Name：Tianhao Qiu
PID: A92066783
Email： t6qiu@ucsd.edu
BIMM143 Project

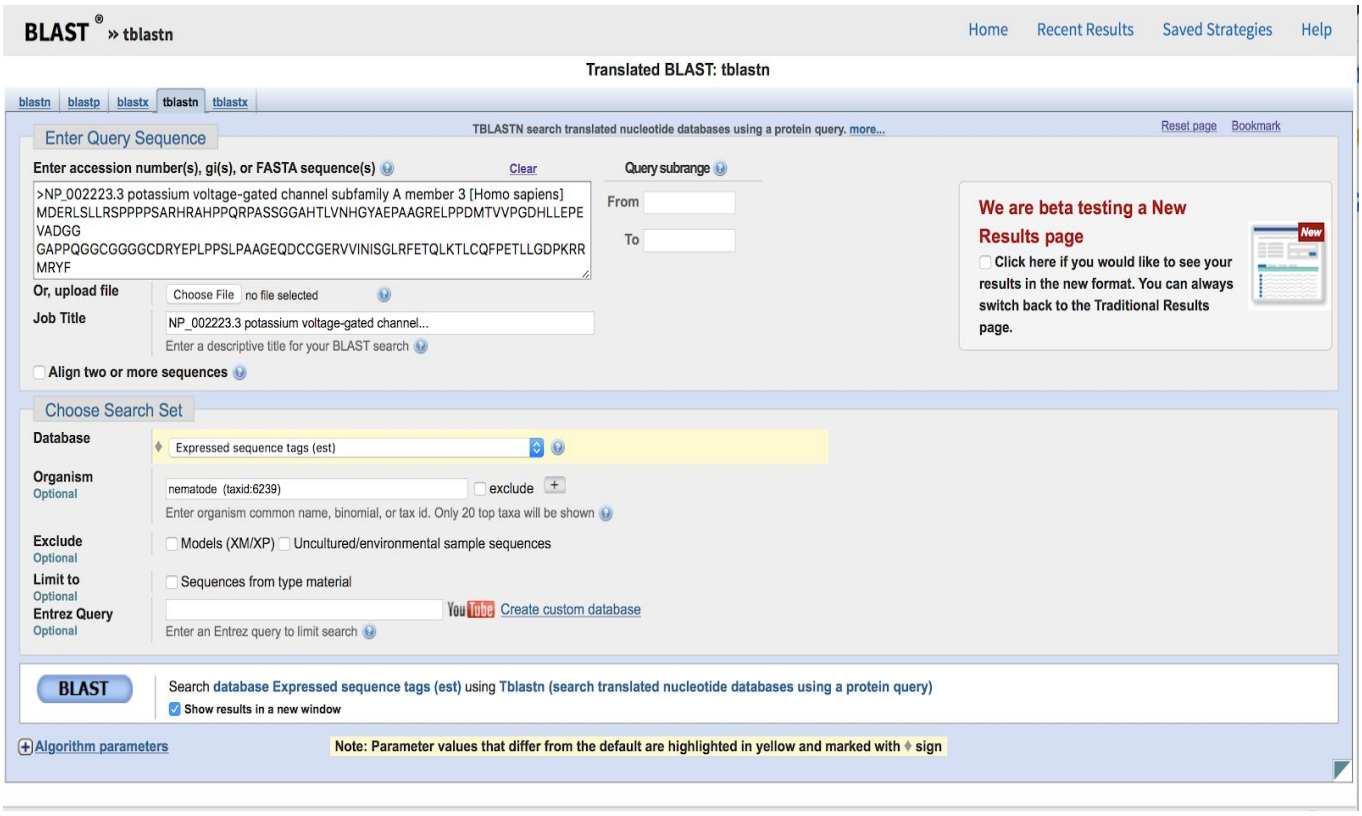
Questions:

[**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 it's function is known.

Name: Potassium voltage-gated channel subfamily A member 3
Accession: `NP_002223.3`
Species: Homo sapiens

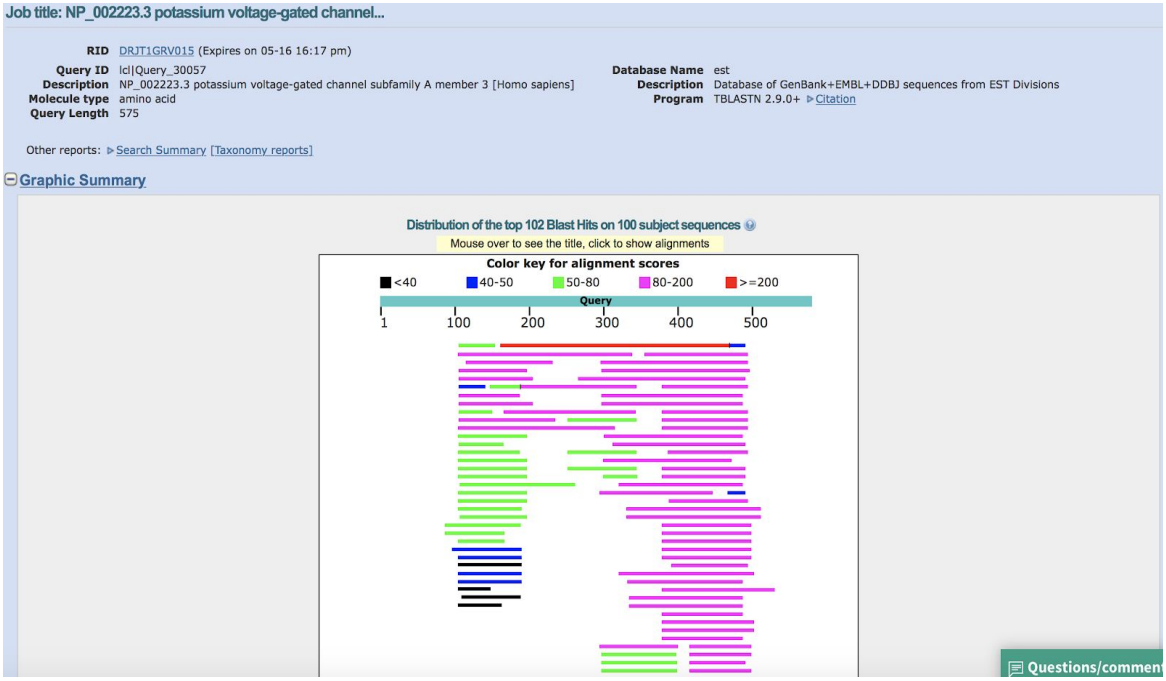
[**Q2**] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

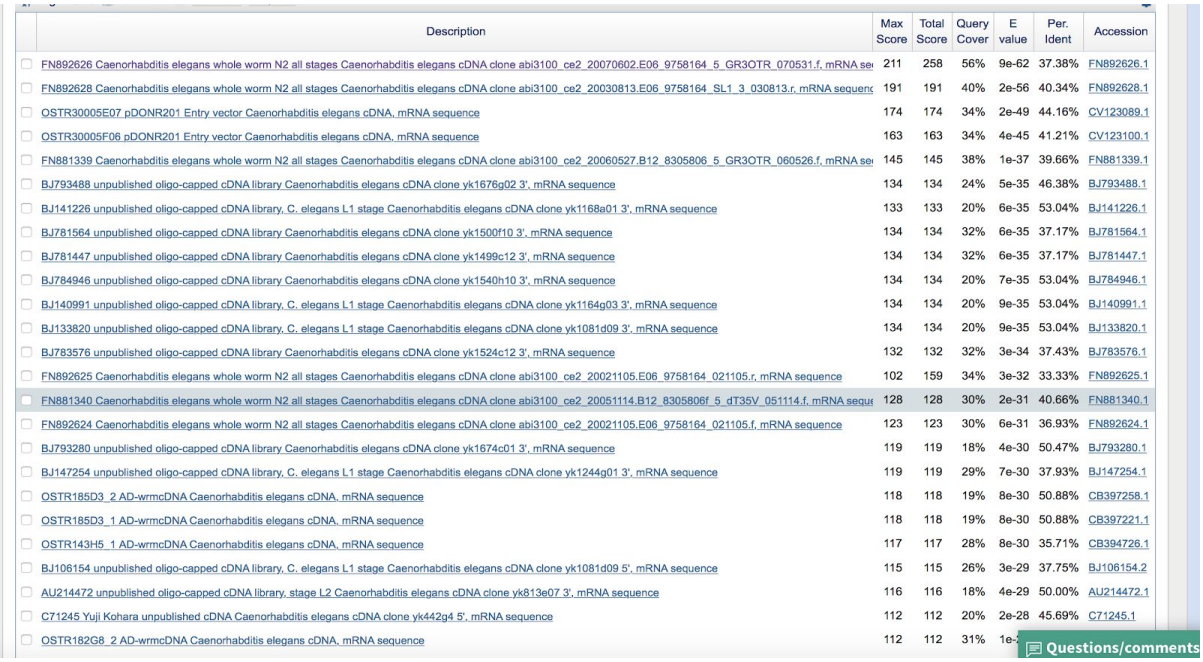
Method : TBLASTN (2.9.0) **Database: Expressed Sequence Tags (est)**
Organism: Nematodes(Taxid: 6239)

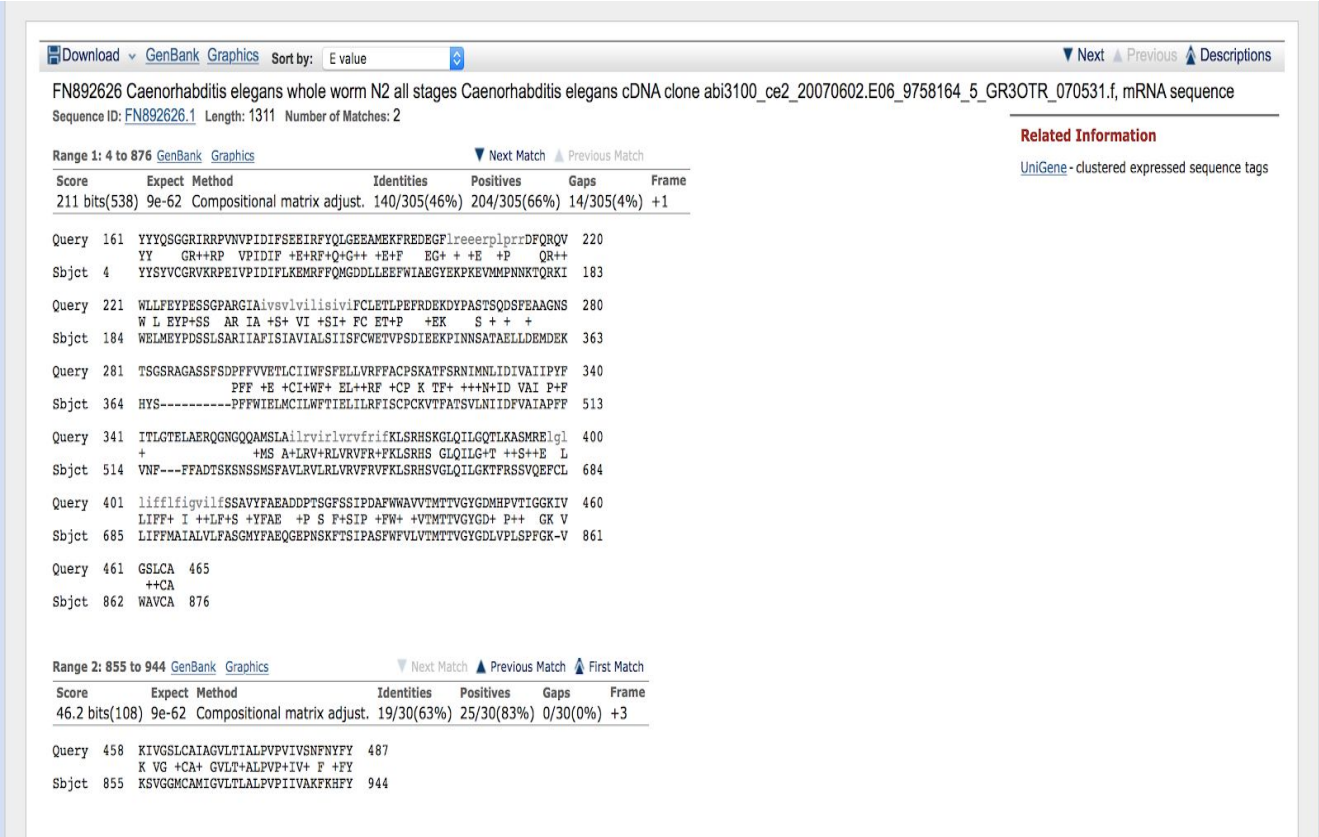


On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or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

Chosen match: Accession FN892626.1, a 1311 base pair clone from *C.Elegans* See below for alignment details.







Download GenBank Graphics Sort by: E value Next Previous Descriptions

FN892626 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

Sequence ID: FN892626.1 Length: 1311 Number of Matches: 2

Related Information

UniGene - clustered expressed sequence tags

Range 1: 4 to 876 GenBank Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
211 bits(538)	9e-62	Compositional matrix adjust.	140/305(46%)	204/305(66%)	14/305(4%)	+1

```
Query  161  YYYQSGGRIRRPVNVPIDIFSEEIRFYQLGEEAMEKFREDEGFlreeerplprrDFQRQV  220
            YY    GR++RP  VPIDIF +E+RF+Q+G++ +E+F   EG+ + +E  +P    QR++
Sbjct  4    YYSYVCGRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKI  183

Query  221  WLLFEYPESSGPARGIAivsvlvilisiviFCLETLPEFRDEKDYPASTSQDSFEAAGNS  280
            W L EYP+SS  AR IA +S+ VI +SI+ FC ET+P   +EK    S + +  +
Sbjct  184  WELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEK  363

Query  281  TSGSRAGASSFSDPFFVVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIPYF  340
                      PFF +E +CI+WF+ EL++RF +CP K TF+ +++N+ID VAI P+F
Sbjct  364  HYS----------PFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFF  513

Query  341  ITLGTELAERQGNGQQAMSLAilrvirlvrvfrifKLSRHSKGLQILGQTLKASMRElgl  400
            +                +MS A+LRV+RLVRVFR+FKLSRHS GLQILG+T ++S++E  L
Sbjct  514  VNF---FFADTSKSNSSMSFAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCL  684

Query  401  lifflfigvilfSSAVYFAEADDPTSGFSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIV  460
            LIFF+ I ++LF+S +YFAE  +P S F+SIP +FW+ +VTMTTVGYGD+ P++  GK V
Sbjct  685  LIFFMAIALVLFASGMYFAEQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK-V  861

Query  461  GSLCA  465
              ++CA
Sbjct  862  WAVCA  876
```

Range 2: 855 to 944 GenBank Graphics Next Match Previous Match First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
46.2 bits(108)	9e-62	Compositional matrix adjust.	19/30(63%)	25/30(83%)	0/30(0%)	+3

```
Query  458  KIVGSLCAIAGVLTIALPVPVIVSNFNYFY  487
            K VG +CA+ GVLT+ALPVP+IV+ F +FY
Sbjct  855  KSVGGMCAMIGVLTLALPVPIIVAKFKHFY  944
```

Alignment Details :

FN892626 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

Length = 1311,
Score = 211 bits (538) for Range 1 and 46.2 bits (108) for Range 2
Expect = 9e-62
Method = Compositional matrix adjust.
Identities = 140/305(46%) for Range 1 and 19/30(62%) for Range 2
Positives = 204/305(66%) for Range 1 and 25/30(83%) for Range 2
Gaps = 14/3055(3%) for Range 1 and 0/30(0%) for Range 2
Frame = +1 for Range 1 and +3 for Range 2

Range 1:
Query 161 YYYQSGGRIRRPVNVPIDIFSEEIRFYQLGEEAMEKFREDEGFlreeerplprrDFQRQV 220
YY GR++RP VPIDIF +E+RF+Q+G++ +E+F EG+ + +E +P QR++
Sbjct 4 YYSYVCGRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKI 183

Query 221 WLLFEYPESSGPARGIAivsvlvilisiviFCLETLPEFRDEKDYPASTSQDSFEAAGNS 280
W L EYP+SS AR IA +S+VI+SI+ FC ET+P +EK S + + +
Sbjct 184 WELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEK 363

Query 281 TSGSRAGASSFSDPFFVVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIPYF 340
PFF +E +CI+WF+ EL++RF +CP K TF+ +++N+ID VAI P+F
Sbjct 364 HYS----------PFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFF 513

Query 341 ITLGTELAERQGNGQQAMSLAilrvirlvrvfrifKLSRHSKGLQILGQTLKASMREIgl 400
+ +MS A+LRV+RLVRVFR+FKLSRHS GLQILG+T ++S++E L
Sbjct 514 VNF---FFADTSKSNSSMSFAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCL 684

Query 401 lifflfigvilfSSAVYFAEADDPTSGFSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIV 460
LIFF+ I ++LF+S +YFAE +P S F+SIP +FW+ +VTMTTVGYGD+ P++ GK V
Sbjct 685 LIFFMAIALVLFASGMYFAEQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK-V 861

Query 461 GSLCA 465
++CA
Sbjct 862 WAVCA 876

Range 2:
Query 458 KIVGSLCAIAGVLTIALPVPVIVSNFNYFY 487
K VG +CA+ GVLT+ALPVP+IV+ F +FY
Sbjct 855 KSVGGMCAMIGVLTLALPVPIIVAKFKHFY 944

[**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.

Chosen Sequence (From Emboss TransSeqResult):

>FN892626.1_1 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

HYYSYVCGRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRK

IWELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDE

KHYSPFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTS

KSNSSMSFAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLF

ASGMYFAEQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVWAVCAQ*LVS*HW

HFQFLLLLPNSNTSTDQKNGLASMKSKGDDANNDFS*RGEGL*NKCETERTPFAPKARHP

LGGTFGLLSSIKPFQPLTISRALQLCFPILKVFYFPTPFSPNFFFLKKRGGKTPCGPPGN

SPPPPPPKKILFSPSPP

>FN892626.1_2 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

IIIHMCAEE*NDLKLCRLTYF*KKCDFSKWVMIFWKNSG*LRVMKNQKK**CQIIKPKEK

YGN*WNIRIVHFPRELLHSFQLLLLHYPSSHSAGKQYHLTSKKNQSIIQQLRNYWMKWTK

NIIPHFSGLSSCVFCGLLLN*FFGLFHVLVK*HLPRLC*ILLTLWQLLRFLLTFSSRTQA

NPIHRCRLQCSVCSDWFEYSEFSNLVDIPLVFKFLEKLSDPLSKSFVFSFSSWQLLWFSL

HQECILRSKGSQTQSSLQFQQASGLYWSQ*QQSDMEIWFR*VHLEKCGRYVRNDWCLDIG

TSSSYYCCQIQTLLPTKKTVLLQ*NQRVTMQTTTSHDGVKAFETSVKPKEPHLPRKLAIP

WGEPSDC*APSSPFSPSPSPGPSNFVFQF*KFFIFQPLFPPIFFFLKKGGGKPLVAPPEI

PPPPPLQKKFFSLPLPP

>FN892626.1_3 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

LLFICVRKSETT*NCAD*HIFERNAIFPNG**SSGRILDS*GL*KTKRSDDAK**NPKKN

MGTNGISG*FTFRENYCIHFNCCYCIIHHLILLGNSTI*HRRKTNQ*FSNCGTIG*NGRK

TLFPIFLD*AHVYFVVYY*INSSVYFMSL*SDICHVCVEYY*LCGNCSVFC*LFLRGHKQ

IQFIDVVCSAPCAPIGSSIQSFQT*STFRWSSNSWKNFPILCPRVLSSHFLHGNCSGSLC

IRNVFCGARGAKLKVHFNSSKLLVCTGHNDNSRIWRFGSAESIWKSVGGMCAMIGVLTLA

LPVPIIVAKFKHFYRPKKRSCFNEIKG*RCKQRLLMTG*RPLKQV*NRKNPICPESSPSP

GGNLRIVKLHQALSAPHHLQGPPTLFSNFKSFLFSNPFFPQFFFS*KKGGENPLWPPRKF

PPPPPSKKNSFLSLSPX

>FN892626.1_4 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GGRGREKNFFWRGGGGGISGGATRGFPPPFFKKKKIGGKRGWKIKNF*NWKTKLEGPGDG

EGLKGLDGA*QSEGSPQGMASFRGKWGSFGFTLVSKAFTPS*EVVVCIVTL*FH*SKTVF

LVGRSV*IWQQ**ELEVPMSRHQSLRTYRPHFSKWTQRNQISISDCCHCDQYKPEACWN*

SEL*VWLPLLRKIHS*CKENQSNCHEENEKTKLLDRGSESFSKNLKTNGMSTKFENSEYS

NQSEHTEHCKRHR*IGFACVREEKVNKKRSNCHKVNNIQHRRGKCHFTRT*NKPKN*FNS

KPQNTHELNPEKWGIMFFVHFIQ*FRSC*IIDWFFFDVRWYCFPAE*DDG*CNNSN*NEC

NNSRGK*TIRIFH*FPYFSLGFIIWHHHFFWFFITLSYPEFFQKIITHLEKSHFFQKYVN

RHNFRSFHSSAHI*IIM

>FN892626.1_5 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GERERKEFFLEGGGGGNFRGGHKGFSPPFF*EKKNWGKKGLENKKLLKLENKVGGPWRW*

GAERA*WSLTIRRFPPGDGELSGQMGFFRFHTCFKGLHPVMRSRCLHRHPLISLKQDRFF

GR*KCLNLATIIGTGSANVKTPIIAHIPPTLFQMDSAEPNLHIRLLSL*PVQTRSLLELK

*TLSLAPLAPQNTFLMQREPEQLP*RK*EDKTLGQRIGKFFQEFEDQRNVD*V*KL*ILE

PIGAHGALQTTSMNWICLCPRRKS*QKTEQLPQSQ*YSTQTWQMSLYKDMK*TEELIQ**

TTKYT*AQSRKMGNNVFRPFHPIVPQLLNY*LVFLRCQMVLFPSRMR*WIMQ*QQLK*MQ

***FSRKVNYPDIPLVPIFFFGFYYLASSLLLVFHNPQLSRILPEDHHPFGKIAFLSKICQS**

AQFQVVSLFRTHMNNNX

>FN892626.1_6 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GGEGEKRIFFGGGGGGEFPGGPQGVFPPLFLRKKKLGEKGVGK*KTFKIGKQSWRALEMV

RG*KGLMELNNPKVPPRGWRAFGANGVLSVSHLFQRPSPRHEKSLFASSPFDFIEARPFF

WSVEVFEFGNNNRNWKCQCQDTNHCAHTAHTFPNGLSGTKSPYPTVVIVTSTNQKLAGIE

VNFEFGSPCSAKYIPDAKRTRAIAMKKMRRQNSWTEDRKVFPRI*RPTECRLSLKTLNTR

TNRSTRSTANDIDELDLLVSAKKKLTKNGAIATKSIIFNTDVANVTLQGHEINRRINSIV

NHKIHMSSIQKNGE*CFSSISSNSSAVAELLIGFSSMSDGTVSQQNEMMDNAITAIEMNA

IILAESELSGYSISSHIFLWVLLFGIITSFGFS*PSAIQNSSRRSSPIWKNRISFKNMSI
GTISGRFTLPHTYEX**

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 of Protein: Not found

Species Caenorhabditis elegans

Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Rhabditina; Rhabditomorpha; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

[**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.

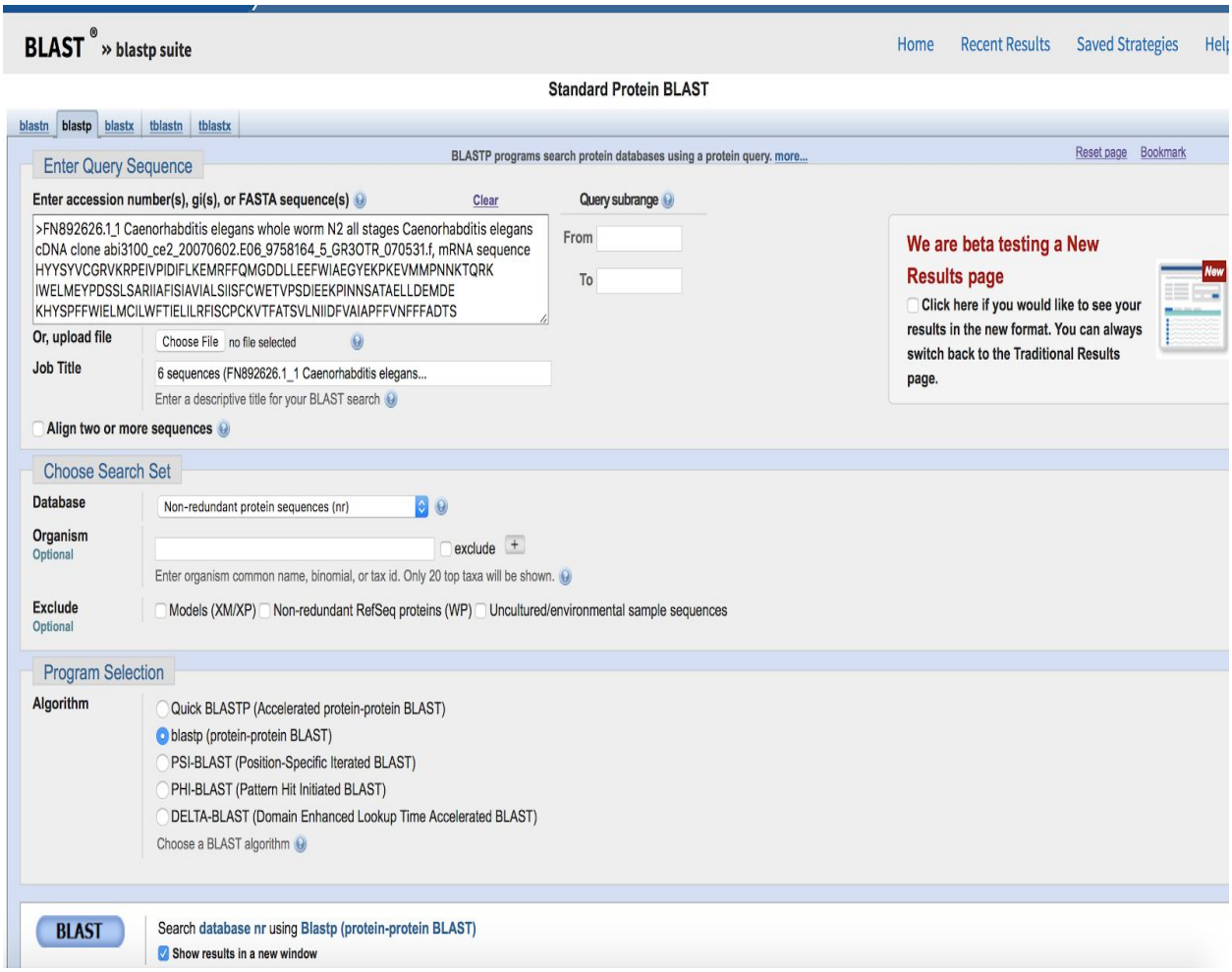
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.

- 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.

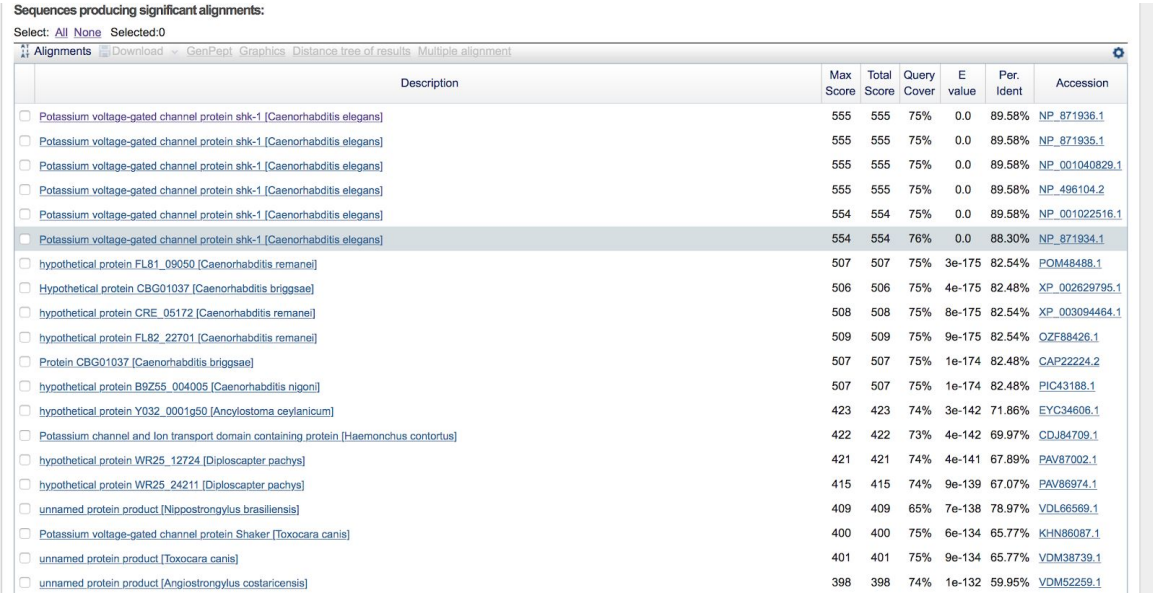
Details:

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

See additional screenshots below for top hits and selected alignment details



The top result is to a protein from Caenorhabditis elegans with identity lower than 100%, see second screenshot below for alignment details:

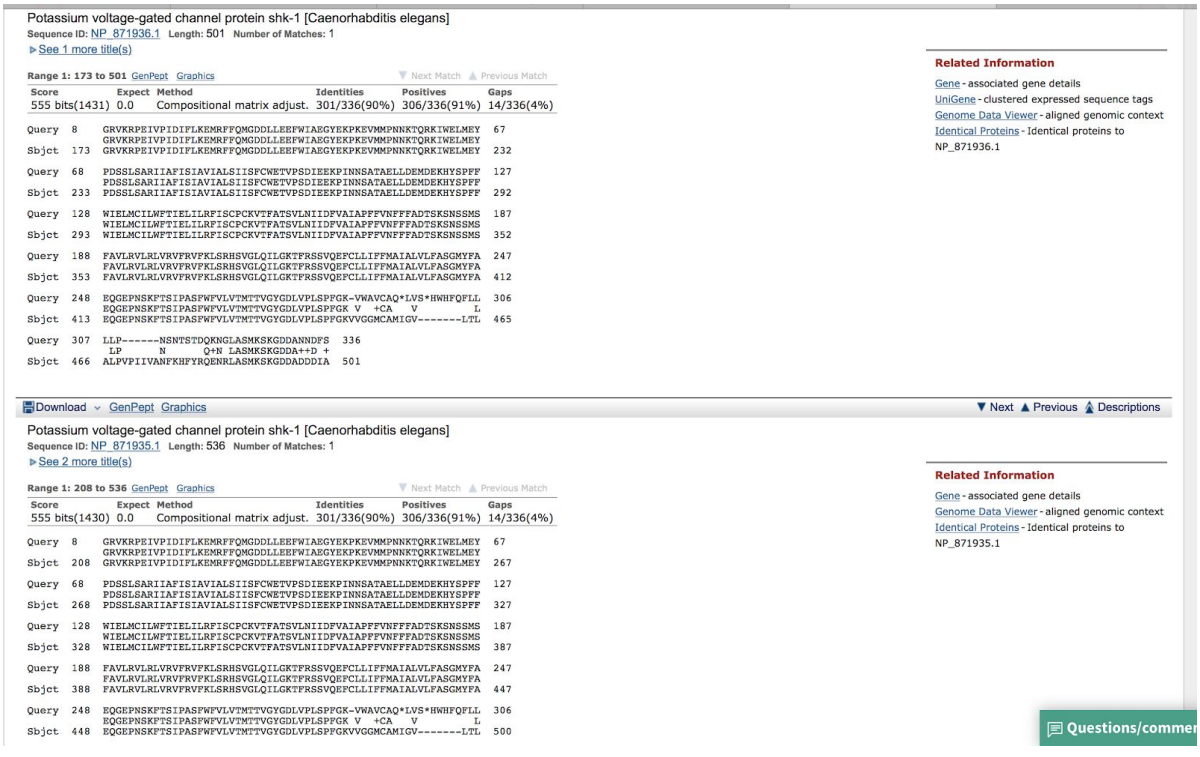


Sequences producing significant alignments:

Select: All None Selected:0

Alignments | Download | GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_871936.1
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_871935.1
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_001040829.1
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_496104.2
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	554	554	75%	0.0	89.58%	NP_001022516.1
Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	554	554	76%	0.0	88.30%	NP_871934.1
hypothetical protein FL81_09050 [Caenorhabditis remanei]	507	507	75%	3e-175	82.54%	POM48488.1
Hypothetical protein CBG01037 [Caenorhabditis briggsae]	506	506	75%	4e-175	82.48%	XP_002629795.1
hypothetical protein CRE_05172 [Caenorhabditis remanei]	508	508	75%	8e-175	82.54%	XP_003094464.1
hypothetical protein FL82_22701 [Caenorhabditis remanei]	509	509	75%	9e-175	82.54%	OZF88426.1
Protein CBG01037 [Caenorhabditis briggsae]	507	507	75%	1e-174	82.48%	CAP22224.2
hypothetical protein B9Z55_004005 [Caenorhabditis nigoni]	507	507	75%	1e-174	82.48%	PIC43188.1
hypothetical protein Y032_0001g50 [Ancylostoma ceylanicum]	423	423	74%	3e-142	71.86%	EYC34606.1
Potassium channel and Ion transport domain containing protein [Haemonchus contortus]	422	422	73%	4e-142	69.97%	CDJ84709.1
hypothetical protein WR25_12724 [Diploscapter pachys]	421	421	74%	4e-141	67.89%	PAV87002.1
hypothetical protein WR25_24211 [Diploscapter pachys]	415	415	74%	9e-139	67.07%	PAV86974.1
unnamed protein product [Nippostrongylus brasiliensis]	409	409	65%	7e-138	78.97%	VDL66569.1
Potassium voltage-gated channel protein Shaker [Toxocara canis]	400	400	75%	6e-134	65.77%	KHN86087.1
unnamed protein product [Toxocara canis]	401	401	75%	9e-134	65.77%	VDM38739.1
unnamed protein product [Angiostrongylus costaricensis]	398	398	74%	1e-132	59.95%	VDM52259.1



Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]

Sequence ID: NP_871936.1 Length: 501 Number of Matches: 1

See 1 more title(s)

Range 1: 173 to 501 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
555 bits(1431)	0.0	Compositional matrix adjust.	301/336(90%)	306/336(91%)	14/336(4%)

```
Query  8    GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY  67
            GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY
Sbjct  173  GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY  232

Query  68   PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF  127
            PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF
Sbjct  233  PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF  292

Query  128  WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS  187
            WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS
Sbjct  293  WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS  352

Query  188  FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA  247
            FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA
Sbjct  353  FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA  412

Query  248  EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK-VWAVCAQ*LVS*HWHFQFLL  306
            EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK V  +CA   V          L
Sbjct  413  EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVVGGMCAMIGV-------LTL  465

Query  307  LLP------NSNTSTDQKNGLASMKSKGDDANNDFS  336
              LP      N       Q+N LASMKSKGDDA++D +
Sbjct  466  ALPVPIIVANFKHFYRQENRLASMKSKGDDADDDIA  501
```

Related Information

Gene - associated gene details
UniGene - clustered expressed sequence tags
Genome Data Viewer - aligned genomic context
Identical Proteins - Identical proteins to NP_871936.1

Download GenPept Graphics ▼ Next ▲ Previous Descriptions

Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]

Sequence ID: NP_871935.1 Length: 536 Number of Matches: 1

See 2 more title(s)

Range 1: 208 to 536 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
555 bits(1430)	0.0	Compositional matrix adjust.	301/336(90%)	306/336(91%)	14/336(4%)

```
Query  8    GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY  67
            GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY
Sbjct  208  GRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEY  267

Query  68   PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF  127
            PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF
Sbjct  268  PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEKHYSPFF  327

Query  128  WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS  187
            WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS
Sbjct  328  WIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSSMS  387

Query  188  FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA  247
            FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA
Sbjct  388  FAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFA  447

Query  248  EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK-VWAVCAQ*LVS*HWHFQFLL  306
            EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGK V  +CA   V          L
Sbjct  448  EQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVVGGMCAMIGV-------LTL  500
```

Related Information

Gene - associated gene details
Genome Data Viewer - aligned genomic context
Identical Proteins - Identical proteins to NP_871935.1

Questions/commer

[**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.

Potassium channel and Ion transport domain containing protein [Haemonchus contortus]	422	422	73%	4e-142	69.97%	CDJ84709.1
hypothetical protein WR25_24211 [Diploscapter pachys]	415	415	74%	9e-139	67.07%	PAV86974.1
unnamed protein product [Nippostrongylus brasiliensis]	409	409	65%	7e-138	78.97%	VDL66569.1
Potassium voltage-gated channel protein Shaker [Toxocara canis]	400	400	75%	6e-134	65.77%	KHN86087.1
unnamed protein product [Toxocara canis]	401	401	75%	9e-134	65.77%	VDM38739.1
unnamed protein product [Angiostrongylus costaricensis]	398	398	74%	1e-132	59.95%	VDM52259.1
transporter, cation channel family protein [Ancylostoma caninum]	388	388	70%	3e-130	70.35%	RCN37415.1
unnamed protein product [Anisakis simplex]	395	395	65%	1e-129	73.43%	VDK46358.1
unnamed protein product [Enterobius vermicularis]	380	380	64%	7e-127	70.71%	VDD95068.1
unnamed protein product [Thelazia callipaeda]	379	379	65%	2e-124	71.43%	VDN08188.1
K+ channel tetramerization domain protein [Onchocerca flexuosa]	368	368	66%	2e-120	67.58%	OZC11969.1
unnamed protein product [Onchocerca flexuosa]	368	368	66%	3e-120	67.92%	VDP12905.1
unnamed protein product [Haemonchus placei]	355	355	73%	1e-116	63.66%	VDO32825.1
unnamed protein product [Onchocerca ochengi]	353	353	66%	2e-115	68.94%	VDK62583.1
unnamed protein product [Brugia timori]	346	346	66%	2e-112	69.62%	VDO32192.1
BMA-SHK-1, isoform e [Brugia malayi]	346	346	66%	3e-112	69.62%	CDP90796.1
unnamed protein product [Wuchereria bancrofti]	342	342	62%	7e-112	70.80%	VDM12676.1
BMA-SHK-1, isoform g [Brugia malayi]	347	347	66%	9e-112	69.62%	CDP90793.1
Potassium voltage-gated channel protein Shaker [Strongyloides ratti]	346	346	65%	7e-111	62.50%	XP_024509009.1
hypothetical protein LOAG_01893 [Loa loa]	342	342	66%	7e-111	68.94%	XP_020303708.1
unnamed protein product [Dracunculus medinensis]	337	337	62%	9e-110	68.50%	VDN55395.1
BMA-SHK-1, isoform c [Brugia malayi]	340	340	66%	1e-109	68.00%	CDP90795.1

(Figure shows the clicked links corresponding to the different species selected)

```
>Human_K+_Channel_Protein | NP_002223.3 | potassium voltage-gated channel
subfamily A member 3 | [Homo sapiens]
MDERLSLLRSPPPPSARHRAHPPQRPASSGGAHTLVNHGYAEPAAGRELPPDMTVVPGDHLLEPEVADGG
GAPPQGGCGGGGCDRYEPLPPSLPAAGEQDCCGERVVINISGLRFETQLKTLCQFPETLLGDPKRRMRYF
DPLRNEYFFDRNRPSFDAILYYYQSGGRIRRPVNVPIDIFSEEIRFYQLGEEAMEKFREDEGFLREEERP
LPRRDFQRQVWLLFEYPESSGPARGIAIVSVLVILISIVIFCLETLPEFRDEKDYPASTSQDSFEAAGNS
TSGSRAGASSFSDPFFVVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIPYFITLGTELAER
QGNGQQAMSLAILRVIRLVRVFRIFKLSRHSKGLQILGQTLKASMRELGLLIFFLFIGVILFSSAVYFAE
ADDPTSGFSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIVGSLCAIAGVLTIALPVPVIVSNFNYFYHRE
TEGEEQSQYMHVGSCQHLSSSAEELRKARSNSTLSKSEYMVIEEGGMNHSAFPQTPFKTGNSTATCTTNN
NPNSCVNIKKIFTDV
```

>C.Elegans K+_Protein(Shaker-1) | NP_871936.1 | Potassium voltage-gated channel protein shk-1 | [Caenorhabditis elegans]
MIDSLIRRLEQVLNVHPEKYQNGCSSQRPSRQPSPVEQLAPIRSHSADFGRAAEMASGNSEEGKGMLLRN
GDDRIRLPSPQRGTPDTSSTQGHPYTDQIVTINVSGMRFQTFESTLSRYPNSLLGDRNKRQHFFVSDTNE
FFFDRHRTTSSSFTFEIRNYLFESILYIYQSGGRVKRPEIVPIDIFLKEMRFFQMGDDLLEEFWIAEGYE
KPKEVMMPNNKTQRKIWELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELL
DEMDEKHYSPFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFFADTSKSNSS
MSFAVLRVLRLVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFAEQGEPNSK
FTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVVGGMCAMIGVLTLALPVPIIVANFKHFYRQENRLASMK
SKGDDADDDIA

>Barber's_pole_worm K+_Channel_Protein | CDJ84709.1 | Potassium channel and Ion transport domain containing protein |[Haemonchus contortus]
MWRLRRYRGPPILDLDEANGNLLVDDSTRNASPQRGTPQGDERLRSSDLVTINVSGLRFQTFERTLARFP
NTLLGCKSKRERYYMQDTNEYFFDRHRSTFESILYIYQSGGRVKRPESVPIDIFLREMRFFQMGEQLIEE
FWIAEGYEKPEEAILPKNLSQRRLWELMEYPDSSLAARIIAFISVFVIVISIVSFCWETVPSGDVRDTNG
TTGVTSPDKDEGRLVWNPFFWLELVCIVWFTIELTLRFISCPSKVTFMMSFLNIIDFVAIAPFFVNLIWA
DANKSSGSMSFAVLRVLRLVRVFRIFKLSRHSAGLQILGKTFRASVQEFCLLIFFMAIALVLFSSGMYFA
EQGEANSKFTSIPASFWFVLVTMTTVGYGDLVPLSTQGKIVGSMCALIGVLTLALPVPIIVANFKHFYRQ
ENRLATMRSAGKDLETAESIDDS

>Dog_roundworm K+_Channel_Protein | KHN86087.1 | Potassium voltage-gated channel protein Shaker [Toxocara canis]
MIPEEEAVNVIPPTDGISSNMSENNSIVTINVSGLKFQTFNSTLSRYPETLLGDPVKRARFWNGHDGEYF
FDRHRGSFESILYIYQSHGIVKRPEAVPIDIFLKEMKFFEMTDDILEAFWISEGYEKPKEAKMPQNIHQR
RIWELMEYPDSSLSARILAFVSIIVITVSIVSFCLETLPSLKPADIEGVRDWSNPFLWVEFVCIIWFTIE
LLLRFASCPCKLTFMRSIPNIIDFIAIAPFFVNLVWSDVSKSNSSMSFAVLRVLRLVRVFRIFKLSRHSV
GLQILGKTFRASIQEFCLLIFFMAIALVLFASGMYFAEQGEPDTKFTSIPASFWFVLVTMTTVGYGDLTP
TGVYGKLVGSVCALIGVLTLALPVPIIVANFKHFYRQETRLATMRANADEEDEECSQDSSRT

>Deer_roundworm K+_Channel_Protein | OZC11969.1 | K+ channel tetramerization domain protein, partial [Onchocerca flexuosa]
RQIEHLLNVHPEKYRSRSGQSIETSTSVHLNQMGCHHRSRSVDYLSPMMDASNKEINEEKSNNGSIFPMI
TEEESSNIMYNDTSLTKRRERVTINVSGLRFQTYESTLRRYPLTLLGNSFKRNRFWDPKNKEYFFDRHRT
SFESILYTYQSGGVVKRPESVPIDMFIKELKFFEMGEELLEKFWISEGYEKPKEIQMPKNVLLRKLWELV
EYPDSSLFARIFALLSIFVISISIISFCLETLPSMKVNPGDVRDWSNPFFHIELFCIIWFTTELLLRFIS
CPNKLSFLRSALNIIDFIAIAPFFGNLLWMDSAKSSMSFAILRVLRLVRVFRIFKLSRHSVGLQILGKTF
RASIQEFCLLIFFMIIALVLFSSGIYFAEQNEPNTKFTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVG
GLCALIGVLTLALPVPIIVANFKHFYRQETRLAQMRASVEDDETASEKSSKSP

>Rat_roundworm K+Channel Protein | XP_024509009.1 | Potassium voltage-gated channel protein Shaker [Strongyloides ratti]
MNKKTFSCEGNDDISFSFRQKIEQILHVHPEKYRARGGVGSRRSTPGKRGSYDDRLHVRSASEELRQRSL
QSSSSDSPSINDSNNSKQLEELSIPIPVECINSGVNKSNDSEYRNPLTSHHNHIPDNGDIKWKNEEVITI
NVSGLRFQTYESTLARFPNSLLGHPAKRAKYWNNKTKEYFLDHSRYCFESILNIYQTNGDVRRPELVPLE
IYLKELKFYEMNEEVMENFWLSEGYEKPIEHIMPNNFIQKYIWELMEYPDSSVPARVLAFISISVIVTSI
VSFCLETLPDFREASETLTEIGYSKWNNPFLILESLCIVWFSIELFLRFISCPSKSLFMKSFLNIVDFIA
IAPFFINLILSEGNKNNSSTTSFAVLRVLRLVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLVFFMVI
ALVLFSSAIFYVEQGEADTRFTSIPASFWFVVATMTTVGYGDLVPTGTYGKLVGSLCALLGVLTLALPVP
IIVANFKHFYRQECRLAAMRVSREIDDASRYNESDEEYTT

>Human_Lymphatic_Roundworm Shaker 1 Protein | CDP90795.1 | BMA-SHK-1, isoform c [Brugia malayi]
MDTPSNREVSDEKPNNGSIFPMITEEEPNNIMFNDTSSAKQLERNLITINVSGLRFQTYESTLERYPLTL
LGNSFKRNRFWDPKNEEYFFDRHRTSFESILYAYQSGGIMKRPESVPIDMFVKELKFFEVNPKLQKMGEE
LLEKFWISEGYEKPKEIQMPQNVLQRQLWELVEYPDSSLFARIFALLSIFVISISIISFCLETLPSMKEN
PGDVRDWSNPFFYIELFCIIWFTIELLLRFISCPNKFSFLRSVLNIIDFVAIAPFFGNLMWMDSTKSSSS
MSFAVLRVLRLVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMVIALVLFSSGVYFAEQNEPNTK
FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGGLCALIGVLTLALPVPIIVANFKHFYRQETRLAQMR
ASAEEDETNSEHSSKSP

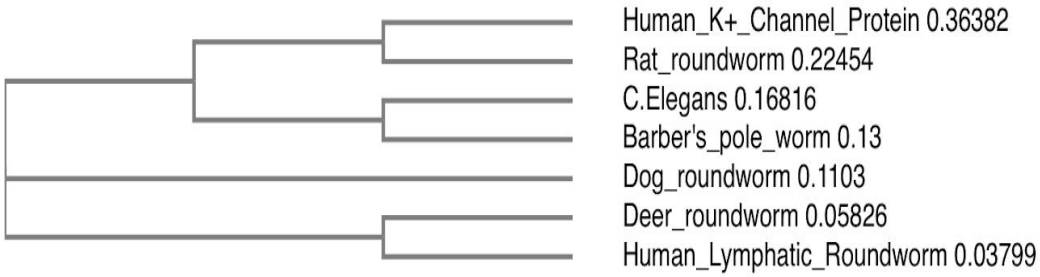
Alignment:
(Obtained using MUSCLE (version 3.8) at EBI:

```
Human_K+_Channel_Protein       -----------------MDERLSLLRSPPPPSARHRAHPPQRPASSGGAHTLVNHGYAEP
Rat_roundworm                  MNKKTFSCEGNDDISFSFRQKIEQILHVHPEKYRARGGVGSRRSTPGKRGSYDDRLHVRS
Dog_roundworm                  ------------------------------------------------------------
Deer_roundworm                 -------------------RQIEHLLNVHPEKYRSRSGQSIETSTSVHLNQMGCHHRSRS
Human_Lymphatic_Roundworm      ------------------------------------------------------------
C.Elegans                      -------------MIDSLIRRLEQVLNVHPEKYQNGCSSQRPSRQPSPVEQLAPIRSHS-
Barber's_pole_worm             ------------------------------------------------------------


Human_K+_Channel_Protein       AAGRELPPDMTVVPGDHLLEPEVADGG-----GAPPQGGCGGGGCDR---------YEPL
Rat_roundworm                  ASEELRQRSLQSSSSDSPSINDSNNSKQLEELSIPIPVECINSGVNKSNDSEYRNPLTSH
Dog_roundworm                  --------------------------------MIPE------EEAVN---------VIPP
Deer_roundworm                 V---DYLSPMMDASNKEINEEKSNNGS-----IFPMITE---EESSN---------IMYN
Human_Lymphatic_Roundworm      ---------MDTPSNREVSDEKPNNGS-----IFPMITE---EEPNN---------IMFN
C.Elegans                      ---------ADFGRAAEMASGNSEEGK-----GMLLRNG---DDRIR---------LPSP
Barber's_pole_worm             ------MWRLRRYRGPPILDLDEANGN-----LLV-------DDSTR---------NASP
                                                                                .


Human_K+_Channel_Protein       PPSLPAAGE---QDCCGERVVINISGLRFETQLKTLCQFPETLLGDPKRRMRYFDPLRNE
Rat_roundworm                  HNHIPDNGD--IKWKNEEVITINVSGLRFQTYESTLARFPNSLLGHPAKRAKYWNNKTKE
Dog_roundworm                  TDGISSNMS-----ENNSIVTINVSGLKFQTFNSTLSRYPETLLGDPVKRARFWNGHDGE
Deer_roundworm                 DTSLTKRRE---------RVTINVSGLRFQTYESTLRRYPLTLLGNSFKRNRFWDPKNKE
Human_Lymphatic_Roundworm      DTSSAKQLE-------RNLITINVSGLRFQTYESTLERYPLTLLGNSFKRNRFWDPKNEE
C.Elegans                      QRGTPDTSSTQGHPYTDQIVTINVSGMRFQTFESTLSRYPNSLLGDRNKRQHFFVSDTNE
Barber's_pole_worm             QRGTPQGDE---RLRSSDLVTINVSGLRFQTFERTLARFPNTLLGCKSKRERYYMQDTNE
                                   .   .          :.**:**:.*:*   ** .:* :***   .* .::     *


Human_K+_Channel_Protein       YFFDRNRPS------------FDAILYYYQSGGRIRRPVNVPIDIFSEEIRFYQ------
Rat_roundworm                  YFLDHSRYC------------FESILNIYQTNGDVRRPELVPLEIYLKELKFYE------
Dog_roundworm                  YFFDRHRGS------------FESILYIYQSHGIVKRPEAVPIDIFLKEMKFFE------
Deer_roundworm                 YFFDRHRTS------------FESILYTYQSGGVVKRPESVPIDMFIKELKFFE------
Human_Lymphatic_Roundworm      YFFDRHRTS------------FESILYAYQSGGIMKRPESVPIDMFVKELKFFEVNPKLQ
C.Elegans                      FFFDRHRTTSSSFTFEIRNYLFESILYIYQSGGRVKRPEIVPIDIFLKEMRFFQ------
Barber's_pole_worm             YFFDRHRST------------FESILYIYQSGGRVKRPESVPIDIFLREMRFFQ------
                               :*:*. *              *::**  **: * :.**  **::::  *:.*::


Human_K+_Channel_Protein       -LGEEAMEKFREDEGFLREEERPLPRRDFQRQVWLLFEYPESSGPARGIAIVSVLVILIS
Rat_roundworm                  -MNEEVMENFWLSEGYEKPIEHIMPNNFIQKYIWELMEYPDSSVPARVLAFISISVIVTS
Dog_roundworm                  -MTDDILEAFWISEGYEKPKEAKMPQNIHQRRIWELMEYPDSSLSARILAFVSIIVITVS
Deer_roundworm                 -MGEELLEKFWISEGYEKPKEIQMPKNVLLRKLWELVEYPDSSLFARIFALLSIFVISIS
Human_Lymphatic_Roundworm      KMGEELLEKFWISEGYEKPKEIQMPQNVLQRQLWELVEYPDSSLFARIFALLSIFVISIS
C.Elegans                      -MGDDLLEEFWIAEGYEKPKEVMMPNNKTQRKIWELMEYPDSSLSARIIAFISIAVIALS
Barber's_pole_worm             -MGEQLIEEFWIAEGYEKPEEAILPKNLSQRRLWELMEYPDSSLAARIIAFISVFVIVIS
                                : :: :* *.  **: .  *  :*..   . :* *.***:**  ** :*::*: **  *


Human_K+_Channel_Protein       IVIFCLETLPEFRDEKDYPASTSQDSFEAAGNSTSGSRAGASSFSDPFFVVETLCIIWFS
```

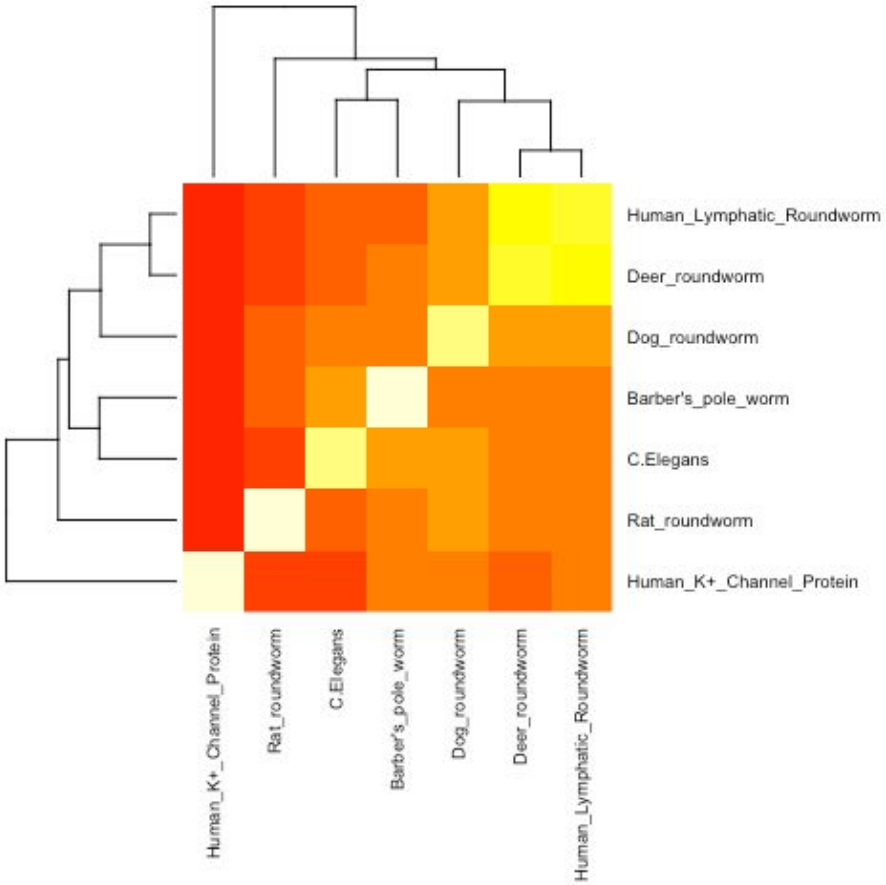
```
Rat_roundworm                IVSFCLETLPDFRE-----------------ASETLTEIGYSKWNNPFLILESLCIVWFS
Dog_roundworm                IVSFCLETLPSLK---------------------PADIEGVRDWSNPFLWVEFVCIIWFT
Deer_roundworm               IISFCLETLPSMKV--------------------NPGD---VRDWSNPFFHIELFCIIWFT
Human_Lymphatic_Roundworm    IISFCLETLPSMKE--------------------NPGD---VRDWSNPFFYIELFCIIWFT
C.Elegans                    IISFCWETVPSDIE---------EKPINNSATAELLDEMDEKHYS-PFFWIELMCILWFT
Barber's_pole_worm           IVSFCWETVPSGDV---------RDTNGTTGVTSPDKDEGRLVWN-PFFWLELVCIVWFT
                             *: ** **:*.                                 :. **: :* .**:**:

Human_K+_Channel_Protein     FELLVRFFACPSKATFSRNIMNLIDIVAIIPYFITLGTELAERQGNGQQAMSLAILRVIR
Rat_roundworm                IELFLRFISCPSKSLFMKSFLNIVDFIAIAPFFINL--ILSEGNKNNSSTTSFAVLRVLR
Dog_roundworm                IELLLRFASCPCKLTFMRSIPNIIDFIAIAPFFVNL--VWSDVSKSNSS-MSFAVLRVLR
Deer_roundworm               TELLLRFISCPNKLSFLRSALNIIDFIAIAPFFGNL--LWMDSAKSS---MSFAILRVLR
Human_Lymphatic_Roundworm    IELLLRFISCPNKFSFLRSVLNIIDFVAIAPFFGNL--MWMDSTKSSSS-MSFAVLRVLR
C.Elegans                    IELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNF--FFADTSKSNSS-MSFAVLRVLR
Barber's_pole_worm           IELTLRFISCPSKVTFMMSFLNIIDFVAIAPFFVNL--IWADANKSSGS-MSFAVLRVLR
                              ** :** :** *  *  .  *::*::** *:* .:     :   ..    *:*:***:*

Human_K+_Channel_Protein     LVRVFRIFKLSRHSKGLQILGQTLKASMRELGLLIFFLFIGVILFSSAVYFAEADDPTSG
Rat_roundworm                LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLVFFMVIALVLFSSAIFYVEQGEADTR
Dog_roundworm                LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMAIALVLFASGMYFAEQGEPDTK
Deer_roundworm               LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMIIALVLFSSGIYFAEQNEPNTK
Human_Lymphatic_Roundworm    LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMVIALVLFSSGVYFAEQNEPNTK
C.Elegans                    LVRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFAEQGEPNSK
Barber's_pole_worm           LVRVFRIFKLSRHSAGLQILGKTFRASVQEFCLLIFFMAIALVLFSSGMYFAEQGEANSK
                             ******:******* ******:*:.:*:.*: **:**: *.::**:*.:::.* .:. :

Human_K+_Channel_Protein     FSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIVGSLCAIAGVLTIALPVPVIVSNFNYFY
Rat_roundworm                FTSIPASFWFVVATMTTVGYGDLVPTGTYGKLVGSLCALLGVLTLALPVPIIVANFKHFY
Dog_roundworm                FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGSVCALIGVLTLALPVPIIVANFKHFY
Deer_roundworm               FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGGLCALIGVLTLALPVPIIVANFKHFY
Human_Lymphatic_Roundworm    FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGGLCALIGVLTLALPVPIIVANFKHFY
C.Elegans                    FTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVVGGMCAMIGVLTLALPVPIIVANFKHFY
Barber's_pole_worm           FTSIPASFWFVLVTMTTVGYGDLVPLSTQGKIVGSMCALIGVLTLALPVPIIVANFKHFY
                             *:*** :**:.:.*********: *    **:**.:**: ****:*****:**:**::**

Human_K+_Channel_Protein     HRETEGEEQSQYMHVGSCQHLSSSAEELRKARSNSTLSKSEYMVIEEGGMNHSAFPQTPF
Rat_roundworm                RQEC------------RLAAMRVSREIDDASRYNESDEEYTT------------------
Dog_roundworm                RQET------------RLATMRANADEEDEECSQDSSRT----------------------
Deer_roundworm               RQET------------RLAQMRASVEDDETASEKSSKSP----------------------
Human_Lymphatic_Roundworm    RQET------------RLAQMRASAEEDETNSEHSSKSP----------------------
C.Elegans                    RQEN------------RLASMKSKGDDADDDIA----------------------------
Barber's_pole_worm           RQEN------------RLATMRSAGKDLETAESIDDS-----------------------
                             ..*               :    .

Human_K+_Channel_Protein     KTGNSTATCTTNNNPNSCVNIKKIFTDV
Rat_roundworm                -----------------------------
Dog_roundworm                -----------------------------
Deer_roundworm               -----------------------------
Human_Lymphatic_Roundworm    -----------------------------
C.Elegans                    -----------------------------
Barber's_pole_worm           -----------------------------
```

[**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” online from the EBI, we got this tree:



[**Q7**] Generate a sequence identity based **heatmap** of your aligned sequences using R.

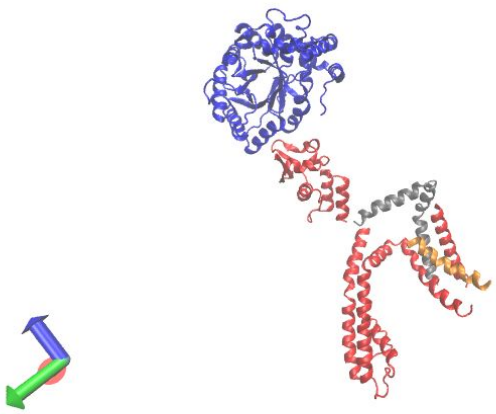


[**Q8**] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

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).

ID	Technique	Resolution	Source	Evalue	Identity
2A79_B	X-RAY DIFFRACTION	2.9	Rattus norvegicus	6.64e-140	51.523
3LUT_B	X-RAY DIFFRACTION	2.9	Rattus norvegicus	7.08e-140	51.523
2R9R_B	X-RAY DIFFRACTION	2.4	Rattus norvegicus	3.98e-134	49.744

[**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?



2A79 is chosen to generate the VMD PDB structures (Rendered in internal Tachyon, saved as tga file and then converted to JPG in order to be put in the google doc). Based on sequence similarity (~50%) and the rule of thumb mentioned in the lecture, since they have more than 45% identity, they should have similar structures.

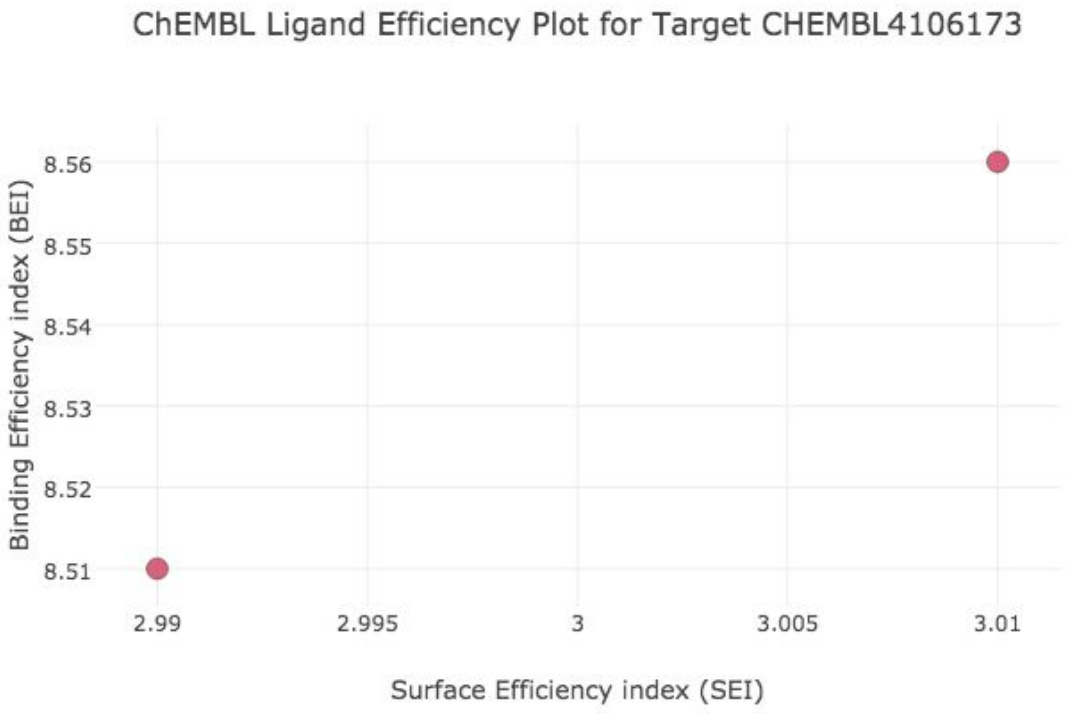
[**Q10**] Perform a “Target” search of ChEMBEL (https://www.ebi.ac.uk/chembl/) with your novel sequence. Are there any **Target Associated Assays** and **ligand efficiency data** reported that may be useful starting points for exploring potential inhibition of your novel protein?

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL4106173/

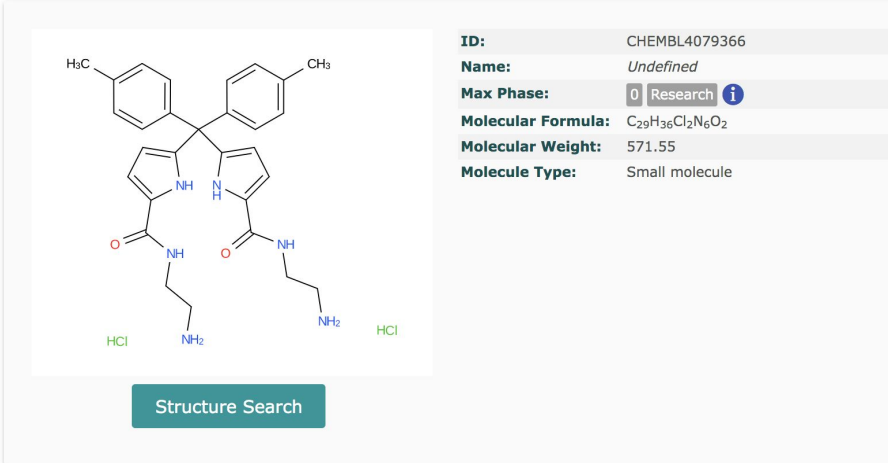
CHEMBL Details 8 Binding Assays

ChEMBL ID	Description	Organism	Compounds	Document	BAO Format	Source
CHEMBL4009564	Inhibition of recombinant rat Kv1.1/1.1/1.2/1.2 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009568	Inhibition of recombinant rat Kv(1.1)3/Kv1.2 expressed in HEK293 cells assessed as time constant for activation of Kv1.1 channel current at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method (Rvb = 3.1 +/- 0.2 ms)	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009563	Inhibition of recombinant rat Kv1.2/1.1/1.2/1.1 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009567	Inhibition of recombinant rat Kv(1.1)2/Kv(1.2)2 expressed in HEK293 cells assessed as time constant for activation of Kv1.1 channel current at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method (Rvb = 2.8 +/- 0.4 ms)	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009561	Inhibition of recombinant rat Kv1.1/Kv(1.2)3 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009560	Inhibition of recombinant rat Kv(1.1)2/Kv(1.2)2 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009565	Inhibition of recombinant rat Kv1.2/1.2/1.1/1.1 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
CHEMBL4009559	Inhibition of recombinant rat Kv(1.1)3/Kv1.2 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature

Link to one of the binding assay linked articles:
https://pubs.acs.org/doi/10.1021/acs.jmedchem.6b01262



Ligand Efficiency session has the following plots. Both points are associated with molecule CHEMBL4079366 (Different standard values) that is possible to inhibit the novel protein.

Information about this molecule:

Name And Classification