. : .:..:
Symbiodinium_microadriaticum	TLRAALPIFSVLPWEITCNVMQVHMTNLR-----LQALGWRLTAKFCHIMAAIGC
Nothobranchius_rachovii	KLEQCTADLDEKHKRLEETSKDLQQTKELVEEEFICSEVTSVHESLYNTAGQLLSTADA
Hydra_vulgaris	KLETTQHLQWIKQERDETKVLVSKH-----VETEQQFLNQASELLDTVEE
Phanerochaete_chrysosporium	EFEQSMALLMRDGEKLETKERLQKKETE-----LKATEGKLE-----
Zonotrichia_albicollis	ELEQCKTDLQIKEKELEETQKDLQETKVQLAEEYVVSVLESTEQELHDTASQLTTVEE

Cricetulus_griseus	ELDQCKSDLQTKTQLETTQKHLQETKLQLVKEEYVSSALERTEEKLHDAASKLLSTVKE
Leptonychotes_weddellii	ELDQCKSDLQNKTKQLETTQKHLQETKLQLVEEYITSALESTEELHDAASRLLTVEE
Homo_sapiens	ELDQCKSDLQNKTKQLETTQKHLQETKLQLVKEEYITSALESTEELHDAASKLLNTVEE
	: : . *
Symbiodinium_microadriaticum	SFLLGGLVILTSRISISEFGKEGNTVERV-----GAPASFLLFFSF
Nothobranchius_rachovii	S--TRDVSGLHDKLDRKKKVESHNKQIQQSFAERMDGALSSMQRCVHRHGAKHNELLSNC
Hydra_vulgaris	T--INHKNILHSSLDNRKNLHVNNLSSQKFKENAHAKVSSLNESIINGRHSSFAYHQDI
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	T--TRDVSGLHAKLERKRAVDQHNAAVQNTFAGQMNASFSKIQDSITENSLKQQQMLTTY
Cricetulus_griseus	T--TRDVSGLHDKLDRKKRAIDEHNAEAQDI-----
Leptonychotes_weddellii	T--TKDVSGLHDKLDRKKRAIDQHNAEAQDIFGKNLNSLFSNMEELIKDSSSKQKAMLETH
Homo_sapiens	T--TKDVSGLHDKLDRKKAVDQHNAEAQDIFGKNLNSLFNNMEELIKDGSQKQKAMLEHV
Symbiodinium_microadriaticum	FCCWTTLFTQCGALCQAGSKAMVAGRADLDLDLKLTAFLYLNALLVLLGPPALGFW----
Nothobranchius_rachovii	SQAVDGLLMMNEVALKGTVTMTESFVGGV--RSVVAEGVARCREKLQQHESVCLQ----
Hydra_vulgaris	VTELNNFKSEKASSYNQLSKIINISVEEIKCLEIARTKQNEIKSECTKKDLLFSHNVET
Phanerochaete_chrysosporium	-----VVKGAL-----
Zonotrichia_albicollis	TNCIGDLLSTSSSTADMFASVVSASFACL--KELVSTEVSHISEKITQHENLSLD----
Cricetulus_griseus	---FGNLMSSSVSALDVTVTALGTLMSI---PQNV SARVTQISDMILKEQSLAAQ----
Leptonychotes_weddellii	KTLFGNLLSSSVSALDTITTTALGSLTSV---PENVSIRVSQISNRILKEQSLAAE----
Homo_sapiens	KTLFGNLLSSSVSALDTITTTALGSLTSI---PENVSTHVSQIFNMILKEQSLAAE----
Symbiodinium_microadriaticum	SWAQFVRRIEMTWTIDICVQIFTALALSGMMGPRNPM-----DAFRKLADL----
Nothobranchius_rachovii	DKESLLQLLEEHOQDMEDVLAARTLMGLSAVNELSDVL-----RSSVEKQRALADKV
Hydra_vulgaris	CKNETLQAINNYHKEIFTNALKEFHECILRMQECDERLVNSVNAMTEHFCETTKGFIIEKQ
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	CKAELLRLIEEHQTGLGRAVNS-LTPMVEFVLGLNCQF-----QSNMKKYSADVADQM
Cricetulus_griseus	SKTVLRGLINGLETDLGLSLKTIAPGVVSILNINRQL-----QHIFRTSLTM----
Leptonychotes_weddellii	SKTVLQTLI-----VSKSLKVFFK-----CYFAKPSKTM----
Homo_sapiens	SKTVLQELINVLTDLSSLEMLSPVVSILKINSQL-----KHIFKTSLTVDAKI
Symbiodinium_microadriaticum	-----
Nothobranchius_rachovii	EE---SVGLFFNSLCRDLDSLREEAVHRLSSLQDEQDKLEDRIQAQERQRLGMOQTIQC
Hydra_vulgaris	TVELRDLGQTVDEHICQKEAIEKVNVSVEKTVQNQDKMQKELDKGFIKSLQA-----SI
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	EDHKKEMDTFFADLSLTLLKKIQEQTAGGFAQLQHNCDSLKEEVEMMRLAHRKSAAELMSS
Cricetulus_griseus	-----
Leptonychotes_weddellii	-----
Homo_sapiens	EDQKKELDGFLSILCNNLHELQENTICSLVESQKQCGNLTEDLTKIKQTHSQE-----
Symbiodinium_microadriaticum	-----SGYGFASKRITFPGHINERAAKCIVSFPGVYNKLWDKAVK
Nothobranchius_rachovii	LQDQLNLLMSQSQDYADLRSASEALKAPALSLQESICSGCSSMKDQATTQADLHLTTSS
Hydra_vulgaris	LDVVEKNFSNLFTQQTQQLQEDTQMICSELSCQTENESNFKNSISSKLKDMNLTAISTVD
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	LQSQLDLFAQETQKSLTDVLTRNGSLKTTITAMQENVHLKTTDLVSSSTNSHNKFAASLD
Cricetulus_griseus	-----AEKELSQVLVSSWTERFCA-----
Leptonychotes_weddellii	-----TSKDIISKTAHSTQFCADCD
Homo_sapiens	----LCKLMNLWTERFCALEEKECENIQKPLSSVQENIQKSKDIVNKMFFHSQKFCADSD
Symbiodinium_microadriaticum	AVRDDFVTEEDL-----CSLACVFLTDSASGLQHSMN-----PDTPGKCWCHA-
Nothobranchius_rachovii	NLTSSLHQTTQH---VEEMNGYCADLHSSISGLVDRDLRWNLGVRDHTERQTREHLVS-
Hydra_vulgaris	GITTLAYQNLQYATENIDALKTCANSLKTKLDDLEKNETSFENVLIDNSNKYHDMTLKSY
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	NFSQELRSINAENKAMLEESNDHCQHLLTNLKNVAQHTNTWGEFTTAQMVNFTNQHLLS-
Cricetulus_griseus	-----LEKKCENIQKSLNKIQENTEH-----LFSEQWASC-
Leptonychotes_weddellii	GLSQELRCCNQEGTKLVEESVKHCDKLSSNLEVVSQETEQRCALDASTLAFSEQWVSC-
Homo_sapiens	GFSQELRNFNQEGTKLVEESVKHSDKLNGNLEKISQETEQRCESLNRTRTVYFSEQWVSS-
Symbiodinium_microadriaticum	-----IYGQIRASAYLRVVEVDPCADSIESREKL RHEVADADAMNQVLLIKSEQ
Nothobranchius_rachovii	-----MEKLSAEAAASLDQDVDTCTNQLQTAEKELSGHREEVKQALLAIQNQTSM
Hydra_vulgaris	KELVTLSDNLISENHKNINELQTLHNSLCDQCLSNI----NTQVLSCDSVIDGLTEKMKQ
Phanerochaete_chrysosporium	-----
Zonotrichia_albicollis	-----FKDEKQQFYQLQKKNEENCDAIAEAIADHIGSQKAAEEKVLNGLLDQIKV
Cricetulus_griseus	-----LSKSKHEELQNLMEFINECEATGSEITAKQSVHAAAIENQHTSFIVQITS
Leptonychotes_weddellii	-----LNKREEELQNLLEVINQGCKAASSEITEKLNHGKAANENQHNTFFGQITTT
Homo_sapiens	-----LNEREQELHNLLLEVVSQCEASSDITEKSDGRKAAHEKQHNIPLDQMTI

```

Symbiodinium_microadriaticum      NDLEWQKQYARALRTARDLGQENGGRAPWGCQWFEVWKKNVDRVELKQELHVFYFQGRK
Nothobranchius_rachovii            DQTILEQQQAELEDHVEV-----NQQLVHKFLQSELQQDIPTGTTTQQET-----
Hydra_vulgaris                     NQADSMNHSLSLLENGHL-----LENDIIEFVEEELKQDSPTGLTPSQKEYNYPRLQR
Phanerochaete_chrysosporium        EEEVVVRQAYQENETVLD-----G-----VATGL-----
Zonotrichia_albicollis             DQEILVEQKLALREQVQH-----GLTQVNGFLQEDLKVDVPTGTTTQPKDYSPVTLVR
Cricetulus_griseus                 DDEKLKAGNLKLDETIKT-----GLTKLNCFLQEDLKLDIPTGMTPERKNYLYPSTLVR
Leptonychotes_weddellii            DEEKLMAQSLELNETIKM-----GLTKLNCFLQEDLKLDIPTGTTTQPKNYLYPSTLVR
Homo_sapiens                       DEDKLIAQNLELNETIKI-----GLTKLNCFLQEDLKLDIPTGTTTQPKSYLYPSTLVR
                                     ::
                                     .

Symbiodinium_microadriaticum      GQGKLSWEDLSSDAAKNRVRPE--SGLGASQTAEVA-----
Nothobranchius_rachovii            SKSN--ESLATELSFDE-----NLVFNESKRVP-----
Hydra_vulgaris                     TQEH---EQIINEYRQRHDIPPLPSLLEYESDENHD-----GSFLQQSEDKQKESEF
Phanerochaete_chrysosporium        -----
Zonotrichia_albicollis             TEPR---QLLLEQLRQKQ--PNLDAMLSSVGKEMEDSAGQDLLLEEGVLQEPSESLACDKY
Cricetulus_griseus                 TQPR---EQLLDQLQKKQ--HN---QLNCSENDKEA-----
Leptonychotes_weddellii            TEPR---EQLLDQLKRKQ--PELLMMLDCSENNKEEISQDLDEEKS SVLGHSVEELPGQEP
Homo_sapiens                       TEPR---EHLLDQLKRKQ--PELLMMLDCSENNKEETIPDV DVEEAVLGQYTEEPLSQEP

Symbiodinium_microadriaticum      -----YLLKMG I-----PFKEHDVQDFLSFLSSESS-----
Nothobranchius_rachovii            -----FFKQKKGKMTKIPTRS---KVSENNDLQTTPQKTRAPLHCVN-
Hydra_vulgaris                     LIPRSPDETGN-----EVL SIDTASDKENR--NALLNNYSSASKIGNNRPRIPLRSSNT
Phanerochaete_chrysosporium        -----
Zonotrichia_albicollis             SMDTINVYCHTNGGIPFFQH KRS LKKGKENKSAAPLE-NKMEDMTEELLQSKXHLPLRLN-
Cricetulus_griseus                 -----FQHKKPHGKDKENRGINPVEKSKVEETSEYSVTKSRLPLRAS-N
Leptonychotes_weddellii            SIDASVDCSSSGGIPFFQH KKS H GKDKENRGINSVERSKVEETTEHSVTRSRLPLRAQIN
Homo_sapiens                       SVDAGVDCSSIGGVPPFFQH KKS H GKDKENRGINTLERSKVEETTEHLVTKSRLPLRAQIN

Symbiodinium_microadriaticum      --
Nothobranchius_rachovii            --
Hydra_vulgaris                     LS
Phanerochaete_chrysosporium        --
Zonotrichia_albicollis             --
Cricetulus_griseus                 L-
Leptonychotes_weddellii            L-
Homo_sapiens                       L-

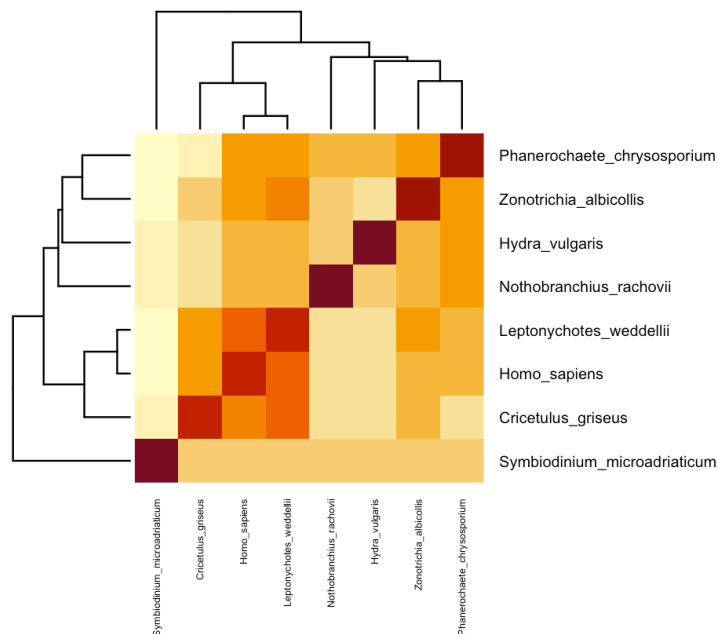
```

[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.

Using “simple phylogeny” from EBI, the phylogeny options are shown below:

STEP 2 - Set your Phylogeny options

TREE FORMAT	DISTANCE CORRECTION	EXCLUDE GAPS	CLUSTERING METHOD	P.I.M.
Default ▾	off ▾	off ▾	Neighbour-joining ▾	off ▾



[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 Evalue and sequence identity to your query. Please also add annotation details of these structures. For example include the annotation terms PDB identifier (structureId), Method used to solve the structure (experimentalTechnique), 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.

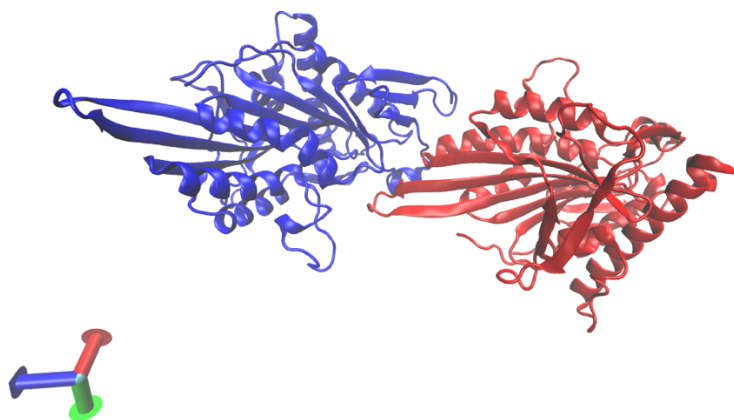
Since the consensus sequence from my alignment has lots of gaps. I chose to use single sequence for blast search. Based on my sequence identity matrix, the highest identity to all other sequence is from *Leptonychotes weddellii*. So, I proceed blast search using Bio3D with *Leptonychotes weddellii* sequence.

ID	Technique	Resolution	Source	Evalue	Identity
3HQD_A	X-RAY DIFFRACTION	2.19	Homo sapiens	0.00E+00	99.128
1II6_A	X-RAY DIFFRACTION	2.1	Homo sapiens	0.00E+00	99.125
3WPN_A	X-RAY DIFFRACTION	2.8	Homo sapiens	0.00E+00	99.128

[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?

The structure of 3HQD (Human kinesin Eg5 motor domain in complex with AMPPNP and Mg²⁺) is shown here.

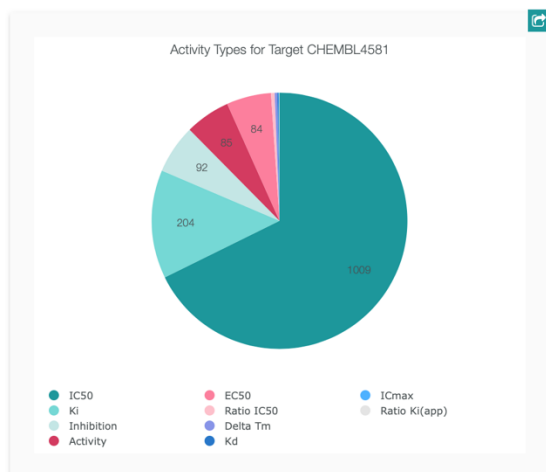


The similarity between my novel protein sequence and chain A sequence from PDBid: 3HQD is 20.8%. So, it is highly unlikely that the structure presented here is similar to the real structure of my novel protein.

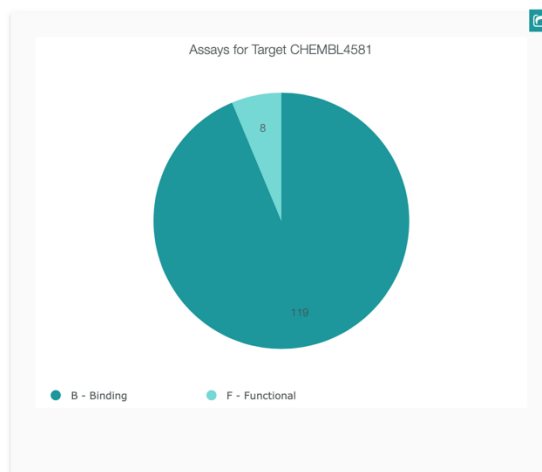
[Q10] Perform a “Target” search of ChEMBEL (<https://www.ebi.ac.uk/chembl/>) with your novel sequence. Are there any **Target Associated Assays** and **ligand efficiency data** reported that may be useful starting points for exploring potential inhibition of your novel protein?

After searching with my novel protein sequence, ChEMBEL feathered 119 binding assays and 8 functional assays. Most of them target kinesin-mediated microtubule function. (https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL4581/)

Associated Bioactivities



Associated Assays



For example, in the review paper reported by *Visanath et al.* (DOI: 10.1021/jm401071u), a allosteric human KIF11 inhibitor SB743921 along with inhibitors AZD4877, MK0731, ARQ621, and ARRY520 are making huge progress towards clinical trials. Considering the similarity between by novel protein and human KIF11, it is highly likely that those inhibitors might work on this novel protein from *Phanerochaete chrysosporium* as well.