**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)

A screenshot of a computer

Description automatically generated

Chosen match: Accession JZ594105.1, H06\_f\_PDF\_8\_025 Panagrolaimus davidi 4 degree Panagrolaimus davidi cDNA, mRNA sequence

(the fourth hit)

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

Alignment details:

A screenshot of a computer

Description automatically generated

**Q3.**

Chosen sequence:

>Panagrolaimus davidi Ddx6 (translated using EMBOSS Transeq, and ORF extracted)

MAVEAQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKGIFEKGWERPSPIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQTSQICVELGKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLMEKGVAEVQDCKMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESHMKNPYEINLMEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC

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.*

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Q5.**

>NP\_851841.2 probable ATP-dependent RNA helicase DDX6 [Mus musculus]

MSTARTENPVIMGLSSQNGQLRGPVKASAGPGGGGTQPQPQLNQLKNTSTINNGTPQQAQSMAATIKPGD

DWKKTLKLPPKDLRIKTSDVTSTKGNEFEDYCLKRELLMGIFEMGWEKPSPIQEESIPIALSGRDILARA

KNGTGKSGAYLIPLLERLDLKKDNIQAMVIVPTRELALQVSQICIQVSKHMGGAKVMATTGGTNLRDDIM

RLDDTVHVVIATPGRILDLIKKGVAKVDHVQMIVLDEADKLLSQDFVQIMEDIILTLPKNRQILLYSATF

PLSVQKFMNSHLQKPYEINLMEELTLKGVTQYYAYVTERQKVHCLNTLFSRLQINQSIIFCNSSQRVELL

AKKISQLGYSCFYIHAKMRQEHRNRVFHDFRNGLCRNLVCTDLFTRGIDIQAVNVVINFDFPKLAETYLH

RIGRSGRFGHLGLAINLITYDDRFNLKSIEEQLGTEIKPIPSNIDKSLYVAEYHSEPAEDEKP

>Panagrolaimus davidi Ddx6 (translated using EMBOSS Transeq, and ORF extracted)

MAVEAQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKGIFEKGWERPSPIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQTSQICVELGKHIGLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLMEKGVAEVQDCKMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESHMKNPYEINLMEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC

>KAE9554242.1 hypothetical protein FO519\_002541 [Halicephalobus sp. NKZ332]

MSTEIATHSDADWKASLNLPAKDGRFKTQDVTNTRGLEFEDFSLSRDLLKGIFEKGWERPSPIQEASIGI

ALSGHDILARAKNGTGKTGAYCIPVIGRIDTSKNEIQALIVVPTRELALQTSQICIELGKHMHLKVMVTT

GGTDLRDDILRLNSVVHLIVATPGRILDLMEKDVAVVDKCKMLVLDEADKLLSQDFQGVLDRLVDRLPLG

RQIMLYSATFPRTVASFMQKHMHHPHEINLMEELTLLGVTQFYAFVQERQKVHCLNTLFRKLQINQSIIF

CNSTQRVELLAKKITELGYSCYYIHSKMAQHHRNRVFHDFREGHCRNLVCSDLLTRGIDIQAVNVVINFD

FPRNSETYLHRIGRSGRFGHLGIAINLITYEDRHNLRRIEKELRTRIEPIPAEVDPKLYVAEYQQDLKDE

DVQSSK

>CAD2123926.1 unnamed protein product [Meloidogyne enterolobii]

MSTAVSPALSMDDWKAELNLPTKDLRFKTADVTATKGIEFEEFGLSRDLLKGIFEKGWEKPSPIQEASIA

IALSGQDVLARAKNGTGKTGAYCIPCIEKIDPSREYIQAMVVVPTRELALQTSQICVELSKHLKIKVMVT

TGGTDLRDDILRLNGVVHLVVATPGRILDLMEKGVANVTRCKTIVLDEADKLLSQDFQGVLDRLVHFMPA

DRQIMLYSATFPRTVESFMEKHMQRPYEINLMEELTLLGITQFYAYVQEKQKVHCLNTLFRKLQINQSII

FCNSTQRVELLAKKITELGYSCYYIHSKMAQQHRNRVFHEFRQGHCRNLVCSDLLTRGIDIQAVNVVINF

DFPRNSETYLHRIGRSGRFGHLGIAINLITYEDRYNLKRIEKELRTHIEPIPKEVDPKLYVAEHQME

>KAF7637891.1 hypothetical protein Mgra\_00002596 [Meloidogyne graminicola]

MSTAVSPTLSMEDWKAELNLPTKDLRFKTADVTATKGTEFEEFGLSRDLLKGIFEKGWEKPSPIQEASIA

IALSGQDILARAKNGTGKTGAYCIPCIEKIDPSREYIQAMIVVPTRELALQTSQICVELSKHLKIKVMVT

TGGTDLRDDILRLNGVVHLVVATPGRILDLMEKGVANVSRCKTIVLDEADKLLSQDFQGVLDRLVHFMPS

DRQIMLYSATFPRTVETFMEKHMQRPYEINLMEELTLLGITQFYAYVQEKQKVHCLNTLFRKLQINQSII

FCNSTQRVELLAKKITELGYSCYYIHSKMAQQHRNRVFHEFRQGHCRNLVCSDLLTRGIDIQAVNVVINF

DFPRNSETYLHRIGRSGRFGHLGIAINLITYEDRYNLKRIEKELRTHIEPIPKEVDPKLYVAEHQME

>VDO27004.1 unnamed protein product [Haemonchus placei]

MSVADTKVAAVTTDENWKAGLVLPAKDKRFKTADVTDTKGVEFEDFCLSRDLLMGIFEKGWEKPSPIQEA

SIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPSIKKIQALVIVPTRELALQTSQICVELSKHIKLKI

MVTTGGTDLRDDIMRLNGSVHMVIATPGRILDLMEKGVADMSNCKMLVLDEADKLLSQDFQGILDRLISF

LPKERQIMLYSATFPMTVTEFMQKHMRKPYEINLMEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQ

SIIFCNSTQRVELLAKKITEIGYSCYYIHSKMAQNHRNRVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVV

INFDFPRNAETYLHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIV

DDEDNRANAN

>CDJ84064.1 DNA RNA helicase domain containing protein [Haemonchus contortus]

MSVADTKVAAVTTDENWKAGLVLPAKDKRFKTADVTDTKGVEFEDFCLSRDLLMGIFEKGWEKPSPIQEA

SIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPSIKKIQALVIVPTRELALQTSQICVELSKHIKLKI

MVTTGGTDLRDDIMRLNGSVHMVIATPGRILDLMEKGVADMSNCKMLVLDEADKLLSQDFQGILDRLISF

LPKERQIMLYSATFPMTVTEFMQKHMRKPYEINLMEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQ

SIIFCNSTQRVELLAKKITEIGYSCYYIHSKMAQNHRNRVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVV

INFDFPRNAETYLHRIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIV

DDEDNNRANAN

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

NP\_851841.2 MSTARTENPVIMGLSSQNGQLRGPVKASAGPGGGGTQPQPQLNQLKNTSTINNGTPQQAQ

VDO27004.1 MSVADTK-----------------------------------------------------

CDJ84064.1 MSVADTK-----------------------------------------------------

CAD2123926.1 MSTA--------------------------------------------------------

KAF7637891.1 MSTA--------------------------------------------------------

Panagrolaimus MAVE--------------------------------------------------------

KAE9554242.1 MSTE--------------------------------------------------------

\*:.

NP\_851841.2 SMAATIKPGDDWKKTLKLPPKDLRIKTSDVTSTKGNEFEDYCLKRELLMGIFEMGWEKPS

VDO27004.1 --VAAVTTDENWKAGLVLPAKDKRFKTADVTDTKGVEFEDFCLSRDLLMGIFEKGWEKPS

CDJ84064.1 --VAAVTTDENWKAGLVLPAKDKRFKTADVTDTKGVEFEDFCLSRDLLMGIFEKGWEKPS

CAD2123926.1 --VSPALSMDDWKAELNLPTKDLRFKTADVTATKGIEFEEFGLSRDLLKGIFEKGWEKPS

KAF7637891.1 --VSPTLSMEDWKAELNLPTKDLRFKTADVTATKGTEFEEFGLSRDLLKGIFEKGWEKPS

Panagrolaimus ---AQNKTNDDWKAQLQLPAKDMRYKTLDVTNTRGLDFEEFSLSRDLLKGIFEKGWERPS

KAE9554242.1 ---IATHSDADWKASLNLPAKDGRFKTQDVTNTRGLEFEDFSLSRDLLKGIFEKGWERPS

. :\*\* \* \*\*.\*\* \* \*\* \*\*\* \*.\* :\*\*:: \*.\*:\*\* \*\*\*\* \*\*\*.\*\*

NP\_851841.2 PIQEESIPIALSGRDILARAKNGTGKSGAYLIPLLERLDLKKDNIQAMVIVPTRELALQV

VDO27004.1 PIQEASIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPSIKKIQALVIVPTRELALQT

CDJ84064.1 PIQEASIGVALSGQDILARAKNGTGKTGAYCIPVIEKINPSIKKIQALVIVPTRELALQT

CAD2123926.1 PIQEASIAIALSGQDVLARAKNGTGKTGAYCIPCIEKIDPSREYIQAMVVVPTRELALQT

KAF7637891.1 PIQEASIAIALSGQDILARAKNGTGKTGAYCIPCIEKIDPSREYIQAMIVVPTRELALQT

Panagrolaimus PIQEASIGIALSGQDILARAKNGTGKTGAYCIPVIERLDTSKDCIQALIIVPTRELALQT

KAE9554242.1 PIQEASIGIALSGHDILARAKNGTGKTGAYCIPVIGRIDTSKNEIQALIVVPTRELALQT

\*\*\*\* \*\* :\*\*\*\*.\*:\*\*\*\*\*\*\*\*\*\*:\*\*\* \*\* : .:: . . \*\*\*:::\*\*\*\*\*\*\*\*\*.

NP\_851841.2 SQICIQVSKHMGGAKVMATTGGTNLRDDIMRLDDTVHVVIATPGRILDLIKKGVAKVDHV

VDO27004.1 SQICVELSKHI-KLKIMVTTGGTDLRDDIMRLNGSVHMVIATPGRILDLMEKGVADMSNC

CDJ84064.1 SQICVELSKHI-KLKIMVTTGGTDLRDDIMRLNGSVHMVIATPGRILDLMEKGVADMSNC

CAD2123926.1 SQICVELSKHL-KIKVMVTTGGTDLRDDILRLNGVVHLVVATPGRILDLMEKGVANVTRC

KAF7637891.1 SQICVELSKHL-KIKVMVTTGGTDLRDDILRLNGVVHLVVATPGRILDLMEKGVANVSRC

Panagrolaimus SQICVELGKHI-GLKVMVTTGGTDLRDDILRLNSAVHLIVATPGRIIDLMEKGVAEVQDC

KAE9554242.1 SQICIELGKHM-HLKVMVTTGGTDLRDDILRLNSVVHLIVATPGRILDLMEKDVAVVDKC

\*\*\*\*:::.\*\*: \*:\*.\*\*\*\*\*:\*\*\*\*\*:\*\*:. \*\*:::\*\*\*\*\*\*:\*\*::\*.\*\* :

NP\_851841.2 QMIVLDEADKLLSQDFVQIMEDIILTLPKNRQILLYSATFPLSVQKFMNSHLQKPYEINL

VDO27004.1 KMLVLDEADKLLSQDFQGILDRLISFLPKERQIMLYSATFPMTVTEFMQKHMRKPYEINL

CDJ84064.1 KMLVLDEADKLLSQDFQGILDRLISFLPKERQIMLYSATFPMTVTEFMQKHMRKPYEINL

CAD2123926.1 KTIVLDEADKLLSQDFQGVLDRLVHFMPADRQIMLYSATFPRTVESFMEKHMQRPYEINL

KAF7637891.1 KTIVLDEADKLLSQDFQGVLDRLVHFMPSDRQIMLYSATFPRTVETFMEKHMQRPYEINL

Panagrolaimus KMLVLDEADKLLSQDFQGVLDRLVSFLPAGRQIMLYSATFPRTVASFMESHMKNPYEINL

KAE9554242.1 KMLVLDEADKLLSQDFQGVLDRLVDRLPLGRQIMLYSATFPRTVASFMQKHMHHPHEINL

: :\*\*\*\*\*\*\*\*\*\*\*\*\* ::: :: :\* \*\*\*:\*\*\*\*\*\*\* :\* \*\*:.\*:..\*:\*\*\*\*

NP\_851841.2 MEELTLKGVTQYYAYVTERQKVHCLNTLFSRLQINQSIIFCNSSQRVELLAKKISQLGYS

VDO27004.1 MEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITEIGYS

CDJ84064.1 MEELTLLGVTQYYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITEIGYS

CAD2123926.1 MEELTLLGITQFYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITELGYS

KAF7637891.1 MEELTLLGITQFYAYVQEKQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITELGYS

Panagrolaimus MEELTLLGVTQFYAFVTEKQKVHCLNTLFRKLQINQSIIFC-------------------

KAE9554242.1 MEELTLLGVTQFYAFVQERQKVHCLNTLFRKLQINQSIIFCNSTQRVELLAKKITELGYS

\*\*\*\*\*\* \*:\*\*:\*\*:\* \*.\*\*\*\*\*\*\*\*\*\* .\*\*\*\*\*\*\*\*\*\*

NP\_851841.2 CFYIHAKMRQEHRNRVFHDFRNGLCRNLVCTDLFTRGIDIQAVNVVINFDFPKLAETYLH

VDO27004.1 CYYIHSKMAQNHRNRVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETYLH

CDJ84064.1 CYYIHSKMAQNHRNRVFHDFRQGNCRNLVCSDLLTRGIDIQAVNVVINFDFPRNAETYLH

CAD2123926.1 CYYIHSKMAQQHRNRVFHEFRQGHCRNLVCSDLLTRGIDIQAVNVVINFDFPRNSETYLH

KAF7637891.1 CYYIHSKMAQQHRNRVFHEFRQGHCRNLVCSDLLTRGIDIQAVNVVINFDFPRNSETYLH

Panagrolaimus ------------------------------------------------------------

KAE9554242.1 CYYIHSKMAQHHRNRVFHDFREGHCRNLVCSDLLTRGIDIQAVNVVINFDFPRNSETYLH

NP\_851841.2 RIGRSGRFGHLGLAINLITYDDRFNLKSIEEQLGTEIKPIPSNIDKSLYVAEYHSEPAED

VDO27004.1 RIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIVDDED

CDJ84064.1 RIGRSGRFGHLGVAINLITYEDRHTLRRIEQELRTQIEPIPKAVDPKLYVADFQIVDDED

CAD2123926.1 RIGRSGRFGHLGIAINLITYEDRYNLKRIEKELRTHIEPIPKEVDPKLYVAEHQME----

KAF7637891.1 RIGRSGRFGHLGIAINLITYEDRYNLKRIEKELRTHIEPIPKEVDPKLYVAEHQME----

Panagrolaimus ------------------------------------------------------------

KAE9554242.1 RIGRSGRFGHLGIAINLITYEDRHNLRRIEKELRTRIEPIPAEVDPKLYVAEYQQDLKDE

NP\_851841.2 EKP----

VDO27004.1 NRANAN-

CDJ84064.1 NNRANAN

CAD2123926.1 -------

KAF7637891.1 -------

Panagrolaimus -------

KAE9554242.1 DVQSSK-

**Q6.**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

Phylogenetic tree file:

(

NP\_851841.2:0.20849,

(

VDO27004.1:0.00581,

CDJ84064.1:0.00581)

:0.08312,

(

(

CAD2123926.1:0.01005,

KAF7637891.1:0.00914)

:0.06476,

(

Panagrolaimus:0.08334,

KAE9554242.1:0.06257)

:0.03015)

:0.00775);

**Q7.**

A close-up of a computer screen

Description automatically generated

**Q8.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Technique | Resolution | Source | Identity | Evalue |
| 4CT4\_B | X-ray | 2.3 | Homo sapiens | 68.984 | 1.32E-150 |
| 4BRU\_A | X-ray | 3.245 | Saccharomyces cerevisiae S288C | 62.703 | 6.22E-132 |
| 1S2M\_A | X-ray | 2.1 | Saccharomyces cerevisiae | 61.856 | 9.18E-132 |

**Q9.**

4CT4 protein (*Homo sapiens* CNOT1 MIF4G domain - DDX6 complex)

Given the identity is 69% similar to consensus sequence, it is likely to be somewhat similar to the novel *Panagrolaimus davidi* Ddx6 protein.

A close-up of several colorful spirals

Description automatically generated