

BIOBANCOS



Trabalho Final 1

- Alinhamentos Locais e Globais
- Matrizes de substituição
- Predição de exons
- Uso de diferentes ferramentas:
 - EMBOS-Needle
 - BLASTp, BLASTn e BLASTx
 - Global Align (BLAST)
 - CLUSTAL OMEGA
 - GENSCAN



Mulheres em
Bioinformática
& Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 1

Objetivos:

- Aprender a fazer e interpretar alinhamentos globais de sequências proteicas usando o Needle do EMBOSS;
- Aprender a usar diferentes matrizes de comparação para melhorar o score de alinhamentos globais de sequências proteicas.

EXERCÍCIO 1

Calcule o alinhamento ideal para as seguintes sequências:

Sequência 1: MTPARGSALS

Sequência 2: MTPVRRSLS

Use o aplicativo EMBOSS Needle:

https://www.ebi.ac.uk/Tools/psa/emboss_needle/ para fazer isso.

Calcule as pontuações para as matrizes de similaridade BLOSUM62, PAM250 e PAM30 usando uma penalidade de GAP opening e uma penalidade de GAP extension de 1.0.

As matrizes de similaridade sugeridas levam a alinhamentos semelhantes ou existem diferenças?

EXERCÍCIO 1

1. Entrar no site: <https://www.ebi.ac.uk/Tools/psa/>

2. Escolher a opção Neddle (EMBOSS)

The screenshot shows the EMBL-EBI website with a dark header bar. On the left, there's a navigation menu with links to 'EMBL-EBI', 'Services', 'Research', 'Training', 'Industry', 'About us', and a search icon. On the right, it shows 'EMBL-EBI' with a green circular logo and 'Hinxton ▾'. Below the header, a large teal section features the title 'Pairwise Sequence Alignment' in white. In the bottom right corner of this teal area, there's a small 'Feedback' link with a speech bubble icon. At the very bottom of the page, a white footer bar contains the text 'Tools > Pairwise Sequence Alignment'.

Tools > Pairwise Sequence Alignment

Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid).

By contrast, **Multiple Sequence Alignment (MSA)** is the alignment of three or more biological sequences of similar length. From the output of MSA applications, homology can be inferred and the evolutionary relationship between the sequences studied.

Global Alignment

Global alignment tools create an end-to-end alignment of the sequences to be aligned.

Needle (EMBOSS)

EMBOSS Needle creates an optimal global alignment of two sequences using the Needleman-Wunsch algorithm.

Launch 

EXERCÍCIO 1

3. Escolher a opção "PROTEIN" e colar a sequência proteica

The screenshot shows the EMBL-EBI website with a dark header bar containing links for EMBL-EBI, Services, Research, Training, Industry, About us, and a search icon. To the right is the EMBL-EBI Hinxton logo. Below the header is a teal banner with the text 'EMBOSS Needle'. Underneath the banner is a navigation bar with links for Input form (which is highlighted), Web services, Help & Documentation, Bioinformatics Tools FAQ, and Feedback. The main content area shows the 'Pairwise Sequence Alignment' page for EMBOSS Needle. It includes a sub-header 'Pairwise Sequence Alignment', a brief description stating 'EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.', and a 'STEP 1 - Enter your protein sequences' section. In this section, there is a dropdown menu labeled 'Enter a pair of' with 'PROTEIN' selected, and a text input field below it containing the sequence 'MTPARGSALS'. A red box highlights the word 'PROTEIN' in the dropdown menu, and another red box highlights the sequence 'MTPARGSALS' in the text input field.

EMBL-EBI Services Research Training Industry About us

EMBL-EBI Hinxton ▾

EMBOSS Needle

Input form Web services Help & Documentation Bioinformatics Tools FAQ Feedback

Tools > Pairwise Sequence Alignment > EMBOSS Needle

Pairwise Sequence Alignment

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

STEP 1 - Enter your protein sequences

Enter a pair of

PROTEIN

sequences. Enter or paste your first **protein** sequence in any supported format:

MTPARGSALS

 Mulheres em Bioinformática & Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 1

4. Colar a segunda sequência proteica na segunda caixa e clicar em "More options"

Or, upload a file: Choose File No file chosen

Use a example sequence | Clear sequence | See more example inputs

AND

Enter or paste your second **protein** sequence in any supported format:

MTPVRRSLS

Or, upload a file: Choose File No file chosen

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

The default settings will fulfill the needs of most users.

More options...

(Click here, if you want to view or change the default settings.)



EXERCÍCIO 1

5. Escolher as opções BLOSUM62, GAP OPEN 1 e GAP EXTEND 1
6. Clicar em "Submit"

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
BLOSUM62	1	1.0	false	10	0.5

STEP 3 - Submit your job

Be notified by email (*Tick this box if you want to be notified by email when the results are available*)

Submit

EXERCÍCIO 1

```
#####
# Program: needle
# Rundate: Sun 17 Apr 2022 12:54:47
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220417-130953
#   -bsequence emboss_needle-I20220417-130953
#   -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
#
#-----#
#-----#
```



EXERCÍCIO 1

7. Escolher as opções PAM250, GAP OPEN 1 e GAP EXTEND 1
8. Clicar em "Submit"

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
PAM 250	1	1.0	false	10	0.5

STEP 3 - Submit your job

Be notified by email (*Tick this box if you want to be notified by email when the results are available*)

EXERCÍCIO 1

```
#####
# Program: needle
# Rundate: Sun 17 Apr 2022 12:58:50
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220417-131
#   -bsequence emboss_needle-I20220417-131
#   -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

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

EXERCÍCIO 1

9. Escolher as opções PAM30, GAP OPEN 1 e GAP EXTEND 1
10. Clicar em "Submit"

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
PAM 30	1	1.0	false	10	0.5

STEP 3 - Submit your job

Be notified by email (*Tick this box if you want to be notified by email when the results are available*)

Submit

EXERCÍCIO 1

```
#####
# Program: needle
# Rundate: Sun 17 Apr 2022 13:12:57
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220417-13
#   -bsequence emboss_needle-I20220417-13
#   -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: EMBOS_001
# 2: EMBOS_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
#
#=====
#-----
```

EXERCÍCIO 1

A introdução de mais GAPs (na matriz PAM30) fez com que aumentasse o escore do alinhamento.

Característica	BLOSUM62	PAM250	PAM30
Identidade	$7/10 = 70\%$	$6/10 = 60\%$	$7/11 = 63.6\%$
Similaridade	$7/10 = 70\%$	$7/10 = 70\%$	$7/11 = 63.6\%$
GAPs	$1/10 = 10\%$	$1/10 = 10\%$	$3/11 = 27.3\%$
Score	31.0	29.0	48.0

EXERCÍCIO 2

Objetivo: Aprender a fazer buscas de sequências proteicas em formato FASTA no NCBI-Protein para logo serem usadas.

EXERCÍCIO 2

Procure o registro do banco de dados Swiss-Prot para o receptor humano 5-hidroxitriptamina 2A no banco de dados de proteínas NCBI [ncbi] e salve a sequência de proteínas no formato FASTA.

EXERCÍCIO 2

National Library of Medicine
National Center for Biotechnology Information

Protein

Create alert Advanced

Species
Animals (8)
Customize ...

Source databases
PDB (0)
RefSeq (0)
 UniProtKB / Swiss-Prot (8)
Customize ...

Sequence length
Custom range...

Molecular weight
Custom range...

Release date
Custom range...

Revision date
Custom range...

[Clear all](#)
[Show additional filters](#)

Summary ▾ 20 per page ▾ Sort by Default order ▾ Send to: ▾

See [2a replication protein 2a](#) in the Gene database
[2a reference sequences Protein \(1\)](#)

See the [results of this search \(6 items\)](#) in our new [Identical Protein Groups](#) database.

Items: 8

Filters activated: UniProtKB / Swiss-Prot. [Clear all](#)

RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; [Short=5-HT-2A](#); AltName:
1. [Full=Serotonin receptor 2A](#)
471 aa protein
Accession: P28223.2 GI: 543727
[PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

RecName: Full=Metabotropic glutamate receptor 2; Short=mGluR2; Flags: Precursor
2. 872 aa protein
Accession: Q14416.2 GI: 76803802
[PubMed](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

1.Ir no site do NCBI e escolher o biobanco "Protein".

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?
PROGRAM=blastp&PAGE_TYPE=BlastSearch
&LINK_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

2.Escrever o nome da proteína de interesse.

3.Escolher a opção Swiss-Prot.

4. Escolher a opção 5HT2A_HUMAN.

EXERCÍCIO 2

5. Verificar que o código da proteína seja
Swiss-Prot AN P28223; ID 5HT2A_HUMAN.

6. Salvar o arquivo.

FASTA ▾

Send to: ▾

**RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; Short=5-HT-2A;
AltName: Full=Serotonin receptor 2A**

UniProtKB/Swiss-Prot: P28223.2

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>sp|P28223.2|5HT2A_HUMAN RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-  
2; Short=5-HT-2A; AltName: Full=Serotonin receptor 2A  
MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCLSLH  
LQEKNWSALLTAVVIIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIADMLLGFLVMPVSMLTILYGYR  
WPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPI  
PVFGLQDDSKVFKEGSCLLADDNFVLIGSFVSFFIPLTIMVITYFLTIKSILQKEATLCVSDLGTRAKLAS  
FSFLPQSSLSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLFVVMMCPFFITNIMAVICK  
ESCNEVEDVIGALLNVFWIWIGYLSSAVNPLVYTLFNKYRSAFSRYIQCQYKENKKPLQLILVNTIPALAYK  
SSQLQMGQKKNSKQDAKTTNDNSMVALGKQHSEEASKDNSDGVNEKVSCV
```



EXERCÍCIO 3

Objetivo: Aprender a fazer e interpretar alinhamentos de sequências proteicas com a base de dados do NCBI usando o BLASTp.

EXERCÍCIO 3

Com a sequência salva do 7 Exercício 3.2, execute uma pesquisa BLASTp para sequências semelhantes no banco de dados de proteínas não redundantes do NCBI.

Faça isso indo para a página NCBI-BLAST [ncbi-blast]:

<https://blast.ncbi.nlm.nih.gov/Blast.cgi?>

PAGE=Proteins&PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=

Quantas sequências semelhantes são encontradas?

Quais informações podem ser extraídas do gráfico na página de resultados?

EXERCÍCIO 3

1. Verificar que o biobanco seja BLASTp.
2. Colar a sequência na caixa principal ou escolher o arquivo.

The screenshot shows the National Library of Medicine BLAST search interface. At the top, the NIH logo and "National Library of Medicine" are displayed, along with the subtitle "National Center for Biotechnology Information". Below this, the title "BLAST® » blastp suite" is shown. A navigation bar at the top has five tabs: "blastn", "blastp" (which is highlighted with a red box), "blastx", "tblastn", and "tblastx". To the right of the tabs, the text "Standard Protein BLAST" is displayed. Below the tabs, a description of the BLASTP program is provided: "BLASTP programs search protein databases using a protein query. [more...](#)". The main search area is titled "Enter Query Sequence". It contains a text input field with the sequence: >sp|P28223.2|5HT2A_HUMAN RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; Short=5-HT-2A; AltName: Full=Serotonin receptor 2A MDILCEENTSLSSTTNQLNDDTRLYSNDNSGEANTSDAFNWTVDSENRT NLSCEGCLSPSCLSLH. This sequence is also highlighted with a red box. To the right of the sequence input, there are fields for "Query subrange" with "From" and "To" inputs, both currently empty. Below the sequence input, there are options to "Or, upload file" (with a "Choose File" button showing "No file chosen") and a "Job Title" input field. A note below the job title says "Enter a descriptive title for your BLAST search". At the bottom left, there is a checkbox for "Align two or more sequences".

EXERCÍCIO 3

3. Verificar que o database seja "non-redundant protein sequences"

4. Clicar em BLAST.

Standard

Database Non-redundant protein sequences (nr)

Organism exclude

Optional

Exclude Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Optional

Entrez Query

Optional

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 using **Blastp (protein-protein BLAST)** Show results in a new window



EXERCÍCIO 3

5. Ler em que consiste o "All non-redundant Protein Sequences GenBank"

Title: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Molecule Type: Protein
Update date: 2022/04/18
Number of sequences: 467912207

EXERCÍCIO 3

6. Observar os resultados

- O BLAST encontrou mais de 100 sequências semelhantes;
- O comprimento da query é de 471 aa;

National Library of Medicine
National Center for Biotechnology Information [Log in](#)

BLAST® » blastp suite » results for RID-5T2CUUPM013 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title	sp P28223.2 5HT2A_HUMAN RecName: Full=5-hydroxytrypt	Filter Results
RID	5T2CUUPM013	Search expires on 04-19 05:41 am Download All
Program	BLASTP	Citation
Database	nr	See details
Query ID	lcl Query_33350	
Description	sp P28223.2 5HT2A_HUMAN RecName: Full=5-hydroxytryp...	
Molecule type	amino acid	
Query Length	471	
Other reports	Distance tree of results Multiple alignment MSA viewer	

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ [Add organism](#)

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)



EXERCÍCIO 3

7. Escolher a primeira opção

Descriptions		Graphic Summary	Alignments	Taxonomy										
Sequences producing significant alignments										Download	Select columns	Show 100	?	
										GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer
	Description		Scientific Name		Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len				
					▼	▼	▼	▼	▼	▼				
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform 1 [Homo sapiens]		Homo sapiens		868	868	100%	0.0	100.00%	471	NP_000612.1			
<input checked="" type="checkbox"/>	5-HT2A receptor [synthetic construct]		synthetic construct		868	868	100%	0.0	99.79%	471	QAR20718.1			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform X1 [Pan paniscus]		Pan paniscus		868	868	100%	0.0	99.79%	471	XP_003811462.1			
<input checked="" type="checkbox"/>	5-HT2A receptor [synthetic construct]		synthetic construct		867	867	100%	0.0	99.79%	471	QAR20717.1			
<input checked="" type="checkbox"/>	HTR2A isoform 2 [Pongo abelii]		Pongo abelii		867	867	100%	0.0	99.79%	471	PNJ60471.1			
<input checked="" type="checkbox"/>	5-HT2A receptor [synthetic construct]		synthetic construct		865	865	100%	0.0	99.58%	471	QAR20719.1			
<input checked="" type="checkbox"/>	PREDICTED: 5-hydroxytryptamine receptor 2A isoform X1 [Colobus angolensis palliatus]	Colobus angolensis palliatus			864	864	100%	0.0	99.36%	471	XP_011795149.1			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A [Pongo abelii]		Pongo abelii		863	863	100%	0.0	99.58%	471	NP_001126918.1			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform X1 [Macaca fascicularis]		Macaca fascicularis		863	863	100%	0.0	99.36%	471	XP_005585885.2			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform X1 [Chlorocebus sabaeus]		Chlorocebus sabaeus		863	863	100%	0.0	99.15%	471	XP_007958563.2			
<input checked="" type="checkbox"/>	PREDICTED: 5-hydroxytryptamine receptor 2A isoform X1 [Cercocebus atys]		Cercocebus atys		862	862	100%	0.0	99.15%	471	XP_011919716.1			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A [Papio anubis]		Papio anubis		862	862	100%	0.0	99.15%	471	XP_021785301.2			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A [Macaca mulatta]		Macaca mulatta		862	862	100%	0.0	99.15%	471	NP_001028138.1			
<input checked="" type="checkbox"/>	PREDICTED: 5-hydroxytryptamine receptor 2A isoform X1 [Macaca fascicularis]		Macaca fascicularis		862	862	100%	0.0	99.15%	471	XP_005585885.1			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform X1 [Saimiri boliviensis boliviensis]		Saimiri boliviensis boliviensis		861	861	100%	0.0	98.73%	471	XP_010328359			
<input checked="" type="checkbox"/>	5-hydroxytryptamine receptor 2A isoform X1 [Piliocolobus tephrosceles]		Piliocolobus tephrosceles		860	860	100%	0.0	98.94%	471	XP_023043159			

EXERCÍCIO 3

8. Observar os resultados:

- 471/471 de identidades e 0 GAP (100% de identidade);
- Corresponde à mesma entrada da mesma sequência (**NP_001365853.1 isoforma 1 do receptor 2A de 5-hidroxitriptamina [Homo sapiens]**).

[Download](#) ▾ [GenPept](#) [Graphics](#) ▾ [Next](#) [Previous](#) [Descriptions](#)

5-hydroxytryptamine receptor 2A isoform 1 [Homo sapiens]
Sequence ID: [NP_000612.1](#) Length: 471 Number of Matches: 1
[See 8 more title\(s\)](#) ▾ [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 471 [GenPept](#) [Graphics](#) ▾ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
868 bits(2244)	0.0	Compositional matrix adjust.	471/471(100%)	471/471(100%)	0/471(0%)

Query 1 MDILCEEntsstsnnslmqlnDDTRLYSNDNSGEANTSDAFNWTVDSENRTN1scegc 60
Sbjct 1 MDILCEENTSLSSTTNQLNDDTRLYSNDNSGEANTSDAFNWTVDSENRTNLSCEGC 60

Query 61 lpsclsllh1QEKNWSSLTAVVIIITIAGNLVIMAVSLEKKLQNATNYFLMSLAIAD 120
Sbjct 61 LSPSCLSLLLQEKNWSSLTAVVIIITIAGNLVIMAVSLEKKLQNATNYFLMSLAIAD 120

Query 121 MLLGFLVMPVSMLTILYGYRWPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNP 180
Sbjct 121 MLLGFLVMPVSMLTILYGYRWPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNP 180

Query 181 IHHSRFNSRTKAFLKIIAVWTISVGISMPIPVGFLQDDSKVFKEGSCLLADDNFVLIGSF 240
Sbjct 181 IHHSRFNSRTKAFLKIIAVWTISVGISMPIPVGFLQDDSKVFKEGSCLLADDNFVLIGSF 240

Query 241 VSFFIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASfsflpqsslsseklfqrsih 300
Sbjct 241 VSFFIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLASFSFLPQSSLSSEKLFRSH 300

Query 301 REPGSYTGRRTMQSISNEQACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEVIGA 360
Sbjct 301 REPGSYTGRRTMQSISNEQACKVLGIVFFLFVVMWCPFFITNIMAVICKESCNEVIGA 360

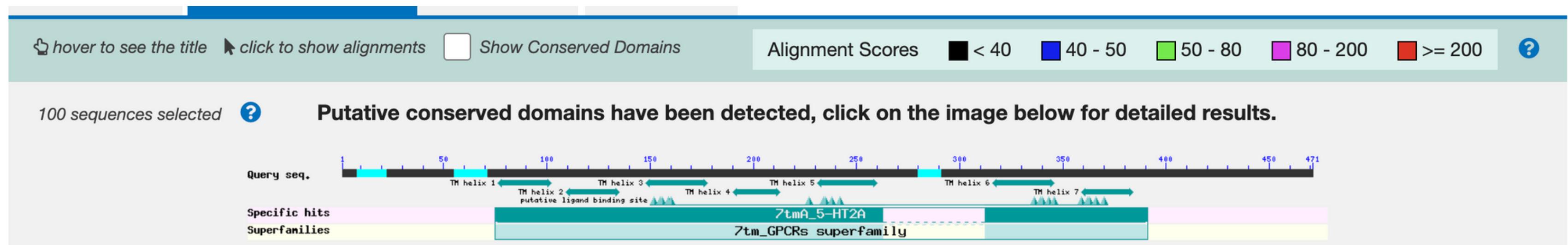
Query 361 LLNVFWIWGYLSSAVNPLVYTLFNKYRSASFYSRYIQCQYKENKKPLQLILVNTIPALAYK 420
Sbjct 361 LLNVFWIWGYLSSAVNPLVYTLFNKYRSASFYSRYIQCQYKENKKPLQLILVNTIPALAYK 420

Query 421 SSQLOMGQKKNNSKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV 471
Sbjct 421 SSQLOMGQKKNNSKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV 471

EXERCÍCIO 3

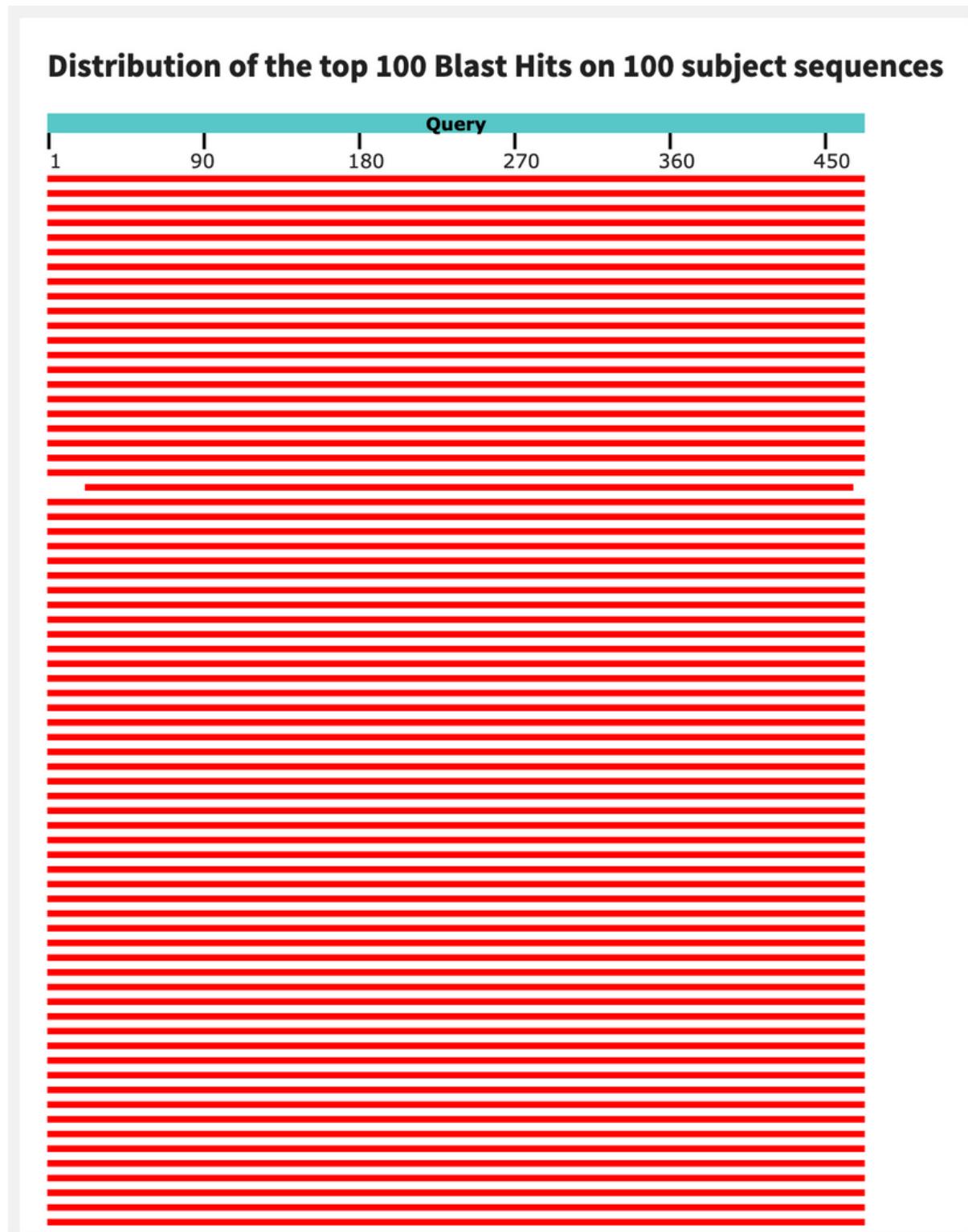
9. Observar os resultados:

- Possui 7 estruturas helicoidais, 3 sítios putativos conservados (receptor de serotonina subtipo 2^a), 7 receptores transmembrana (família da rodopsina e proteína semelhante ao receptor de quimiocina CC) e 4 sítios de ligação de ligantes putativos.



EXERCÍCIO 3

10. Observar a qualidade do alinhamento da query com todas as sequências encontradas.



EXERCÍCIO 4

Objetivo: Aprender a fazer e interpretar alinhamentos de sequências nucleotídicas a partir do BLASTn e do BLASTx.

EXERCÍCIO 4

No banco de dados de nucleotídeos do NCBI, procure a entrada com o AB037513 e salve a sequência de nucleotídeos no formato FASTA. A sequência codifica um receptor 5HT2 humano.

Em seguida, execute pesquisas BLAST usando BLASTn e BLASTx contra o banco de dados do genoma do organismo *Drosophila melanogaster*.

BLASTn: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

tBLASTx: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastx&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab&LAST_PAGE=blastn

Quantas sequências semelhantes são encontradas em cada caso?

O que se pode afirmar sobre a qualidade dos acessos?

Quais são as diferenças entre os dois programas BLASTn e BLASTx?

Como os respectivos resultados de pesquisa se originam?

EXERCÍCIO 4

1. No NCBI - Nucleotide, escrever o nome da sequência nucleotídica AB037513.

2. Clicar em FASTA.

The screenshot shows the NCBI Nucleotide search results for sequence AB037513. The search bar at the top has 'Nucleotide' selected and 'AB037513' entered. Below the search bar, the sequence information is displayed:

Homo sapiens HTR2A gene for 5-hydroxytryptamine (serotonin) receptor 2A, partial cds

GenBank: AB037513.1

FASTA [Graphics](#)

Go to: ▾

LOCUS AB037513 736 bp DNA linear PRI 26-JUL-2016

DEFINITION Homo sapiens HTR2A gene for 5-hydroxytryptamine (serotonin) receptor 2A, partial cds.

ACCESSION AB037513

VERSION AB037513.1

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Kitano,T., Kobayakawa,H. and Saitou,N.

TITLE Silver Project

JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 736)

AUTHORS Kitano,T., Kobayakawa,H. and Saitou,N.

TITLE Direct Submission

On the right side, there are several links and options:

- Send to: ▾
- Change region shown
- Customize view
- Analyze this sequence
- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence
- Articles about the HTR2A gene
 - Association of 5-HTR2A T102C and A-1438G polymorphisms with clinical [Neurosci Lett. 2022]
 - Is there an association of genetic polymorphisms of the catechol-O-methyltransferase [Arch Oral Biol. 2022]
 - Serotonin 5-HT_{2A} receptor activity mediates adipocyte differentiation [Sci Rep. 2021]
- See all...



EXERCÍCIO 4

3. Salvar em formato FASTA a sequência escolhida.

Homo sapiens HTR2A gene for 5-hydroxytryptamine (serotonin) receptor 2A, partial cds

GenBank: AB037513.1

[GenBank](#) [Graphics](#)

```
>AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytryptamine (serotonin) receptor
2A, partial cds
CAGGACGATTCGAAGGTCTTAAGGAGGGAGTTGCTTACTCGCCGATGATAACTTGTCTGATCGGCT
CTTTGTGTCACTTTCATTCCCTAACCATCATGGTGATCACCTACTTCTAACTATCAAGTCACTCCA
GAAAGAAGCTACTTGTGTGAAGTGATCTGGCACACGGGCCAAATTAGCTTCTTCAGCTTCCTCCCT
CAGAGTTCTTGTCTTCAGAAAAGCTCTCCAGCGGTGATCCATAGGGAGCCAGGGTCTACACAGGCA
GGAGGACTATGCAGTCCATCAGCAATGAGCAAAAGGCATGCAAGGTGCTGGGCATCGTCTTCTTCTGTT
TGTGGTATGTGGTGCCTTCTTCATCACAAACATCATGGCCGTATCTGCAAAGAGTCCTGCAATGAG
GATGTCATTGGGCCCTGCTCAATGTGTTGGATCGTTATCTCTCTTCAGCAGTCAACCCACTAG
TCTACACACTGTTCAACAAGACCTATAGGTCAGCCTTCACGGTATATTCAAGTGTCAAGTACAAGGAAAA
CAAAAAACCATTGCAGTTAATTAGTGAACACAATACGGCTTGGCTACAAGTCTAGCCAACTTCAA
ATGGGACAAAAAAAGAATTCAAAGCAAGATGCCAAGACAACAGATAATGACTGCTCAATGGTTGCTCTAG
GAAAGCAGCATTCTGAAGAGGCTCTAAAGACAATA
```



EXERCÍCIO 4

4. Verificar o biobanco BLASTp e colar a sequência nucleotídica na caixa ou adicionar o arquivo
5. Escolher a opção "RefSeq Genome Database"

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite

Home Recer

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) AGTCTAGCCAACTTCAA
ATGGGACAAAAAAAAGAATTCAAAGCAAGATGCCAAGACAACAGATAATGACT
GCTCAATGGTGCTCTAG
GAAAGCAGCATTCTGAAGAGGCTCTAAAGACAATA

Query subrange

From
To

Or, upload file No file chosen

Job Title AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytryptamine...
Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Standard databases (nr etc.) rRNA/ITS databases Genomic + transcript databases Betacoronavirus

Organism RefSeq Genome Database (refseq_genomes)

Exclude Enter organism name or id--completions will be suggested exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Models (XM/XP) Uncultured/environmental sample sequences

EXERCÍCIO 4

6. Ler em que consiste o "RefSeq Genome Database"

Title: RefSeq Genome Database

Description: This database contains NCBI Refseq genomes across all taxonomy groups. It contains only the top-level sequences, i.e., only the longest sequences representing any given part of the genomes are included (for example, while the assembled chromosomal sequence is included, any shorter sequences such as contigs that are used to assemble this chromosome sequence are not included).

Molecule Type: Genomic

Update date: 2022/04/17

Number of sequences: 39977935



EXERCÍCIO 4

7. Escolher a opção "*Drosophila melanogaster*" como organismo de busca.

8. Clicar em BLAST.

Organism
Optional

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

[YouTube](#) Create custom database

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST

Search **database Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

Show results in a new window

EXERCÍCIO 4

9. Repetir os mesmos passos para BLASTx.

10. Selecionar a opção *Nonredundant protein sequences*

BLAST® » blastx

Translated BLAST: blastx

blastn blastp **blastx** tblastn tblastx

BLASTX search protein databases using a translated nucleotide query. [more...](#)

Enter Query Sequence

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

>AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytryptamine (serotonin) receptor 2A, partial cds
CAGGACCGATTCAAAGGTCTTAAGGAGGGAGTTGCTTACTCGCCGATGATA
ACTTTGTCCTGATCGGCT

Query subrange [?](#)

From
To

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

Genetic code

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

Align two or more sequences [?](#)

Choose Search Set

Database

Organism [Optional](#) Drosophila melanogaster (taxid:7227) 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

BLAST

Search database nr using Blastx (search protein databases using a translated nucleotide query)
 Show results in a new window



EXERCÍCIO 4

11. Observar que não houveram resultados significativos para BLASTn.

National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite » results for RID-5T58X0A6013

Home Recent Results Saved Strategies Help

< Edit Search Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Your search is limited to records that include: Drosophila melanogaster (taxid:7227)

Job Title AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytrypt Filter Results

RID 5T58X0A6013 Search expires on 04-19 06:30 am Download All

Program Citation

Database nt See details

Query ID lcl|Query_10627

Description AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytry ...

Molecule type dna

Query Length 736

Other reports

Percent Identity [] to [] E value [] to [] Query Coverage [] to []

Filter Reset

No significant similarity found. For reasons why, click here



EXERCÍCIO 4

12. Observar os resultados com BLASTx

National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastx » results for RID-5T5KM2CT013

Home Recent Results Saved Strategies Help

◀ Edit Search Save Search Search Summary ▾

ⓘ How to read this report? 🎥 BLAST Help Videos ⏪ Back to Traditional Results Page

ⓘ Your search is limited to records that include: Drosophila melanogaster (taxid:7227)

Job Title AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytrypt...

RID 5T5KM2CT013 Search expires on 04-19 06:36 am Download All ▾

Program BLASTX ⓘ Citation ▾

Database nr See details ▾

Query ID Icl|Query_94748

Description AB037513.1 Homo sapiens HTR2A gene for 5-hydroxytrypt...

Molecule type dna

Query Length 736

Other reports ⓘ

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity E value Query Coverage

to to to to

Filter Reset



EXERCÍCIO 4

13. Observam-se mais de 100 resultados positivos com BLASTx

Sequences producing significant alignments										Download	Select columns	Show 100	?
										GenPept	Graphics		
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession				
<input checked="" type="checkbox"/>	HL07802p [Drosophila melanogaster]	Drosophila melanogaster	103	103	39%	6e-26	45.36%	295	AAL28587.1				
<input checked="" type="checkbox"/>	MIP06036p [Drosophila melanogaster]	Drosophila melanogaster	103	103	39%	2e-25	46.39%	393	ACN32192.1				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2B, isoform F [Drosophila melanogaster]	Drosophila melanogaster	103	103	41%	7e-25	45.28%	904	NP_001287238.1				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2B, isoform E [Drosophila melanogaster]	Drosophila melanogaster	103	103	41%	7e-25	45.28%	947	NP_001262373.1				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2A, isoform B [Drosophila melanogaster]	Drosophila melanogaster	103	103	39%	1e-24	45.36%	930	NP_730859.1				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2A, isoform F [Drosophila melanogaster]	Drosophila melanogaster	103	103	39%	1e-24	45.36%	916	NP_001163506.2				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2A, isoform H [Drosophila melanogaster]	Drosophila melanogaster	103	103	39%	1e-24	45.36%	917	NP_001097684.2				
<input checked="" type="checkbox"/>	serotonin receptor 5-HT2 subtype [Drosophila melanogaster]	Drosophila melanogaster	97.1	97.1	37%	2e-22	46.15%	868	CAA57429.1				
<input checked="" type="checkbox"/>	5-hydroxytryptamine (serotonin) receptor 2A, isoform A [Drosophila melanogaster]	Drosophila melanogaster	97.1	97.1	37%	2e-22	46.15%	868	NP_524223.2				
<input checked="" type="checkbox"/>	octopamine beta2 receptor, isoform F [Drosophila melanogaster]	Drosophila melanogaster	87.8	87.8	71%	2e-19	27.72%	630	NP_001163596.2				
<input checked="" type="checkbox"/>	octopamine beta2 receptor, isoform A [Drosophila melanogaster]	Drosophila melanogaster	86.7	86.7	71%	5e-19	27.72%	536	NP_001034049.1				
<input checked="" type="checkbox"/>	G-protein coupled octopamine receptor [Drosophila melanogaster]	Drosophila melanogaster	83.2	83.2	71%	7e-18	27.17%	536	CAI56430.1				
<input checked="" type="checkbox"/>	RH07250p1 [Drosophila melanogaster]	Drosophila melanogaster	79.3	79.3	67%	2e-16	27.98%	604	AEO17899.1				
<input checked="" type="checkbox"/>	5-HT1B [Drosophila melanogaster]	Drosophila melanogaster	78.2	78.2	35%	2e-16	37.93%	320	AAS74383.1				
<input checked="" type="checkbox"/>	5-HT1B [Drosophila melanogaster]	Drosophila melanogaster	78.2	78.2	35%	2e-16	37.93%	318	AAS74456.1				
<input checked="" type="checkbox"/>	5-HT1B [Drosophila melanogaster]	Drosophila melanogaster	78.2	78.2	35%	2e-16	37.93%	316	AAS74470.1				

EXERCÍCIO 4

14. Observar o primeiro resultado: Mostra 3 ranges

[Download](#) ▾ [GenBank](#) [Graphics](#) Sort by: E value ▾

▼ Next ▲ Previous ◀ Descriptions

Drosophila melanogaster 5-hydroxytryptamine (serotonin) receptor 2A (5-HT2A), transcript variant F, mRNA

Sequence ID: [NM_001170035.2](#) Length: 3541 Number of Matches: 3

Range 1: 2853 to 2993 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps	Frame
69.3 bits(145)	2e-28	21/47(45%)	38/47(80%)	0/47(0%)	+1/+3

Query 436 LLNVFWIWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENKKPLQ 576
++N +W+GY+SS +NP++YT+FN+T+R+AF R ++C K + +PL+
Sbjct 2853 VVNNTCLWLGYVSSTINPIIYTIFNRTFRAAFIRLLKCNCRKSGRPLR 2993

Related Information

[Gene](#) - associated gene details
[PubChem BioAssay](#) - bioactivity screening
[Genome Data Viewer](#) - aligned genomic context

Range 2: 2721 to 2834 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Positives	Gaps	Frame
58.8 bits(122)	2e-28	22/38(58%)	28/38(73%)	0/38(0%)	+1/+3

Query 295 SISNEQKACKVLGIVFFLFVVMWCPFFITNIMAVICKE 408
+++ EQKA KVLG+VFF FV+ W PFFI NI+ C E
Sbjct 2721 AVATEQKATKVLGLVFFTFLVLCWSPFFILNIIFAACPE 2834

Range 3: 1725 to 1841 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Positives	Gaps	Frame
45.5 bits(93)	2e-28	18/39(46%)	28/39(71%)	0/39(0%)	+1/+3

Query 34 CLLADDNFVLIGSFVSFFIPLTIMVITYFLTIKSILQKEA 150
C++ + F + GS V+F+IP+ +MV TY LTI L+K+A
Sbjct 1725 CVINNRAFFVFGSLVAFYIPMLMMVTTYALTIPLLRKKA 1841



EXERCÍCIO 4

15. Interpretar os resultados obtidos:

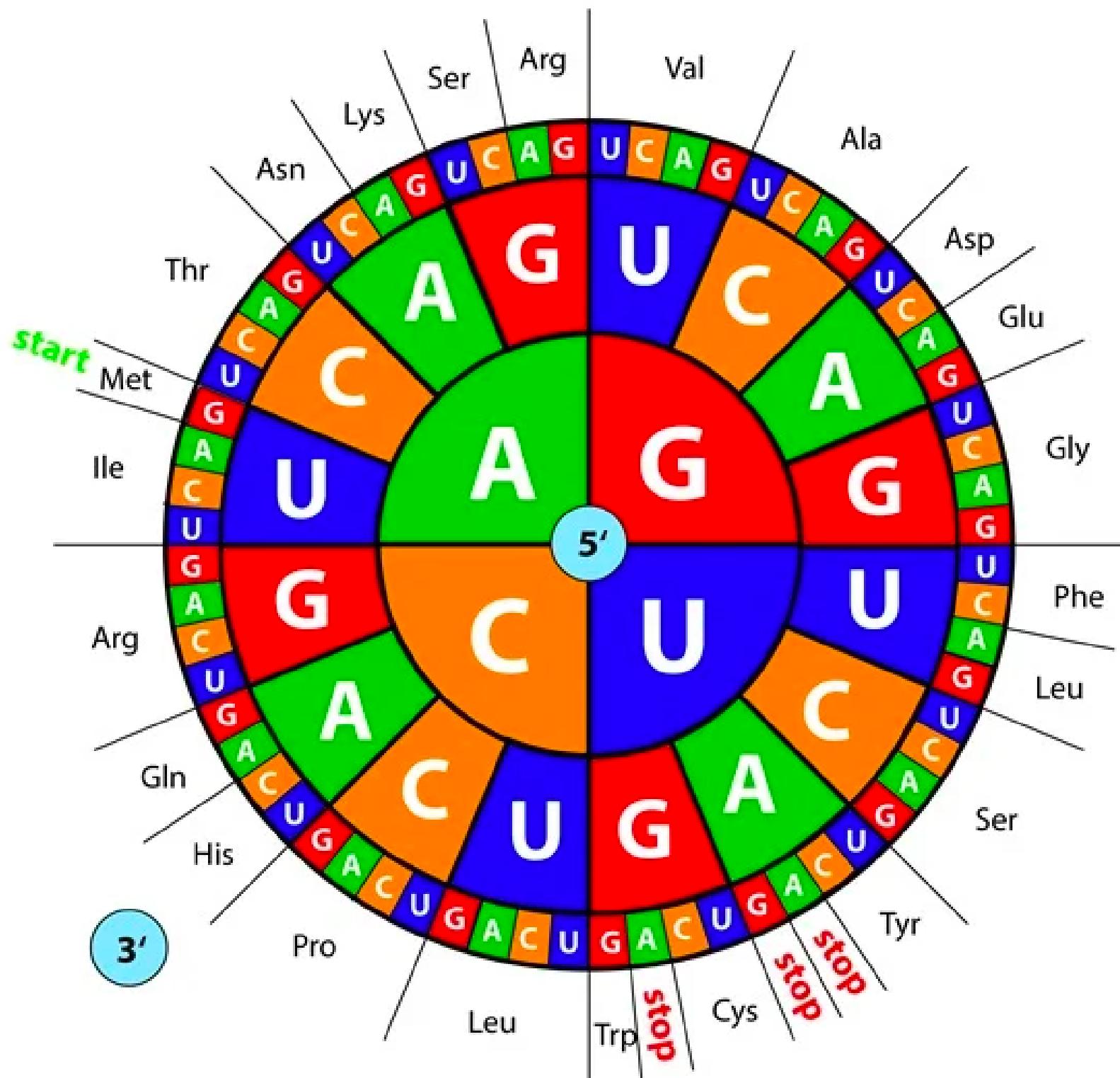
A busca BLASTn mostra de que nenhuma similaridade significativa foi encontrada.

Com BLASTx, no entanto, mais de 100 registros foram encontrados e alguns deles apresentam uma alta significância.

A discrepância entre os resultados se deve a diferenças em como o BLASTn e o BLASTx executam buscas e o uso de códons entre as duas espécies (*Drosophila melanogaster* e *Homo sapiens*). Enquanto o BLASTn realiza uma comparação simples a nível de nucleotídeos, o BLASTx trabalha primeiramente a nível de proteína traduzindo a sequência nucleotídica de consulta em todos os 6 quadros de leitura e depois comparando essas 6 proteínas teóricas com um banco de dados de proteínas.

Como o código genético é degenerado, um aminoácido pode ser codificado por diferentes trincas de códons. O uso de códons entre *D. melanogaster* e *H. sapiens* é tão diferente que nenhuma boa concordância foi encontrada a nível de nucleotídeos.

EXERCÍCIO 4



©magnetix / Shutterstock.com

EXERCÍCIO 4

Característica	BLASTn	<u>BLASTx</u>
Sequências similares	Nenhuma semelhança significativa encontrada	38/47 (80%)
Qualidade dos hits	-	Alta
Diferenças	<p>Significado: n = nucleotídeo Entrada: Sequência de nucleotídeos Banco de dados: Sequências de nucleotídeos Frequência de uso: Mais comum</p>	<p>Significado: x = sequência de nucleotídeos traduzida pesquisada contra sequências de proteínas Entrada: Sequências de nucleotídeos Banco de dados: Sequências de proteínas (traduzidas automaticamente para uma sequência proteica antes da pesquisa) Frequência de uso: Menos comum Vantagens:</p> <ul style="list-style-type: none"> - Se uma sequência de nucleotídeos for traduzida antes da busca é mais provável que encontre acertos mais precisos do que apenas uma busca em BLASTn (as sequências de proteínas são evolutivamente mais conservadas do que as sequências de nucleotídeos); - Os acertos muito provavelmente são proteínas já foram anotados (função da proteína do gene sequenciado)

EXERCÍCIO 5

Objetivo:

Aprender a fazer e interpretar alinhamentos globais usando o *Global Align* do BLAST

EXERCÍCIO 5

Realize um alinhamento global das sequências de proteínas gi|543727 e gi|10726392 usando *Global Align* em *Specialized Searches* na página NCBI-BLAST [blast].

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?
PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln
&LINK_LOC=BlastHomeLink](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink)

Os ANs podem ser inseridos diretamente após a seleção de *Protein*, para que não sejam necessárias mais consultas ao banco de dados.

As duas sequências são o já mencionado receptor 5HT2 humano e seu ortólogo em *Drosophila melanogaster*.

Como o resultado pode ser interpretado?

EXERCÍCIO 5

1. Escolher a opção de Global Align

Standalone and API BLAST



Download BLAST

Get BLAST databases and executables



Use BLAST API

Call BLAST from your application



Use BLAST in the cloud

Start an instance at a cloud provider

Specialized searches

SmartBLAST



Find proteins highly similar to your query

Primer-BLAST



Design primers specific to your PCR template

Global Align



Compare two sequences across their entire span (Needleman-Wunsch)

CD-search



Find conserved domains in your sequence

IgBLAST



Search immunoglobulins and T cell receptor sequences

VecScreen



Search sequences for vector contamination

CDART



Find sequences with similar conserved domain architecture

Multiple Alignment



Align sequences using domain and protein constraints

MOLE-BLAST



Establish taxonomy for



Mulheres em
Bioinformática
& Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 5

2.Escolher a opção "Protein" - 3. Adicionar os códigos das sequências - 4. Clicar em BLAST

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® > Global Alignment

Needleman-Wunsch Global Align Protein Sequences

Nucleotide Protein

Enter Query Sequence

Enter accession number, gi, or FASTA sequence Clear

Query subrange From
 To

Or, upload file No file chosen

Job Title

Enter a descriptive title for your BLAST search

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence Clear

Subject subrange From
 To

Or, upload file No file chosen

Show results in a new window

EXERCÍCIO 5

5. Clicar na opção do alinhamento obtido.

BLAST® » Global Alignment » results for RID-5TBEWCBC114

Home Recent Results Saved Strategies

[Edit Search](#) Save Search Search Summary ▾ [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title	gi 543727
RID	5TBEWCBC114 Search expires on 04-19 08:16 am Download All ▾
Program	Needleman-Wunsch alignment of two sequences Citation ▾
Query ID	P28223.2 (amino acid)
Query Descr	RecName: Full=5-hydroxytryptamine receptor 2A; Short=F...
Query Length	471
Subject ID	AAF54255.2 (amino acid)
Subject Descr	CG8007, isoform A [Drosophila melanogaster]
Subject Length	416

Descriptions Graphic Summary Alignments Dot Plot

Sequences producing significant alignments Download Manage columns Show 100 ?

select all 1 sequences selected GenPept Graphics Multiple alignment MSA Viewer

Description	Score	Percent Ident	Accession
CG8007, isoform A [Drosophila melanogaster]	204	20.00%	AAF54255.2

Mulheres em Bioinformática & Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 5

6. Observar e interpretar o resultado

NW Score	Identities	Positives	Gaps
204	121/607(20%)	181/607(29%)	327/607(53%)
Query 1	MDILCEENTSLSSTTNSLMQLNDDTRLYSNDNSGEANTSDAFNWTVDSENRTNLSCUGC M M-----	+NR S G KNRFFFSTG-	60
Sbjct 1			11
Query 61	LSPSCLSLLHLQEKNWSALLTAVVIIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIAD YF	YF	120
Sbjct 12	-----	YF-----	13
Query 121	MLLGFLVMPVSMLTILYGYRWPILPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNP PL S+ C WI LDVLF TASIMHLC IS+DRY+++ P PLGSEHCLTWICLDVLFCASIMHLCTISVDRLYLSRYP		180
Sbjct 14	-----		52
Query 181	IHHSRFNSRTKAFLKIIIAVWTISVGISMPIPVFGLOQDDSKVFKEGSCLLAADDNFVLIGSF + R +R + LKI+ VW +S+ +S+P+ + ++ + V G+C + D + L+GS MRFRNKTRRRVTLLKIVFWLSSIAMSLPLSLMYSKHNHASVLVNGTCQIPDPVYKLVGSI		240
Sbjct 53	-----		112
Query 241	VSFFIPLTIMVITYFLTICKSLQKE V F+IPL +M++TY LT++ L ++		264
Sbjct 113	VCFYIPLGVMLLTYCLTVRLLARQRQNLGGGQQTAAATPGWASGWLQQAPALGGPAGAAP		172
Query 265	-----ATLCVSDLGTRAKLASFSFLPQ-----SSLSSEKLFQRSIHRE ++L V G+ A + P S +S + S HR		302
Sbjct 173	ATVGSTLSPHSAHNQGSSLTVQSDGSGYLAAPGTPCPGRRKLSISKASVVTWDSSRRRR		232
Query 303	PGSYTG-----RR--TMQSIQN----- GS G RR T++S N		317
Sbjct 233	RGSSFGGVRTSLLLPTKTATTNSNSTPLRRSATLRSHQNMNYQGAGECGGKTRTTSSP		292
Query 318	-----EQKACKVLGIVFFLFVVMWCFFITNI EQKA KVLG+VFF FV+W PFF+ N+		344
Sbjct 293	CMLQRQQTVRSHHSRNSSVISRNSSRHGRIIRLEQKATKVLGVVFTFVILWSPFFVLNL		352
Query 345	MAVICKESCNEDVIGALLNVFWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENKK + +C E C E + + +V W+GY SS VNP+ YT+FNK +R AF + + C+Y		404
Sbjct 353	LPTVCAE-CEERISHWVFDVVTWLGYASSMVNPIFYTIFNKVFRQAFKKVLLCRYSTS-		410
Query 405	PLQLILVNTIPALAYKSSQLQMGQKKNSKQDAKTTDNDCSMVALGKQHSEEASKDNSDGV A++ S+		464
Sbjct 411	-----AWRPSR		416
Query 465	NEKVSCV 471		

O resultado mostra que nas duas sequências estão presentes duas regiões com identidade superior a 40%.

No receptor de serotonina humano, as duas regiões estão próximas umas das outras, enquanto estão separadas por mais de 200 aminoácidos na sequência de *Drosophila melanogaster*.

A visão geral gráfica mostra muito bem o arranjo espacial dessas regiões de sequência.

No entanto, esta visão geral não deve ser considerada definitiva porque contém poucas informações sobre a qualidade do alinhamento.

EXERCÍCIO 6

**Objetivo: Aprender a fazer e interpretar alinhamentos de sequências proteicas múltiplas
usando o CLUSTAL OMEGA**

EXERCÍCIO 6

Realize um alinhamento múltiplo com as sequências de proteínas
gi|543727, gi|7296517 e NP_649806 usando Clustal Omega [clustalomega]:

<https://www.ebi.ac.uk/Tools/msa/clustalo/>

Como o resultado pode ser interpretado?

EXERCÍCIO 6



Clustal Omega

- Latest version of Clustal - fast and scalable (can align hundreds of thousands of sequences in hours), greater accuracy due to new HMM alignment engine
- Command line/web server only (GUI public beta available soon)



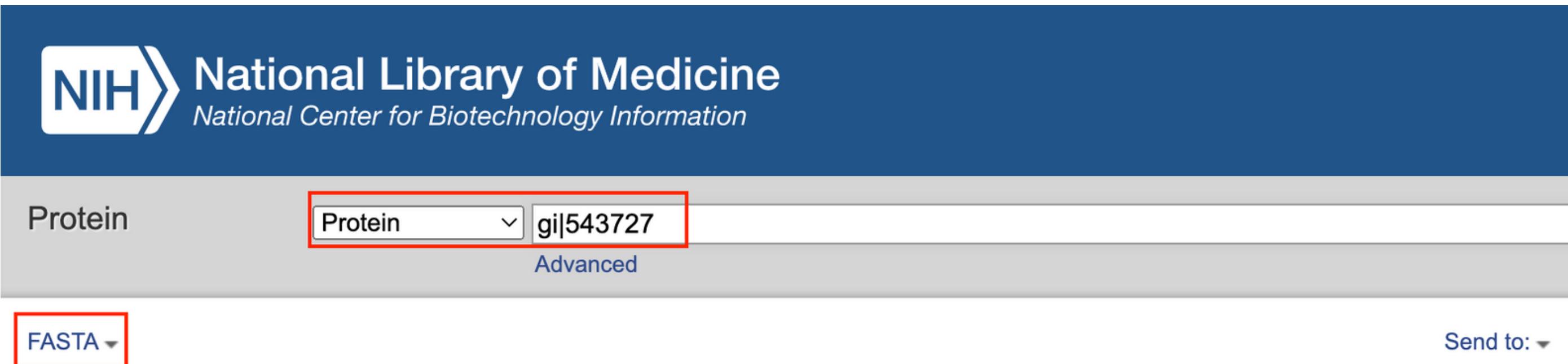
ClustalW/ClustalX

- "Classic Clustal"
- GUI (ClustalX), command line (ClustalW), web server versions available

<http://www.clustal.org/>

EXERCÍCIO 6

1. Buscar as sequências proteicas no NCBI - Protein



The screenshot shows the NCBI Protein search interface. In the search bar, 'gi|543727' is entered. Below the search bar, there are two buttons: 'FASTA' (highlighted with a red box) and 'Send to:'. The main content area displays protein information for P28223.2, including its UniProtKB/Swiss-Prot ID, RecName, AltName, and a partial sequence in FASTA format.

NIH National Library of Medicine
National Center for Biotechnology Information

Protein Protein gi|543727 Advanced

FASTA Send to:

RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; Short=5-HT-2A;
AltName: Full=Serotonin receptor 2A

UniProtKB/Swiss-Prot: P28223.2

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>sp|P28223.2|5HT2A_HUMAN RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; Short=5-HT-2A; AltName: Full=Serotonin receptor 2A
MDILCEENTSLSSTTNSLMQLNDDTRLYSNDFNSGEANTSDAFNWTVDSENRTNLSCEGCLSPSCLSLH
LQEKNWSALLTAVVIIILTIAGNILVIMAVSLEKKLQNATNYFLMSLAIADMILLGFLVMPVSMLTILYGYR
WPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPI
PVFGLQDDSKVFKEGSCLLADDNFVLIGSFVSFFIPLTIMVITYFLTIKSQKEATLCVSDLGTRAKLAS
FSFLPQSSLSSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLVVMCPFFITNIMAVICK
ESCNEDVIGALLNVFWIGYLSSAVNPLVYTLFNKYRSAFSRYIQCQYKENKKPLQLILVNTIPALAYK
SSQLQMGOKKNSKQDAKTTDNDCSMVALGKQHSEEASKDNSDGVNEKVSCV
```



EXERCÍCIO 6

1. Buscar as sequências proteicas no NCBI - Protein

NIH National Library of Medicine
National Center for Biotechnology Information

Protein Protein Advanced

FASTA

Octopamine-Tyramine receptor, isoform A [Drosophila melanogaster]

GenBank: AAF51802.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>AAF51802.1 Octopamine-Tyramine receptor, isoform A [Drosophila melanogaster]
MPSADQILFVNVTAAALTAAAATGGNAARGYTDSDDDAGMGTEAVANISGSLVEGLTTVT
AALSTAQADKDSAGECEGAVEELHASILGLQLAVPEWEALLTALVLSVIIVLTIIGNILVILSVFTYKPL
RIVQNFFIVSLAVADLTVALVLPFNVAYSILGRWEFGIHLCKLWLTCVLCCCTSSILNLCAIALDRYWA
ITDPINYAQKRTVGRVLLISGVWLLSLLISSPPLIGWNDWPDEFTSATPCELTSQRGYVISSLGSFFI
PLAIMTIVYIEIFVATRRRLRERARANKLNTIALKSTELEPMANSSPVAASNNSGSKSRLLASWLCCGRDR
AQFATPMIQNDQESISSETHQPQDSSKAGPHGNSDPQQQHVVVLVKKSRAKTKDSIKHGKTRGGRKSQS
SSTCEPHGEQQLLPAGGDGGSCPGGGHSGGGKSDAEISTESGSDPKGCIQVCVTQADEFQTSLKLTTPQS
STGVAAVSVTPLQKKTSGVNQFIEEKQKISLSKERRAARTLGIIMGVFVICWLPFFLMYVILPFCQTCCP
TNKFKNFIWLGYINSGLNPVIYTIFNLDYRRAFKRLGLN
```



EXERCÍCIO 6

1. Buscar as sequências proteicas no NCBI - Protein

The screenshot shows the NCBI Protein search interface. In the search bar, 'NP_649806' is entered. Below the search bar, the protein name '5-hydroxytryptamine (serotonin) receptor 2B, isoform D [Drosophila melanogaster]' is displayed. The FASTA sequence is shown below:

```
>NP_649806.2 5-hydroxytryptamine (serotonin) receptor 2B, isoform D [Drosophila melanogaster]
MEEDVYASLGAYNDGGDDWSSSEHVLWEEDETQRTTANATSRHNQLHVARWNATGNATISATFEDVPF
DANNYWALLALVLVLGTAAGNILVCLAIAWERRLQNVTNYFLMSLAITDLMVALVMPLGILTLVKGYFP
LGSEHCLTWICLDVLFCTASIMHLCTISVDRYLSLRPMRFGRNKTRRRVTLKIVFVWLSSIAMSPLSL
MYSKNHASVLVNGTCQIPDPVYKLVGSIVCFYIPLGVMLTYCLTVRLRARQRQNLLGGQQTAAATPGWA
SGWLGQAPALERRCTWRRLKPGPGNASSVLAHSANSTDLDSTLDNHELWLPDSSIKEPTPTMTALH
QFGAEMLKLSRGLESVASSSTGSPKSEFSLSNLQLQYPSSPQRYAATHQQSHHHHQHPAGVYHQ
ASPKGRHGTTVLGLSTTLGIERESTRNSLASSRMGEQSDGTLSQLSQRRLRAYKRRRASSAVPGREKR
AGHEDADEDVETPTSTLRRHKRHSNLPKNALYPRHTVQESLDDDEEEESAQQKDQQTKSKICHSDTE
LDPPQKAKQCHIAGQNHQSDYLQLPSVCTCPYFGDRPLQNCVSAEVKIISAFRVTTTTAVSSSPSEM
ELLMCSSGNKKSLTSSVSAGITGGPAGAAPATVGSTLSPHSAHNQGSSLTVQSDGSGYLAAPGTPCPGRR
KLSISKTAHSVVTWDSSRHRRGSSFGGVRTSLLLPTKTATTNSSTPLRRSATLRSQNMNYQGAGECG
GKTRTTTSSPCMLQRQQTVRSHHSRNSSVISRNSSRHGRIIRLEQKATKVLGVVFVILWSPFFVLNL
LPTVCAECEERISHWVFDVVTWLGYASSMVNPIFYTIFNKVFRQAFKKVLLCRYSSTSARPSR
```



EXERCÍCIO 6

2. Copiar as 3 sequências em formato FASTA no CLUSTAL OMEGA

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

```
>sp|P28223.2|5HT2A_HUMAN RecName: Full=5-hydroxytryptamine receptor 2A; Short=5-HT-2; Short=5-HT-2A; AltName: Full=Serotonin receptor 2A  
MDILCEENTSLSSTTNSLMQLNDDTRLYSNDNSGEANTSDAFNWTVDSENRTNLSCSEGCLSPSCLSLH  
LQEKNWSALLTAVVIIAGNILVIMAVSLEKKLQNATNYFLMSLAIADMLLGFLVMPVSMLTILYGYR  
WPLPSKLCAVWIYLDVLFSTASIMHLCAISLDRYVAIQNPIHHSRFNSRTKAFLKIIAVWTISVGISMPI  
PVFGLQDDSKVFKEGSCLLADDNFVLIGSFVSFFIPLTIMVITYFLTIKSLQKEATLCVSDLGTRAKLAS  
FSFLPQSSLSEKLFQRSIHREPGSYTGRRTMQSISNEQKACKVLGIVFFLVVMWCPFFITNIMAVICK  
ESCNEVDIGALLNVFWIWIGYLSSAVNPLVYTLFNKYRSAFSRYIQCQYKENKKPLQLILVNTIPALAYK  
SCOLOMCOKKNSKODAKTIDNDSMVALCKOUEEASKDNDPOVNEKVSQV
```

Or, upload a file: No file chosen

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)



EXERCÍCIO 6

3. Submeter o trabalho

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

[More options...](#)

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

[Submit](#)



EXERCÍCIO 6

4. Observar o resultado

EXERCÍCIO 6

5. Observar o resultado em cores (visualização mais simples)

EXERCÍCIO 6

6. Observar a árvore filogenética

Results for job clustalo-l20220418-014808-0946-81671636-p2m

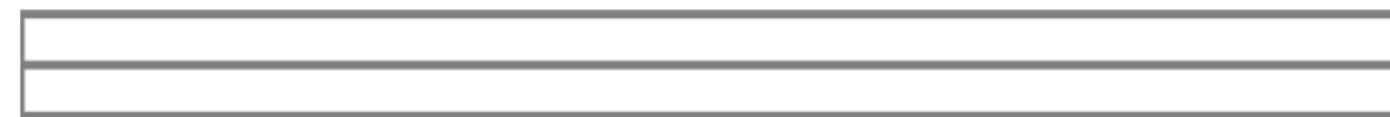
[Alignments](#) [Result Summary](#) [Guide Tree](#) [Phylogenetic Tree](#) [Results Viewers](#) [Submission Details](#)

[Download Phylogenetic Tree Data](#)

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: Cladogram Real



AAF51802.1 0.36364
splP28223.2I5HT2A_HUMAN 0.35842
NP_649806.2 0.34934



Mulheres em
Bioinformática
& Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 6

7. Interpretar os resultados

O alinhamento múltiplo das três proteínas mostra um número baixo de correspondências, enquanto duas sequências mostram aminoácidos idênticos em áreas amplas e ainda mais quando são considerados aminoácidos conservados.

Aminoácidos idênticos em todas as três sequências ocorrem muito raramente. Na visualização da Árvore Filogenética, isso pode ser visto claramente, pois todas as sequências têm distâncias semelhantes.

O alinhamento pode ser armazenado como um arquivo de texto simples com a extensão. **clustal** para que o alinhamento possa ser visualizado com outro software.

EXERCÍCIO 7

**Objetivo: Aprender a fazer e interpretar alinhamentos de sequências proteicas múltiplas
usando o CLUSTAL OMEGA**

EXERCÍCIO 7

Execute um alinhamento múltiplo com as seguintes sequências de forma análoga ao exercício anterior e calcule uma árvore filogenética para as proteínas Q28944.1, P25975.3, NP_081182.2, NP_640355.1, NP_001903.1, AAH12612.1.

Como o resultado pode ser interpretado?

A que tipo de proteínas pertencem as sequências?

EXERCÍCIO 7

1. Buscar as 6 sequências proteicas no NCBI - Protein e colcar na caixa de busca do CLUSTAL OMEGA

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

```
>sp|Q28944.1|CATL1_PIG RecName: Full=Procathepsin L; AltName: Full=Cathepsin L1; Contains: RecName: Full=Cathepsin L; Contains: RecName: Full=Cathepsin L heavy chain; Contains: RecName: Full=Cathepsin L light chain; Flags: Precursor MKPSLFLTALCLGIASAAPKLDQNLDADWYWKATHGRLYGMNEEGWRRRAVWEKNMKMIELHNQEYSQGK HGFSMAMNAFGDMTNEEFRQVMNGFQNQKHKGKVHESLVLEVPKSVDWREKGYVTAVKNQGQCGSCWA FSATGALEGQMFRKTGKLVSLSEQNLVDCSRPQGNQGCNGGLMDNAFQYVKDNGGLDEESYPYLGRETN SCTYKPECSAANDTGFVDIPQREKALMAVATVGPIVAIDAGHSSFQFYKSGIYYDPDCSSKDLDHGVL VVGYGFEGTDSNSKFWIVKNSWGPEWGWNGYVKMAKDQNNHCGISTAASYPTV
```

Or, upload a file: No file chosen

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)



EXERCÍCIO 7

2. Submeter o trabalho de alinhamento

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

More options...

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit



Mulheres em
Bioinformática
& Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 7

3. Observar o resultado do alinhamento

NP_001903.1	MNPTLILAAFCLGIASATLTDFHSLEAQWTKWKAMHNRLYGMNEEGWRRRAVWEKNMKMIE	60
AAH12612.1	MNPTLILAAFCLGIASATLTDFHSLEAQWTKWKAMHNRLYGMNEEGWRRRAVWEKNVKMIE	60
sp Q28944.1 CATL1_PIG	MKPSLFLTALCLGIASAAPKLDQNLDQWYKWKATHGRLYGMNEEGWRRRAVWEKNMKMIE	60
sp P25975.3 CATL1_BOVIN	MNPSFFLTVLCLGVASAAPKLDPNLDAHWHQWKATHRRLYGMNEEWRRRAVWEKNKKIID	60
NP_081182.2	MTPAVFLAILWLVMASSSPSPDPILDAEWQKWKIKYGKTYSLEEEGQKRAVWEENMKKIK	60
NP_640355.1	MTPAVFLVILCLGVVPGASALDLSLDVQWQEWEKIKYEKLYSPEEVLKRVVWEENVKKIE	60
	.:.*. : * : . . : * *:.* : ** : : * . :** :* . ***: * * *	
NP_001903.1	LHNQEYREGKHSFTMAMNAFGDMTSEEFRQVMNGFQNRKPRKGKVQEP-----	109
AAH12612.1	LHNQEYREGKHSFTMAMNAFGDMTSEEFRQVMNGFQNRKPRKGKVQEP-----	109
sp Q28944.1 CATL1_PIG	LHNQEYSQGKHGFSMAMNAFGDMTNEEFRQVMNGFQNQKHKKGKVHES-----	109
sp P25975.3 CATL1_BOVIN	LHNQEYSEGKHGFRMAMNAFGDMTNEEFRQVMNGFQNQKHKKGKLHEP-----	109
NP_081182.2	LHNGENGLGKHGFTMEMNAFGDMTLEEFRKEMIEIPVPTVKKGKSVQKR-----	109
NP_640355.1	LHNRENSLGKNTYTMEINDFADMTDEEFKDMIIGFQLPVHNTEKRLWKRALGSFFPNSWN	120
	*** * * * : : * : * * . * * * * : . : : . . * . :	
NP_001903.1	LFYEAPRSVDWREKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDC	169
AAH12612.1	LFYEAPRSVDWREKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDC	169
sp Q28944.1 CATL1_PIG	LVLEVPKSVDWREKGYVTAVKNQGQCGSCWAFSATGALEGQMFRKTGKLVSLSEQNLVDC	169
sp P25975.3 CATL1_BOVIN	LLVDVPKSVDWTKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGKLVSLSEQNLVDC	169
NP_081182.2	LSVNLPKFINWKKRGYVTPVRTQIACNSCWAIISVTGAIEGQMFRKTGQLIPLSVQNLVDC	169
NP_640355.1	WRDALPKFVDWRNEGVTVRKQGGCSSCWAFPVGTGAIEGQMFKTGKLIPLSVQNLIDC	180
	* : : * : . * * * * : . * : * . * * * : . * * : * * * : * : * * * :	



EXERCÍCIO 7

3. Observar o resultado do alinhamento

EXERCÍCIO 7

5. Interpretar os resultados:

O alinhamento múltiplo mostra que as sequências são muito **SEMELHANTES**.

Os aminoácidos são idênticos ou trocados de forma conservada em regiões amplas.

A sequência NP_640355.1 tem uma **INSERÇÃO** de aprox. 10 AMINOÁCIDOS.

Por causa da alta identidade, pode-se supor que são sequências **HOMÓLOGAS**.

De fato, as sequências são **PROTEASES** da família das **CATEPSINAS** de diferentes espécies.

EXERCÍCIO 7

Sequências:

Q28944.1 Precursor da catepsina L de *Sus scrofa* (suíno)

P25975.3 Precursor da catepsina L de *Bos taurus* (ovelha)

NP_081182.2 Precursor da catepsina 3 de *Mus musculus* (camundongo)

NP_640355.1 Catepsina Q de *Rattus norvegicus* (rato)

NP_001903.1 Pré-proteína catepsina L de *Homo sapiens* (humano)

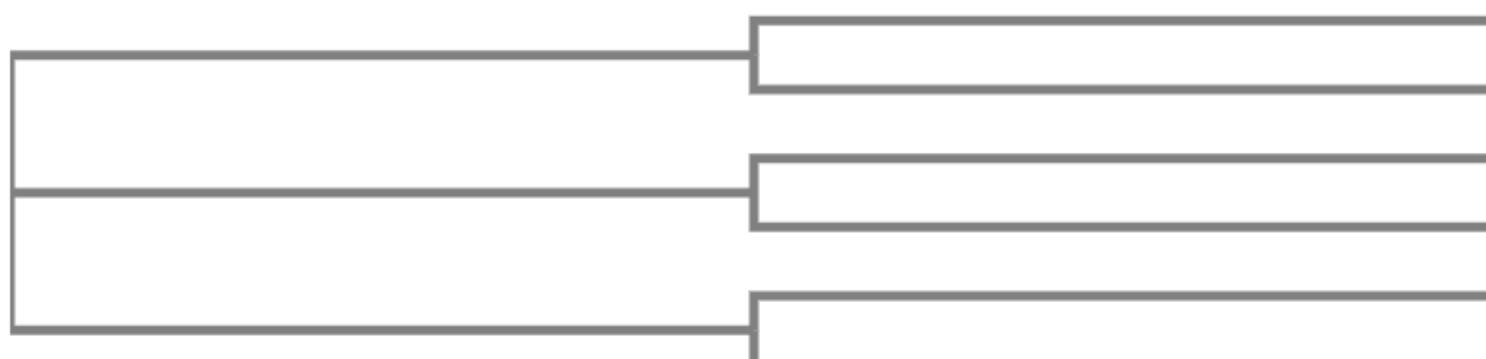
AAH12612.1 semelhante a Catepsina L de *Homo sapiens* (humano)

EXERCÍCIO 7

A árvore filogenética indica a relação entre as seis sequências.

É calculada uma relação próxima entre as duas sequências humanas, bem como entre as sequências de bovinos e suínos.

De acordo com esta análise, as sequências de camundongo e rato estão mais distantes das sequências de humano.



NP_001903.1 0.001 *Homo sapiens* (humano)
AAH12612.1 0.002 *Homo sapiens* (humano)
splQ28944.1|CATL1_PIG 0.04715
splP25975.3|CATL1_BOVIN 0.05465
NP_081182.2 0.21001 *Mus musculus* (camundongo)
NP_640355.1 0.21856 *Rattus norvegicus* (rato)

EXERCÍCIO 8

Objetivo: Aprender a predizer presença de exons a partir de sequências nucleotídicas usando o GENSCAN

EXERCÍCIO 8

Encontre uma entrada para um cosmídeo eucariótico no banco de dados de nucleotídeos NCBI, por exemplo, AN: AC012088, e exiba a sequência no formato FASTA.

Em uma segunda janela do navegador, vá para o servidor Genscan [genscan] e copie e cole a sequência na janela correspondente.

<https://bio.tools/genscan>

Em seguida, execute o Genscan.

Tente interpretar o resultado.

Procure outras sequências de cosmídeos de diferentes espécies e repita o exercício.

EXERCÍCIO 8

1. Cosmídeo AC012088 em formato FASTA

National Library of Medicine
National Center for Biotechnology Information

Nucleotide Nucleotide ▾ AC012088
Advanced

FASTA ▾ Send to: ▾

Homo sapiens Chromosome X Cosmid C232E9 (Lawrence Livermore LL0XNC01 Human Cosmid Libraries) complete sequence

GenBank: AC012088.2

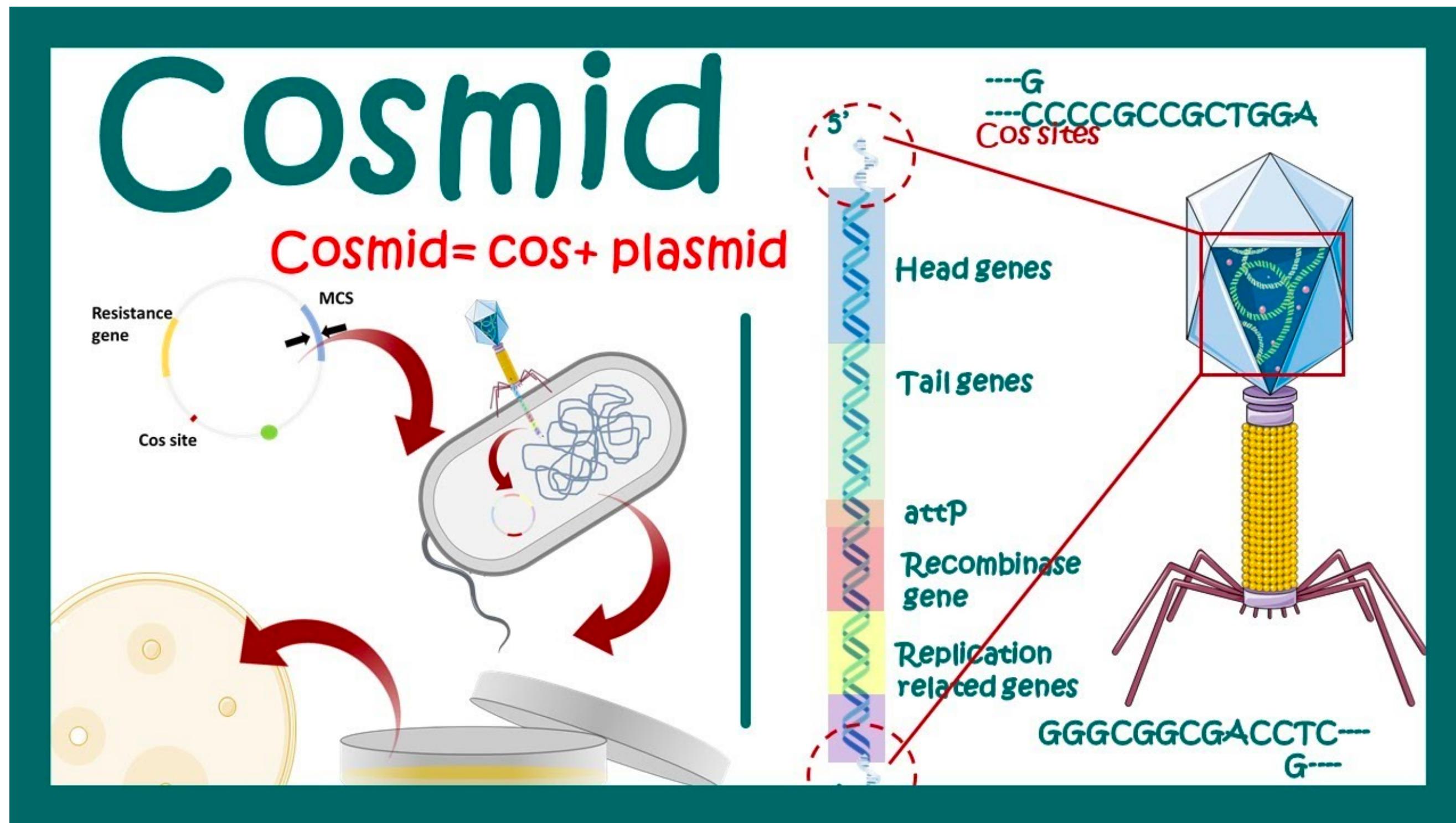
[GenBank](#) [Graphics](#)

```
>AC012088.2 Homo sapiens Chromosome X Cosmid C232E9 (Lawrence Livermore LL0XNC01
Human Cosmid Libraries) complete sequence
GGATCTGTGAGCACAGGATATCTTCCATTTTGTCGCCTTAGTTCTTCATCAATGTGTTATAGT
TTTCAGTGTACAGATCTTCATCTCTGGTAAATTATTCTTAAGTATTATTGTCTTGATGCTAT
TATAAATAGGATTGTTCTCTTTAGATAATTCAATTGTTAATGAATAAAAATGCAACTAATTTT
TGTACGTTGATTTGATTCTGCAACTTATTGAATTCAATTAGTTCTAACAGTTCTGGTGGAG
TCTTAGGGTCTCTGATATAAGATTGTCATCTCAAACAGAGAAAATGTTACTTTCTGATTGGG
ATGTCTTTATTTCTTGCTTAATTGCTCTGTCTAGGACTCTAGTAGTATGTTGAATAGAAGTGGT
GAGAGTGGGCATCCTGTGTTCTGATCTTAGAGAAAAGAATTCACTTCCCCTATTGAGTATAA
TGTAGAGGGGGCTTGTATATGACCTTATTATGTTGAGGTACATCTACACCAACTTGTGAC
AGCTTTATTATGAAAAAAATGTTGAATTGTTAAATGCTTTCTGCCTCTGTTGAGATGCTCATATGA
TTTTATATTCATTTGCAAATGTTGATATCACATTATTGATTACATGTCAAACTCTCCTTGCATC
CCAGAGACAAATCTCACTGATCATGGTATATAATCTTGTGATGTGCTGTTGAATTGGTTACTAGTA
TTTGTGCATATTTCACTTCTGAGTTGCTGAGGTTATTGTTATGGCCAGAATATGGCTATTAA
ATTCAATGTTCATCATACTTAGTGAATATTCACTATGTTCACTACAGAAAATTCATATGATTATTCT
TGGGGCTCCCTAGTCTTAATGTTGACACACAGTATATGTTGACATATAATTCTAAAATTGAA
AAACTGTGAATTCTGAAATACATCTCAGGGATTGGGATAAGAAATTCTGGTTATTATCATC
CTTGTGATCTAGAATAAGTTATATTATGTTGACATGACAAGAAAACAATAAATTAAAT
TTCATAGAAGGACTTGCAGACTCAAAGAATATTCCATAGACCATGGGGCTGTTCCCAGCTAAA
AACTTAGCAATGATGAATTACAGTATAATGATGTTGAGTTCTTGGCTCTAGCTGCTTCAGCAA
```



EXERCÍCIO 8

Um cosmídeo é um tipo de plasmídeo híbrido que contém uma sequência cos do vírus bacteriófago Lambda.



EXERCÍCIO 8

2. Colar a sequência nucleotídica no site do GENSCAN, escolher a opção vertebrado (porque o inserto é de humano), trocar o exon cutoff para 0.01 e rodar a análise

[?](#)

[For information about Genscan, click here](#)

Server update, November, 2009: We've been recently upgrading the GENSCAN webserver hardware, which resulted in some problems in the output of GENSCAN. We apologize for the inconvenience. These output errors were resolved.

This server provides access to the program Genscan for predicting the locations and exon-intron structures of genes in genomic sequences from a variety of organisms.

This server can accept sequences up to 1 million base pairs (1 Mbp) in length. If you have trouble with the web server or if you have a large number of sequences to process, request a local copy of the program (see instructions at the bottom of this page).

Organism: **Vertebrate** Suboptimal exon cutoff (optional): **0.01**

Sequence name (optional):

Print options: **Predicted peptides only**

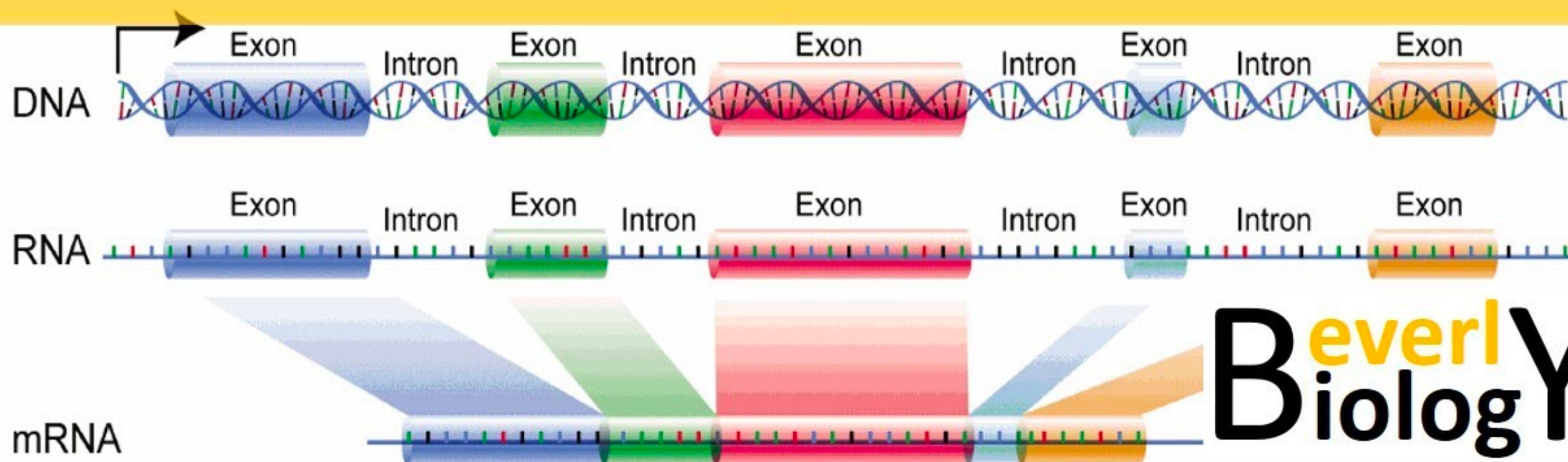
Upload your DNA sequence file (upper or lower case, spaces/numbers ignored): No file chosen

Or paste your DNA sequence here (upper or lower case, spaces/numbers ignored):
>AC012088.2 Homo sapiens Chromosome X Cosmid C232E9 (Lawrence Livermore
LL0XNC01 Human Cosmid Libraries) complete sequence
GGATCTGTGAGCACAGGATATCTTCCATTTCGTTGCGCTTAGTTCTTCATCAATGTGTTATAGT
TTTCAGTGTACAGATCTTCATCTCTTGGTTAAATTATTCCTAAGTATTGTTATTGTCTTGATGCTAT
TATAAAATAGGATTGTTCTTCTTTAGATAATTCAATTGTTAATGAATAAAAATGCAACTAATTTT
TGTACGTTGATTGTTGATTCTGCAACATTATTGAAATTCAATTATTAGTTCTAACAGTTCTGGTGGAG
TCTTTAGGGTTCTCTTGATATAAGATTGTCATCTTCAAACAGAGAAAATGTTACTTTCTGATTGGG
ATGTCTTTATTGTTCTTGCTTAATTGCTCTGTTAGGACTCTAGTAGTATGTTGAATAGAAGTGGT
GAGAGTGGGCATCCTTGTGTTGTTCTGATCTTAGAGAAAAAGAATTCACTTCCCCTATTGAGTATAA
TGTTAGAGGTGGGCTTGTATATATGACCTTTATTATGTTGAGGTACATCTACACCAACTTGTGAC
AGCTTTTATTATGAAAAAAATGTTGAATTGTTAAATGCTTTCTGCCTCTGTTGAGATGCTCATATGA
TTTTTATATTCACTTGCCTGAAATGTGGTATATCACATTATTGATTACATGTCAAACACTCTGTCATC



EXERCÍCIO 8

Introns vs Exons



Beverly
BiologY



Mulheres em
Bioinformática
& Data Science LA
Promovendo a colaboração entre mulheres

EXERCÍCIO 8

3. Analisar o resultado

Para o cosmídeo AC012088, duas proteínas são previstas. A primeira é codificada por um único gene-exon, o que significa que o gene consiste em um único exon sem íntrons.

```
>/tmp/04_17_22-21:42:05.fasta|GENSCAN_predicted_peptide_1|144_aa  
MTLARCIRSGDEEPEGMLVSLTLKRNCLLLSDGCSKCQGARMVEPQGRRIMGPLVITRR  
QVTMGIHLLCLCSHRPAQLLGIVRLITVFSDCGYRKQGLEYPGMSNNKEHVDKQSQEQQ  
ADKKFQEGSSLSSWGAKKYSPPPV  
  
>/tmp/04_17_22-21:42:05.fasta|GENSCAN_predicted_peptide_2|115_aa  
MATEAKIDKWDLIKLKSFCTAKETTIRVNGQPTEWEKIFAVYSSDKGLISRIYNELKQIY  
KKKTNHPIKKWVKDMNRRFSKEVIYAANRHEKMLIITGHQRNANQNHNEIPSHAS
```



REFERÊNCIAS

EMBOSS-Needle:

https://www.ebi.ac.uk/Tools/psa/emboss_needle/

BLASTp:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

BLASTn:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

BLASTx:

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab&LAST_PAGE=blastn

Global Align - BLAST:

[https://blast.ncbi.nlm.nih.gov/Blast.cgi?
PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink)

CLUSTAL OMEGA:

<https://www.ebi.ac.uk/Tools/msa/clustalo/>
<http://www.clustal.org/>

GENSCAN:

<https://bio.tools/genscan>