

Análise de tempo em ferramentas de Bioinformática

Anderson Ferreira Sepulveda

Universidade de São Paulo
Instituto de Física

Objetivos

- Comparar a eficiência do algoritmo BLAST;
 - Usarei as ferramentas BLASTp do NCBI e BLAST do UniProt.
- Há horas mais “tranquilas”?
- Há dias mais “tranquilos”?
- Quem ganha? NCBI ou Uniprot?



X



Metodologia

- Três sequências de proteínas:
 - Canal de sódio (com 1682 aa);
 - Dopamina (com 578 aa);
 - Insulina (com 110 aa);

Metodologia

Protein BLAST: search prot... x +

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&QUERY=AAA59172.1&LINK_LOC=protein&PAGE_TYPE=BlastSe Pesquisar

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® » blastp suite Home Recent Results Saved Strategies Help

Standard Protein BLAST

blastn blastp **blastx** tblastn tblastx

Enter Query Sequence

BLASTp programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

1 malxmrlpl lallalvwd paaafvngbl sashivealy lvsgeroffv
tpktresed
61 laqqvvelg aggaqalpl alegalqkg lvsasctalc slvqlenyn

Or, upload file [Selecionar arquivo...](#) Nenhum arquivo selecionado. [Job Title](#) Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database Non-redundant protein sequences (nr) [Organism](#) Optional 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

☐ Exclude [Entrez Query](#) Optional Enter an Entrez query to limit search [YouTube](#) [Create custom database](#)

Program Selection

Algorithm

- ☒ 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

O Firefox automaticamente envia alguns dados para a Mozilla de forma que possamos aprimorar sua experiência. Escolher o que compartilhar

14:41 17/11/2016

Metodologia

The screenshot shows the UniProt BLAST search page. The browser address bar displays `http://www.uniprot.org/blast/`. The UniProt logo is in the top left, and a search bar with a dropdown menu set to 'UniProtKB' is in the top right. Below the logo, navigation links for 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search' are visible, along with 'Help' and 'Contact' links.

BLAST

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g. P00750 or A4_HUMAN or UP10000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

[Help](#) [BLAST help video](#) [Other tutorials and videos](#) [Downloads](#)

```
1 malmmlpl lallalwgd paaafvnghl gghlvealy lvcgeraffv tptkrreadd
61 lvggveidg gggagldpl aleggakrg ivegcctalc elyqlenyn
```

Target databaseⁱ: UniProtKB | E-Thresholdⁱ: 10 | Matrixⁱ: BLOSUM-62 | Filteringⁱ: None | Gappedⁱ: yes | Hitsⁱ: 100

☐ Run BLAST in a separate window.

[Clear](#) [Run BLAST](#)

Tools	Core data	Supporting data	Information
BLAST	Protein knowledgebase (UniProtKB)	Literature citations	About UniProt
Align	Sequence clusters (UniRef)	Taxonomy	Help
Retrieve/ID mapping	Sequence archive (UniParc)	Keywords	FAQ
Peptide search	Proteomes	Subcellular locations	UniProtKB manual
		Cross-referenced databases	Technical corner

© O Firefox automaticamente envia alguns dados para a Mozilla de forma que possamos aprimorar sua experiência.

[Escolher o que compartilhar](#)

14:44
17/11/2016

Metodologia

- Parâmetros:
 - Bancos de dados:
 - nr (BLAST);
 - UniProtKb
 - Valor esperado: $E = 10$;
 - Tabela BLOSUM-62;
 - Retorno de sequências: 100.

Metodologia

NCBI Blast:Protein Sequen... x +

https://blast.ncbi.nlm.nih.gov/Blast.cgi

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST » **blastp suite** » RID-2VRU76CG014 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[\[Formatting options\]](#)

Format Request Status

Job Title: Protein Sequence (110 letters)

Request ID	2VRU76CG014
Status	Searching
Submitted at	Thu Nov 17 11:48:07 2016
Current time	Thu Nov 17 11:48:13 2016
Time since submission	00:00:05

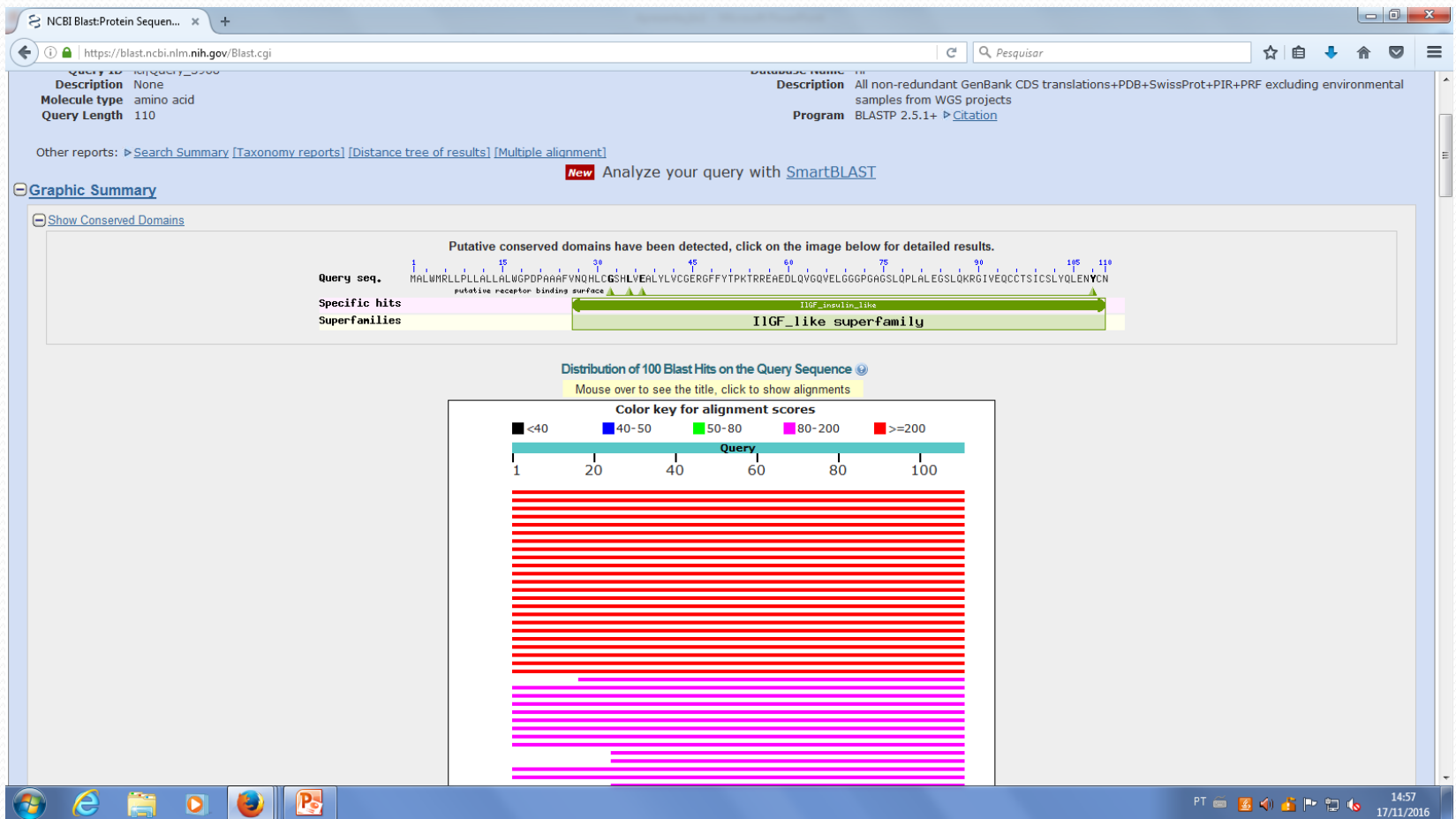
This page will be automatically updated in 2 seconds

BLAST is a registered trademark of the National Library of Medicine.

[Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#) | [Support center](#) NCBI | NLM | NIH | DHHS

PT 14:50 17/11/2016

Metodologia



Metodologia

The screenshot shows the UniProt BLAST web interface. The browser's address bar displays the URL: `www.uniprot.org/blast/uniprot/B2016111814483A1C7ED25EE8374758DF3FD545FD4804F8Y`. The UniProt logo is in the top left, and a search bar with 'UniProtKB' is in the top right. Below the search bar, there are navigation links: 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. The 'BLAST' section is highlighted. A 'Basket' button is in the top right corner. A 'How to use this tool' section provides instructions on using BLAST. Below this, the 'Job status: RUNNING' is displayed. A red circle highlights the 'Running blastp job against UNIPROTKB for 7s' text, with a red arrow pointing to it. The 'Job information' section shows the query sequence, job identifier, and program used.

UniProt

UniProtKB

Advanced Search

BLAST Align Retrieve/ID mapping Peptide search

Help Contact

Basket

How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g. P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

Help BLAST help video Other tutorials and videos Downloads

Job status: RUNNING

Running **blastp** job against **UNIPROTKB** for 7s

Job information

Query sequence	malwmrllpl lallalwgpdpaaafvnqhl cgshlvealy lvcgergffy tpktrreaed lqvqqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn
Job identifier	B2016111814483A1C7ED25EE8374758DF3FD545FD4804F8Y (jobs are stored for 7 days)
Program	blastp

Metodologia

sequences as well as help identify members of gene families.

3. Click the **Run BLAST** button.

[Help](#) [BLAST help video](#) [Other tutorials and videos](#) [Downloads](#)

100 80 60 40 20 0
Identity %

Filter byⁱ

- Reviewed (25)
Swiss-Prot
- Unreviewed (75)
TrEMBL
- With 3D structure (5)
- Proteomes (67)

Popular organisms

- Human (5)
- Mouse (5)
- Rat (2)
- Bovine (3)
- GORGO (1)

Map to

- UniProtKB
- UniRef

[Edit and resubmit](#) Order by: **Score**

Overview

[Show all 100](#)

Entry	Protein names	Match hit	Identity
I3WAC9	Insulin (Homo sapiens)		100.0%
P01308	Insulin (Homo sapiens)		100.0%
Q6YK33	Insulin (Gorilla gorilla gorilla)		100.0%
Q8HXV2	Insulin (Pongo pygmaeus)		99.1%

Alignments

[BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#)

1 to 25 of 100 Show 25

Entry	Alignment overview	Info	Status
Query: B2016111814483A1C7ED25EE8374758DF3FD545FD4804F8Y			
I3WAC9	I3WAC9_HUMAN - Insulin Homo sapiens (Human) - View alignment	E-value: 7.7e-76	

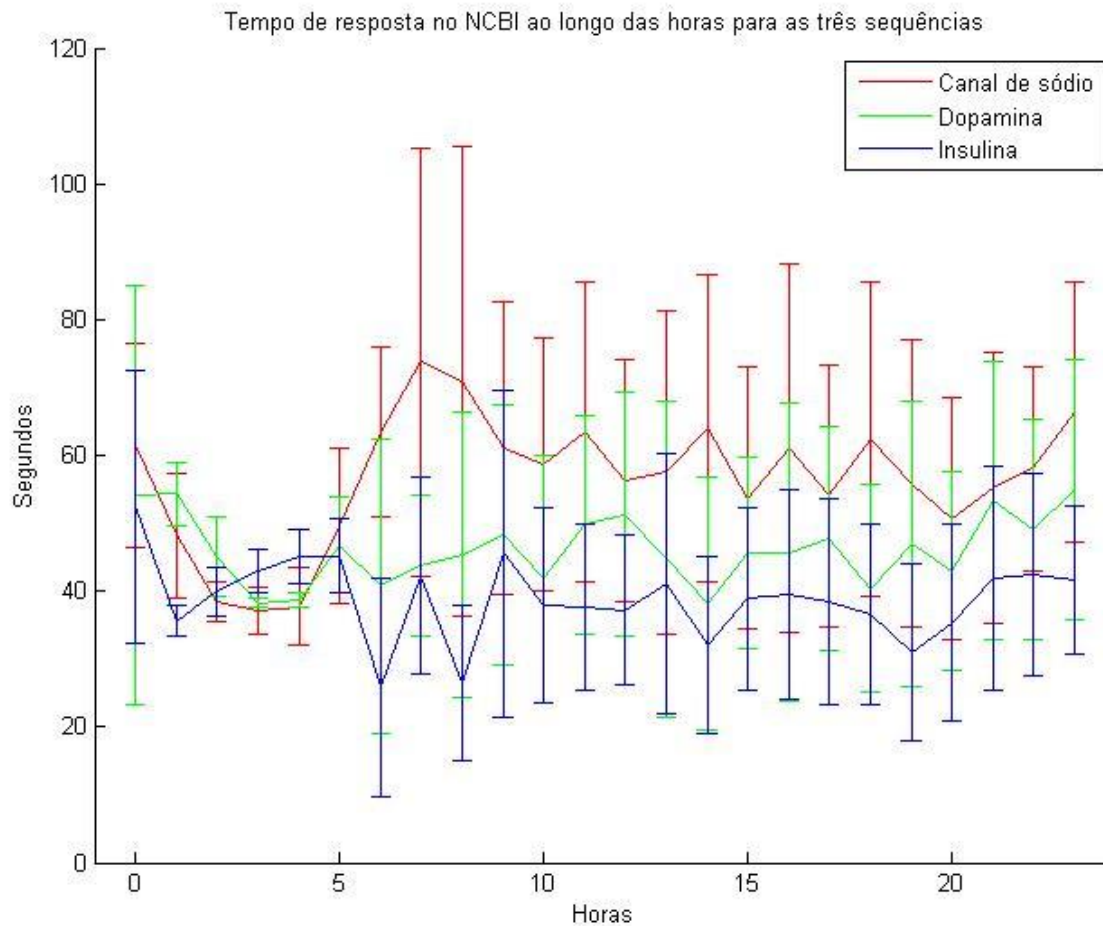
Pergunte-me alguma coisa

22:12 17/11/2016

O que espero?

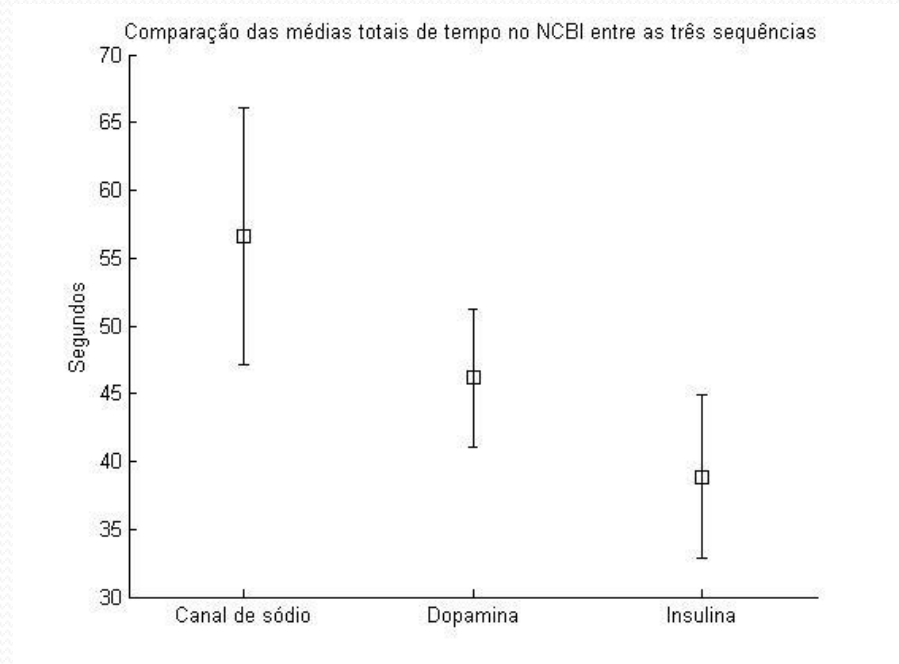
- O pico de uso é no meio do dia;
- Finais de semana o tempo de resposta é menor;
- Sequências maiores demoram mais para responder;
- NCBI é mais eficiente que UniProt.

Resultados

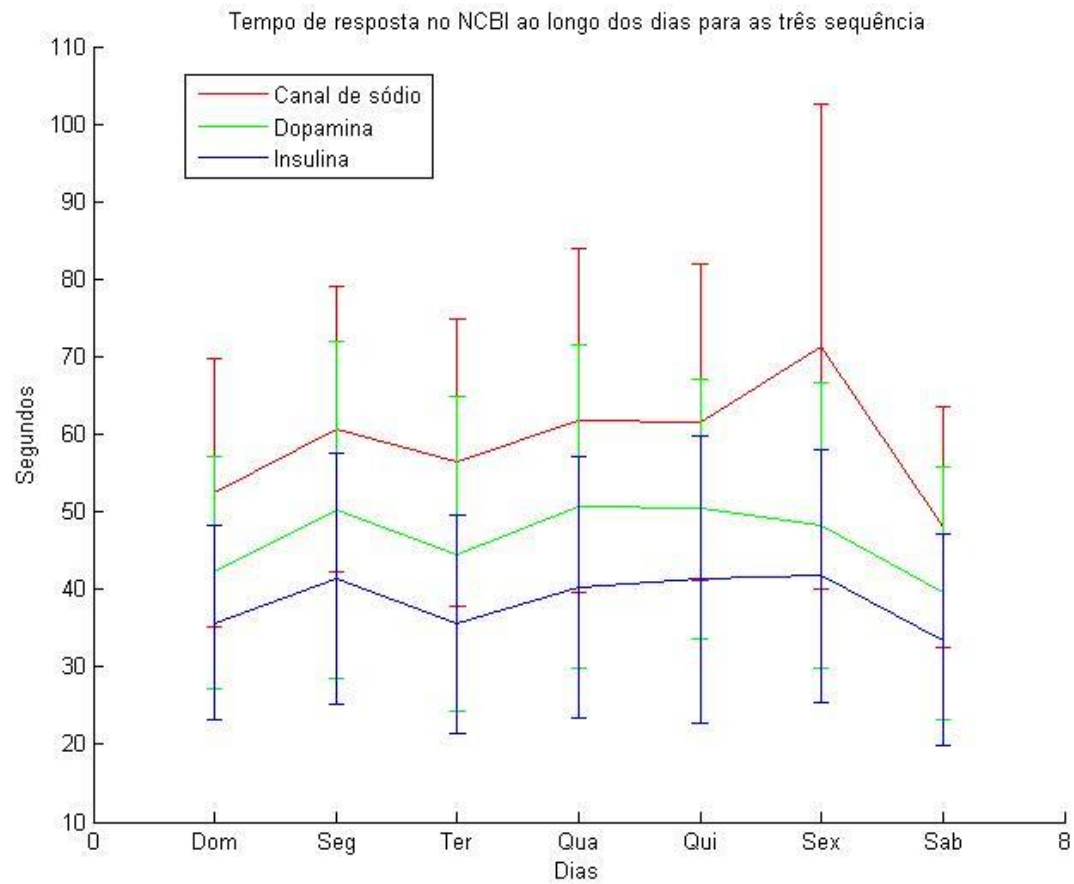


Resultados

- Quanto maior a sequência, maior a média de tempo;
- Percebe-se que durante a madrugada, as médias de tempo são mais baixas que no resto dia;
- Grande variação de tempo de resposta do NCBI



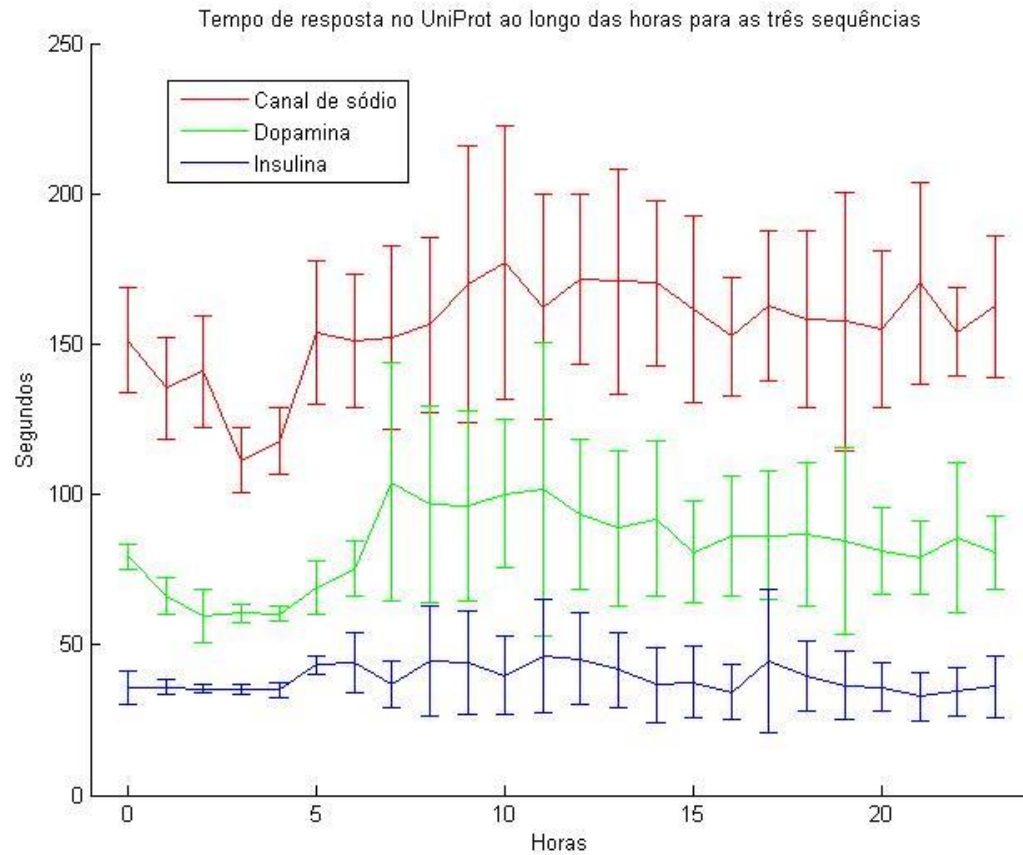
Resultados



Resultados

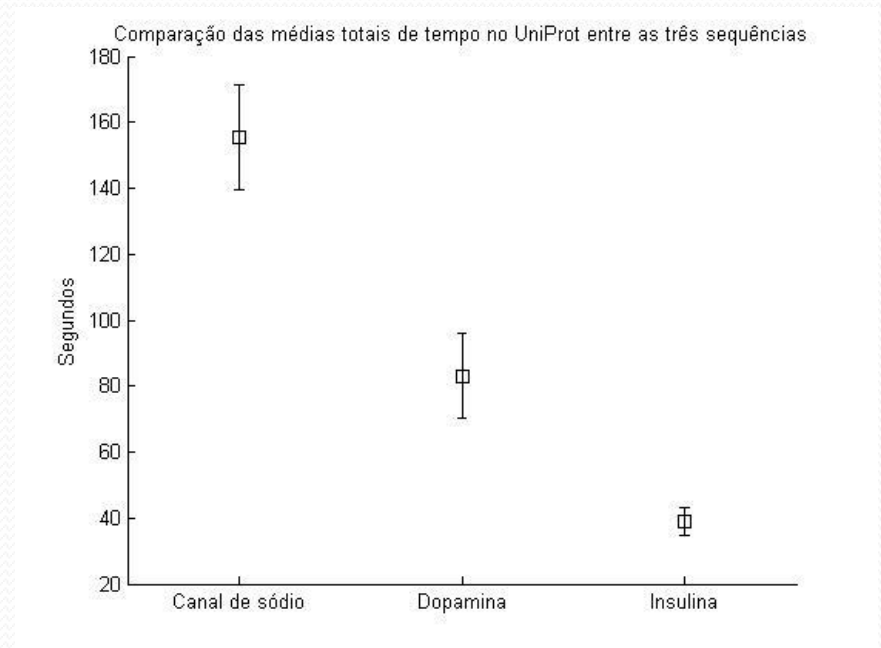
- É claro que o tamanho da frequência influi na média do tempo;
- É confirmada a grande variação de tempo;
- Sábado e domingo são dias com médias mais baixas;

Resultados

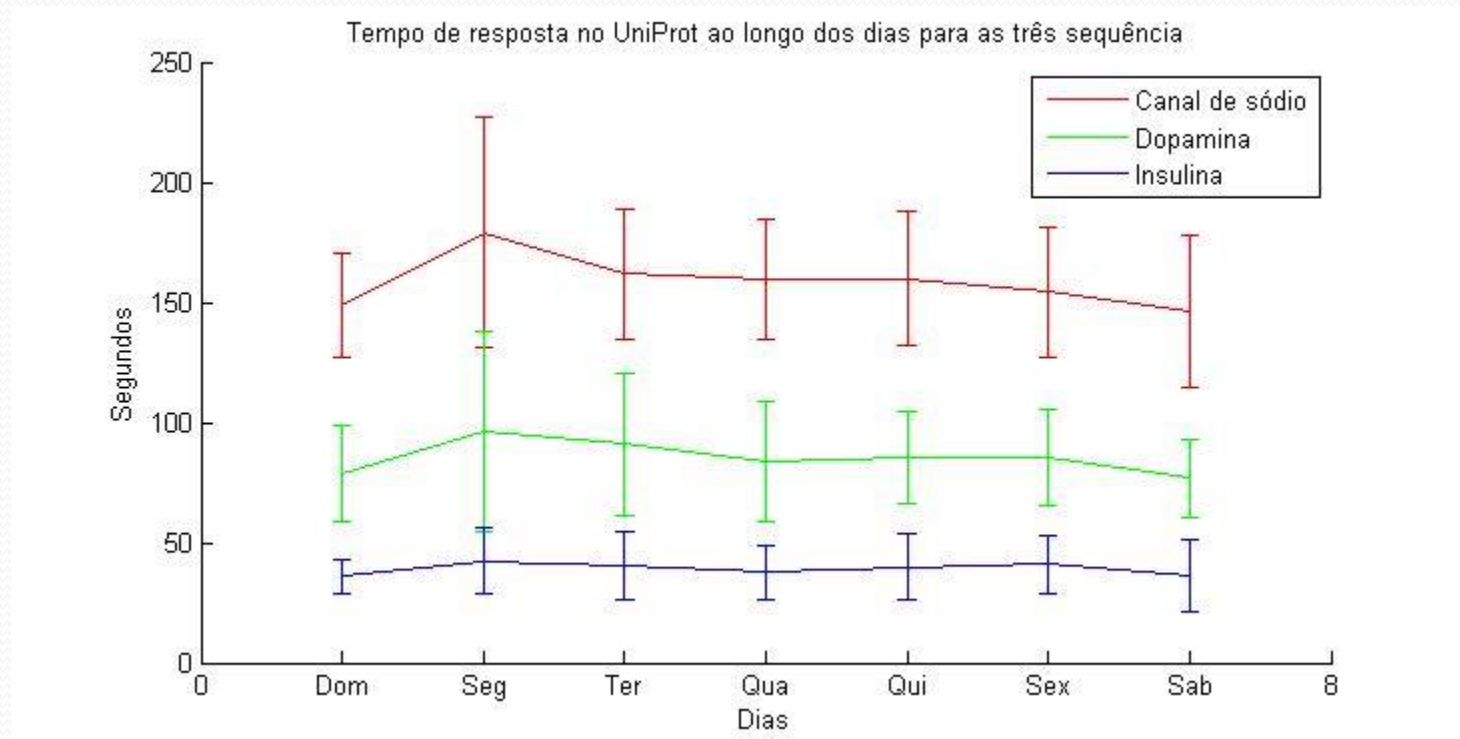


Resultados

- Quanto maior a sequência maior a média do tempo de resposta;
- Madrugadas têm as médias mais baixas;
- Em comparação com NCBI, UniProt tem as variações de tempo mais comportadas e definidas.



Resultados



Resultados

- Quanto menor a sequência, mais estável o tempo de resposta, isto é, as médias de tempo são praticamente as mesmas em todos os dias e com menor variação de tempo.

O que obtive?

- O menor tempo médio foi na madrugada, mas não existe um pico (ou picos) em um momento específico do dia;
- Finais de semana realmente tem a menor média, mas no caso do UniProt quanto menor a sequência menos visível é a variação da média ao longo da semana;
- O tamanho da sequência é realmente importante para o BLAST;
- A média do tempo de resposta do NCBI é mais baixa que do UniProt.

Conclusões

- O tempo de resposta do algoritmo BLAST também é afetado por vários fatores: a velocidade da internet, a qualidade do computador, o uso do banco de dados dos sites...
- A manipulação dos parâmetros pode alterar o tempo de resposta.