**Tiantai Ma**

[**t2ma@ucsd.edu**](mailto:t2ma@ucsd.edu)

**PID: A53308792**

**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 it's 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](https://www.ncbi.nlm.nih.gov/gene/3832))

**Accession number:** NP\_004514

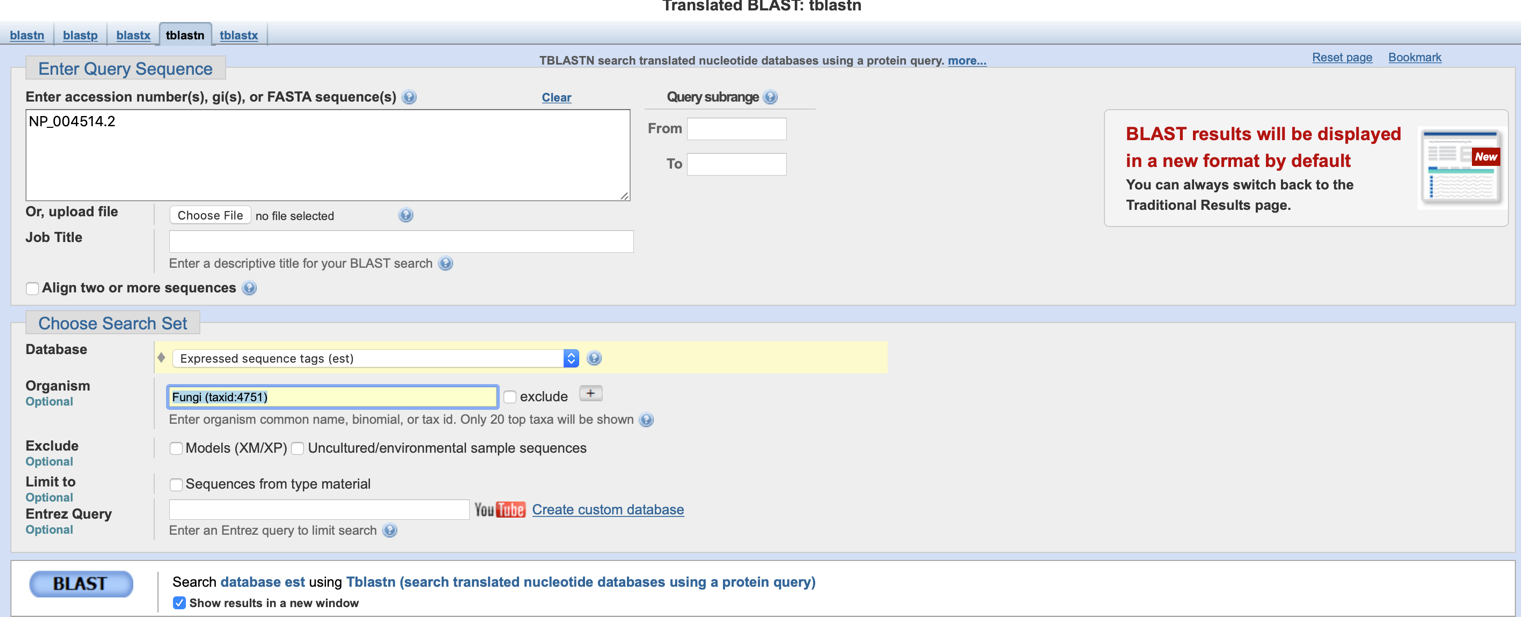
**Species:** Homo sapiens

[**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

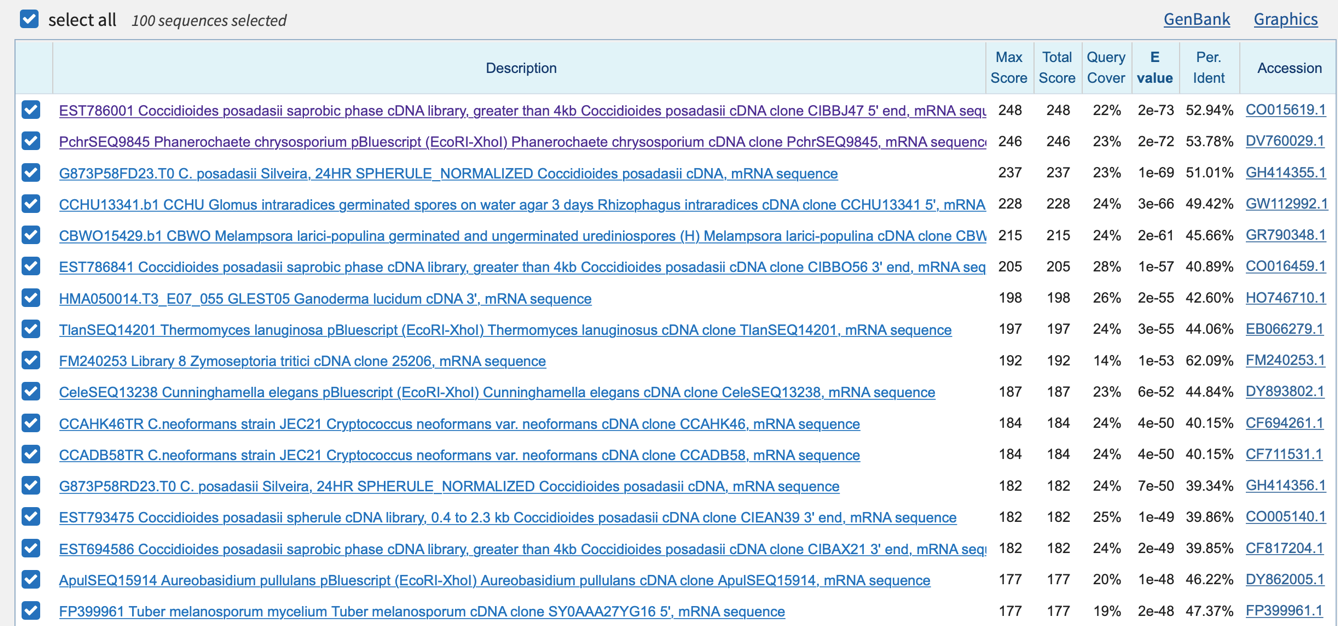
**Database:** Expressed sequence tag (est)

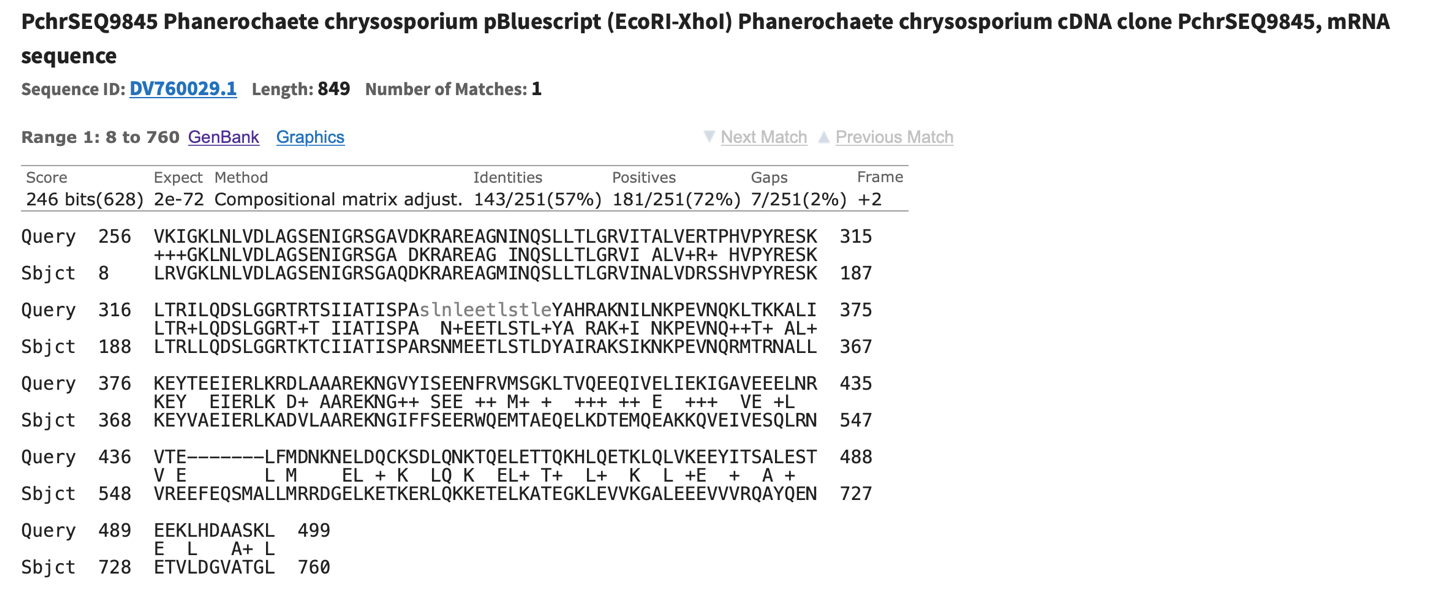
**Organism:** Fungi (taxid:4751)



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](https://www.ncbi.nlm.nih.gov/nucleotide/DV760029.1?report=genbank&log$=nucltop&blast_rank=2&RID=WWH07X67015), 849 base pairs cDNA clone from *Phanerochaete chrysosporium.*





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

# Sequence ID: [DV760029.1](https://www.ncbi.nlm.nih.gov/nucleotide/DV760029.1?report=genbank&log$=nuclalign&blast_rank=2&RID=WWH07X67015) Length: 849 Number of Matches: 1

# Range 1: 8 to 760[GenBank](https://www.ncbi.nlm.nih.gov/nucleotide/DV760029.1?report=genbank&log$=nuclalign&blast_rank=2&RID=WWH07X67015&from=8&to=760)[Graphics](https://www.ncbi.nlm.nih.gov/nuccore/DV760029.1?report=graph&rid=WWH07X67015%5bDV760029.1%5d&tracks=%5bkey:sequence_track,name:Sequence,display_name:Sequence,id:STD1,category:Sequence,annots:Sequence,ShowLabel:true%5d%5bkey:gene_model_track,CDSProductFeats:false%5d%5bkey:alignment_track,name:other%20alignments,annots:NG%20Alignments|Refseq%20Alignments|Gnomon%20Alignments|Unnamed,shown:false%5d&v=0:797&appname=ncbiblast&link_loc=fromHSP)Next 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 VKIGKLNLVDLAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESK 315

+++GKLNLVDLAGSENIGRSGA DKRAREAG INQSLLTLGRVI ALV+R+ HVPYRESK

Sbjct 8 LRVGKLNLVDLAGSENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESK 187

Query 316 LTRILQDSLGGRTRTSIIATISPAslnleetlstleYAHRAKNILNKPEVNQKLTKKALI 375

LTR+LQDSLGGRT+T IIATISPA N+EETLSTL+YA RAK+I NKPEVNQ++T+ AL+

Sbjct 188 LTRLLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL 367

Query 376 KEYTEEIERLKRDLAAAREKNGVYISEENFRVMSGKLTVQEEQIVELIEKIGAVEEELNR 435

KEY EIERLK D+ AAREKNG++ SEE ++ M+ + +++ ++ E +++ VE +L

Sbjct 368 KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDTEMQEAKKQVEIVESQLRN 547

Query 436 VTE-------LFMDNKNELDQCKSDLQNKTQELETTQKHLQETKLQLVKEEYITSALEST 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

LRVGKLNLVDLAGSENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESK

LTRLLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALL

KEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDTEMQEAKKQVEIVESQLRN

VREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALEEEVVVRQAYQEN

ETVLDGVATGL

**Name:** *Phanerochaete chrysosporium*, hypothetical protein

**Species:** [*Phanerochaete chrysosporium*](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=5306)

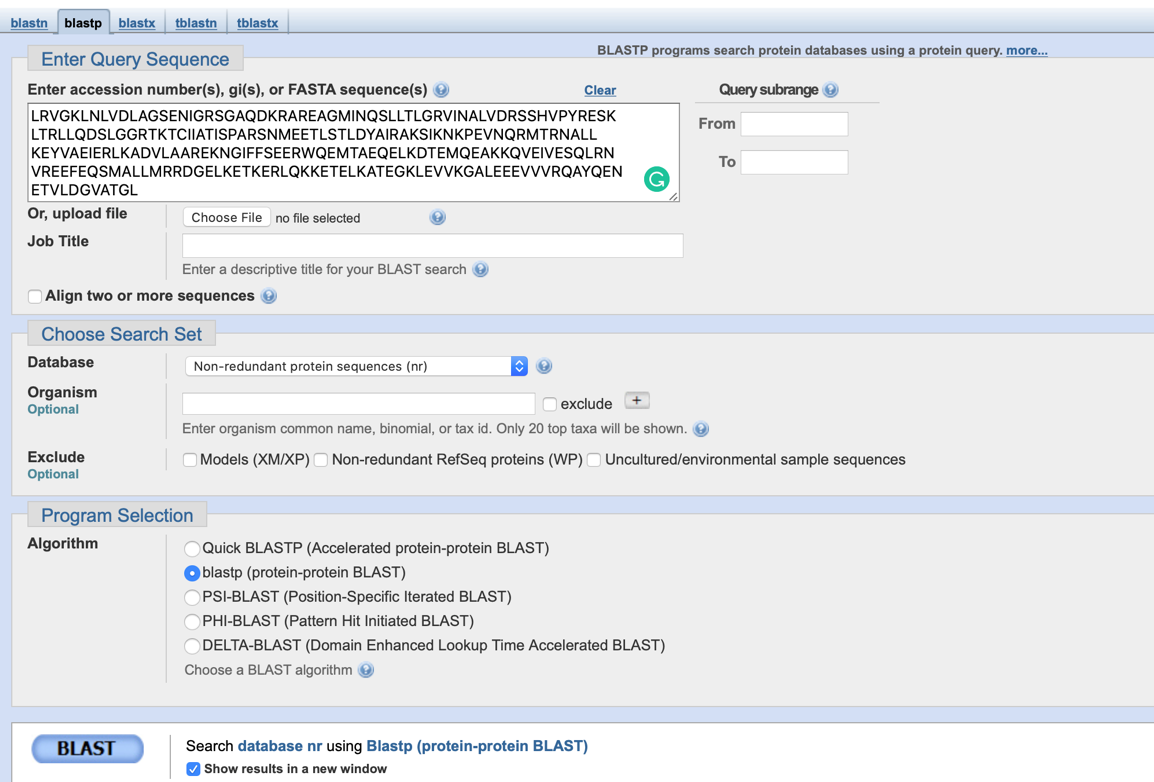
*Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;*

*Agaricomycetes; Polyporales; Phanerochaetaceae; Phanerochaete*

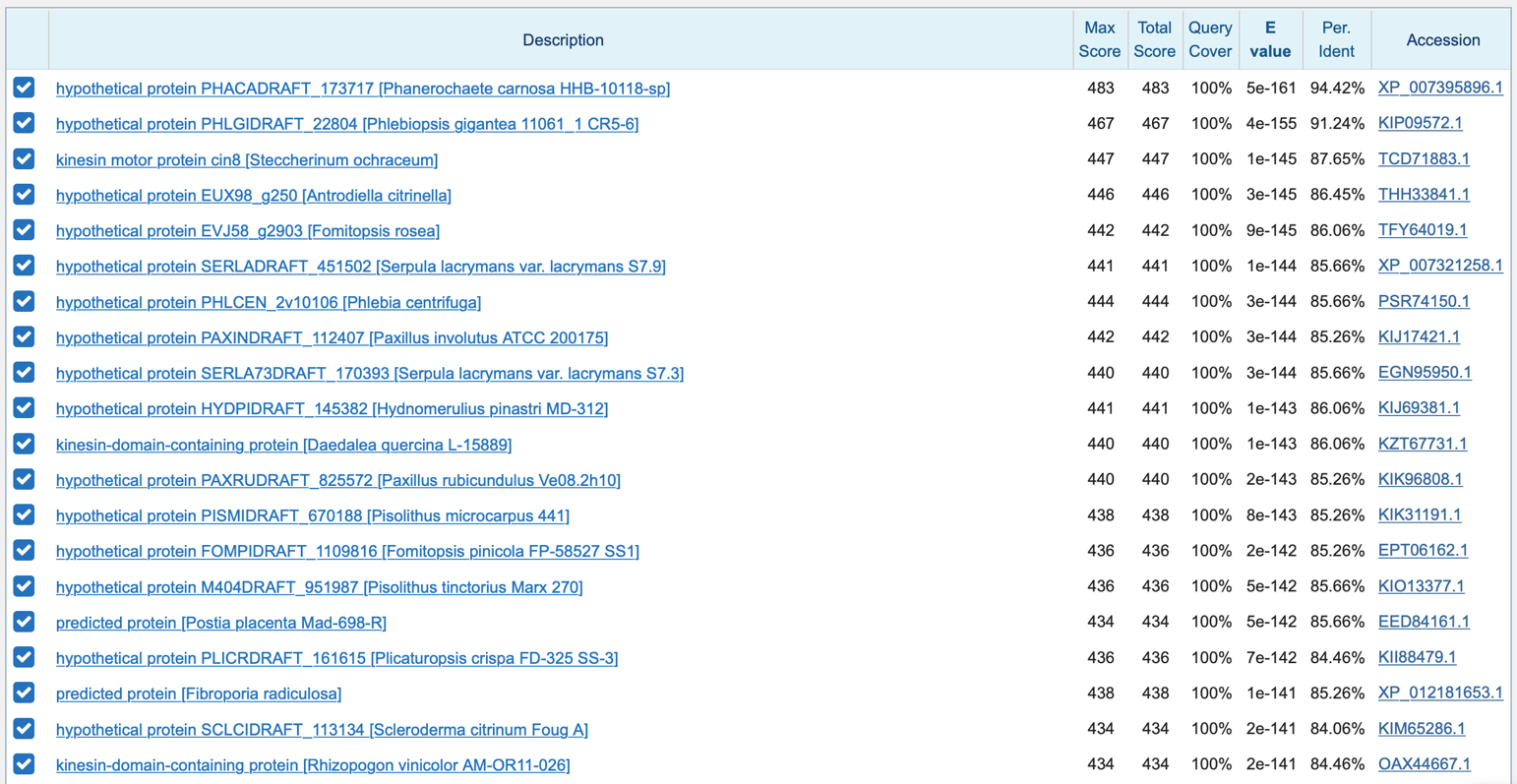
[**Q4**] Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

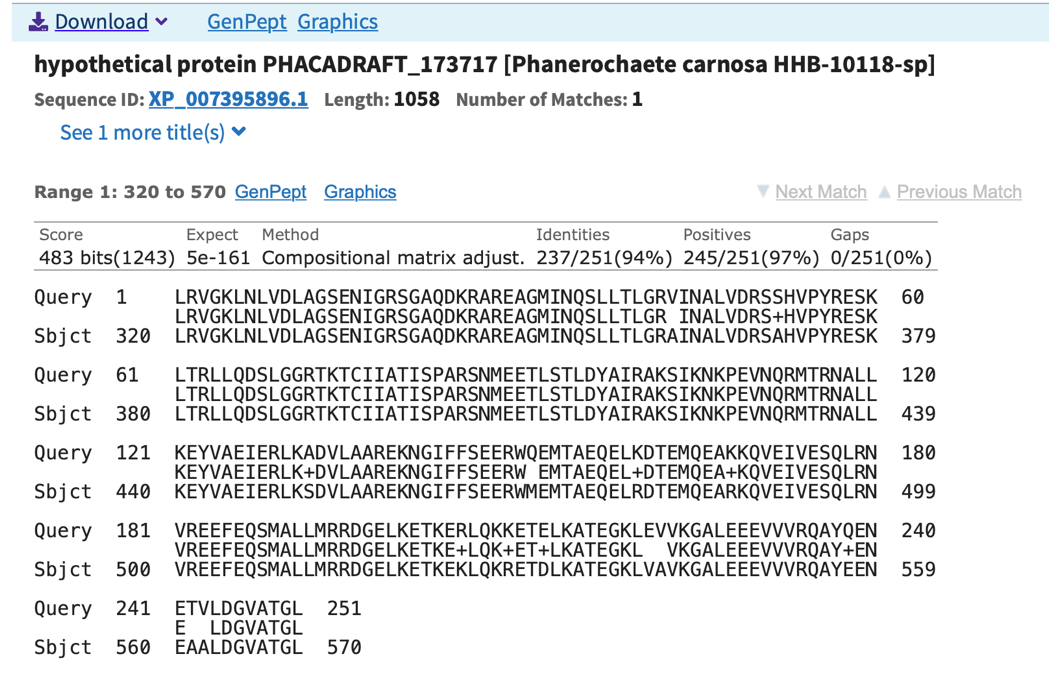
* If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
* If the 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:



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





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]

MNGKLTVQEEQIVESAEKIGALEEELSKVTGLFVDSKNELDQCKSDLQTKTQELETTQKHLQETKLQLVK

EEYVSSALERTEEKLHDAASKLLSTVKETTRDVSGLHSKLDRKRAIDEHNAEAQDIFGNLMSSSVSALDT

VTTTALGTLMSIPQNVSARVTQISDMILKEQSLAAQSKTVLRGLINGLETDLLGSLKTILAPGVVSILNI

NRQLQHIFRTSLTMAEKELSQLVSSWTERFCALEKKCENIQKSLNKIQENTEHLFSEQWASCLSKSKEEL

QNLMEFINECCEATGSEITAKQSVHAAAIENQHTSFIVQITSDDEKLKAGNLKLDETIKTGLTKLNCFLQ

EDLKLDIPTGMTPERKNYLYPSTLVRTQPREQLLDQLQKKQHNQLNCSENDKEAFQHKKPHGKDKENRGI

NPVEKSKVEETSEYSVTKSRLPLRASNL

>Leptonychotes\_weddellii XP\_006733941.1 kinesin-like protein KIF11 [Leptonychotes weddellii]

MSGRPFNLAERKANAHSVVECDHVRKEVSVRTGGLADKSSRKTYTFDMVFGASTKQIDVYRSVVCPILDE

VIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTLHQIFEKLTDNGTEFSVKVSLLE

IYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEVYQILEKGAAKRTTAATLMNAYS

SRSHSVFSVTIHMKETTIDGEELVKIGKLNLVDLAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITAL

VERTPHVPYRESKLTRILQDSLGGRTRTSIIATISPASLNLEETLSTLEYAHRAKNILNKPEVNQKLTKR

ALIKEYTEEIERLKRDLAAAREKNGVYISEENFRAMSGKLTVQEEQIVELIEKIGAIEEELNRVTELFMD

NKNELDQCKSDLQNKTQELETTQKHLQETKLQLVEEEYITSALESTEEKLHDAASRLLTTVEETTKDVSG

LHSKLDRKKAIDQHNAEAQDIFGKNLNSLFSNMEELIKDSSSKQKAMLETHKTLFGNLLSSSVSALDTIT

TTALGSLTSVPENVSIRVSQISNRILKEQSLAAESKTVLQTLIVSKSLKVFFKCYFAKPSKTMTSKDIIS

KTASHSTQFCADCDGLSQELRCCNQEGTKLVEESVKHCDKLSSNLEVVSQETEQRCEALDASTLAFSEQW

VSCLNKREEELQNLLEVINQGCKAASSEITEKLNGHKAANENQHNTFFGQITTDEEKLMAQSLELNETIK

MGLTKLNCFLQQDLKLDIPTGTTPQRKNYLYPSTLVRTEPREQLLDQLKRKQPELLMMLDCSENNKEEIS

QDLDEEKSVLGHSVEELPGQEPSIDASVDCSSSGGIPFFQHKKSHGKDKENRGINSVERSKVEETTEHSV

TRSRLPLRAQINL

>Symbiodinium\_microadriaticum OLP75816.1 Kinesin-like protein KIF11 [Symbiodinium microadriaticum]

MHEPDGCPDNRGTACRRQARSRPATPTGASNAEIPVTNVRVVCRIRPMNDREKKADPQCIEYAGATPAAT

ASTERREVAVVRSTFHFDDVLTSFSSQEDVFRATLQPLVGQVHVARAMLVLAVVSLPGWRANWTQRKAHP

EFTESTVTVSYLEIYNEELSDLLATAERHPKLDLKDIGSGRGVCCQGLSEALPHKDASKAVHGETLTPLE

RPRSLCAMPDGASPAPTVELRLLSTAAAPSPEAEHHGFLEVPSARRGHRRSSLARLTELEHQLHRDKGHL

AEKKLAKALALVLYTLARISANLVSIDYENHIRIVNALVRPLVTGLGVGMFLIREERFSTKSRLLIWSRR

TVFVLAIFVFDESMSTLRAALPIFSVLPWEITCNVMQVHMTNLRLQALGWRLTAKFCHIMAAIGCSFLLG

GLVILTSRISISEFGKEGNLTVERVGAPASFLLFFSFFCCWTTLFTQCGALCQAGSKAMVAGRADLDLDL

KLTACFLYLNALLVLLGPALGFWSWAQFVRRIEMTWLTIDICVQIFTALALSGMMGPRNPMDAFRKLADL

SGYGFASKRITFPGHINERAAKCIVSFPGVYNKLWDKAVKAVRDDFVTEEDLCSLACVFLTDSASGLGQH

SMNPDTPGKCWCHAIYGQIRASAYLRVVEVDPCEADSIESREKLRHEVADADAMNQVLLIKSEQNDLEWQ

KQYARALRTARDLGQENGGRAPWGCQWFEVWKKNVDRAVELKQELHVFYFQGRKGQGKLSWEDLSSDAAK

NRVRPESGLGASQTAEVAYLLKMGIPFKEHDVQDFLSFLSSESS

>Homo\_sapiens NP\_004514.2 kinesin-like protein KIF11 [Homo sapiens]

MASQPNSSAKKKEEKGKNIQVVVRCRPFNLAERKASAHSIVECDPVRKEVSVRTGGLADKSSRKTYTFDM

VFGASTKQIDVYRSVVCPILDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTL

HQIFEKLTDNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKDEV

YQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTIDGEELVKIGKLNLVDLAGSENIGRSGAVDK

RAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSLGGRTRTSIIATISPASLNLEETLSTL

EYAHRAKNILNKPEVNQKLTKKALIKEYTEEIERLKRDLAAAREKNGVYISEENFRVMSGKLTVQEEQIV

ELIEKIGAVEEELNRVTELFMDNKNELDQCKSDLQNKTQELETTQKHLQETKLQLVKEEYITSALESTEE

KLHDAASKLLNTVEETTKDVSGLHSKLDRKKAVDQHNAEAQDIFGKNLNSLFNNMEELIKDGSSKQKAML

EVHKTLFGNLLSSSVSALDTITTVALGSLTSIPENVSTHVSQIFNMILKEQSLAAESKTVLQELINVLKT

DLLSSLEMILSPTVVSILKINSQLKHIFKTSLTVADKIEDQKKELDGFLSILCNNLHELQENTICSLVES

QKQCGNLTEDLKTIKQTHSQELCKLMNLWTERFCALEEKCENIQKPLSSVQENIQQKSKDIVNKMTFHSQ

KFCADSDGFSQELRNFNQEGTKLVEESVKHSDKLNGNLEKISQETEQRCESLNTRTVYFSEQWVSSLNER

EQELHNLLEVVSQCCEASSSDITEKSDGRKAAHEKQHNIFLDQMTIDEDKLIAQNLELNETIKIGLTKLN

CFLEQDLKLDIPTGTTPQRKSYLYPSTLVRTEPREHLLDQLKRKQPELLMMLNCSENNKEETIPDVDVEE

AVLGQYTEEPLSQEPSVDAGVDCSSIGGVPFFQHKKSHGKDKENRGINTLERSKVEETTEHLVTKSRLPL

RAQINL

>Phanerochaete\_chrysosporium 8-760\_1 PchrSEQ9845 Phanerochaete chrysosporium pBluescript (EcoRI-XhoI) Phanerochaete chrysosporium cDNA clone PchrSEQ9845, mRNA sequence

LRVGKLNLVDLAGSENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESKLTRLLQDSLGGRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALLKEYVAEIERLKADVLAAREKNGIFFSEERWQEMTAEQELKDTEMQEAKKQVEIVESQLRNVREEFEQSMALLMRRDGELKETKERLQKKETELKATEGKLEVVKGALEEEVVVRQAYQENETVLDGVATGL

>Nothobranchius\_rachovii SBS17602.1 kinesin family member 11 [Nothobranchius rachovii]

MAGASVKVAVRVRPFNSRETEKESKCIIQMSGNTTTIINPKQAKDNKSFNFDFSYWSHTSPEDVNYASQM

RVYKDIGEEMLLHAFEGYNVCIFAYGQTGAGKSYTMMGKQDVKDQQGIIPLLCEDLFTKINDNTNNSMSY

SVEVSYMEIYCERVRDLLNPKNKGNLRVREHPLMGPYVEDLSKLAVTSYNDIQDLMDSGNKARTVAATNM

NETSSRSHAVFNIIFTQKRYDAETDNTSEKVSKISLVDLAGSENIGRSGAVDKRAREAGNINQSLLTLGR

VITALVEKRPHIPYRESKLTRILQDSLGGRTKTSIIATVSPSSSNMEETMSTLEYACRAKNIMNKPEVNQ

KLTKRTLIKEYTEEIERLKRDLIAVREKNGVYLSSENYESMMTQIMAHEEQIAEYVDKIIIMEEELRKVT

ELFEDSQTKLEQCTADLDEKHKRLEETSKDLQQTKEKLVEEEFICSEVTSVHESLYNTAGQLLSTADAST

RDVSGLHDKLDRKKKVESHNKQIQQSFAERMDGALSSMQRCVHRHGAKHNELLSNCSQAVDGLLMMNEVA

LKGTVTTMESFVGGVRSVVAEGVARCREKLQQHESVCLQDKESLLQLLEEHQQDMEDVLAARTLMGLSAV

NELSDVLRSSVEKQRALADKVEESVGLFFNSLCRDLDSLREEAVHRLSSLQDEQDKLEDRISQAQERQRL

GMQQTIQCLQDQLNLLSMQSQQDYADLRSASEALKAPALSLQESICSGCSSMKDQATTQADLHLTTSSNL

TSSLHQTQQHVEEMNGYCADLHSSISGLVDRDLRWNLGVRDHTERQTREHLSVMEKLSAEAASLDQDVDT

CCTNQLQTAEKELSGHREEVKQALLAIQNQTSMDQTILEQQQAELEDHVEVNQQLVHKFLQSELQQDIPT

GTTPQQETSKSNESLATELSFDENLVFNESKRVPFFKQKKGKMTKIPTRSKVSENNDSLQTTPQKTRAPL

HCVN

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

MGKPILKLDKNQNIQVAVRCRPRNSQEIKAGSADVVDINNQTKEITVRQDMNYMDKGNNKTFSFDKVFGP

KSKQIDIYQSMVCPVIDEVLQGYNCTIFAYGQTGTGKTFTMEGDRLDSEEFSWQDDPLAGIIPRAMHQLF

ELLNIMEECAEFSVRVSFLEIYNEELFDLLGTSLDSQKLRLFEDTTKKGSVVIQGLEEVIVHSRNEVYHI

LERGAARRQTAATLLNAQSSRSHSLFMVTIHMKENNINGEEFLKTGKLNLVDLAGSENIGRSGAVEKRAR

EAGTINQSLLTLGRVITALVENAPHVPYRESKLTRLLKDSLGGRTKTSIIATISPAACNLEETLSTLDYA

QRAKNITNKPEINQKLTKKALIKEYTEEIEKLKKDLFAAREKNGIFIAEDNYILMQNQLSSQKSSLREYV

EKIQFMEEEMKKTEQLFASMQQKLEKTTQHLQWIKQERDETKVLVSKHVETEQQLFNQASELLDTVEETI

NHKNILHSSLDRNKNLHVVNLLSSQKFKENAHAKVSSLNESIINGRHSSFAYHQDIVTELNNFKSEKASS

YNQLSKIINISVEEIYKCLEIARTKQNEIKSECTKKDLLFSHNVETCKNETLQAINNYHKEIFTNALKEF

HECILRMQECDERLVNSVNAMTEHFCETTKGFIEKQTVELRDLGQTVDEHICCQKEAIEKVNSYVEKTVQ

NQDKMQKELKDGFIKSLQASILDVVEKNFSNLFQTQQTQLQEDTQMICSELSCQTENESNFKNSISSKLK

DMNLTAISTVDGITTLAYQNLQYATENIDALKTCANSLKTKLDDLEKNETSFENVLIDNSNKYHDMTLKS

YKELVTLSDNLISENHKNINELQTLHNSLCDQCLSNINTQVLSCDSVIDGLTEKMKQNQADSMNHSLSLE

NGLHLLENDIIEFVEEELKQDSPTGLTPSQKEYNYPRSLQRTQEHEQIINEYRQRHDIPPLPSLLEYESD

ENHDGSFLQQSEDKQKESEFLIPRSPDETGNEVLSIDTASDKENRNALLNNYYSSASKIGNNRPRIPLRS

SNTLS

>Zonotrichia\_albicollis XP\_014131087.1 kinesin-like protein KIF11 isoform X2 [Zonotrichia albicollis]

MVFGAQAKQIDVYRSVVCPILDEVIMGYNCTVFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRT

LHQIFEKLTENGTEFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKNQ

VYQILERGAAKRTTAATYMNAYSSRSHSVFSITIHMKETTVDGEELVKIGKLNLVDLAGSENIGRSGAVD

KRAREAGNINQSLLTLGRVITALVERAPHIPYRESKLTRILQDSLGGRTKTSIIATISPASVNLEETLST

LEYAHRAKNIMNKPEVNQKLTKKALIKEYTEEIERLKRDLAAAREKNGVYISAENYEALNGKLTVQEEQI

TEYIDKISVMEEEVKRVTELFRVSKNELEQCKTDLQIKEKELEETQKDLQETKVQLAEEEYVVSVLESTE

QELHDTASQLLTTVEETTRDVSGLHAKLERKRAVDQHNAAVQNTFAGQMNASFSKIQDSITENSLKQQQM

LTYYTNCIGDLLSTSSSTADMFASVVSASFACLKELVSTEVSHISEKITQHENLSLDCKAELLRLIEEHQ

TGLGRAVNSLTPMVEFVLGLNCQFQSNMKKYSAVADQMEDHKKEMDTFFADLSLTLKKIQEQTAGGFAQL

QHNCDSLKEEVEMMRLAHRKSAAELMSSLQSQLDLFAQETQKSLTDVLTRNGSLKTTITAMQENVHLKTT

DLVSSTNSNHNKFAASLDNFSQELRSINAENKAMLEESNDHCQHLLTNLKNVAQHTNTWGEFTTAQMVNF

TNQHLLSFKDEKQQFQYLQKKNEENCDQAIAEIADHIGSQKAAEEKVLNGLLDQIKVDQEILVEQKLALR

EQVQHGLTQVNGFLQEDLKVDVPTGTTPQRKDYSYPVTLVRTEPRQLLLEQLRQKQPNLDAMLSSVGKEM

EDSAGQDLLEEGVLQEPSESLACDKYSMDTNVYCHTNGGIPFFQHKRSLKKGKENKSAAPLENKMEDMTEELLQKSKHPLRLLN

Alignment using MUSCLE at EBI:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Symbiodinium\_microadriaticum MHEPDGCPDNRGTACRRQARSRPATPTGASNAEIPVTNVRVVCRIRPMNDREKKADPQCI

Nothobranchius\_rachovii ---------------------------------MAGASVKVAVRVRPFNSRETEKESKCI

Hydra\_vulgaris -------------------------MGKPILKLDKNQNIQVAVRCRPRNSQEIKAGSADV

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis ------------------------------------------------------------

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii ------------------------------------------MSGRPFNLAERKANAHSV

Homo\_sapiens --------------------MASQPNSSAKKKEEKGKNIQVVVRCRPFNLAERKASAHSI

Symbiodinium\_microadriaticum -EYAGATPAATASTE-RREVAVVRSTFHFDDVLTSFSSQED--------VFRATLQPLVG

Nothobranchius\_rachovii IQMSGNTTTII-----NPKQAKDNKSFNFDFSYWSHTSPEDVNYASQMRVYKDIGEEMLL

Hydra\_vulgaris VDINNQTKEITVRQDMNYMDKGNNKTFSFDKVFGPKSKQID--------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 HAFEGYNVCIFAYGQTGAGKSYTMMGKQDVKDQ---Q-----GIIPLLCEDLFTKINDNT

Hydra\_vulgaris EVLQGYNCTIFAYGQTGTGKTFTMEGDRLDSEEFSWQDDPLAGIIPRAMHQLFELLNIME

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis EVIMGYNCTVFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTLHQIFEKLTENG

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii EVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTLHQIFEKLTDNG

Homo\_sapiens EVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEEDPLAGIIPRTLHQIFEKLTDNG

Symbiodinium\_microadriaticum ----ESTVTVSYLEIYNEELSDLLATAER-HPKLDL-KDIGSGRGVCCQGLSEALPH---

Nothobranchius\_rachovii NNSMSYSVEVSYMEIYCERVRDLLNPKNKGNLRVR----EHPLMGPYVEDLSKLAVTSYN

Hydra\_vulgaris ECA-EFSVRVSFLEIYNEELFDLLGTSLD-SQKLRLFEDTTKKGSVVIQGLEEVIVHSRN

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis T---EFSVKVSLLEIYNEELFDLLNPTPDVGERLQMFDDPRNKRGVIIKGLEEVTVHNKN

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii T---EFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKD

Homo\_sapiens T---EFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDDPRNKRGVIIKGLEEITVHNKD

Symbiodinium\_microadriaticum -------KDASKAVHGETLTPLERPRSLCAMP----------DGASPA-PTVELRLLSTA

Nothobranchius\_rachovii DIQDLMDSGNKARTVAATNMNETSSRSHAVFNIIFTQKRYDAETDNTSEKVSKISLVDLA

Hydra\_vulgaris EVYHILERGAARRQTAATLLNAQSSRSHSLFMVTIHMKENNINGEEFL-KTGKLNLVDLA

Phanerochaete\_chrysosporium -----------------------------------------------L-RVGKLNLVDLA

Zonotrichia\_albicollis QVYQILERGAAKRTTAATYMNAYSSRSHSVFSITIHMKETTVDGEELV-KIGKLNLVDLA

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii EVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTIDGEELV-KIGKLNLVDLA

Homo\_sapiens EVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTIDGEELV-KIGKLNLVDLA

Symbiodinium\_microadriaticum AAPSPEAEHHGFLEVPSARRGHRRSSLARLTELEHQLHRDKGHLA--EKKLAKAL-----

Nothobranchius\_rachovii G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHIPYRESKLTRILQDSLG

Hydra\_vulgaris G--SENIGRSGAVEKRAREAGTINQSLLTLGRVITALVENAPHVPYRESKLTRLLKDSLG

Phanerochaete\_chrysosporium G--SENIGRSGAQDKRAREAGMINQSLLTLGRVINALVDRSSHVPYRESKLTRLLQDSLG

Zonotrichia\_albicollis G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERAPHIPYRESKLTRILQDSLG

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSLG

Homo\_sapiens G--SENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSLG

Symbiodinium\_microadriaticum ----ALVLYTLARISAN----LVSIDYENHIRIV------------NALVRPLVTGLGV-

Nothobranchius\_rachovii GRTKTSIIATVSPSSSNMEETMSTLEYACRAKNIMNKPEVNQKLTKRTLIKEYTEEIERL

Hydra\_vulgaris GRTKTSIIATISPAACNLEETLSTLDYAQRAKNITNKPEINQKLTKKALIKEYTEEIEKL

Phanerochaete\_chrysosporium GRTKTCIIATISPARSNMEETLSTLDYAIRAKSIKNKPEVNQRMTRNALLKEYVAEIERL

Zonotrichia\_albicollis GRTKTSIIATISPASVNLEETLSTLEYAHRAKNIMNKPEVNQKLTKKALIKEYTEEIERL

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii GRTRTSIIATISPASLNLEETLSTLEYAHRAKNILNKPEVNQKLTKRALIKEYTEEIERL

Homo\_sapiens GRTRTSIIATISPASLNLEETLSTLEYAHRAKNILNKPEVNQKLTKKALIKEYTEEIERL

Symbiodinium\_microadriaticum -----------GMFLIREERFSTKSRLLIWSRRTV-FVLAIFVFDESM-----------S

Nothobranchius\_rachovii KRDLIAVREKNGVYLSSENYESMMTQIMAHEEQIAEYVDKIIIMEEELRKVTELFEDSQT

Hydra\_vulgaris KKDLFAAREKNGIFIAEDNYILMQNQLSSQKSSLREYVEKIQFMEEEMKKTEQLFASMQQ

Phanerochaete\_chrysosporium KADVLAAREKNGIFFSEERWQEMTAEQELKDTEMQEAKKQVEIVESQLRNV-------RE

Zonotrichia\_albicollis KRDLAAAREKNGVYISAENYEALNGKLTVQEEQITEYIDKISVMEEEVKRVTELFRVSKN

Cricetulus\_griseus ----------------------MNGKLTVQEEQIVESAEKIGALEEELSKVTGLFVDSKN

Leptonychotes\_weddellii KRDLAAAREKNGVYISEENFRAMSGKLTVQEEQIVELIEKIGAIEEELNRVTELFMDNKN

Homo\_sapiens KRDLAAAREKNGVYISEENFRVMSGKLTVQEEQIVELIEKIGAVEEELNRVTELFMDNKN

.. : .:..:

Symbiodinium\_microadriaticum TLRAALPIFSVLPWEITCNVMQVHMTNLR----------LQALGWRLTAKFCHIMAAIGC

Nothobranchius\_rachovii KLEQCTADLDEKHKRLEETSKDLQQTKEKLVEEEFICSEVTSVHESLYNTAGQLLSTADA

Hydra\_vulgaris KLEKTTQHLQWIKQERDETKVLVSKH--------------VETEQQLFNQASELLDTVEE

Phanerochaete\_chrysosporium EFEQSMALLMRRDGELKETKERLQKKETE----------LKATEGKLE------------

Zonotrichia\_albicollis ELEQCKTDLQIKEKELEETQKDLQETKVQLAEEEYVVSVLESTEQELHDTASQLLTTVEE

Cricetulus\_griseus ELDQCKSDLQTKTQELETTQKHLQETKLQLVKEEYVSSALERTEEKLHDAASKLLSTVKE

Leptonychotes\_weddellii ELDQCKSDLQNKTQELETTQKHLQETKLQLVEEEYITSALESTEEKLHDAASRLLTTVEE

Homo\_sapiens ELDQCKSDLQNKTQELETTQKHLQETKLQLVKEEYITSALESTEEKLHDAASKLLNTVEE

: : . : \*

Symbiodinium\_microadriaticum SFLLGGLVILTSRISISEFGKEGNLTVERV------------------GAPASFLLFFSF

Nothobranchius\_rachovii S--TRDVSGLHDKLDRKKKVESHNKQIQQSFAERMDGALSSMQRCVHRHGAKHNELLSNC

Hydra\_vulgaris T--INHKNILHSSLDRNKNLHVVNLLSSQKFKENAHAKVSSLNESIINGRHSSFAYHQDI

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis T--TRDVSGLHAKLERKRAVDQHNAAVQNTFAGQMNASFSKIQDSITENSLKQQQMLTYY

Cricetulus\_griseus T--TRDVSGLHSKLDRKRAIDEHNAEAQDI------------------------------

Leptonychotes\_weddellii T--TKDVSGLHSKLDRKKAIDQHNAEAQDIFGKNLNSLFSNMEELIKDSSSKQKAMLETH

Homo\_sapiens T--TKDVSGLHSKLDRKKAVDQHNAEAQDIFGKNLNSLFNNMEELIKDGSSKQKAMLEVH

Symbiodinium\_microadriaticum FCCWTTLFTQCGALCQAGSKAMVAGRADLDLDLKLTACFLYLNALLVLLGPALGFW----

Nothobranchius\_rachovii SQAVDGLLMMNEVALKGTVTTMESFVGGV---RSVVAEGVARCREKLQQHESVCLQ----

Hydra\_vulgaris VTELNNFKSEKASSYNQLSKIINISVEEIYKCLEIARTKQNEIKSECTKKDLLFSHNVET

Phanerochaete\_chrysosporium --------------------VVKGAL----------------------------------

Zonotrichia\_albicollis TNCIGDLLSTSSSTADMFASVVSASFACL---KELVSTEVSHISEKITQHENLSLD----

Cricetulus\_griseus ---FGNLMSSSVSALDTVTTTALGTLMSI---PQNVSARVTQISDMILKEQSLAAQ----

Leptonychotes\_weddellii KTLFGNLLSSSVSALDTITTTALGSLTSV---PENVSIRVSQISNRILKEQSLAAE----

Homo\_sapiens KTLFGNLLSSSVSALDTITTVALGSLTSI---PENVSTHVSQIFNMILKEQSLAAE----

Symbiodinium\_microadriaticum SWAQFVRRIEMTWLTIDICVQIFTALALSGMMGPRNPM---------DAFRKLADL----

Nothobranchius\_rachovii DKESLLQLLEEHQQDMEDVLAARTLMGLSAVNELSDVL--------RSSVEKQRALADKV

Hydra\_vulgaris CKNETLQAINNYHKEIFTNALKEFHECILRMQECDERLVNSVNAMTEHFCETTKGFIEKQ

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis CKAELLRLIEEHQTGLGRAVNS-LTPMVEFVLGLNCQF--------QSNMKKYSAVADQM

Cricetulus\_griseus SKTVLRGLINGLETDLLGSLKTILAPGVVSILNINRQL--------QHIFRTSLTM----

Leptonychotes\_weddellii SKTVLQTLI------VSKSLKVFFK---------------------CYFAKPSKTM----

Homo\_sapiens SKTVLQELINVLKTDLLSSLEMILSPTVVSILKINSQL--------KHIFKTSLTVADKI

Symbiodinium\_microadriaticum ------------------------------------------------------------

Nothobranchius\_rachovii EE---SVGLFFNSLCRDLDSLREEAVHRLSSLQDEQDKLEDRISQAQERQRLGMQQTIQC

Hydra\_vulgaris TVELRDLGQTVDEHICCQKEAIEKVNSYVEKTVQNQDKMQKELKDGFIKSLQA-----SI

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis EDHKKEMDTFFADLSLTLKKIQEQTAGGFAQLQHNCDSLKEEVEMMRLAHRKSAAELMSS

Cricetulus\_griseus ------------------------------------------------------------

Leptonychotes\_weddellii ------------------------------------------------------------

Homo\_sapiens EDQKKELDGFLSILCNNLHELQENTICSLVESQKQCGNLTEDLKTIKQTHSQE-------

Symbiodinium\_microadriaticum --------------------SGYGFASKRITFPGHINERAAKCIVSFPGVYNKLWDKAVK

Nothobranchius\_rachovii LQDQLNLLSMQSQQDYADLRSASEALKAPALSLQESICSGCSSMKDQATTQADLHLTTSS

Hydra\_vulgaris LDVVEKNFSNLFQTQQTQLQEDTQMICSELSCQTENESNFKNSISSKLKDMNLTAISTVD

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis LQSQLDLFAQETQKSLTDVLTRNGSLKTTITAMQENVHLKTTDLVSSTNSNHNKFAASLD

Cricetulus\_griseus ---------------------------------------AEKELSQLVSSWTERFCA---

Leptonychotes\_weddellii ---------------------------------------TSKDIISKTASHSTQFCADCD

Homo\_sapiens ----LCKLMNLWTERFCALEEKCENIQKPLSSVQENIQQKSKDIVNKMTFHSQKFCADSD

Symbiodinium\_microadriaticum AVRDDFVTEEDL-------CSLACVFLTDSASGLGQHSMN--------PDTPGKCWCHA-

Nothobranchius\_rachovii NLTSSLHQTQQH----VEEMNGYCADLHSSISGLVDRDLRWNLGVRDHTERQTREHLSV-

Hydra\_vulgaris GITTLAYQNLQYATENIDALKTCANSLKTKLDDLEKNETSFENVLIDNSNKYHDMTLKSY

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis NFSQELRSINAENKAMLEESNDHCQHLLTNLKNVAQHTNTWGEFTTAQMVNFTNQHLLS-

Cricetulus\_griseus -------------------LEKKCENIQKSLNKIQENTEH----------LFSEQWASC-

Leptonychotes\_weddellii GLSQELRCCNQEGTKLVEESVKHCDKLSSNLEVVSQETEQRCEALDASTLAFSEQWVSC-

Homo\_sapiens GFSQELRNFNQEGTKLVEESVKHSDKLNGNLEKISQETEQRCESLNTRTVYFSEQWVSS-

Symbiodinium\_microadriaticum ----------IYGQIRASAYLRVVEVDPCEADSIESREKLRHEVADADAMNQVLLIKSEQ

Nothobranchius\_rachovii ----------MEKLSAEAASLDQDVDTCCTNQLQTAEKELSGHREEVKQALLAIQNQTSM

Hydra\_vulgaris KELVTLSDNLISENHKNINELQTLHNSLCDQCLSNI----NTQVLSCDSVIDGLTEKMKQ

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis ----------FKDEKQQFQYLQKKNEENCDQAIAEIADHIGSQKAAEEKVLNGLLDQIKV

Cricetulus\_griseus ----------LSKSKEELQNLMEFINECCEATGSEITAKQSVHAAAIENQHTSFIVQITS

Leptonychotes\_weddellii ----------LNKREEELQNLLEVINQGCKAASSEITEKLNGHKAANENQHNTFFGQITT

Homo\_sapiens ----------LNEREQELHNLLEVVSQCCEASSSDITEKSDGRKAAHEKQHNIFLDQMTI

Symbiodinium\_microadriaticum NDLEWQKQYARALRTARDLGQENGGRAPWGCQWFEVWKKNVDRAVELKQELHVFYFQGRK

Nothobranchius\_rachovii DQTILEQQQAELEDHVEV------NQQLVHKFLQSELQQDIPTGTTPQQET---------

Hydra\_vulgaris NQADSMNHSLSLENGLHL------LENDIIEFVEEELKQDSPTGLTPSQKEYNYPRSLQR

Phanerochaete\_chrysosporium EEEVVVRQAYQENETVLD------G---------------VATGL---------------

Zonotrichia\_albicollis DQEILVEQKLALREQVQH------GLTQVNGFLQEDLKVDVPTGTTPQRKDYSYPVTLVR

Cricetulus\_griseus DDEKLKAGNLKLDETIKT------GLTKLNCFLQEDLKLDIPTGMTPERKNYLYPSTLVR

Leptonychotes\_weddellii DEEKLMAQSLELNETIKM------GLTKLNCFLQQDLKLDIPTGTTPQRKNYLYPSTLVR

Homo\_sapiens DEDKLIAQNLELNETIKI------GLTKLNCFLEQDLKLDIPTGTTPQRKSYLYPSTLVR

:: .

Symbiodinium\_microadriaticum GQGKLSWEDLSSDAAKNRVRPE--SGLGASQTAEVA------------------------

Nothobranchius\_rachovii SKSN---ESLATELSFDE-------NLVFNESKRVP------------------------

Hydra\_vulgaris TQEH---EQIINEYRQRHDIPPLPSLLEYESDENHD--------GSFLQQSEDKQKESEF

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis TEPR---QLLLEQLRQKQ--PNLDAMLSSVGKEMEDSAGQDLLEEGVLQEPSESLACDKY

Cricetulus\_griseus TQPR---EQLLDQLQKKQ--HN---QLNCSENDKEA------------------------

Leptonychotes\_weddellii TEPR---EQLLDQLKRKQ--PELLMMLDCSENNKEEISQDLDEEKSVLGHSVEELPGQEP

Homo\_sapiens TEPR---EHLLDQLKRKQ--PELLMMLNCSENNKEETIPDVDVEEAVLGQYTEEPLSQEP

Symbiodinium\_microadriaticum ----------------YLLKMGI----------PFKEHDVQDFLSFLSSESS--------

Nothobranchius\_rachovii ---------------FFKQKKGKMTKIPTRS----KVSENNDSLQTTPQKTRAPLHCVN-

Hydra\_vulgaris LIPRSPDETGN-----EVLSIDTASDKENR--NALLNNYYSSASKIGNNRPRIPLRSSNT

Phanerochaete\_chrysosporium ------------------------------------------------------------

Zonotrichia\_albicollis SMDTNVYCHTNGGIPFFQHKRSLKKGKENKSAAPLE-NKMEDMTEELLQKSKHPLRLL-N

Cricetulus\_griseus ----------------FQHKKPHGKDKENRGINPVEKSKVEETSEYSVTKSRLPLRAS-N

Leptonychotes\_weddellii SIDASVDCSSSGGIPFFQHKKSHGKDKENRGINSVERSKVEETTEHSVTRSRLPLRAQIN

Homo\_sapiens SVDAGVDCSSIGGVPFFQHKKSHGKDKENRGINTLERSKVEETTEHLVTKSRLPLRAQIN

Symbiodinium\_microadriaticum --

Nothobranchius\_rachovii --

Hydra\_vulgaris LS

Phanerochaete\_chrysosporium --

Zonotrichia\_albicollis --

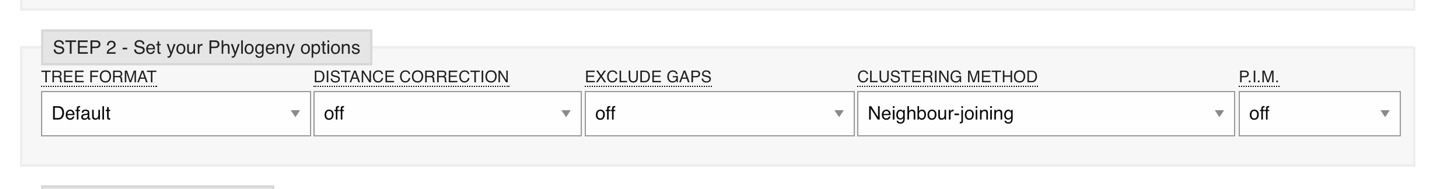
Cricetulus\_griseus L-

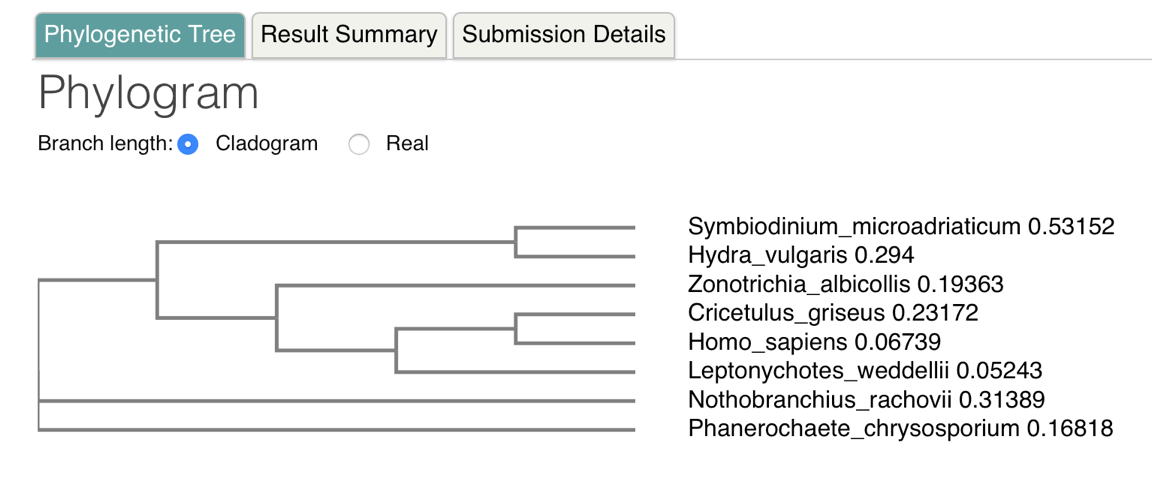
Leptonychotes\_weddellii L-

Homo\_sapiens L-

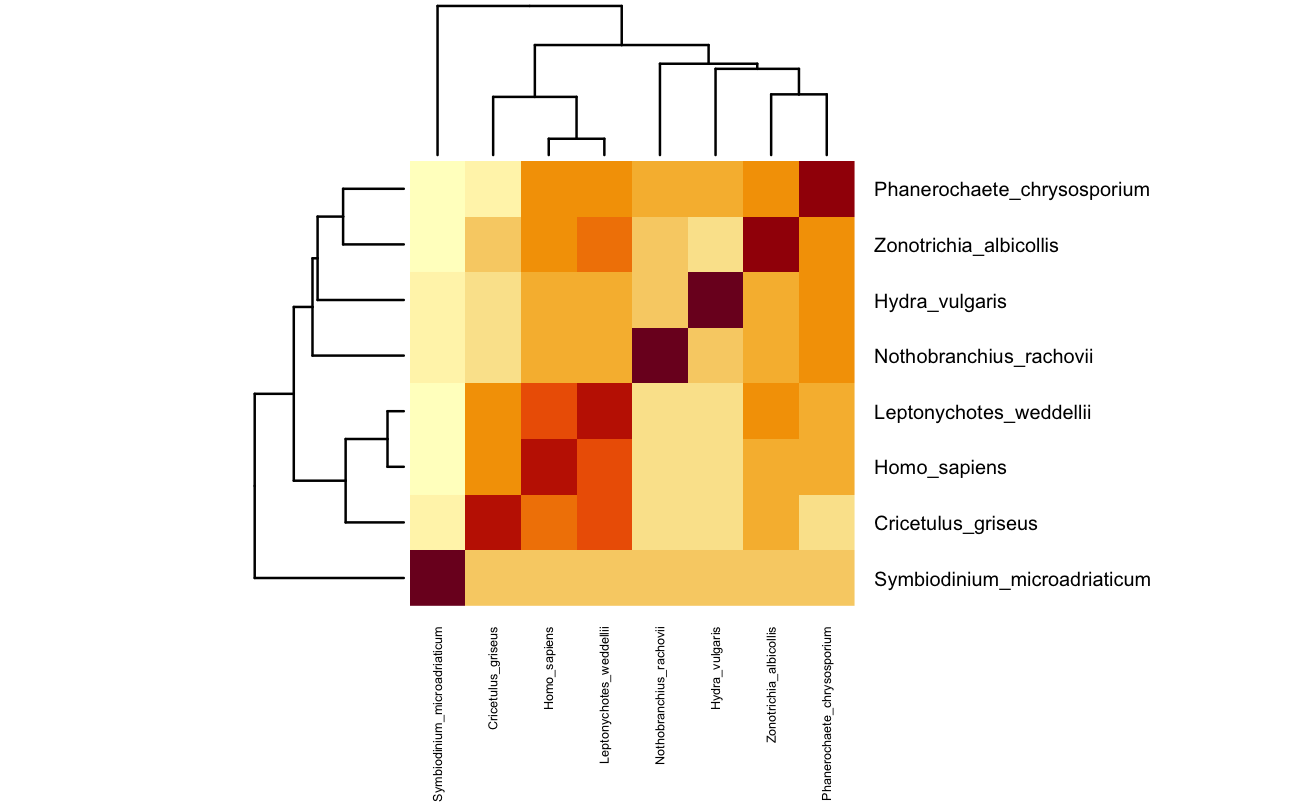
[**Q6**] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your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

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





[**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 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.

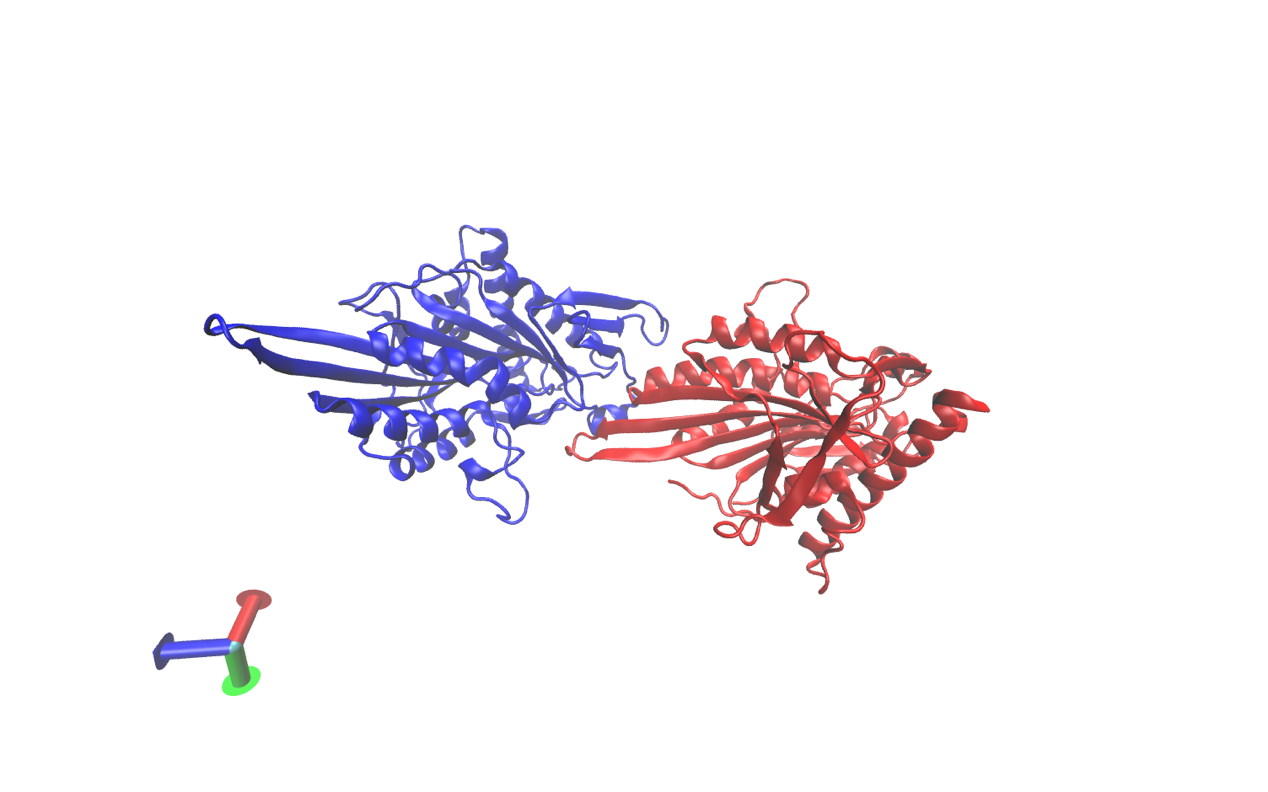
Since the consensus sequence from my alignment has lots of gaps. I chose to use single sequence for blast search. Based on my sequence identity matrix, the highest identity to all other sequence is from *Leptonychotes weddellii.*  So, I proceed blast search using Bio3D with *Leptonychotes weddellii*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

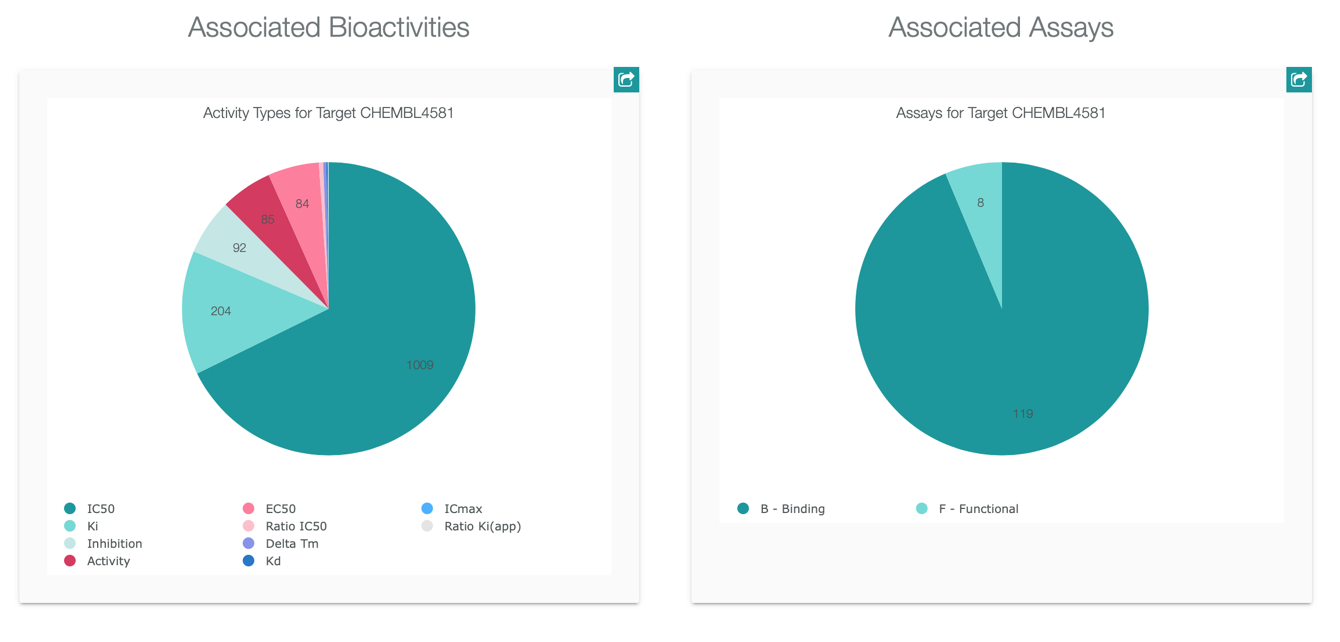
The structure of 3HQD (Human kinesin Eg5 motor domain in complex with AMPPNP and Mg2+) is shown here.



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/)



For example, in the review paper reported by *Visanath et al*. (DOI: [10.1021/jm401071u](http://dx.doi.org/10.1021%2Fjm401071u)), 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.