BGGN-213: FOUNDATIONS OF BIOINFORMATICS

The find-a-gene project assignment

Alexander Liu (PID: 69026918)

Q1.

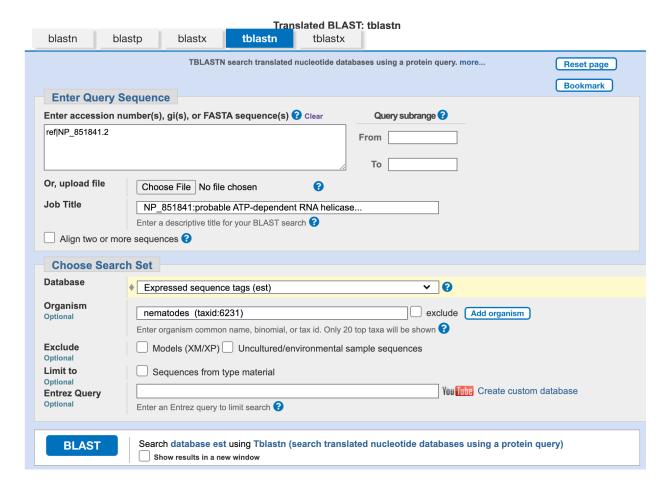Name: probable ATP-dependent RNA helicase DDX6Organism: Mus musculus
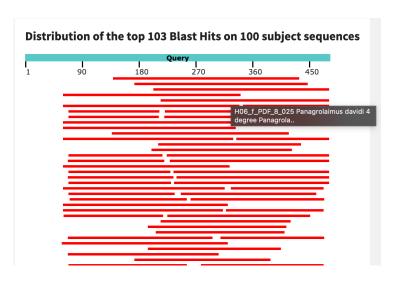Accession: NP_851841.2
Species: *Mus musculus*

Q2.

Method: TBLASTN search against nematodes ESTs
Database: Expressed Sequence Tags (est)
Organism: Nematodes (Taxid: 6231)

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

Distribution of the top 103 Blast Hits on 100 subject sequences

	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   SMAATIKPGDDWKKTLKLPPKDLRIKTSDVTSTKGNEFEDYCLKRELLMGI
            ++ A  K  DDWK  L+LP KD+R KT DVT+T+G +FE++ L R+LL GI
Sbjct  26   AVEAQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKGI

Query  121  PIQEESIPIALSGRDILARAKNGTGKSGAYLIPLLERLDLKKDNIQAMVIV
            PIQE SI IALSG+DILARAKNGTGK+GAY IP++ERLD  KD IQA++IV
Sbjct  206  PIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIV

Query  181  SQICIQVSKHMGGAKVMATTGGTNLRDDIMRLDDTVHVVIATPGRILDLIK
            SQIC+++ KH+G  KVM TTGGT+LRDDI+RL+  VH+++ATPGRI+DL++
Sbjct  386  SQICVELGKHIG-LKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLME

Query  241  QMIVLDEADKLLSQDFVQIMEDIILTLPKNRQILLYSATFPLSVQKFMNSH
            +M+VLDEADKLLSQDF +++ ++ LP  RQI+LYSATFP +V  FM SH
Sbjct  563  KMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESH

Query  301  MEELTLKGVTQYYAYVTERQKVHCLNTLFSRLQINQSIIFC      341
            MEELTL GVTQ+YA+VTE+QKVHCLNTLF +LQINQSIIFC
Sbjct  743  MEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC      865
```

Q3.

Chosen sequence:
```
>Panagrolaimus davidi Ddx6 (translated using EMBOSS Transeq, and ORF
extracted)
MAVEAQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKGIFEKGWERPSPIQ
EASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQTSQICVEL
GKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLMEKGVAEVQDCKMLVLDEADKLL
SQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESHMKNPYEINLMEELTLLGVTQFYAFV
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.*

(see below)

| blastn | **blastp** | blastx | tblastn | tblastx |

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

```
IIVPTRELALQTSQICVELGKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRII
DLMEKGVAEVQDCKMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFP
RTVASFMESHMKNPYEINLMEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQ
SIIFC
```

Query subrange ❓

From [   ]

To [   ]

Or, upload file   [Choose File] No file chosen ❓

Job Title   [unnamed protein product   ]

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   [Non-redundant protein sequences (nr)   ▾] ❓

Organism
Optional   [Enter organism name or id--completions will be suggested] ☐ exclude [Add organism]

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 ❓

Feedback

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 ▾
☑ hypothetical protein FO519_002541 [Halicephalobus sp. NKZ332]	Halicephal...	497	497	100%	8e-174	85.41%
☑ unnamed protein product [Meloidogyne enterolobii]	Meloidogyn...	483	483	96%	4e-168	82.66%
☑ hypothetical protein Mgra_00002596 [Meloidogyne graminicola]	Meloidogyn...	479	479	96%	1e-166	82.29%
☑ unnamed protein product [Haemonchus placei]	Haemonch...	474	474	97%	1e-164	80.59%
☑ 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   ▼ Next Match ▲ Previous Match

Related Information

AlphaFold Structure - 3D structure displays

Score	Expect	Method	Identities	Positives	Gaps
497 bits(1280)	8e-174	Compositional matrix adjust.	240/281(85%)	256/281(91%)	0/281(0%)

```
Query  1    MAVEAQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKG
            M+ E    ++ DWKA L LPAKD R+KT DVTNTRGL+FE+FSLSRDLLKG
Sbjct  1    MSTEIATHSDADWKASLNLPAKDGRFKTQDVTNTRGLEFEDFSLSRDLLKG

Query  61   SPIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALII
            SPIQEASIGIALSG DILARAKNGTGKTGAYCIPVI R+DTSK+ IQALI+
Sbjct  61   SPIQEASIGIALSGHDILARAKNGTGKTGAYCIPVIGRIDTSKNEIQALIV

Query  121  TSQICVELGKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLME
            TSQIC+ELGKH+ LKVMVTTGGTDLRDDILRLNS VHLIVATPGRI+DLME
Sbjct  121  TSQICIELGKHMHLKVMVTTGGTDLRDDILRLNSVVHLIVATPGRILDLME

Query  181  KMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESH
            KMLVLDEADKLLSQDFQGVLDRLV  LP GRQIMLYSATFPRTVASFM+ H
Sbjct  181  KMLVLDEADKLLSQDFQGVLDRLVDRLPLGRQIMLYSATFPRTVASFMQKH

Query  241  MEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC    281
            MEELTLLGVTQFYAFV E+QKVHCLNTLFRKLQINQSIIFC
Sbjct  241  MEELTLLGVTQFYAFVQERQKVHCLNTLFRKLQINQSIIFC    281
```