Bioinformatics Project.

LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

Human protein sequences:

MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAIS
LSVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSATD
PKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVEIF
GSAIGLFGVIVAILQTSRVKMGD

The protein accession number of the orthologs in multiple species:
Species Human : NP_004038
Species Mouse : NP_291095
Species Rat : NP_001100151
Species Fish : NP_955855
Species Fly : NP_652010

Pre-steps:

- Combine all the above protein accession sequences in a .txt file. This file can be used in all bioinformatics tool.

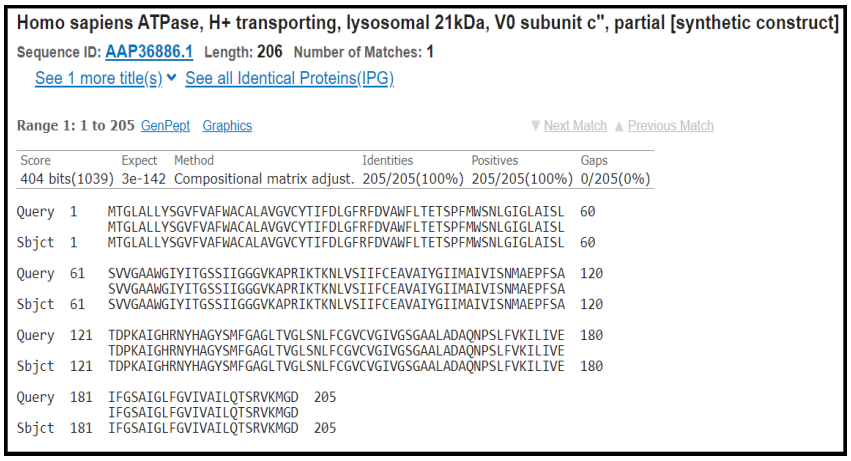
ANSWERS

1. Please find the protein/gene name by searching human protein sequences with proper bioinformatics methods.

    - The method used for finding the name of the protein/gene is BLAST.
    - The step are as follows:
        - Insert the sequence in BLAST.
        - Look for the closest alignment to the sequence.
        - From the above method, I found the below two results.
    - Homo sapiens ATPase, H+ transporting, lysosomal 21kDa, V0 subunit c", partial [synthetic construct].

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020



Homo sapiens ATPase, H+ transporting, lysosomal 21kDa, V0 subunit c", partial [synthetic construct]

Sequence ID: AAP36886.1 Length: 206 Number of Matches: 1

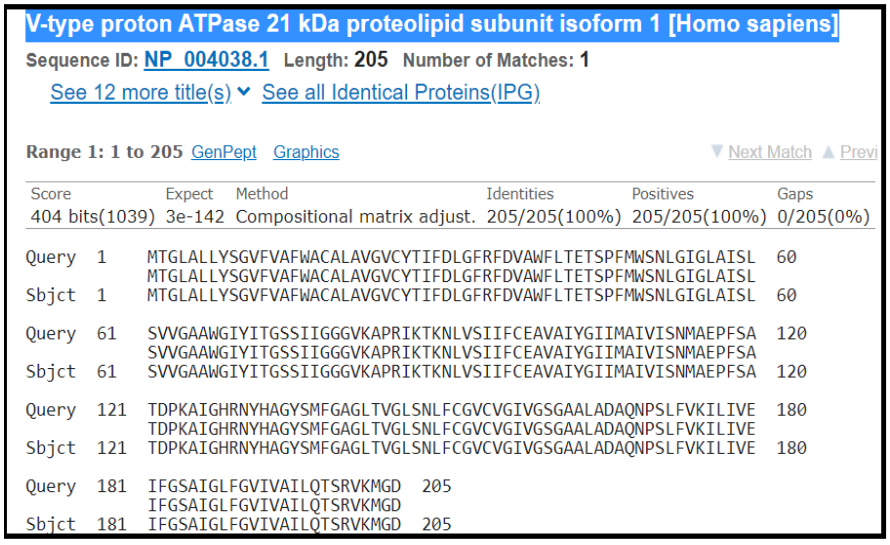
See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 205 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
404 bits(1039)	3e-142	Compositional matrix adjust.	205/205(100%)	205/205(100%)	0/205(0%)

```
Query  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  60
            MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL
Sbjct  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  60

Query  61   SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  120
            SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA
Sbjct  61   SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  120

Query  121  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  180
            TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE
Sbjct  121  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  180

Query  181  IFGSAIGLFGVIVAILQTSRVKMGD  205
            IFGSAIGLFGVIVAILQTSRVKMGD
Sbjct  181  IFGSAIGLFGVIVAILQTSRVKMGD  205
```

- V-type proton ATPase 21 kDa proteolipid subunit isoform 1 [Homo sapiens]



V-type proton ATPase 21 kDa proteolipid subunit isoform 1 [Homo sapiens]

Sequence ID: NP_004038.1 Length: 205 Number of Matches: 1

See 12 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 205 GenPept Graphics ▼ Next Match ▲ Previ

Score	Expect	Method	Identities	Positives	Gaps
404 bits(1039)	3e-142	Compositional matrix adjust.	205/205(100%)	205/205(100%)	0/205(0%)

```
Query  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  60
            MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL
Sbjct  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  60

Query  61   SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  120
            SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA
Sbjct  61   SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  120

Query  121  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  180
            TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE
Sbjct  121  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  180

Query  181  IFGSAIGLFGVIVAILQTSRVKMGD  205
            IFGSAIGLFGVIVAILQTSRVKMGD
Sbjct  181  IFGSAIGLFGVIVAILQTSRVKMGD  205
```

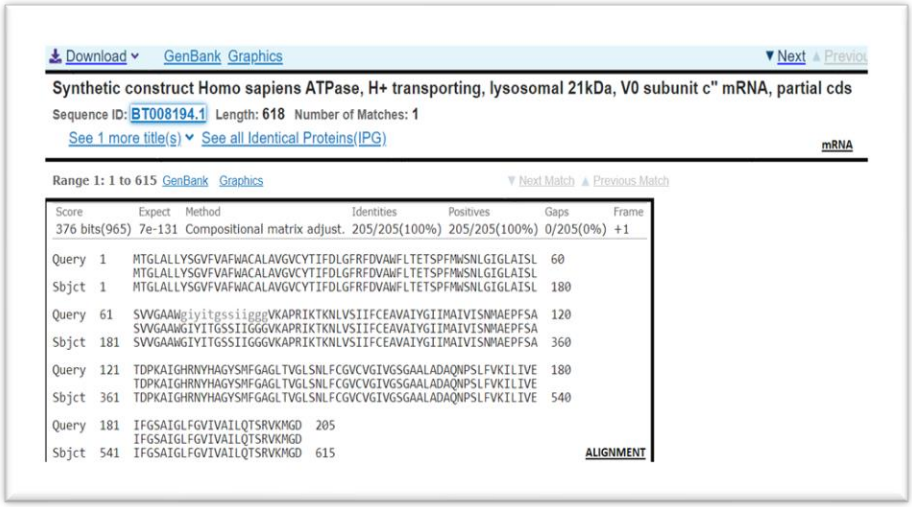
ANSWER: The answer to this question is, there are two protein highly similar to the given sequence. These accession number are as follows:

1. **NP_004038.1➔ I CONSIDER this as the closest protein to the given sequence.**
2. **AAP36886.1**

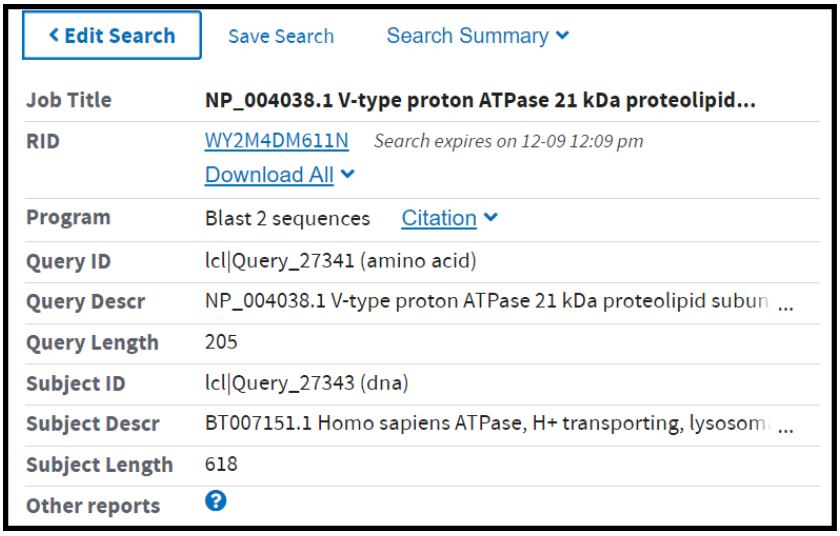
2. Please find the best mRNA sequences (both NCBI accession number and FASTA format of sequence) from which this human protein can be produced. Please also provide the alignment between protein sequence and this mRNA sequence.
   - The best mRNA sequence can be found using the tblastn.
   - From the given sequence, using the tblastn, we can find the mrna sequence.
   - Step: use the obtained accession number in question 1, in tblastn, and observe the results which are closest.

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020



- The closest Mrna to the given sequence in **BT007151.1.**
- For the alignment, I would like to use the blast. The alignment would be implemented on the protein sequence and mrna sequence.



Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

BT007151.1 Homo sapiens ATPase, H+ transporting, lysosomal 21kDa, V0 subunit c" mRNA, complete cds

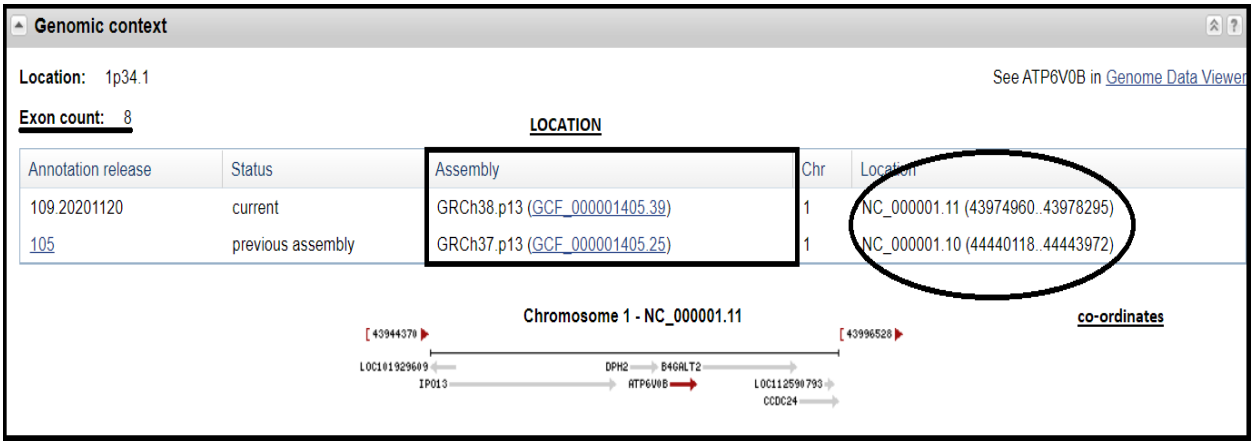
Sequence ID: Query_27343 Length: 618 Number of Matches: 1

Range 1: 1 to 615 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
375 bits(964)	1e-139	Compositional matrix adjust.	205/205(100%)	205/205(100%)	0/205(0%)	+1

```
Query  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  60
            MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL
Sbjct  1    MTGLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISL  180

Query  61   SVVGAAWgiyitgssiigggVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  120
            SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA
Sbjct  181  SVVGAAWGIYITGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSA  360

Query  121  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  180
            TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE
Sbjct  361  TDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVE  540

Query  181  IFGSAIGLFGVIVAILQTSRVKMGD  205
            IFGSAIGLFGVIVAILQTSRVKMGD
Sbjct  541  IFGSAIGLFGVIVAILQTSRVKMGD  615
```

3. Get the coordinates of each exons based on this mRNA sequence. Please generate a table with each row is one exon, including the positions of start and end on mRNA sequence, and on hg38 genomic DNA sequence.
    - The genomic coordinates(hg38) of the exon on mrna BT007151.1, is as follows,



4. extract all protein sequences based on the accession numbers for all species listed above:

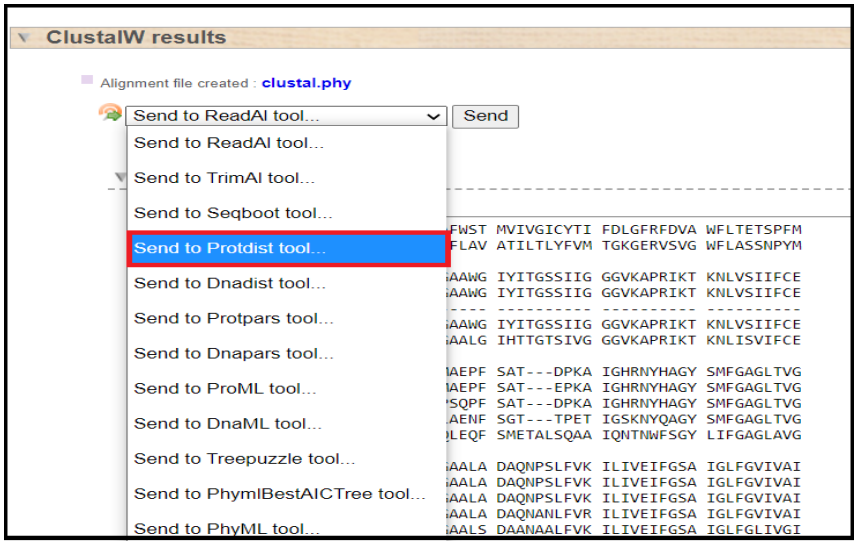
Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

```
GLALLYSGVFVAFWACALAVGVCYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISLSVVGAAWGIY
GSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSATDPKAIGHRNYHAGYSMFGA
.TVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVEIFGSAIGLFGVIVAILQTSRVKMGD
IP_291095.1 V-type proton ATPase 21 kDa proteolipid subunit [Mus musculus]
GLELLYLGIFVAFWACMVVVGICYTIFDLGFRFDVAWFLTETSPFMWSNLGIGLAISLSVVGAAWGIY
GSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNMAEPFSATEPKAIGHRNYHAGYSMFGA
.TVGLSNLFCGVCVGIVGSGAALADAQNPSLFVKILIVEIFGSAIGLFGVIVAILQTSRVKMGD
IP_001100151.1 V-type proton ATPase 21 kDa proteolipid subunit [Rattus norvegicus]
SNNLFCPSQPFSATDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNPSLFVK
.IVEIFGSAIGLFGVIVAILQTSRVKMGD
IP_955855.2 V-type proton ATPase 21 kDa proteolipid subunit [Danio rerio]
NGHAILYTGVTLAFWSTMVIVGICYTIFDLGFRFDVAWFLTETSPFMWANLGIGLAISLSVVGAAWGI
:TGSSIIGGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAIVISNLAENFSGTTPETIGSKNYQAGYSMFG
iLTVGFSNLFCGICVGIVGSGAALADAQNANLFVRILIVEIFGSAIGLFGVIVAILQTSKVKMGN
IP_652010.1 vacuolar H[+] ATPase PPA1 subunit 1, isoform A [Drosophila melanogaster
AQIRTVVSQTFLWLFLAVATILTLYFVMTGKGERVSVGWFLASSNPYMWACLGIGLSVSLSVVGAALG
ITTGTSIVGGGVKAPRIKTKNLISVIFCEAVAIYGLITAIVLSGQLEQFSMETALSQAAIQNTNWFSGY
:FGAGLAVGLVNLFCGIAVGIVGSGAALSDAANAALFVKILIVEIFGSAIGLFGLIVGIYMTSKSKMGD
```

5. perform phylogenetic analysis to construct UPGMA tree of proteins in all 5 species and show the tree with branch length. please use Kimura model to correct the substitution rate

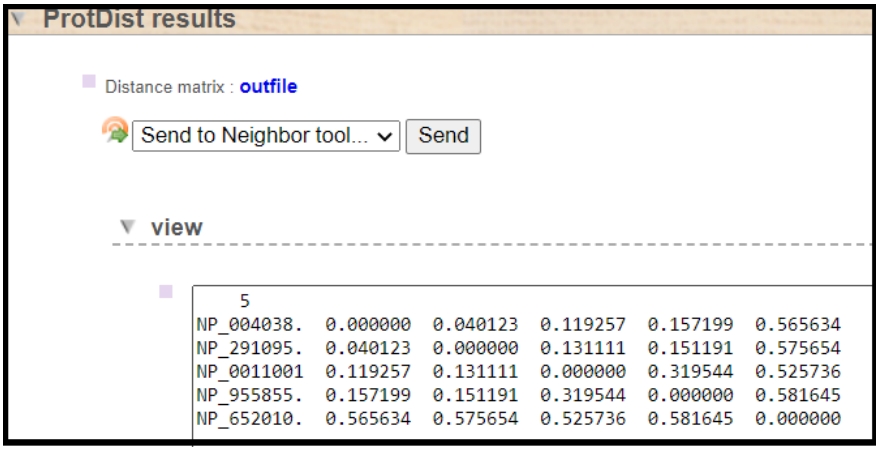
- This question has the following sequence to generate the desired results.
- Phylemon-→ClustalW→prodist→Neighbour→ETE results.
- Step1: Insert all the sequences as file or all sequences together, starting from different line, like in question 4.
- Step2: Insert the obtained MSA results in Prodist.



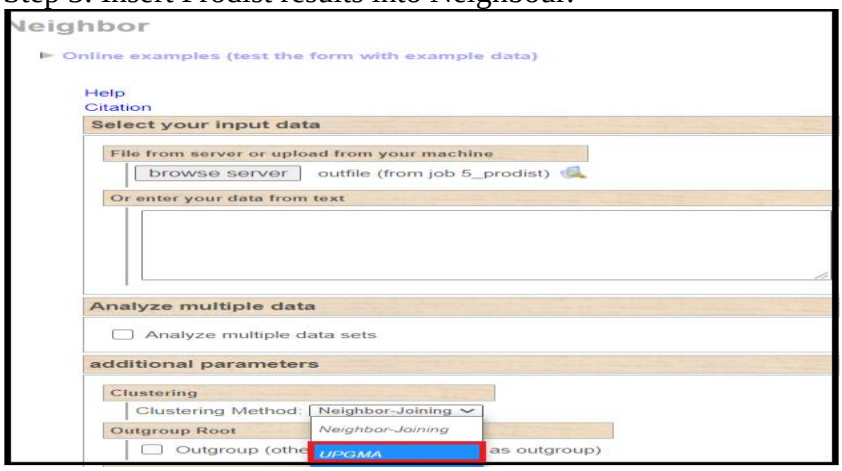
- Results of Prodist:

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020



- Step 3: Insert Prodist results into Neighbour.



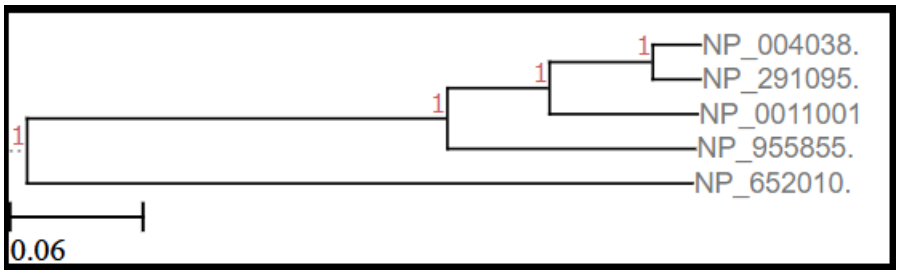
- Neighbor results.

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

```
   5 Populations

Neighbor-Joining/UPGMA method version 3.69


 UPGMA method

 Negative branch lengths allowed


                      +NP_004038.
                   +--1
                +-2   +NP_291095.
                ! !
  +-------------3 +---NP_0011001
  !             !
--4             +-----NP_955855.
  !
  +-----------------NP_652010.


From      To              Length           Height
----      --              ------           ------
   4          3          0.17643          0.17643
   3          2          0.04206          0.21849
   2          1          0.04253          0.26102
   1      NP_004038.     0.02006          0.28108
   1      NP_291095.     0.02006          0.28108
   2      NP_0011001     0.06259          0.28108
   3      NP_955855.     0.10466          0.28108
   4      NP_652010.     0.28108          0.28108
```

- ETE results.



Answer: here the alignment is between the given accession numbers.

6. extract the coding region nucleotide sequences from NCBI/UCSC database for each of these 5 proteins

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

```
>NM_004047.5:82-699 Homo sapiens ATPase H+ transporting V0 subunit b (ATP6V0B), transcript variant 1, mRNA
ATGACGGGGCTAGCACTGCTCTACTCCGGGGTCTTCGTGGCCTTCTGGGCCTGCGCGCTGGCCGTGGGAG
TCTGCTACACCATTTTTGATTTGGGCTTCCGCTTTGATGTGGCATGGTTCCTGACGGAGACTTCGCCCTT
CATGTGGTCCAACCTGGGCATTGGCCTAGCTATCTCCCTGTCTGTGGTTGGGGCAGCCTGGGGCATCTAT
ATTACCGGCTCCTCCATCATTGGTGGAGGAGTGAAGGCCCCCAGGATCAAGACCAAGAACCTGGTCAGCA
TCATCTTCTGTGAGGCTGTGGCCATCTACGGCATCATCATGGCAATTGTCATTAGCAACATGGCTGAGCC
TTTCAGTGCCACAGACCCCAAGGCCATCGGCCATCGGAACTACCATGCAGGCTACTCCATGTTTGGGGCT
GGCCTCACCGTAGGCCTGTCTAACCTCTTCTGTGGAGTCTGCGTGGGCATCGTGGGCAGTGGGGCTGCCC
TGGCCGATGCTCAGAACCCCAGCCTCTTTGTAAAGATTCTCATCGTGGAGATCTTTGGCAGCGCCATTGG
CCTCTTTGGGGTCATCGTCGCAATTCTTCAGACCTCCAGAGTGAAGATGGGTGACTAG
>NM_033617.3:94-711 Mus musculus ATPase, H+ transporting, lysosomal V0 subunit B (Atp6v0b), mRNA
ATGACGGGGCTGGAGTTGCTCTACCTCGGGATCTTTGTGGCCTTCTGGGCCTGCATGGTCGTTGTGGGAA
TCTGCTACACCATCTTTGACCTGGGCTTTCGCTTTGATGTGGCATGGTTCCTGACGGAAACTTCCCCCTT
CATGTGGTCGAATCTGGGCATTGGCCTAGCAATTTCTCTGTCTGTGGTTGGAGCAGCCTGGGGAATCTAT
ATAACCGGCTCATCCATTATTGGGGGTGGGGTGAAAGCCCCCAGAATCAAAACCAAGAACTTGGTTAGTA
TTATCTTCTGTGAAGCGGTGGCCATCTATGGCATCATCATGGCAATTGTCATCAGCAACATGGCTGAGCC
TTTCAGTGCTACGGAGCCCAAGGCCATTGGCCATCGAAACTACCATGCAGGTTACTCCATGTTTGGGGCT
GGCCTCACAGTCGGTCTGTCCAACCTGTTCTGTGGAGTCTGCGTGGGCATCGTGGGCAGTGGGGCCGCCC
TGGCGGATGCACAGAACCCCAGCCTCTTTGTAAAAATTCTCATCGTGGAGATCTTTGGCAGTGCCATTGG
CCTCTTTGGGGTCATTGTTGCAATCCTTCAGACCTCCAGAGTGAAGATGGGTGACTAG
>NM_001106681.1:79-381 Rattus norvegicus ATPase H+ transporting V0 subunit B (Atp6v0b), mRNA
ATGCCTTCTAACAACTTATTCTGCCCCTCACAGCCTTTCAGTGCTACTGACCCCAAGGCCATTGGCCATC
GAAACTACCACGCAGGCTACTCCATGTTTGGGGCTGGCCTCACAGTCGGTCTGTCCAACCTGTTCTGTGG
AGTCTGCGTGGGCATTGTGGGCAGTGGGGCTGCCCTGGCTGACGCACAGAACCCCAGCCTCTTTGTAAAA
ATTCTCATCGTGGAGATCTTTGGCAGTGCCATTGGCCTCTTTGGGGTCATCGTCGCAATCCTTCAGACCT
CCAGAGTGAAGATGGGTGACTAG
>NM_199561.2:191-811 Danio rerio ATPase H+ transporting V0 subunit b (atp6v0b), mRNA
ATGATGAACGGGCACGCGATTTTATACACCGGGGTCACTTTGGCCTTCTGGTCGACTATGGTGATCGTCG
GTATTTGCTATACAATTTTTGACCTTGGATTTCGATTTGATGTAGCATGGTTTTTAACGGAGACTTCTCC
ATTTATGTGGGCTAATCTTGGAATTGGCCTGGCCATTTCTCTGTCTGTTGTTGGAGCTGCATGGGGGATT
TACATCACTGGGTCCAGCATCATTGGTGGTGGAGTCAAAGCTCCAAGAATCAAGACCAAAAATCTTGTCA
GTATTATCTTTTGTGAAGCTGTTGCCATTTATGGGATCATCATGGCAATTGTCATTAGCAATTTGGCAGA
GAACTTCAGTGGCACGACTCCAGAGACTATTGGGTCAAAGAACTACCAAGCGGGCTACTCCATGTTCGGT
GCTGGACTCACGGTTGGCTTTTCAAACCTCTTCTGTGGCATCTGTGTTGGCATTGTGGGCAGTGGTGCTG
CCCTGGCGGATGCTCAGAATGCCAACCTCTTTGTCAGGATCCTTATTGTTGAAATTTTCGGCAGTGCCAT
TGGACTGTTTGGAGTGATTGTAGCCATTTTGCAGACATCGAAAGTAAAAATGGGAAATTAG
>NM_143753.2:131-769 Drosophila melanogaster vacuolar H[+] ATPase PPA1 subunit 1, transcript variant A (VhaPPA1-1), mRNA
ATGGCGGCCCAAATACGCACCGTGGTGTCCCAAACGTTCCTGTGGCTCTTCCTCGCCGTGGCCACCATCC
TGACCCTGTACTTCGTGATGACGGGCAAGGGCGAGCGGGTGAGCGTGGGCTGGTTCCTGGCCTCCTCAAA
CCCGTACATGTGGGCCTGCCTGGGCATCGGACTCTCCGTCTCGCTGTCCGTCGTGGGCGCCGCCCTGGGC
ATCCATACGACGGGCACGAGCATCGTGGGCGGTGGTGTGAAGGCGCCCCGCATCAAGACCAAGAATCTGA
TCTCGGTCATCTTCTGCGAGGCCGTGGCCATCTACGGCCTGATCACCGCCATCGTTCTGTCCGGCCAGCT
GGAGCAGTTCTCGATGGAGACGGCCCTTTCGCAGGCGGCTATTCAGAACACGAACTGGTTCTCCGGCTAC
CTCATCTTCGGTGCTGGCCTGGCTGTCGGCCTGGTCAATCTGTTCTGCGGCATTGCTGTGGGCATTGTGG
GTTCGGGTGCCGCCCTCTCGGACGCCGCCAATGCCGCCCTGTTCGTCAAGATCCTTATTGTGGAGATCTT
CGGTTCGGCCATCGGTCTGTTCGGCCTCATCGTGGGCATCTACATGACCTCCAAGTCCAAGATGGGCGAC
AAGGAGTAG
```

7. perform phylogenetic analysis to construct UPGMA tree of CDS in all 5 species and show the tree with branch length. please also use Kimura 2-parameter model to correct the substitution rate
    - to perform the phylogenetic analysis, we need to take the sequences collected in above question
    - the steps for this phylogenetic analysis is, clustalW→DnaDist→neighbor->Ete.
    - Step1: paste the cds sequeces for ClustalW.
    - Step 2: implement DnaDist to the clustalW results.

```
5
NM_033617. 0.000000 0.069368 0.111964 0.307462 0.648612
NM_0011066 0.069368 0.000000 0.099063 0.373980 0.533106
NM_004047. 0.111964 0.099063 0.000000 0.308149 0.609416
NM_199561. 0.307462 0.373980 0.308149 0.000000 0.676645
NM_143753. 0.648612 0.533106 0.609416 0.676645 0.000000
```

    - Step3: Apply neighbour method to obtained clustalW results.

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

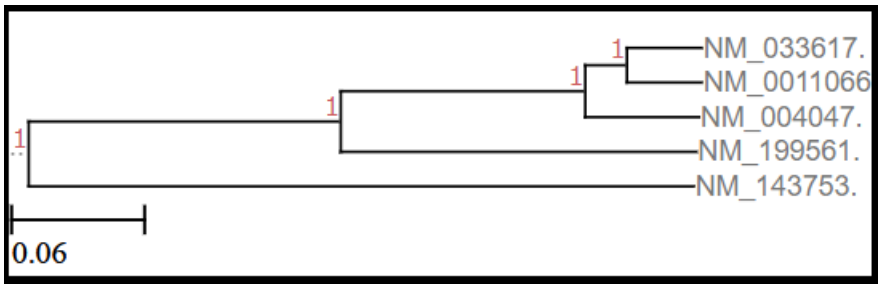
DATE: 12/07/2020

```
5 Populations

Neighbor-Joining/UPGMA method version 3.69


 UPGMA method

 Negative branch lengths allowed


                    +--NM_033617.
                  +-1
            +-----2 +--NM_0011066
            !     !
  +---------3     +---NM_004047.
  !         !
--4         +---------NM_199561.
  !
  +-------------------NM_143753.


From      To              Length          Height
----      --              ------          ------
   4         3           0.14354         0.14354
   3         2           0.11218         0.25572
   2         1           0.01807         0.27379
   1      NM_033617.     0.03468         0.30847
   1      NM_0011066     0.03468         0.30847
   2      NM_004047.     0.05276         0.30847
   3      NM_199561.     0.16493         0.30847
   4      NM_143753.     0.30847         0.30847
```

TREE VIEW:

((((NM_033617.:0.03468,NM_0011066:0.03468):0.01807,NM_004047.:0.05276)
:0.11218,
NM_199561.:0.16493):0.14354,NM_143753.:0.30847);

- Step 4: enter the obtained tree view into ete website.



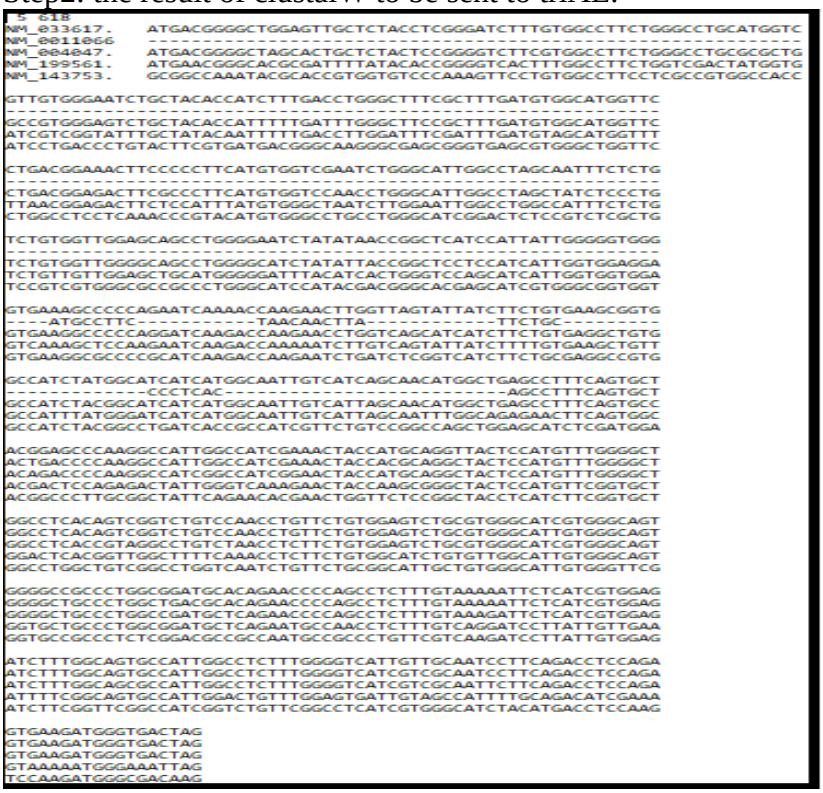
ANSWER: Here the phylogenetic tree is between the cds of the given accession number,

8. perform phylogenetic analysis to construct maximal likelihood tree of CDS in all 5 species by phyML and show the tree with branch length please also use Kimura 2-parameter model (K80) to correct the substitution rate.
    - The question is answered on the basis of knowledge on phylogentic analysis.

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

- The steps implemented are as follows:
  - ClustalW→triAL→phyML→ Ete.
  - Step1: implement the clustalW on the cds sequence.
  - Step2: the result of clustalW to be sent to triAL.

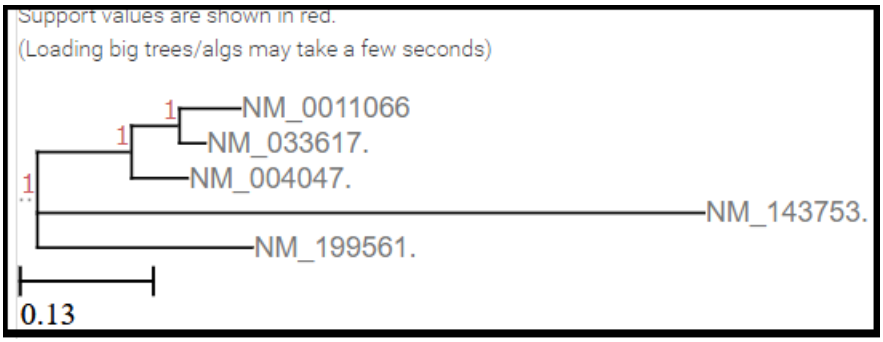


  - Step3: send the results of triAL, into phyML.
  - Set the parameters as follows:
    - Use kimura 2-parameter(K80), substitution model.
    - Data ype: DNA.
    - How to determine character frequencies: Equal Frequencies (A=0.25, C=0.25, G=0.25, T=0.25)
    - Search: Nearest Neighbor Interchange (fast).
    - Tree optimization options: branch length optimization, rate parameters optimization and tree topology optimization.
    - (((NM_0011066:0.0589378099,NM_033617.:0.0249833087)1:0.0440494507,NM_004047.:0.0544535018)1:0.0890309528,NM_143753.:0.6481760000,NM_199561.:0.2089169067);

  - Step 4: Copy the phyML results and paste in the Ete treeviewer.(http://etetoolkit.org/treeview).

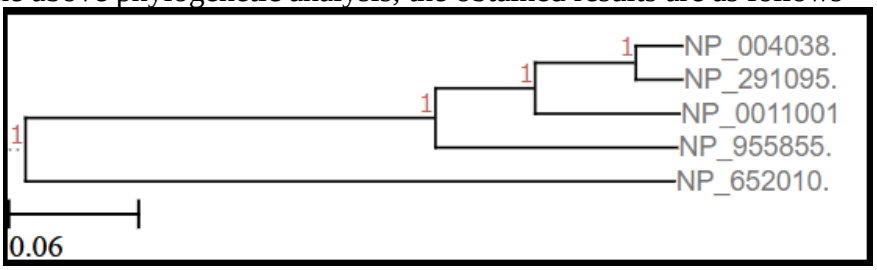
Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020

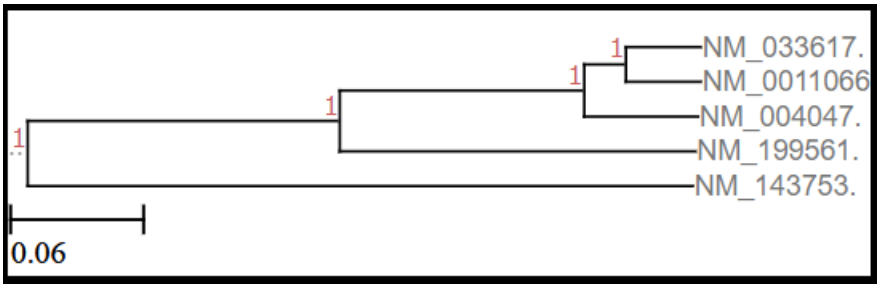


ANSWER: The phylogenetic tree is between the cds of given accession number

9. compare the two CDS tree and one protein tree, and draw your conclusion. (optional) you could also try bootstrapping to test the stability of the trees
    - From the above phylogenetic analysis, the obtained results are as follows



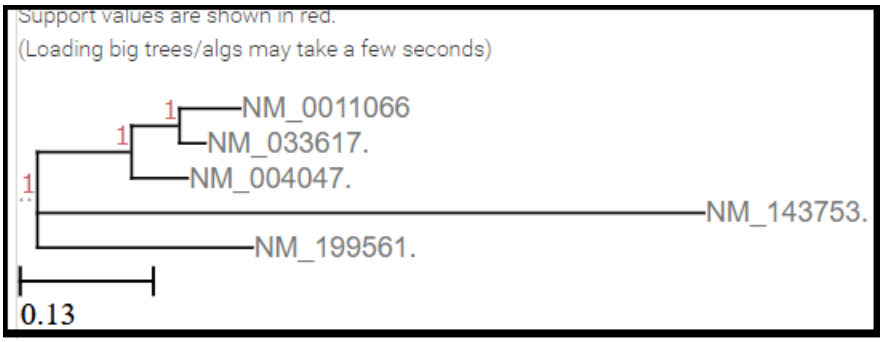
        - HERE, human and mouse are closely related, forming clade 1.
        - Rat is the first closely related to clade 1. Forming clade 2.
        - Lastly, fly is the most distinct from all the species.



        - Here, according to the cds, mouse and rat forms closely related clade.
        - Human becomes the first species related to clade 1.
        - Lastly, fly is the most didtinct of all the species.

Bioinformatics Project.
LIPI VIKRAM THAKKER
LXT190004

DATE: 12/07/2020



- Here, similar to above tree, mouse and tree are closely related, with human close to both of them.
- But here, fish is the most distinct species.

Conclusion: here the tree obtained in 5 and 7 show similar results. However, the tree obtained from the phyML, displays great diversity. Considering the relationship cladewise, each method gives its different relationship, it can be assumed at, as the sequences becomes more conserved and method becomes more intensive, the results become more specific, and the distance remains the same.

From the results, I believe trees obtained from answer 5 and 7 are more accurates, as they match the sequence divergence and give same results even after more intensive method is applied and more conserved sequence is used.