1.

Results for job emboss_needle-I20230511-161623-0921-21399683-p2m

Alignment | Submission Details

View Alignment File

```
########################################
# Program: needle
# Rundate: Thu 11 May 2023 16:16:33
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20230511-161623-0921-21399683-p2m.asequence
#    -bsequence emboss_needle-I20230511-161623-0921-21399683-p2m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 1.0
#    -gapextend 1.0
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1.0
# Extend_penalty: 1.0
#
# Length: 10
# Identity:       7/10 (70.0%)
# Similarity:     7/10 (70.0%)
# Gaps:           1/10 (10.0%)
# Score: 31.0
#
#
#=======================================

EMBOSS_001         1 MTPARGSALS     10
                     |||.|.| ||
EMBOSS_001         1 MTPVRRS-LS      9


#---------------------------------------
#---------------------------------------
```

Results for job emboss_needle-I20230511-161155-0396-77433385-p2m

Alignment | Submission Details

View Alignment File

```
########################################
# Program: needle
# Rundate: Thu 11 May 2023 16:11:58
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20230511-161155-0396-77433385-p2m.asequence
#    -bsequence emboss_needle-I20230511-161155-0396-77433385-p2m.bsequence
#    -datafile EPAM250
#    -gapopen 1.0
#    -gapextend 1.0
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM250
# Gap_penalty: 1.0
# Extend_penalty: 1.0
#
# Length: 10
# Identity:       6/10 (60.0%)
# Similarity:     7/10 (70.0%)
# Gaps:           1/10 (10.0%)
# Score: 29.0
#
#
#=======================================

EMBOSS_001         1 MTPARGSALS     10
                     |||.| .:||
EMBOSS_001         1 MTPVR-RSLS      9


#---------------------------------------
#---------------------------------------
```

Results for job emboss_needle-I20230511-161326-0565-47483407-p2m

Alignment | Submission Details

View Alignment File

```
########################################
# Program: needle
# Rundate: Thu 11 May 2023 16:13:31
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20230511-161326-0565-47483407-p2m.asequence
#    -bsequence emboss_needle-I20230511-161326-0565-47483407-p2m.bsequence
#    -datafile EPAM30
#    -gapopen 1.0
#    -gapextend 1.0
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM30
# Gap_penalty: 1.0
# Extend_penalty: 1.0
#
# Length: 11
# Identity:       7/11 (63.6%)
# Similarity:     7/11 (63.6%)
# Gaps:           3/11 (27.3%)
# Score: 48.0
#
#
#=======================================

EMBOSS_001         1 MTPARG-SALS     10
                     |||.|  | ||
EMBOSS_001         1 MTPVR-RS-LS      9


#---------------------------------------
#---------------------------------------
```

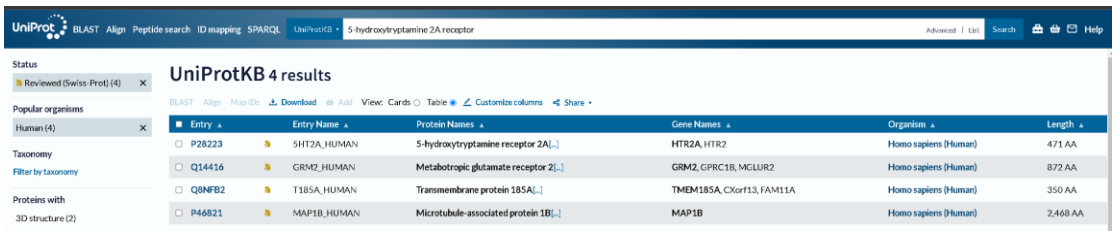
EBLOSUM62 score : 31

PAM250 score : 29

PAM30 score : 48

As the picture we can see that Similarities of EBLOSUM62 and PAM250 are same but PAM30 is different. Scores and Identity are all different.

2.



The protein sequence of P28223 · 5HT2A_HUMAN in FASTA format:

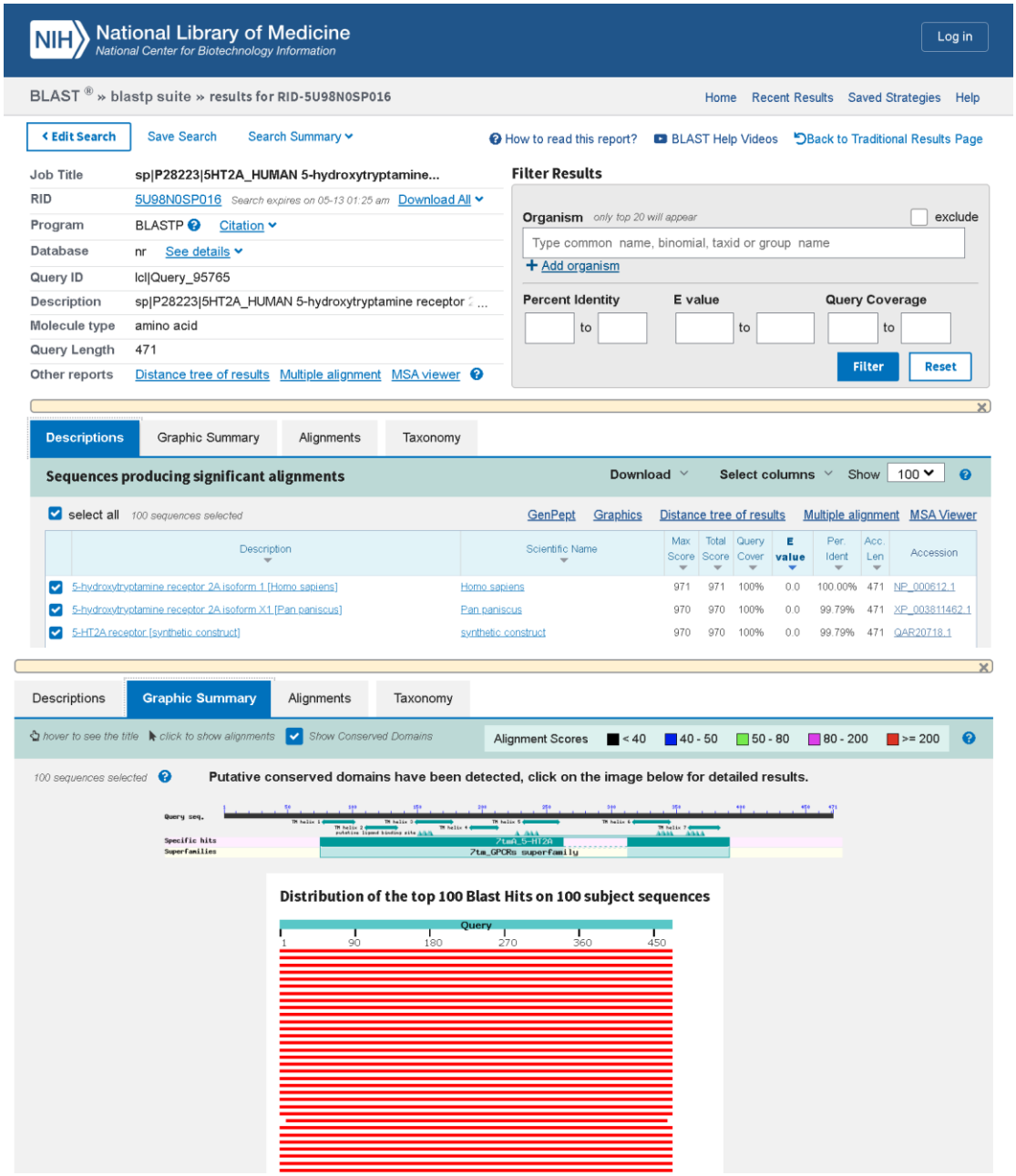
```
>sp|P28223|5HT2A_HUMAN 5-hydroxytryptamine receptor 2A OS=Homo sapiens OX=9606 GN=HTR2A PE=1 SV=2
MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGC
LSPSCLSLLHLQEKNWSALLTAVVIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIAD
MLLGFLVMPVSMLTILYGYRWPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNP
IHHSRFNSRTKAFLKIIAVWTISVGISMPIPVFGLQDDSKVFKEGSCLLADDNFVLIGSF
VSFFIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFQRSIH
REPGSYTGRRTMQSISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEDVIGA
LLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENKKPLQLILVNTIPALAYK
SSQLQMGQKKNSKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
```

3.

We have found 100 similarity sequence.

The results page will contain several sections with different types of information.

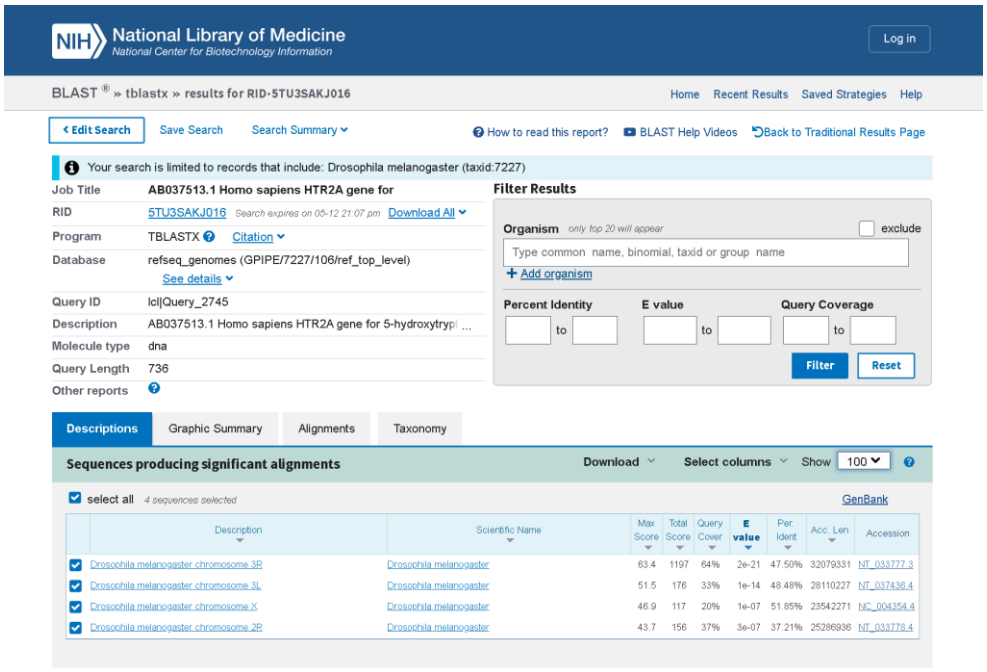
Including Descriptions,Graphic Summary,Alignment and Tazonomy.



4.

blastn: no similarity found

tblastx: we have found 4 similarity cases



The differences between blastn and tblastx lie in the underlying algorithms and the way the searches are performed:

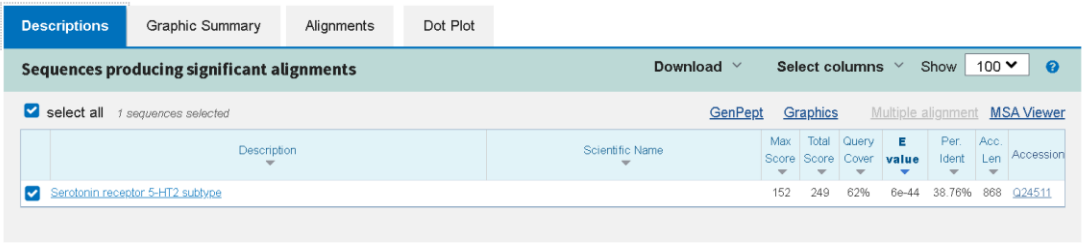
blastn: blastn compares nucleotide sequences directly, matching nucleotide bases.

tblastx: tblastx translates the query sequence in six frames and compares it against the six-frame translations of the target database, matching amino acid sequences.

The respective search results originate from the different algorithms and scoring systems used by blastn and tblastx to evaluate the similarity and significance of the matches. blastn searches for nucleotide sequence similarity, while tblastx searches for translated amino acid sequence similarity.

5.

The BLAST alignment output typically shows the aligned regions of the two sequences. It displays the matching amino acid residues and indicates gaps or insertions. The aligned regions highlight the conserved regions or motifs shared between the two proteins.

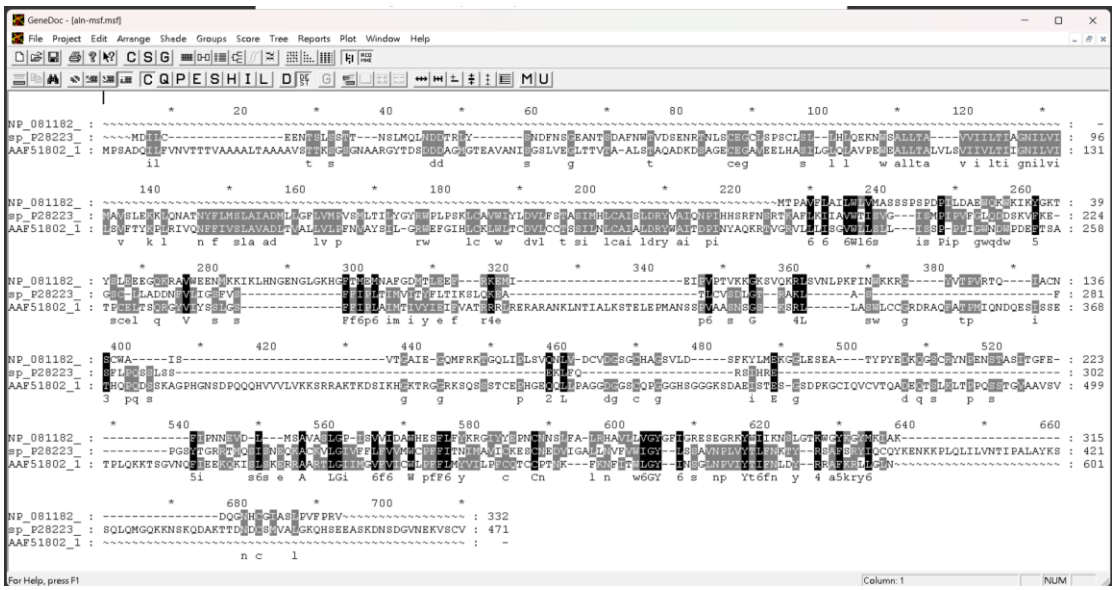


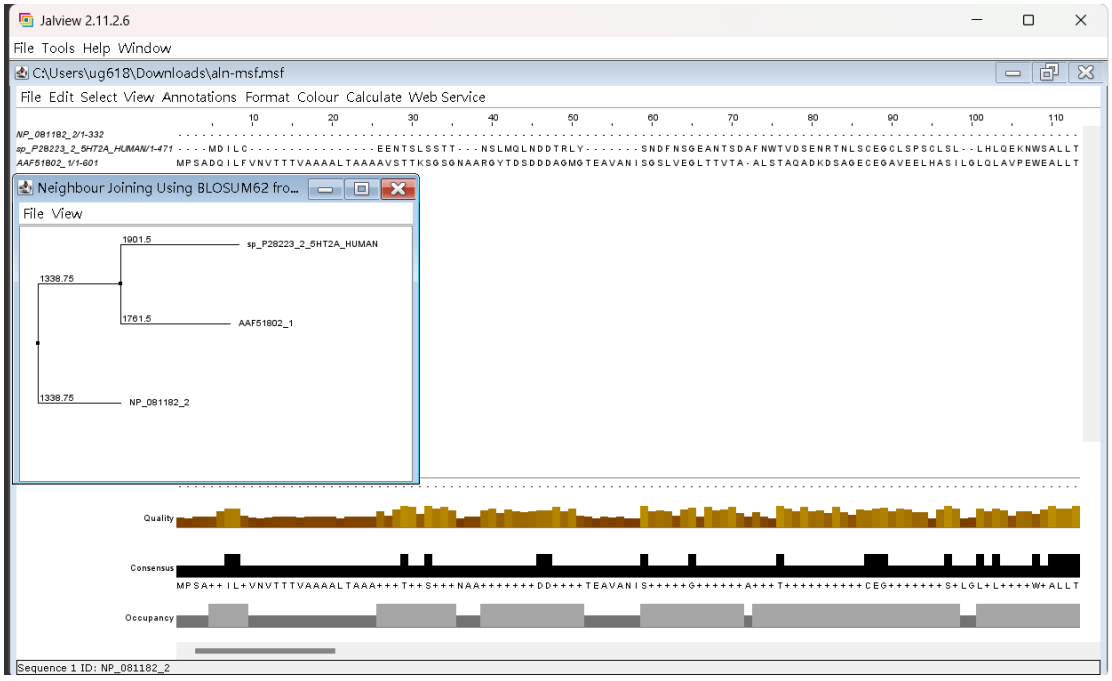
6.

In GeneDoc or similar alignment visualization tools, you can explore and analyze the alignment. You can change sequence names, adjust the color representation of the alignments, and perform various editing or analysis tasks depending on the capabilities of the software.

The multiple alignment obtained from Clustal Omega will show the aligned positions of the three protein sequences, highlighting regions of similarity and conservation. Conserved regions may indicate functional domains or important residues shared among the sequences. Gaps and variations in the alignment can provide insights into sequence differences and potential structural or functional distinctions.

By visualizing the alignment, you can observe the similarities and differences between the sequences. Conserved regions, gaps, and variations in the amino acid or nucleotide sequences can be identified. This information can provide insights into the evolutionary relationships or functional similarities among the sequences.

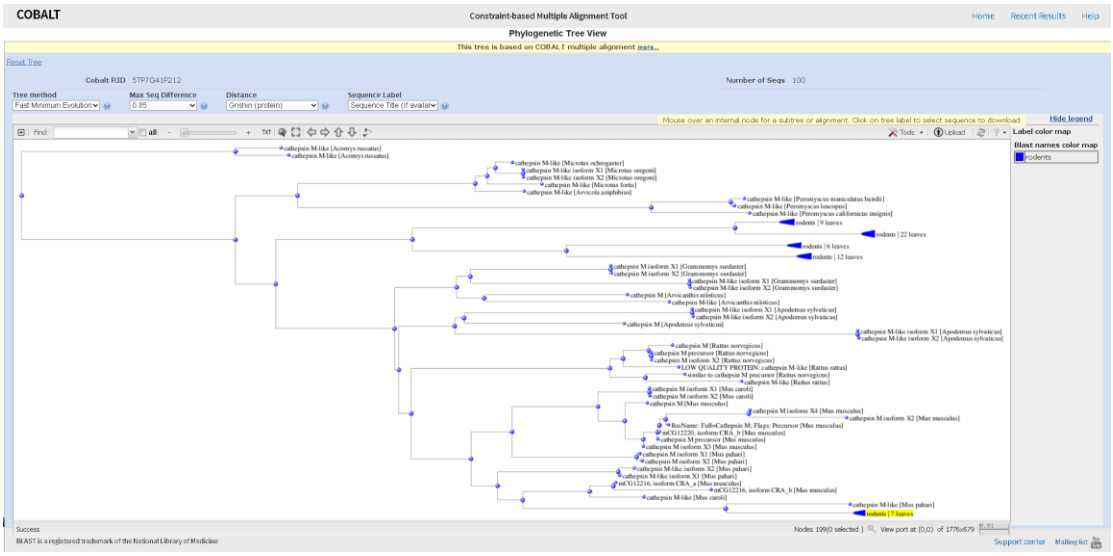


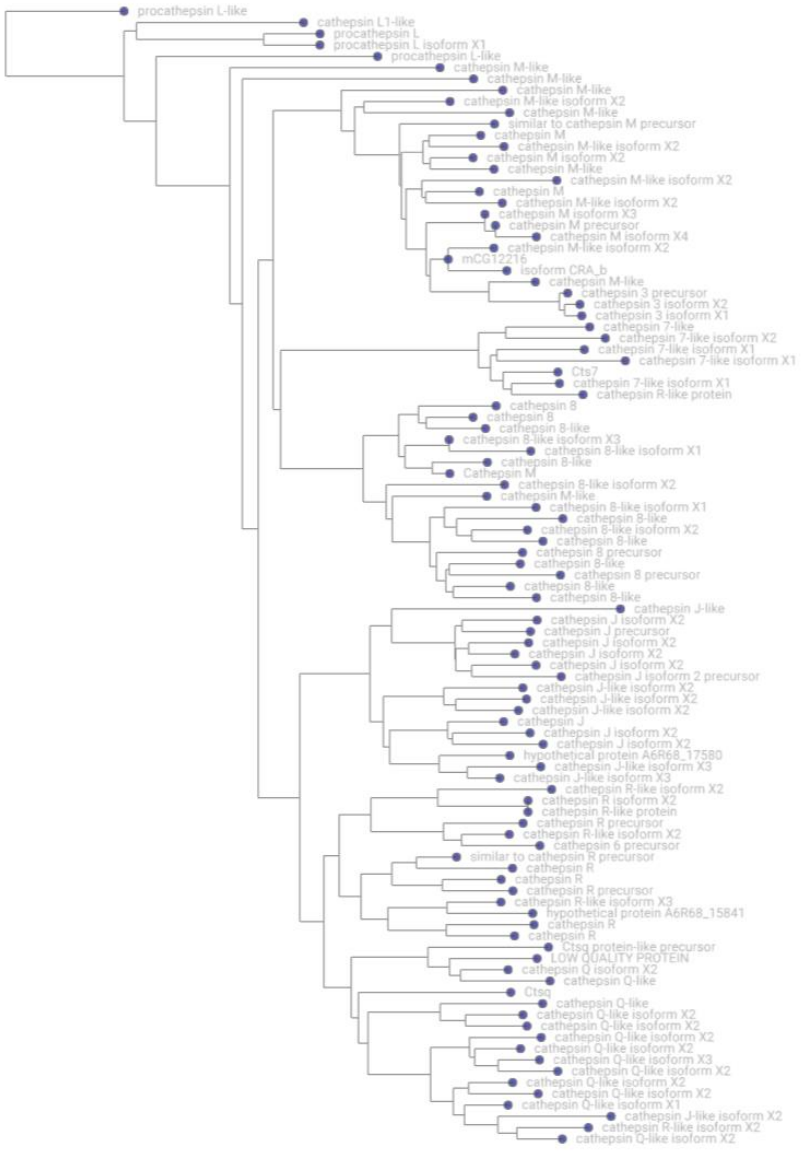


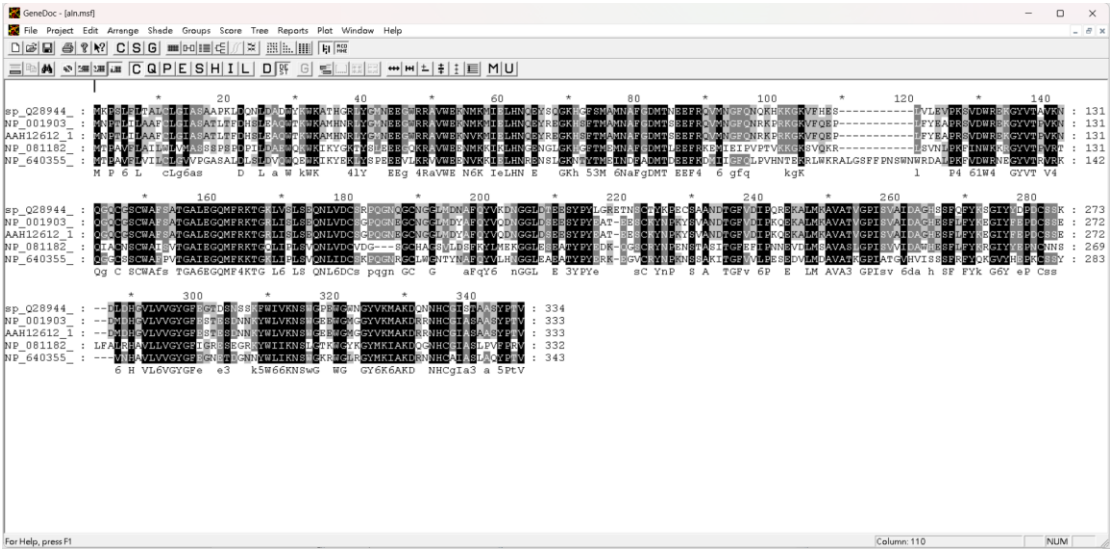
7.

It would depend on the analysis of the alignment and the phylogenetic tree. By examining the alignment, you can observe the degree of similarity between the sequences and identify conserved regions or motifs. The phylogenetic tree will provide insights into the evolutionary relationships among the proteins and their likely evolutionary history.

Based on the clustering of the sequences in the phylogenetic tree, you can infer whether they belong to the same protein family, share a common ancestor, or have diverged into distinct protein groups. Additional information, such as protein annotations or domain analysis, can further help determine the functional characteristics and classification of the proteins.


COBALT
Constraint-based Multiple Alignment Tool
Home
Recent Results
Help
Phylogenetic Tree View
This tree is based on COBALT multiple alignment more...
Reset Tree
Cobalt RID 5TP7G41P212
Number of Seqs 100
Tree method
Fast Minimum Evolution
Max Seq Difference
0.85
Distance
Grishin (protein)
Sequence Label
Sequence Title (if available)
Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download.
Hide legend
Label color map
Blast names color map
rodents
Success
BLAST is a registered trademark of the National Library of Medicine
Support center
Mailing list


procathepsin L-like
cathepsin L1-like
procathepsin L
procathepsin L isoform X1
procathepsin L-like
cathepsin M-like
cathepsin M-like
cathepsin M-like isoform X2
cathepsin M-like
similar to cathepsin M precursor
cathepsin M
cathepsin M-like isoform X2
cathepsin M isoform X2
cathepsin M-like
cathepsin M-like isoform X2
cathepsin M
cathepsin M-like isoform X2
cathepsin M isoform X3
cathepsin M precursor
cathepsin M isoform X4
cathepsin M-like isoform X2
mCG12216
isoform CRA_b
cathepsin M-like
cathepsin 3 precursor
cathepsin 3 isoform X2
cathepsin 3 isoform X1
cathepsin 7-like
cathepsin 7-like isoform X2
cathepsin 7-like isoform X1
cathepsin 7-like isoform X1
Cts7
cathepsin 7-like isoform X1
cathepsin R-like protein
cathepsin 8
cathepsin 8
cathepsin 8-like
cathepsin 8-like isoform X3
cathepsin 8-like isoform X1
cathepsin 8-like
Cathepsin M
cathepsin 8-like isoform X2
cathepsin M-like
cathepsin 8-like isoform X1
cathepsin 8-like
cathepsin 8-like isoform X2
cathepsin 8-like
cathepsin 8 precursor
cathepsin 8-like
cathepsin 8 precursor
cathepsin 8-like
cathepsin 8-like
cathepsin J-like
cathepsin J isoform X2
cathepsin J precursor
cathepsin J isoform X2
cathepsin J isoform X2
cathepsin J isoform 2 precursor
cathepsin J-like isoform X2
cathepsin J-like isoform X2
cathepsin J-like isoform X2
cathepsin J
cathepsin J isoform X2
cathepsin J isoform X2
hypothetical protein A6R68_17580
cathepsin J-like isoform X3
cathepsin J-like isoform X3
cathepsin R-like isoform X2
cathepsin R isoform X2
cathepsin R-like protein
cathepsin R precursor
cathepsin R-like isoform X2
cathepsin 6 precursor
similar to cathepsin R precursor
cathepsin R
cathepsin R
cathepsin R precursor
cathepsin R-like isoform X3
hypothetical protein A6R68_15841
cathepsin R
cathepsin R
Ctsq protein-like precursor
LOW QUALITY PROTEIN
cathepsin Q isoform X2
cathepsin Q-like
Ctsq
cathepsin Q-like
cathepsin Q-like isoform X2
cathepsin Q-like isoform X2
cathepsin Q-like isoform X2
cathepsin Q-like isoform X2
cathepsin Q-like isoform X3
cathepsin Q-like isoform X2
cathepsin Q-like isoform X2
cathepsin Q-like isoform X2
cathepsin Q-like isoform X1
cathepsin J-like isoform X2
cathepsin R-like isoform X2
cathepsin Q-like isoform X2



8.

You can examine the predicted gene structures, their positions, and their annotations. Genscan attempts to identify potential protein-coding genes in a given DNA sequence by using computational methods based on statistical models of gene features. The predicted exons and splice sites can provide insights into the potential gene structure within the cosmid sequence.

To interpret the results, you can look for patterns consistent with known gene structures, such as the presence of start and stop codons, exon-intron boundaries, and conserved regions. Additionally, you can compare the predicted gene structures to known genes or perform further analysis to validate the predictions.

To repeat the exercise with cosmid sequences from different species, you can search for other cosmid entries in the NCBI Nucleotide database using the same procedure outlined above. Obtain the sequence in FASTA format and analyze it using the Genscan server as before.

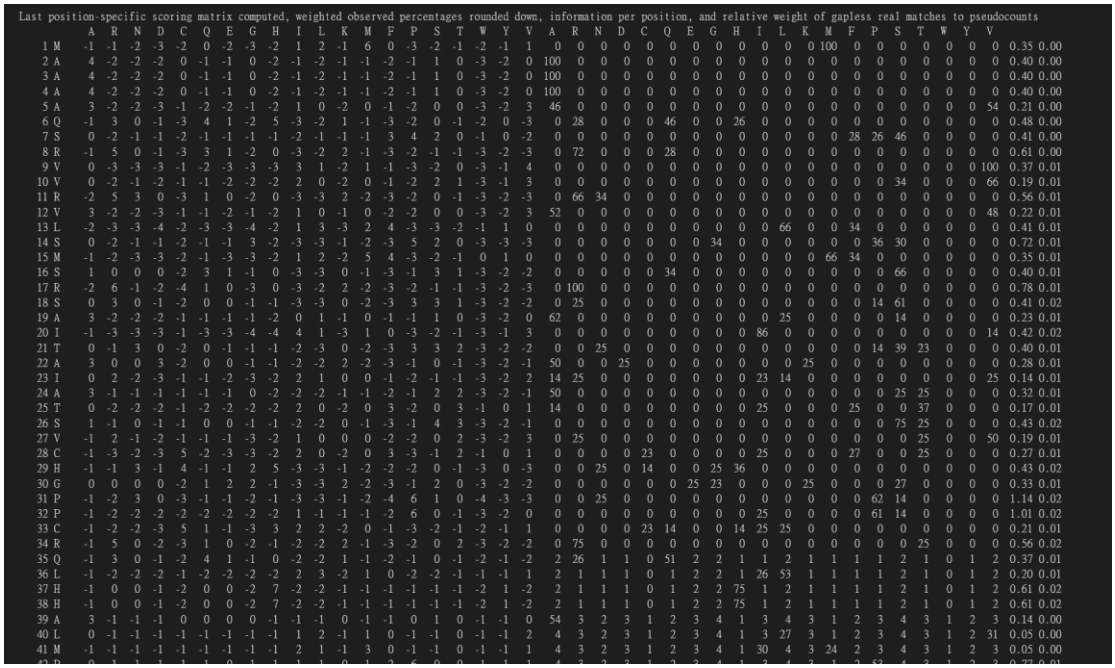
View gene model output: PS | PDF

```
GENSCAN 1.0     Date run: 11-May-123    Time: 10:43:00


Sequence /tmp/05_11_23-10:42:59.fasta : 42435 bp : 38.45% C+G : Isochore 1 ( 0 - 43 C+G%)


Parameter matrix: HumanIso.smat


Predicted genes/exons:


Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..

----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------


 1.01 Init +   7050   7146   97  2  1   55   23   119 0.436   2.62

 1.02 Intr +   8617   8727  111  2  0   74   73    39 0.247   0.43

 1.03 Term +  17174  17400  227  0  2  104   49   125 0.630   6.16

 1.04 PlyA +  20303  20308    6                                1.05


 2.02 PlyA -  22449  22444    6                                1.05

 2.01 Sngl -  34573  34226  348  1  0   65   39   152 0.424   3.79


Suboptimal exons with probability > 1.000


Exnum Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr..

----- ---- - ------ ------ ---- -- -- ---- ---- ----- ----- ------



NO EXONS FOUND AT GIVEN PROBABILITY CUTOFF



Predicted peptide sequence(s):



>/tmp/05_11_23-10:42:59.fasta|GENSCAN_predicted_peptide_1|144_aa

MTLARCIRSGDEEPEGMLVSLTLKRNCLLLDSDGCSKCQGARMVEPQGRRIMGPLVITRR

QVTMGIHLLCLCSHRPAQLLGIVRLITVFSDCGYRKQGLEYPSGMSNNWKEHVDKQSQEQ

ADKKFQEGSSLSWWGAKKYSPPPV


>/tmp/05_11_23-10:42:59.fasta|GENSCAN_predicted_peptide_2|115_aa

MATEAKIDKWDLIKLKSFCTAKETTIRVNGQPTEWEKIFAVYSSDKGLISRIYNELKQIY

KKKTNHPIKKWAKDMNRRFSKEVIYAANRHEKMLIITGHQRNANQNHNEIPSHAS
```

9.

Once the process is complete, we will have PSSM files for each protein sequence, which can be used for further analysis or input into other bioinformatics tools.

Seq1.pssm inside :



Last position-specific scoring matrix computed, weighted observed percentages rounded down, information per position, and relative weight of gapless real matches to pseudocounts

		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V		
1	M	-1	-1	-2	-3	-2	0	-2	-3	-2	1	2	-1	6	0	-3	-2	-1	-2	-1	1	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0.35	0.00
2	A	4	-2	-2	-2	0	-1	-1	0	-2	-1	-2	-1	-1	-2	-1	1	0	-3	-2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.40	0.00
3	A	4	-2	-2	-2	0	-1	-1	0	-2	-1	-2	-1	-1	-2	-1	1	0	-3	-2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.40	0.00
4	A	4	-2	-2	-2	0	-1	-1	0	-2	-1	-2	-1	-1	-2	-1	1	0	-3	-2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.40	0.00
5	A	3	-2	-2	-3	-1	-2	-2	-1	-2	1	0	-2	0	-1	-2	0	0	-3	-2	3	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54	0.21	0.00
6	Q	-1	3	0	-1	-3	4	1	-2	5	-3	-2	1	-1	-3	-2	0	-1	-2	0	-3	0	28	0	0	0	46	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0.48	0.00
7	S	0	-2	-1	-1	-2	-1	-1	-1	-1	-2	-1	-1	-1	3	4	2	0	-1	0	-2	0	0	0	0	0	0	0	0	0	0	0	28	26	46	0	0	0	0	0	0	0.41	0.00
8	R	-1	5	0	-1	-3	3	1	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	0	72	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.61	0.00
9	V	0	-3	-3	-3	-1	-2	-3	-3	-3	3	1	-2	1	-1	-3	-2	0	-3	-1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0.37	0.01
10	V	0	-2	-1	-2	-1	-1	-2	-2	-2	2	0	-2	0	-1	-2	2	1	-3	-1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	66	0.19	0.01
11	R	-2	5	3	0	-3	1	0	-2	0	-3	-3	2	-2	-3	-2	0	-1	-3	-2	-3	0	66	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.56	0.01
12	V	3	-2	-2	-3	-1	-1	-2	-1	-2	1	0	-1	0	-2	-2	0	0	-3	-2	3	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0.22	0.01
13	L	-2	-3	-3	-4	-2	-3	-3	-4	-2	1	3	-3	2	4	-3	-3	-2	-1	1	0	0	0	0	0	0	0	0	0	0	66	0	0	34	0	0	0	0	0	0	0	0.41	0.01
14	S	0	-2	-1	-1	-2	-1	-1	3	-2	-3	-3	-1	-2	-3	5	2	0	-3	-3	-3	0	0	0	0	0	0	0	34	0	0	0	0	0	0	36	30	0	0	0	0	0.72	0.01
15	M	-1	-2	-3	-3	-2	-1	-3	-3	-2	1	2	-2	5	4	-3	-2	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	66	34	0	0	0	0	0	0	0	0.35	0.01
16	S	1	0	0	0	-2	3	1	-1	0	-3	-3	0	-1	-3	-1	3	1	-3	-2	-2	0	0	0	0	0	34	0	0	0	0	0	0	0	0	66	0	0	0	0	0	0.40	0.01
17	R	-2	6	-1	-2	-4	1	0	-3	0	-3	-2	2	-2	-3	-2	-1	-1	-3	-2	-3	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.78	0.01
18	S	0	3	0	-1	-2	0	0	-1	-1	-3	-3	0	-2	-3	3	3	1	-3	-2	-2	0	25	0	0	0	0	0	0	0	0	0	0	0	0	14	61	0	0	0	0	0.41	0.02
19	A	3	-2	-2	-2	-1	-1	-1	-1	-2	0	1	-1	0	-1	-1	1	0	-3	-2	0	62	0	0	0	0	0	0	0	0	25	0	0	0	0	0	14	0	0	0	0	0.23	0.01
20	I	-1	-3	-3	-3	-1	-3	-3	-4	-4	4	1	-3	1	0	-3	-2	-1	-3	-1	3	0	0	0	0	0	0	0	0	0	86	0	0	0	0	0	0	0	0	0	14	0.42	0.02
21	T	0	-1	3	0	-2	0	-1	-1	-1	-2	-3	0	-2	-3	3	3	2	-3	-2	-2	0	0	25	0	0	0	0	0	0	0	0	0	0	0	14	39	23	0	0	0	0.40	0.01
22	A	3	0	0	3	-2	0	0	-1	-1	-2	-2	2	-2	-3	-1	0	-1	-3	-2	-1	50	0	0	25	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0.28	0.01
23	I	0	2	-2	-3	-1	-1	-2	-3	-2	2	1	0	0	-1	-2	-1	-1	-3	-2	2	14	25	0	0	0	0	0	0	0	23	14	0	0	0	0	0	0	0	0	25	0.14	0.01
24	A	3	-1	-1	-1	-1	-1	-1	0	-2	-2	-2	-1	-1	-2	-1	2	2	-3	-2	-1	50	0	0	0	0	0	0	0	0	0	0	0	0	0	25	25	0	0	0	0	0.32	0.01
25	T	0	-2	-2	-2	-1	-2	-2	-2	-2	2	0	-2	0	3	-2	0	3	-1	0	1	14	0	0	0	0	0	0	0	0	25	0	0	0	0	25	0	0	37	0	0	0.17	0.01
26	S	1	-1	0	-1	-1	0	0	-1	-1	-2	-2	0	-1	-3	-1	4	3	-3	-2	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75	25	0	0	0	0.43	0.02
27	V	-1	2	-1	-2	-1	-1	-1	-3	-2	1	0	0	0	-2	-2	0	2	-3	-2	3	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	50	0.19	0.01
28	C	-1	-3	-2	-3	5	-2	-3	-3	-2	2	0	-2	0	3	-3	-1	2	-1	0	1	0	0	0	0	23	0	0	0	0	25	0	0	0	27	0	0	25	0	0	0	0.27	0.01
29	H	-1	-1	3	-1	4	-1	-1	2	5	-3	-3	-1	-2	-2	-2	0	-1	-3	0	-3	0	0	25	0	14	0	0	25	36	0	0	0	0	0	0	0	0	0	0	0	0.43	0.02
30	G	0	0	0	0	-2	1	2	2	-1	-3	-3	2	-2	-3	-1	2	0	-3	-2	-2	0	0	0	0	0	0	0	25	23	0	0	0	25	0	0	0	27	0	0	0	0.33	0.01
31	P	-1	-2	3	0	-3	-1	-1	-2	-1	-3	-3	-1	-2	-4	6	1	0	-4	-3	-3	0	0	25	0	0	0	0	0	0	0	0	0	0	0	62	14	0	0	0	0	1.14	0.02
32	P	-1	-2	-2	-2	-2	-2	-2	-2	-2	1	-1	-1	-1	-2	6	0	-1	-3	-2	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	61	14	0	0	0	0	1.01	0.02
33	C	-1	-2	-2	-3	5	1	-1	-3	3	2	2	-2	0	-1	-3	-2	-1	-2	-1	1	0	0	0	0	0	23	14	0	0	14	25	25	0	0	0	0	0	0	0	0	0.21	0.01
34	R	-1	5	0	-2	-3	1	0	-2	-1	-2	-2	2	-1	-3	-2	0	2	-3	-2	-2	0	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0.56	0.02
35	Q	-1	3	0	-1	-2	4	1	-1	0	-2	-2	1	-1	-2	-1	0	-1	-2	-1	-2	2	26	1	1	0	51	2	2	1	1	2	1	1	1	1	2	1	0	1	2	0.37	0.01
36	L	-1	-2	-2	-2	-1	-2	-2	-2	-2	2	3	-2	1	0	-2	-2	-1	-1	-1	1	2	1	1	1	0	1	2	2	1	26	53	1	1	1	1	2	1	0	1	2	0.20	0.01
37	H	-1	0	0	-1	-2	0	0	-2	7	-2	-2	-1	-1	-1	-1	-1	-1	-2	1	-2	2	1	1	1	0	1	2	2	75	1	2	1	1	1	1	2	1	0	1	2	0.61	0.02
38	H	-1	0	0	-1	-2	0	0	-2	7	-2	-2	-1	-1	-1	-1	-1	-1	-2	1	-2	2	1	1	1	0	1	2	2	75	1	2	1	1	1	1	2	1	0	1	2	0.61	0.02
39	A	3	-1	-1	-1	0	0	0	0	-1	-1	-1	0	-1	-1	0	1	0	-1	-1	0	54	3	2	3	1	2	3	4	1	3	4	3	1	2	3	4	3	1	2	3	0.14	0.00
40	L	0	-1	-1	-1	-1	-1	-1	-1	-1	2	-1	1	0	-1	-1	0	-1	-1	-1	2	4	3	2	3	1	2	3	4	1	3	27	3	1	2	3	4	3	1	2	31	0.05	0.00
41	M	-1	-1	-1	-1	-1	-1	-1	-1	-1	2	1	-1	3	0	-1	-1	0	-1	-1	1	4	3	2	3	1	2	3	4	1	30	4	3	24	2	3	4	3	1	2	3	0.05	0.00

		K	Lambda
Standard Ungapped		0.1395	0.3247
Standard Gapped		0.0410	0.2670
PSI Ungapped		0.1395	0.3247
PSI Gapped		0.0410	0.2670

10.

MMseqs2 (Many-against-Many sequence searching) is a software suite to search and cluster huge protein and nucleotide sequence sets. MMseqs2 can run 10000 times faster than BLAST. At 100 times its speed it achieves almost the same sensitivity. It can perform profile searches with the same sensitivity as PSI-BLAST at over 400 times its speed.

Both PSI-Blast and MMseqs2 are popular bioinformatics tools used for sequence alignment and homology searching. While both tools serve similar purposes, there are differences in their algorithms and performance characteristics. Here's a comparison of their results and execution time:

Results:

PSI-Blast: PSI-Blast is a widely used tool for detecting remote homologs using position-specific scoring matrices (PSSMs). It performs an iterative search process

that builds a PSSM based on the initial alignment results and then searches for additional matches. PSI-Blast is known for its sensitivity in detecting distant homologs, especially when there is low sequence similarity. However, the iterative nature of the algorithm can sometimes lead to false positives.

MMseqs2: MMseqs2 is a versatile sequence search tool that offers various search modes, including sensitive and fast options. It employs a fast indexing algorithm called "minimizer" to quickly identify potential matches. MMseqs2 is known for its speed and scalability, making it suitable for large-scale sequence database searches. It provides competitive sensitivity and can handle diverse sequence datasets efficiently.

Execution Time:

PSI-Blast: The execution time of PSI-Blast can vary depending on several factors, including the size of the database, the number of iterations, and the convergence criteria set. Since PSI-Blast performs iterative searches and builds a PSSM at each step, it generally takes longer to execute compared to other methods. The execution time can range from minutes to hours, depending on the complexity of the search.

MMseqs2: MMseqs2 is designed to be highly efficient and optimized for speed. It utilizes advanced indexing and parallelization techniques to accelerate the search process. As a result, MMseqs2 can achieve significantly faster execution times compared to PSI-Blast. The exact execution time will depend on the specific search mode selected and the size of the database, but it is typically much faster than PSI-Blast, often completing searches within seconds to minutes.

In summary, PSI-Blast is known for its sensitivity in detecting distant homologs but can be slower due to its iterative nature. On the other hand, MMseqs2 offers fast execution times with competitive sensitivity, making it a popular choice for large-scale sequence searches. The choice between the two tools will depend on the specific requirements of your analysis, such as the desired sensitivity, speed, and scale of the search.

The following picture is the pssm output file of the MMseqs2.

1	Query profile of sequence 0																				
2	Pos	Cns	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
3	0	M	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	11	-1	-2	-1	-3	-1	-2	-2	-2	-2
4	1	V	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	-2	-1	-2	-1	-2	-1	-2	7	-2	-2
5	2	T	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	-2	-1	-2	-1	-2	-1	9	-2	-2	-2
6	3	R	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	-2	-1	-2	-1	8	-1	-2	-2	-2	-2
7	4	F	-1	-1	-1	-1	9	-1	-1	-1	-1	-2	-2	-1	-2	-1	-2	-1	-2	-2	-2	-2
8	5	L	-1	-1	-1	-1	-1	-2	-1	-1	-1	7	-2	-1	-2	-1	-2	-1	-2	-2	-2	-2
9	6	G	-2	-1	-1	-1	-1	8	-1	-1	-1	-2	-1	-1	-2	-1	-2	-1	-2	-2	-2	-2
10	7	P	-2	-1	-1	-1	-1	-2	-1	-1	-1	-2	-1	-1	9	-1	-2	-1	-2	-2	-2	-2
11	8	R	-2	-1	-1	-1	-1	-2	-1	-1	-1	-2	-1	-1	-2	-1	8	-1	-2	-2	-2	-2
12	9	Y	-2	-1	-1	-1	-1	-2	-1	-1	-1	-2	-1	-1	-2	-1	-2	-1	-2	-2	-2	10
13	10	R	-2	-1	-1	-1	-1	-2	-1	-1	-1	-2	-1	-1	-2	-1	9	-1	-2	-2	-2	-2