

BGGN-213: FOUNDATIONS OF BIOINFORMATICS

The find-a-gene project assignment

Alexander Liu (PID: 69026918)

Q1.

Name: probable ATP-dependent RNA helicase DDX6 Organism: Mus musculus

Accession: NP_851841.2

Species: *Mus musculus*

Q2.

Method: TBLASTN search against nematodes ESTs

Database: Expressed Sequence Tags (est)

Organism: Nematodes (Taxid: 6231)

blastn

blastp

blastx

tblastn

tblastx

Translated BLAST: tblastn

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

ref|NP_851841.2

Query subrange ?

From

To

Or, upload file

Choose File

No file chosen ?

Job Title

NP_851841:probable ATP-dependent RNA helicase...

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

Expressed sequence tags (est) ?

Organism

Optional

nematodes (taxid:6231) ☐ exclude [Add organism](#)

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

Exclude

Optional

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

Limit to

Optional

☐ Sequences from type material

Entrez Query

Optional

Enter an Entrez query to limit search ?

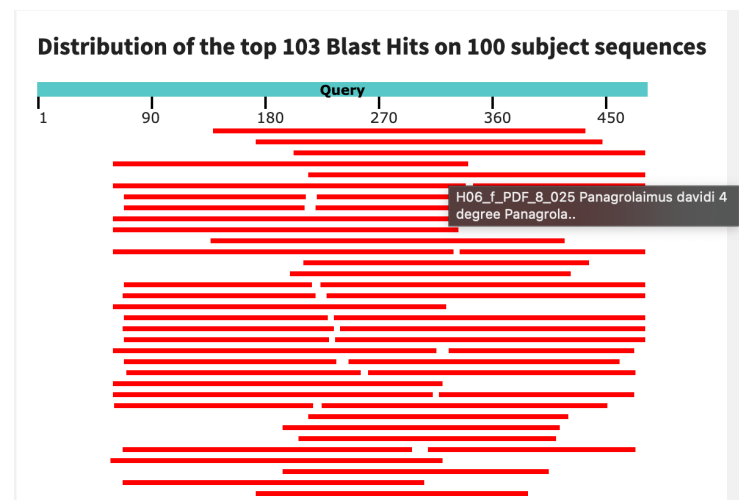
[YouTube](#) [Create custom database](#)

BLAST

Search database est using Tblastn (search translated nucleotide databases using a protein query)

☐ Show results in a new window

Chosen match: Accession JZ594105.1, H06_f_PDF_8_025 Panagrolaimus davidi 4 degree Panagrolaimus davidi cDNA, mRNA sequence (the fourth hit)



	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	Sr_pAMT7_014o23_T7 S. ratti mixed stage pAMP Strongyloides r...	Strongyloi...	469	469	61%	2e-164	73.90%	883	FC819473.1
✓	Sr_pAMT7_016a20_T7 S. ratti mixed stage pAMP Strongyloides r...	Strongyloi...	447	447	56%	9e-156	75.55%	829	FC819862.1
✓	M13_r_PDT_22_051 Panagrolaimus davidi 20 degree Panagrolai...	Panagrola...	407	407	57%	1e-139	67.99%	959	JZ675587.1
✓	H06_f_PDF_8_025 Panagrolaimus davidi 4 degree Panagrolaimu...	Panagrola...	406	406	58%	2e-139	67.62%	865	JZ594105.1
✓	G21_r_PDT_35_089 Panagrolaimus davidi 20 degree Panagrolai...	Panagrola...	389	389	55%	1e-132	67.79%	958	JZ666706.1

Alignment details:

H06_f_PDF_8_025 Panagrolaimus davidi 4 degree Panagrolaimus davidi cDNA, mRNA sequence

Sequence ID: [JZ594105.1](#) Length: 865 Number of Matches: 1

Range 1: 26 to 865 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
406 bits(1043)	2e-139	Compositional matrix adjust.	190/281(68%)	236/281(83%)	1/281(0%)	+2
Query 61	SMAATIKPGDDWKKTLKLPKDLRIKTS DVTSTKGNEFEDYCLKRELLMGI					
Sbjct 26	AVEAQNKTNDDWKAQLQLPAKDMRYKTL DVTNTRGLDFEEFSLSRDLLKGI					
Query 121	PIQEESIPIALSGRDILARAKNGTGKSGAYLIPLERLDLKKDNIQAMVIV					
Sbjct 206	PIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIV					
Query 181	SQICIQVSKHMGAKVMATTGGTNLRDDIMRLDDTVHVVIATPGRILDLIK					
Sbjct 386	SQICVELGKHIG-LKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLME					
Query 241	QMIVLDEADKLLSQDFVQIMEDIILTLPKNRQILLYSATFPLSVQKFMNSH					
Sbjct 563	KMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESH					
Query 301	MEELTLKGV TQYYAYVTERQKVHCLNTLFSRLQINQSIIFC					341
Sbjct 743	MEELTLGVTQ+YA+VTE+QKVHCLNTLF +LQINQSIIFC					865

Q3.

Chosen sequence:

```
>Panagrolaimus davidi Ddx6 (translated using EMBOSS Transeq, and ORF
extracted)
MAVEAQNKTNDDWKAQLQLPAKDMRYKTL DVTNTRGLDFEEFSLSRDLLKGIFEKGWERPSPIQ
EASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQTSQICVEL
GKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLMEKGVAEVQDCKMLVLDEADKLL
SQDFQGVLDRLVSLPAGRQIMLYSATFPRTVASFMESHMKNPYEINLMEELTLLGVTQFYAFV
TEKQKVHCLNTLFRKLQINQSIIFC
```

Name: *Panagrolaimus davidi* Ddx6

Species: *Panagrolaimus davidi*

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Nematoda;
Chromadorea; Rhabditida; Tylenchina; Panagrolaimomorpha; Panagrolaimoidea;
Panagrolaimidae; Panagrolaimus

Q4.

A BLASTP search against NR database yielded a top hit result is to a protein from
Halicephalobus sp. NKZ332.

blastn

blastp

blastx

tblastn

tblastx

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more...

Reset page

Bookmark

Enter Query Sequence

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

IIVPTRELALQTSOICVELGKHIGLVKVMVTTGGTDLRDDILRLNSAVHLIVATPGRII
DLMEKGVAEVQDCKMLVLDEADKLLSQDFOGVLDRLVSFLPAGROIIMLYSATFP
RTVASFMESHMKNPYEINLMEELTLGVTQFYAFVTEKQKVHCLNTLFRKLQINQ
SIIFC

Query subrange [?](#)
From
To

Or, upload file No file chosen [?](#)

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

☐ Align two or more sequences [?](#)

Choose Search Set

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

[< Try experimental clustered nr database](#) [?](#)
For more info see [What is clustered nr?](#)

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

Standard

Database ?

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident
<input checked="" type="checkbox"/>	hypothetical protein FO519_002541 [Halicephalobus sp. NKZ332]	Halicephal...	497	497	100%	8e-174	85.41%
<input checked="" type="checkbox"/>	unnamed protein product [Meloidogyne enterolobii]	Meloidogyn...	483	483	96%	4e-168	82.66%
<input checked="" type="checkbox"/>	hypothetical protein Mgra_00002596 [Meloidogyne graminicola]	Meloidogyn...	479	479	96%	1e-166	82.29%
<input checked="" type="checkbox"/>	unnamed protein product [Haemonchus placei]	Haemonch...	474	474	97%	1e-164	80.59%
<input checked="" type="checkbox"/>	DNA RNA helicase domain containing protein [Haemonchus contortus]	Haemonch...	474	474	97%	1e-164	80.59%

hypothetical protein FO519_002541 [Halicephalobus sp. NKZ332]

Sequence ID: [KAE9554242.1](#) Length: 426 Number of Matches: 1

Range 1: 1 to 281 GenPept Graphics						<div>▼ Next Match ▲ Previous Match</div>		Related Information	
Score	Expect	Method	Identities	Positives	Gaps				
497 bits(1280)	8e-174	Compositional matrix adjust.	240/281(85%)	256/281(91%)	0/281(0%)	AlphaFold Structure - 3D structure displays			
Query	1	MAVEAQNKTNDDWKAQLQLPAKDMRYKTL DVTNTRGLDFEEFSLSRDLLKG							
Sbjct	1	M+ E ++ DWKA L LPAKD R+KT DVTNTRGL+FE+FSLSRDLLKG							
		MSTEIATHSDADWKASLNLPKDG RFKTQDVTNTRGLEFEDFSLSRDLLKG							
Query	61	SPIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALII							
Sbjct	61	SPIQEASIGIALSG DILARAKNGTGKTGAYCIPVI R+DTSK+ IQALI+							
		SPIQEASIGIALSGHDILARAKNGTGKTGAYCIPVIGRIDTSKNEIQALIV							
Query	121	TSQICVELGKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLME							
Sbjct	121	TSQIC+ELGKH+ LKVMVTTGGTDLRDDILRLNS VHLIVATPGRI+DLME							
		TSQICIELGKHMHLKVMVTTGGTDLRDDILRLNSVHLIVATPGRIIDLME							
Query	181	KMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESH							
Sbjct	181	KMLVLDEADKLLSQDFQGVLDRLV LP GRQIMLYSATFPRTVASFM+ H							
		KMLVLDEADKLLSQDFQGVLDRLVDRLPLGRQIMLYSATFPRTVASFMQKH							
Query	241	MEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC			281				
Sbjct	241	MEELTLLGVTQFYAFV E+QKVHCLNTLFRKLQINQSIIFC			281				
		MEELTLLGVTQFYAFVQERQKVHCLNTLFRKLQINQSIIFC			281				

Q5.

Re-labeled sequences

```
>House_mouse_DDX6 | probable ATP-dependent RNA helicase DDX6 [Mus musculus] |
NP_851841.2
MSTARTENPVIMGLSSQNGQLRGPVKASAGPGGGGTQPPQQLNQLKNTSTINNGTPQQAQSMAATIKPGD
DWKKTLLKLPKDLRIKTS DVTSTKGNEFEDYCLKRELLMGIFEMGWEEKPSPIQEESIPIALSGRDILARA
KNGTGKSGAYLIPLLERLDLKKDNIQAMVIVPTRELALQVSQICIQVSKHMGGAQVMATTGGTNLRDDIM
RLDDTVHVVIATPGRIIDLIIKKGVAQVDHVQMIVLDEADKLLSQDFVQIMEDIILTLPKNRQIILYSATF
PLSVQKFMNSHLQKPYEINLMEELTLKGVTQYYAYVTERQKVHCLNTLFSRLQINQSIIFCNSSQRVELL
AKKISQLGYSCFYIHAKMRQEHRNRVFDHFRNGLCRNLVCTDLFTRGIDIQAVNVVINFDFPKLAETYLH
RIGRSGRFGHLGLAINLITYDDRFLNKSIEEQLGTEIKPIPSNIDKSLYVAEYHSEPAEDEKP

>Panagrolaimus_davidi_Ddx6 | (translated using EMBOSS Transeq, and ORF
extracted)
MAVEAQNKTNDDWKAQLQLPAKDMRYKTL DVTNTRGLDFEEFSLSRDLLKGIFEKGWERPSPIQEASIGIALSGQDI
LARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQTSQICVELGKHIGLKVMVTTGGTDLRDDILRLNS
AVHLIVATPGRIIDLMEKGVAEVQDCKMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESH
MKNPYEINLMEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC

>African_eye_worm | ATP-dependent RNA helicase cgh-1 [Loa loa] |
XP_020303523.1
MTSKTESQSDQSWKENLNLPKDLRFKTTDVTDTKGIEFEDFCLKRPLLMGIFEKGWEKPSPIQEASISI
ALSGQDILARAKNGTGKTGAYCIPCIDKVDVEKKYPQSLIIVPTRELAFQTSNICVDLSKHMNLKVMVTT
GGTEL RNDIMRLNGTVHIIIVATPGRIIDLMDKNVADMSHCTTLVLDEADKLLSQDFQGILDRVIKFLPPD
RQIMLYSATFPLTVATFMQKHMRNPYEINLMEELTLVGVTQYYAYVQEKQKVHCLNTLFRKLQINQSIIF
```

CNSTQRVELLAKKITEIGYSCYYIHSRMAQGHRNRVFDHFRKGACRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETYLHRIGRSGRFGHLGIAINLITYEDRFTLRRIEAE LRTRI APIPKSVDPKLYVAEYQIVDENGSEMVGSSKETAA

>dog_roundworm | ATP-dependent RNA helicase cgh-1 [Toxocara canis] | KHN80260.1
MSSEAVSHSDHSWKEGLKVPKADMDRYKTADVTDTKGIEFEDFCLKRDLLMGIFEKGWEKPSPIQEASIAIALSGQDILARAKNGTGKTGAYCIPKIEKIDPELKKIQALIIVPTRELALQTSQICVELSKHLKLKVMVTTGGTDLRDDIMRLNGIVHLVVATPGRILDMDKGVADMSRCATLVLDEADKLLSQDFQGILDRVIKFLPPERQIMLYSATFPLTVATFMQKHMNRNPYEINLMEELTLLGVQTQFYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITEIGYSCYYIHSRMAQNHNRNRVFDHFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETYLHRIGRSGRFGHLGIAINLITYEDRFTLRRIEAE LRTHI APIPKSVDPKLYVAEHQLVDENANENNSVGSSSSKPVAVA

>large_cabbage_white | ATP-dependent RNA helicase me31b isoform X2 [Pieris brassicae] | XP_045532384.1
MMTENRISAINHVGNSINNQKGDVDKTVDDMGWKSCLKLPKDRRMKTS DVTDTRGNEFEFECFLKREILMGIFEKGWEKPSPIQEASIPIALSGKDVLARAKNGTGKTGAYCIPVIEQVDPKKDCIQALIVVPTRELALQTSQICIELAKHTDIRVMVTTGGTNLRDDIMRIYQNVQVIIATPGRIIDLMDKQVAKMDQCRMLVLDEADKLLSQDFKGMLDLVISRLPKERQILLFSATFPLSVKQFMEKHLREPYEINLMEELTLKGVTQYYAFVQERQKVHCLNTLFSKLQINQSIIFCNSTQRVELLAKKITELG YCCYYIHARMAQAHRNRVFDHFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPRMAETYLHRIGRSGRFGHLGIAINLITYEDRFALHRIERELGTEIKPIPKVIDPALYVARVSEEDDGADK

>western_flower_thrips | ATP-dependent RNA helicase me31b [Frankliniella occidentalis] | XP_026291730.1
MTEVHSNHVIGGISMNKDLDDGIGWKSCLKIPPKDQRIKTS DVTDTRGNEFEFECFLKRELLMGIFEKGWEKPSPIQEASIPIALSGKDVLARAKNGTGKTGAYSIPVLEQVDPKKDCIQALVIVPTRELALQTSQICIELAKHMDVKVMVTTGGTNLRDDIMRIYQKVHIIIATPGRIIDLMDKNVANMENCRIILVLDEADKLLSQDFKGM LDHVISRLPKERQILLYSATFPLTVKQFMEKHLKEPYEINLMEELTLKGVTQYYAFVQERQKVHCLNTLFSKLQINQSIIFCNSTQRVELLAKKITELG YCCYYIHAKMAQAHRNRVFDHFRNGLCRNLVCSDLFTRGIDVQAVNVVINFDFPKMAETYLHRIGRSGRFGHLGIAINLITFDDR FALHRIEQELGTEIKPIPKVIDPALYVAKSLDDSHAIEEANNVSK

>pink_bollworm | ATP-dependent RNA helicase me31b [Pectinophora gossypiella] | XP_049868573.1
MMTENRISSSNHVGNSMNNQKGDVDKTIDDIGWKSCLKIPAKDRRIKTS DVTDTRGNEFEFECFLKRELLMGIFEKGWEKPSPIQEASIPIALSGKDVLARAKNGTGKTGAYCIPVLEQVDPKKDCIQALVVVPTRELALQTSQICIELAKHTDIRVMVTTGGTNLRDDIMRIYQNVQVIIATPGRMIDLMDKQVAKMDQCRMLVLDEADKLLSQDFKGMLDMVISRLPKERQILLFSATFPLSVKQFMEKHLREPYEINLMEELTLKGVTQYYAFVQERQKVHCLNTLFSKLQINQSIIFCNSTQRVELLAKKITELG YCCYYIHARMAQAHRNRVFDHFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPRMAETYLHRIGRSGRFGHLGIAINLITYEDRFALHRIEQELGTEIKPIPKVIDPALYVARPDEDDGADK

>migratory_locust | DEAD-box DDX6 [Locusta migratoria] | QOS47384.1
MMTETHVNSNHVSLLPVNSINNQKMDMDVTDDKGWKSCLKIPPKDRRIQTS DVTDTRGNEFEFECFLKRELLMGIFEKGWEKPSPIQEASIPIALSGKDVLARAKNGTGKTGAYSIPVLEQVDPKKDCIQALIIVPTRELALQTSQICIELAKHMKVKVMVTTGGTNLRDDIMRIYQTVHVVIATPGRILDMDKHVANMEHCRMLVLDEADKLLSQDFKGMLDHVIAQLPSE RQILLYSATFPLSVKQFMEKHLRKPYEINLMEELTLKGVTQYYAFVQERQKVHCLNTLFSKLQINQSIIFCNSTQRVELLAKKITELG YCCYYIHAKMAQAHRNRVFDHFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPKMAETYLHRIGRSGRFGHLGIAINLITYEDRFALHRIEQELGTEIKPIPKVIDPSLYVATKIEDAQGIDESNVSK

```
>bovine_lungworm | DEAD/DEAH box helicase [Dictyocaulus viviparus] |
KJH46407.1
MSAADAKVTVTDDENWKSSLVIPAKDKRFKTADVDTDTKGVEFEDFCLSRDLLMGIFEKGWEKPSPIQEAS
IGVALSGQDILARAKNGTGKTGAYCIPVIEKINPAIKKIQALVIVPTRELALQTSQICVELSKHIKLIKIM
VTTGGTDLRDDIMRLNGSVHVMVIATPGRILDLMEKGVADMSHCKMLVLDEADKLLSQDFQGILDRLISFL
LKERQIMLYSATFPMTVTTEFMQKHKMRKPYEINLMEELTLLGVTQYYAFVQEKQKVHCLNTLFRKLQINQS
IIFCNSTQRVELLAKKITEIGYSCYYIHSKMAQNHRNRVFDHFRQGNCRNLVCSDDLTRGIDIQAVNVVI
NFDFFRNSETYLHRIGRSGRFGHLGVAINLITFEDRHTFRRIEQELRTQIEPIPKAIDPKLYFFCSSLVLP
CDLLCVLWMNLCVS
```

```
>New_World_hookworm | DEAD/DEAH box helicase [Necator americanus] |
XP_013307512.1
MLEIQDKRFKTAVLIFCPLGDASKIWIRDYQEIDLRTLQDVTDTDTKGVEFEDFCLSRDLLMGIFEKGWEKPSPIQEASIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPAIKKIQALVIVPTRELALQTSQICVELSKHIKLIKIMVTTGGTDLRDDIMRLNGSVHVMVIATPGRILDLMEKGVADMSNCKMLVLDEADKLLSQDFQGILDRLISFLPKERQIMLYSATFPMTVTTEFMQKHKMRKPYEINLMEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITEIGYSCYYIHSKMAQNHRNRVFDHFRQGNCRNLVCSDDLTRGIDIQAVNVVINFDFFRNAETYLHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIVEDEDNNKTAN
```

```
>barber_pole_worm | DNA RNA helicase domain containing protein [Haemonchus contortus] | CDJ84064.1
MSVADTKVAAVTTDENWKAGLVLPKDKRFKTADVDTDTKGVEFEDFCLSRDLLMGIFEKGWEKPSPIQEASIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPSIKKIQALVIVPTRELALQTSQICVELSKHIKLIKIMVTTGGTDLRDDIMRLNGSVHVMVIATPGRILDLMEKGVADMSNCKMLVLDEADKLLSQDFQGILDRLISFLPKERQIMLYSATFPMTVTTEFMQKHKMRKPYEINLMEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITEIGYSCYYIHSKMAQNHRNRVFDHFRQGNCRNLVCSDDLTRGIDIQAVNVVINFDFFRNAETYLHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIVDDEDNNRANAN
```

Alignment:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```
House_mouse_DDX6      MSTARTENPVMIGLSSQNGQLRGPVKASAGPGGGGTQPQPQLNQLKNTSTINNGTPQQAQ
large_cabbage_white   -----MMTENRISAINHV-----GNSINNQ
pink_bollworm         -----MMTENRISSSNHV-----GNSMNNQ
migratory_locust      -----MMTETHVNS-NHVSLPPVNSINNQ
western_flower_thrips -----MTEVHS-NHV-----IGGI
Panagrolaimus_davidi_Ddx6 -----MAVEAQNK-----
bovine_lungworm       -----MSAADAKVTVT-----
New_World_hookworm    -----MLEIQDKRF-----KTAV
barber_pole_worm      -----MSVADTKVAAVT-----
African_eye_worm      -----MTSKTESQ-----
dog_roundworm         -----MSSEAVSH-----
```

```
House_mouse_DDX6      SMAATIK-PGDDWKTKLPPKDLRIKT-SDVTSTKGNEFEDYCLKRELLMGIFEMGWEK
large_cabbage_white   KGDVDKTVDDMGWKSCLKLPPKDRMKT-SDVTDTRGNEFEFCLKREILMGIFEKGWEK
pink_bollworm         KGDVDKTIDDIGWKSCLKIPAKDRRIKT-SDVTDTRGNEFEFCLKRELLMGIFEKGWEK
migratory_locust      KMDMDVT-DDKGWKSCLKIPPKDRRIQT-SDVTDTRGNEFEFCLKRELLMGIFEKGWEK
western_flower_thrips SMNKDLLD-DGIGWKSCLKIPPKDQRIKT-SDVTDTRGNEFEFCLKRELLMGIFEKGWEK
Panagrolaimus_davidi_Ddx6 -----TNDWKAQLQLPAKDMRYKT-LDVNTNRGLDFEEFSLSRDLLKGIFEKGWER
bovine_lungworm       -----TDENWKSSLVIPAKDKRFKT-ADVDTDTKGVEFEDFCLSRDLLMGIFEKGWEK
New_World_hookworm    LIFCPLGDASKIW----IRDYQEIDLRTLQDVTDTKGVEFEDFCLSRDLLMGIFEKGWEK
```

```

-----TDENWKAGLVLPAKDKRFKT-ADVDTKGVEFEDFCLSRDLLMGIFEKGWEK
-----SDQSWKENLNLPAKDLRFKT-TDVDTKGIEFEDFCLKRPLLMGIFEKGWEK
-----SDHSWKEGLVPKADMRYKT-ADVDTKGIEFEDFCLKRDLLMGIFEKGWEK
          *           :      *          * * * * : : : : * * * * *

```

[illegible][illegible]

HVQMIVLDEADKLLSQDFVQIMEDIILTLFKNRQILLYSATFPLSVQKFMNSHLQKPYEI
 QCRMVLVLEADKLLSQDFKGMLDVISRLPKERQILLFSATFPLSVQKFMEKHLREPYEI
 QCRMVLVLEADKLLSQDFKGMLDMVISRLPKERQILLFSATFPLSVQKFMEKHLREPYEI
 HCRMVLVLEADKLLSQDFKGMLDHVIAQLPSERQILLYSATFPLSVQKFMEKHLRKPYEI
 NCRILVLEADKLLSQDFKGMLDHVISRLPKERQILLYSATFPLTVQKFMEKHLKEPYEI
 DCKMVLVLEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPFTVTFEQMSHMKNPYEI
 HCKMVLVLEADKLLSQDFGILDRILSLFLKPKERQIMLYSATFPMTVSFQMCHKRPYEI
 NCKMVLVLEADKLLSQDFQGILDRILSILPKERQIMLYSATFPMTVTVEFMQCHKRPYEI
 NCKMVLVLEADKLLSQDFQGILDRILSFLPKERQIMLYSATFPMTVTVEFMQCHKRPYEI
 HCTTVLLEADKLLSQDFQGILDRVIFKLPPDRQIMLYSATFPLTVATFMQCHKMKNPYEI
 RCATVLVLEADKLLSQDFGILDRVIKFLPPERQIMLYSATFPLTVATFMQCHKMKNPYEI

[illegible]

YSCFYIHAKMRQEHNRNVFHDFRNGLCRNLVCTDLFTRGIDIQAVNVVINFDFPKLAETY
YCCYYIHARMAQAHNRNVFHDFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPRMAETY
YCCYYIHARMAQAHNRNVFHDFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPRMAETY
YCCYYIHAKMAQAHNRNVFHDFRAGLCRNLVCSDLFTRGIDVQAVNVVINFDFPKMAETY
YCCYYIHAKMAQAHNRNVFHDFRNGLCRNLVCSDLFTRGIDVQAVNVVINFDFPKMAETY

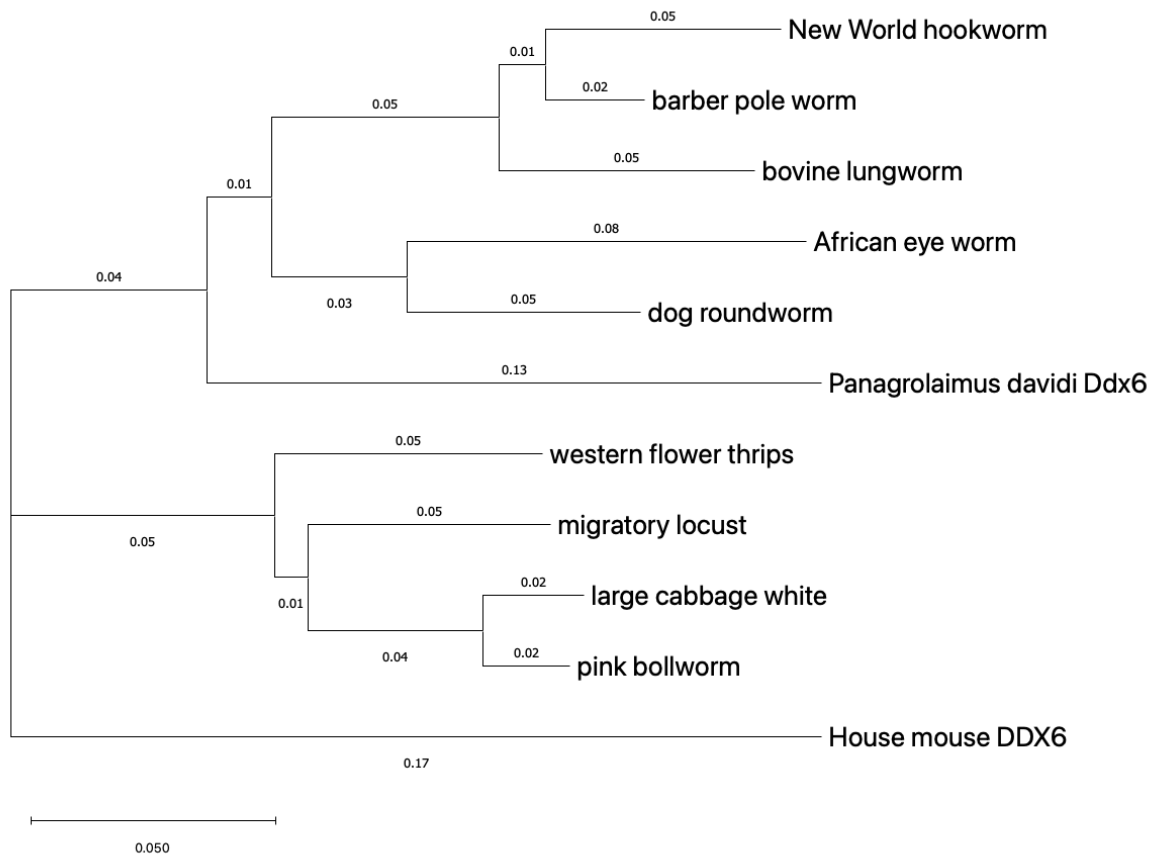
YSCYYIHSKMAQNHRNVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNSETY
YSCYYIHSKMAQNHRNVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETY
YSCYYIHSKMAQNHRNVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETY
YSCYYIHSRMAQGHNRNVFHDFRKGACRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETY
YSCYYIHSRMAQGHNRNVFHDFRKGACRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETY

LHRIGRSGRFGHLGLAINLITYDDRNLKSI EEQLGTEIKPIPSNIDKSLYVAEYHSEPA

large_cabbage_white	LHRIGRSGRFGHLGIAINLITYEDRFALHRIEQLGTEIKPIPKVIDPALYVARVSEEDD
pink_bollworm	LHRIGRSGRFGHLGIAINLITYEDRFALHRIEQLGTEIKPIPKVIDPALYVA-RPDEDD
migratory_locust	LHRIGRSGRFGHLGIAINLITYEDRFALHRIEQLGTEIKPIPKVIDPSLYVATKIEDAQ
western_flower_thrips	LHRIGRSGRFGHLGIAINLITFDDRFALHRIEQLGTEIKPIPKVIDPALYVAKSLDDSH
Panagrolaimus_davidi_Ddx6	-----
bovine_lungworm	LHRIGRSGRFGHLGVAINLITFEDRHTFRRIEQLRTQIEPIPKAIDPKLYFFCSSVLP
New_World_hookworm	LHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQLRTQIEPIPKAVDPKLYVADFQIV
barber_pole_worm	LHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQLRTQIEPIPKAVDPKLYVADFQIVDD
African_eye_worm	LHRIGRSGRFGHLGIAINLITYEDRFTLRRIEAEELRTIAPIPKSVDPKLYVAEQIVDE
dog_roundworm	LHRIGRSGRFGHLGIAINLITYEDRFTLRRIEAEELRTHIAPIPKSVDPKLYVAEQIVDE
House_mouse_DDX6	EDEKP-----
large_cabbage_white	GADK-----
pink_bollworm	GADK-----
migratory_locust	GIDESNVSK-----
western_flower_thrips	AIEEANNVSK-----
Panagrolaimus_davidi_Ddx6	-----
bovine_lungworm	DLLCVLWMNLCVS-----
New_World_hookworm	EDNNKTNAN-----
barber_pole_worm	EDNNRANAN-----
African_eye_worm	NGSEMVGSSKETAA-----
dog_roundworm	NAENENSVGSSSKPVAVA

Q6.

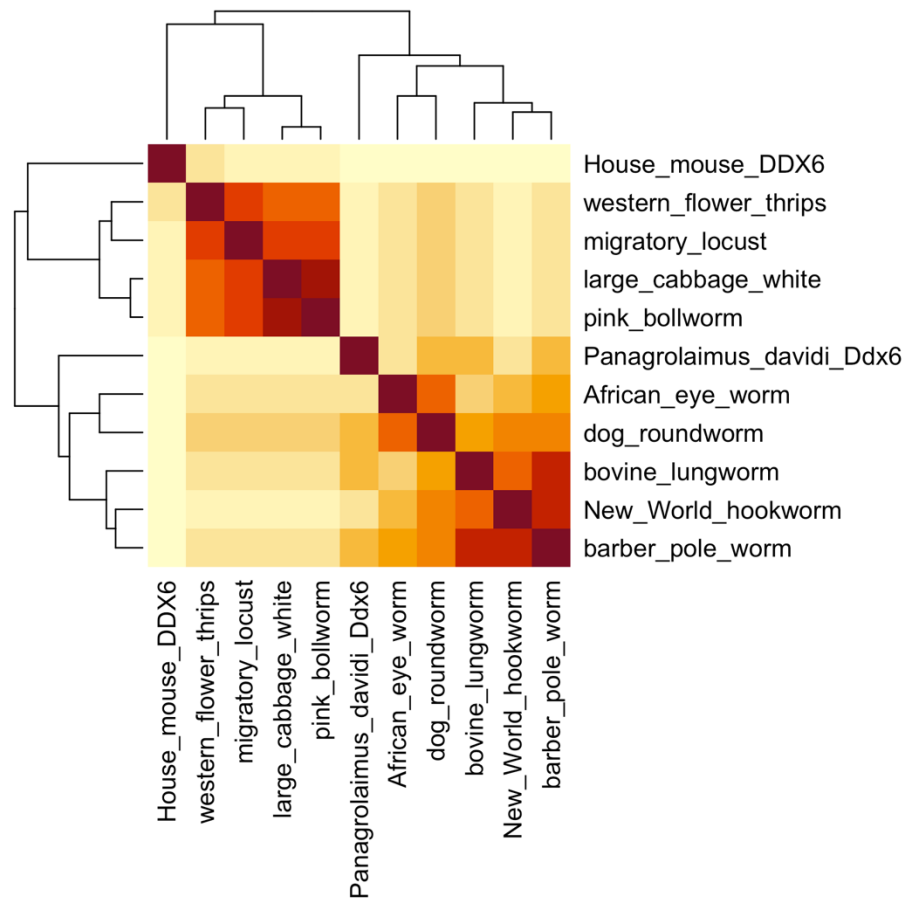
Phylogenetic tree (created by MEGA11):



Analysis summary:

Name	
Application Information	
Version	11.0.13
Build	11220621-x86_64
User Interface	Graphical
Operating System	mac
Data Information	
Data Type	amino acid
No. of Taxa	11
No. of Sites	499
Data File	
Analysis Settings	
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	500
Substitutions Type	Amino acid
Model/Method	p=distance
Rates among Sites	Uniform Rates
Pattern among Lineages	Same (Homogeneous)
Gaps/Missing Data Treatment	Pairwise deletion
Number of Threads	8
Analysis Statistics	
Start Time	7-12-23 14:01:42
End Time	7-12-23 14:01:43
Execution Time	1.815 (seconds)
Peak Memory Used(Working Set)	Not Available
Sum of branch lengths	0.919

Q7.



Q8.

The consensus sequence of the 11 proteins of interest contained too many gaps. Among the 11 proteins of interest, the one with the highest average similarity to all the others is calculated to be “barber pole worm” protein.

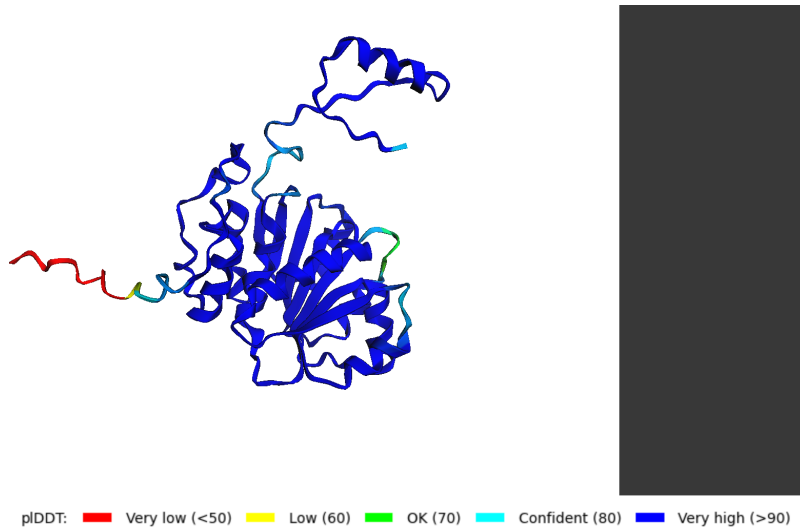
The top 3 proteins yielded from BLAST research against this “barber pole worm” protein are as follow.

ID	Technique	Resolution	Source	Identity	Evalue
4CT4	X-RAY DIFFRACTION	2.3	Homo sapiens	73	0.0
1S2M	X-RAY DIFFRACTION	2.1	Saccharomyces cerevisiae	68	0.0
7DTK	X-RAY DIFFRACTION	1.85	Caenorhabditis elegans	90	5e-135

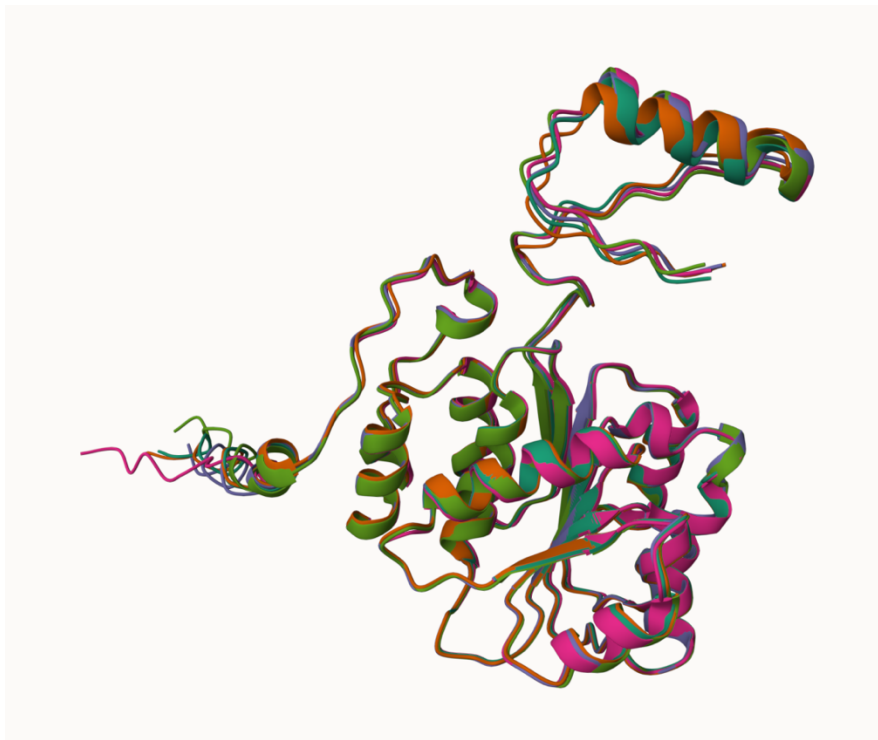
Q9.

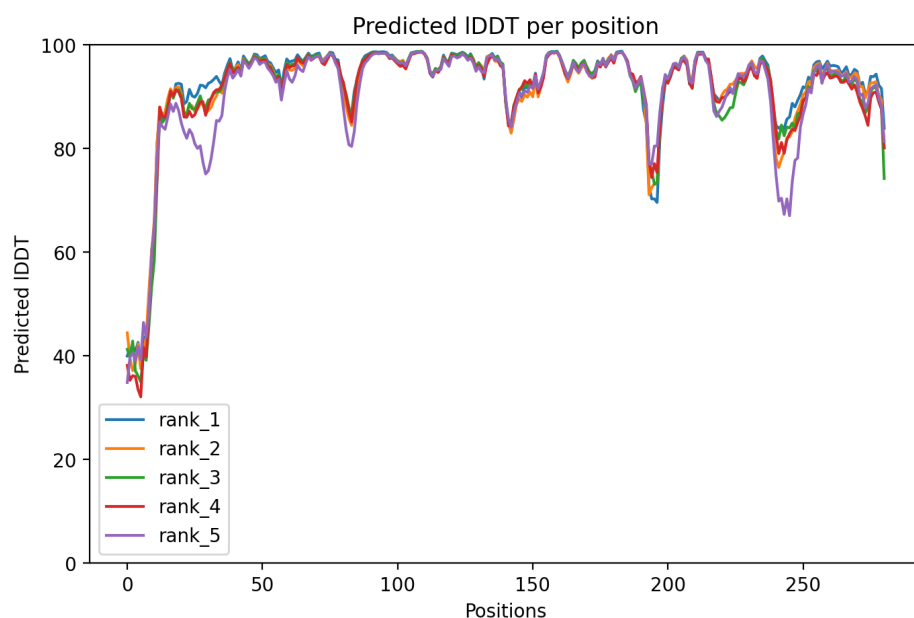
The novel *Panagrolaimus davidi* Ddx6 protein predicted by AlphaFold2.

Rank 1 prediction with pLDDT quality score:



Top five predictions superposed:





Q10.

CHEMBL shows two binding assays (CHEMBL4118655 and CHEMBL3991690) and one ligand efficiency data (CHEMBL2165191).

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL4105783/

Some of the compounds used in the two binding assays shown below may act as a potential inhibitor for the novel *Panagrolaimus davidi* Ddx6 protein.

<input type="checkbox"/>	ChEMBL ID	Assay Type	Description	Organism	Compounds	Document ChEMBL ID	BAO Format	Source
<input type="checkbox"/>	CHEMBL3991690	B	Kinobeads (epsilon), multiple immobilized ATP-competitive broad spectrum kinase inhibitors, used to assess residual binding of ~300 proteins simultaneously from cell lysate in the presence of a compound. Quantitative readout performed by mass spectrometry.	Homo sapiens	236	CHEMBL3991601	subcellular format	Kuster lab chemical proteomics drug profiling
<input type="checkbox"/>	CHEMBL4118655	B	Binding affinity to DDX6 in human NCI-H23 cells at 1 uM by mass spectrometry based pull down assay	Homo sapiens	1	CHEMBL4118021	cell-based format	Scientific Literature