

Laboratórios de Bioinformática

- BLAST

BLAST

BLAST

- A sequência a pesquisar pode ser introduzida em formato fasta, apenas a sequência (sem linha de título) ou usando o código da base de dados (accession number)
- A sequência a introduzir pode ser a sequência completa ou apenas um fragmento

BLAST – Escolher o tipo de BLAST

- **Blastn**
 - procurar uma sequência de DNA na base de dados de DNA
- **Blastp**
 - procurar uma proteína na base de dados de proteínas
- **Blastx**
 - procurar uma sequência de DNA traduzida na base de dados de proteínas
- **Tblastn**
 - procurar uma proteína na base de dados de sequências de DNA traduzida
- **Tblastx**
 - procurar uma sequência de DNA traduzida na base de dados de sequências de DNA traduzidas

BLAST Assembled Genomes

Find Genomic BLAST pages:

Enter organism name or id--completions will be suggested

GO

▣ [Human](#)
▣ [Mouse](#)
▣ [Rat](#)
▣ [Cow](#)
▣ [Pig](#)
▣ [Dog](#)

▣ [Rabbit](#)
▣ [Chimp](#)
▣ [Guinea pig](#)
▣ [Fruit fly](#)
▣ [Honey bee](#)
▣ [Chicken](#)

▣ [Zebrafish](#)
▣ [Clawed frog](#)
▣ [Arabidopsis](#)
▣ [Rice](#)
▣ [Yeast](#)
▣ [Microbes](#)

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#)

Search a nucleotide database using a nucleotide query
Algorithms: blastn, megablast, discontinuous megablast

[protein blast](#)

Search protein database using a protein query
Algorithms: blastp, psi-blast, phi-blast, delta-blast

[blastx](#)

Search protein database using a translated nucleotide query

[tblastn](#)

Search translated nucleotide database using a protein query

[tblastx](#)

Search translated nucleotide database using a translated nucleotide query

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- ▣ Make specific primers with [Primer-BLAST](#)
- ▣ Cluster multiple sequences together with their database neighbors using [MOLE-BLAST](#)
- ▣ Find [conserved domains](#) in your sequence (cds)
- ▣ Find sequences with similar [conserved domain architecture](#) (odart)
- ▣ Search sequences that have [gene expression profiles](#) (GEO)
- ▣ Search [immunoglobulins and T cell receptor sequences](#) (IgBLAST)
- ▣ Screen sequence for [vector contamination](#) (vecscreen)
- ▣ [Align](#) two (or more) sequences using BLAST (bl2seq)



BLAST - input

[Display Settings:](#) ☒ FASTA

hemoglobin subunit beta [Homo sapiens]

NCBI Reference Sequence: NP_000509.1

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

```
>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta [Homo sapiens]  
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLG  
AFSDGLAHLDNLKGTFTALSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAAYQKVAVGVA  
N  
ALAHKYH
```

Copiar fasta

ou

[Send to:](#) ☒

Change region shown

[Analyze this sequence](#)

[Run BLAST](#)

[Identify Conserved Domains](#)

[Highlight Sequence Features](#)


[Find in this Sequence](#)


**Executa um BLAST
da sequência**

The screenshot shows the NCBI BLAST Basic Local Alignment Search Tool interface. At the top, there are tabs for 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. Below this, the 'NCBI/BLAST/blastp suite' is selected. The main section is titled 'Enter Query Sequence' and contains a large text area for entering a sequence. To the right of this area are 'Query subrange' fields for 'From' and 'To'. Below the text area is a 'Procurar...' button. Further down, there are fields for 'Job Title' and a checkbox for 'Align two or more sequences'. The 'Choose Search Set' section includes a 'Database' dropdown set to 'Non-redundant protein sequences (nr)', an 'Organism' field with a suggestion box, and an 'Exclude' section with checkboxes for 'Models (XM/XP)' and 'Uncultured/environmental sample sequences'. At the bottom, there is an 'Entrez Query' field and a 'Program Selection' section with a radio button for 'blastp (protein-protein BLAST)'.



BLAST - UniProt



UniProtKB Advanced 

BLAST [Align](#) [Retrieve/ID Mapping](#) [Help](#) [Contact](#)

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](#) [Tutorials and Videos](#) [Downloads](#)

BLAST

Protein sequence, Nucleotide sequence or UniProt identifier

Target database¹

E-Threshold¹

Matrix¹

Filtering¹

Gapped¹

Hits¹

UniProtKB

10

PAM-30

None

yes

250

☐ Run Blast in a separate window.

Run BLAST

Clear



BLAST - Parâmetros

NCBI/ BLAST/ blastp suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

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

>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPKVKAHGKK
VLGAFSDGLAHLNLKGT FATLSELHCDKLHVDPENFRLLGNVLCVLAHFGKEFTPPVQAAYQKV
VAGVANALAHKYH

Or, upload file [Procurar...](#)

Job Title
Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Organism [Optional](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude [Optional](#)
☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query [Optional](#)

Enter an Entrez query to limit search

[YouTube](#) [Create custom](#)

Bases de dados de nucleótidos

- Nucleotide collection (nr/nt)
- Nucleotide collection (nr/nt)
- RefSeq Select RNA sequences (refseq_select)
- Reference RNA sequences (refseq_ma)
- RefSeq Representative genomes (refseq_representative_genomes)
- RefSeq Genome Database (refseq_genomes)
- Whole-genome shotgun contigs (wgs)
- Expressed sequence tags (est)
- Sequence Read Archive (SRA)
- Transcriptome Shotgun Assembly (TSA)
- High throughput genomic sequences (HTGS)
- Patent sequences(pat)
- PDB nucleotide database (pdb)
- Human RefSeqGene sequences(RefSeq_Gene)
- Genomic survey sequences (gss)
- Sequence tagged sites (dbsts)

Bases de dados de proteínas

- Non-redundant protein sequences (nr)
- Non-redundant protein sequences (nr)
- RefSeq Select proteins (refseq_select)
- Reference proteins (refseq_protein)
- Model Organisms (landmark)
- UniProtKB/Swiss-Prot(swissprot)
- Patented protein sequences(pataa)
- Protein Data Bank proteins(pdb)
- Metagenomic proteins(env_nr)
- Transcriptome Shotgun Assembly proteins (tsa_nr)

Escolher a base de dados

Bases de dados de proteínas

nr = non-redundant (usada mais frequentemente)



BLAST - Parâmetros

NCBI/ BLAST/ blastp suite Standard Protein BLAST

blastn blastp **blastx** tblastn tblastx

BLASTP programs search protein databases using a protein query sequence

Enter Query Sequence

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

```
>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta [Homo sapiens]  
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAVMGNPKVKAH  
VLGAFSDGLAHLNKLGTFTLSELHCDKLHVDPENFRLLGNVLCVLAHFGKEFTPPVQAAY  
VAGVANALAHKYH
```

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Organism Optional ☐ Exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

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

Pode-se limitar a procura a a um determinado organismo ou grupo taxonómico ou, por outro lado, pode-se excluir organismos ou grupos taxonómicos da procura (importante em diversas áreas de investigação)

BLAST - Parâmetros

NCBI/ BLAST/ blastp suite

blastn blastp blastx tblastn tblastx

BLASTP program

Enter Query Sequence

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

>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPKVKRAHGKK
VLGAFSDGLAHLNLRKGTFTLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAYQKV
VAGVANALAHKYH

Query subrange

From

To

Or, upload file [Procurar...](#)

Job Title
Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Organism Optional ☐ Exclude [+](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

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

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

Executar o BLAST

BLAST Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

☐ Show results in a new window

Pode-se limitar a busca
(tamanho, palavra-chave...)

Tipo de algoritmo

Executar o BLAST

BLAST - Parâmetros

Utilizadores podem mudar vários parâmetros no algoritmo:

- o valor-e (expected value)
- tamanho da palavra a procurar
- matriz
- custo de introdução de espaçamentos
- filtragem dos resultados
- output
- ...

The screenshot shows the 'Algorithm parameters' section of the NCBI BLAST web interface. It is divided into three main sections: 'General Parameters', 'Scoring Parameters', and 'Filters and Masking'. Arrows from the text box on the left point to the following settings:

- Expect threshold:** Set to 10 in the 'General Parameters' section.
- Word size:** Set to 3 in the 'General Parameters' section.
- Matrix:** Set to BLOSUM62 in the 'Scoring Parameters' section.
- Gap Costs:** Set to Existence: 11 Extension: 1 in the 'Scoring Parameters' section.
- Filters and Masking:** The section contains checkboxes for 'Low complexity regions', 'Mask for lookup table only', and 'Mask lower case letters', all of which are currently unchecked.

At the bottom of the interface, there is a 'BLAST' button and a checkbox for 'Show results in a new window'.

BLAST - Output

The image shows a screenshot of the BLAST web interface with several annotations in Portuguese. The interface displays the results of a BLAST search for the query sequence 'gi|4504349|ref|NP_000509.1| hemoglobin subunit...'. The search was performed using the BLASTP program against the 'nr' database. The results show a high-scoring hit for 'Hb-beta_like' and 'Globin-like superfamily'. The interface also includes a 'Graphic Summary' section with a sequence logo and a 'Taxonomy reports' section.

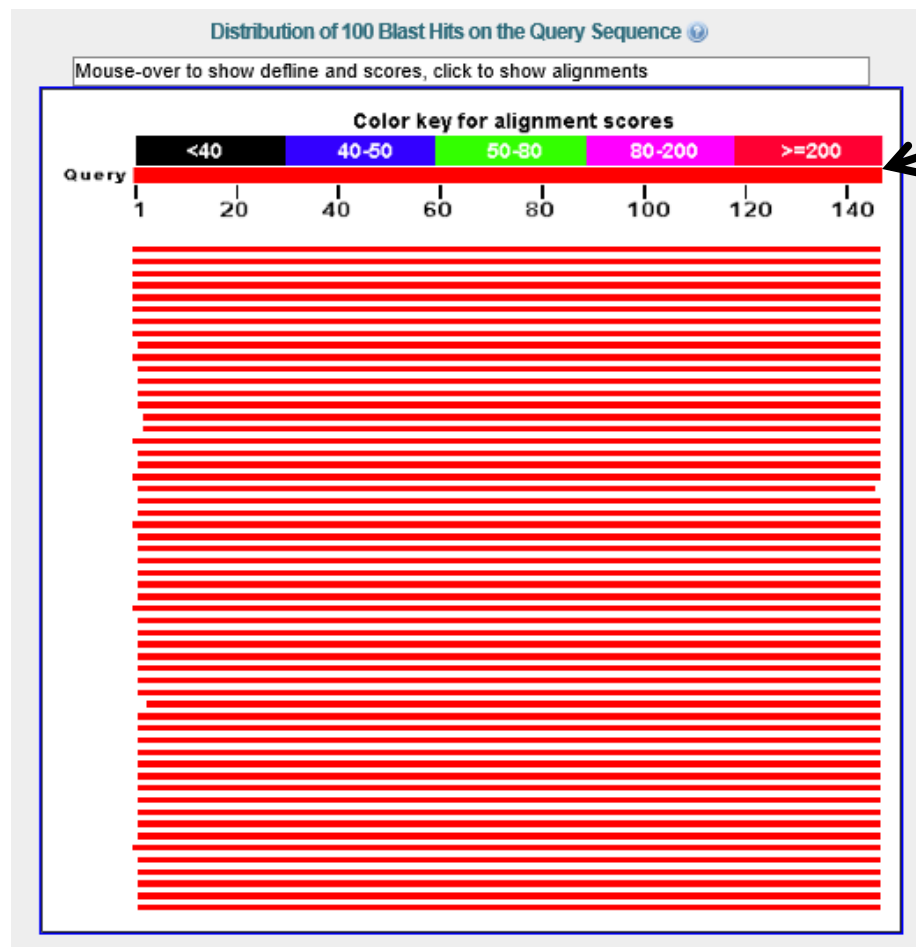
Nome da procura (Search Name): Points to the query sequence 'gi|4504349|ref|NP_000509.1| hemoglobin subunit...'.
Base de dados (Database): Points to the 'Database Name' field, which is 'nr'.
Versão do programa (Program Version): Points to the 'Program' field, which is 'BLASTP 2.2.30+'.
Relatório Taxonómico (Taxonomic Report): Points to the 'Taxonomy reports' link in the 'Other reports' section.
Motivos conservados detectados (Detected conserved motifs): Points to the 'Graphic Summary' section, which displays a sequence logo and identifies conserved domains like 'Hb-beta_like' and 'Globin-like superfamily'.

BLAST - Relatório Taxonómico

Lineage Report

| | | | |
|---|-----|----------|-------------------|
| root | | | |
| synthetic construct | 301 | 10 hits | [other sequences] |
| Homo sapiens (man) | 301 | 576 hits | [primates] |
| Pan troglodytes | 301 | 5 hits | [primates] |
| Pan paniscus (bonobo) | 301 | 4 hits | [primates] |
| Gorilla gorilla gorilla (lowland gorilla) | 300 | 2 hits | [primates] |
| Hylobates lar (white-handed gibbon) | 294 | 2 hits | [primates] |
| Pongo pygmaeus (orang utan) | 293 | 3 hits | [primates] |
| Semnopithecus entellus (Hanuman langur) | 291 | 1 hit | [primates] |
| Colobus polykomos | 291 | 1 hit | [primates] |
| Ateles paniscus | 290 | 2 hits | [primates] |
| Ateles geoffroyi (Central American spide...) | 289 | 1 hit | [primates] |
| Ateles belzebuth (long-haired spider monkey) | 289 | 2 hits | [primates] |
| Macaca fascicularis (long-tailed macaque) | 289 | 4 hits | [primates] |
| Macaca fuscata fuscata | 289 | 1 hit | [primates] |
| Macaca speciosa | 289 | 1 hit | [primates] |
| Macaca nemestrina (pigtail monkey) | 289 | 1 hit | [primates] |
| Cercocebus atys | 289 | 2 hits | [primates] |
| Macaca mulatta (rhesus monkeys) | 289 | 5 hits | [primates] |
| Pithecia pithecia (Guianan saki) | 289 | 2 hits | [primates] |
| Lagothrix lagotricha (common woolly monkey) | 289 | 2 hits | [primates] |
| Chlorocebus aethiops (savanah monkey) | 289 | 2 hits | [primates] |
| Callithrix jacchus (white ear-tufted marmoset) | 288 | 4 hits | [primates] |
| Sapajus apella (brown-capped capuchin) | 288 | 3 hits | [primates] |
| Callicebus moloch (Dusky titi) | 288 | 1 hit | [primates] |
| Callimico goeldii (Goeldi's monkey) | 287 | 2 hits | [primates] |
| Macaca arctoides (bear macaque) | 286 | 1 hit | [primates] |
| Papio anubis (baboon) | 286 | 3 hits | [primates] |
| Ptilocolobus badius (red colobus) | 286 | 1 hit | [primates] |
| Saimiri sciureus (South American squirrel...) | 286 | 2 hits | [primates] |
| Callithrix argentata | 285 | 2 hits | [primates] |
| Mandillus sibiricus | 285 | 2 hits | [primates] |
| hemoglobin beta [synthetic construct] | | | |
| hemoglobin subunit beta [Homo sapiens] >gi 55635219 ref XP | | | |
| hemoglobin subunit beta [Homo sapiens] >gi 55635219 ref XP | | | |
| hemoglobin subunit beta [Homo sapiens] >gi 55635219 ref XP | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| hemoglobin beta | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| hemoglobin subunit beta [Macaca fascicularis] >gi 62901537 | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| PREDICTED: hemoglobin subunit beta [Callithrix jacchus] >gi | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| beta globin (predicted) [Callicebus moloch] | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| hemoglobin beta chain - stump-tailed macaque | | | |
| hemoglobin subunit beta [Papio anubis] >gi 226693534 sp Q9T | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |
| RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-g | | | |

BLAST – Output gráfico



Extensão da sequência
introduzida na procura

As 100 sequências com scores mais
altos apresentam homologia ao
longo de toda a região e com um
“Score” alto (mais de 200)

BLAST - Output

Score

E-value

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

| | Description | Max score | Total score | Query cover | E value | Ident | Accession |
|--------------------------|---|-----------|-------------|-------------|---------|-------|-----------------------------|
| <input type="checkbox"/> | hemoglobin beta [synthetic construct] | 301 | 301 | 100% | 2e-102 | 100% | AAX37051.1 |
| <input type="checkbox"/> | hemoglobin beta [synthetic construct] | 301 | 301 | 100% | 2e-102 | 100% | AAX29557.1 |
| <input type="checkbox"/> | hemoglobin subunit beta [Homo sapiens] | 301 | 301 | 100% | 3e-102 | 100% | NP_000509.1 |
| <input type="checkbox"/> | RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin b | 300 | 300 | 100% | 9e-102 | 99% | P02024.2 |
| <input type="checkbox"/> | beta globin chain variant [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAN84548.1 |
| <input type="checkbox"/> | beta globin [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAZ39780.1 |
| <input type="checkbox"/> | beta-globin [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | ACU56984.1 |
| <input type="checkbox"/> | hemoglobin beta chain [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAD19696.1 |
| <input type="checkbox"/> | Chain B, Structure Of Haemoglobin In The Deoxy Quaternary State With Ligand Bound At The Alpt | 298 | 298 | 99% | 2e-101 | 100% | 1COH_B |
| <input type="checkbox"/> | hemoglobin beta subunit variant [Homo sapiens] | 298 | 298 | 100% | 2e-101 | 99% | AAF00489.1 |

Percentagem de
sequência de procura que
está alinhada

Percentagem de
identidade

- $s' = \frac{\lambda S - \ln K}{\ln 2}$
- $e = m n 2^{-s'}$



+BLAST - Scores

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

| | Description | Max score | Total score | Query cover | E value | Ident | Accession |
|--------------------------|---|-----------|-------------|-------------|---------|-------|-----------------------------|
| <input type="checkbox"/> | hemoglobin beta [synthetic construct] | 301 | 301 | 100% | 2e-102 | 100% | AAX37051.1 |
| <input type="checkbox"/> | hemoglobin beta [synthetic construct] | 301 | 301 | 100% | 2e-102 | 100% | AAX29557.1 |
| <input type="checkbox"/> | hemoglobin subunit beta [Homo sapiens] | 301 | 301 | 100% | 3e-102 | 100% | NP_000509.1 |
| <input type="checkbox"/> | RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin b | 300 | 300 | 100% | 9e-102 | 99% | P02024.2 |
| <input type="checkbox"/> | beta globin chain variant [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAN84548.1 |
| <input type="checkbox"/> | beta globin [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAZ39780.1 |
| <input type="checkbox"/> | beta-globin [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | ACU56984.1 |
| <input type="checkbox"/> | hemoglobin beta chain [Homo sapiens] | 299 | 299 | 100% | 1e-101 | 99% | AAD19696.1 |
| <input type="checkbox"/> | Chain B, Structure Of Haemoglobin In The Deoxy Quaternary State With Ligand Bound At The Alpl | 298 | 298 | 99% | 2e-101 | 100% | 1COH_B |
| <input type="checkbox"/> | hemoglobin beta subunit variant [Homo sapiens] | 298 | 298 | 100% | 2e-101 | 99% | AAF00489.1 |

- “Scores” altos correspondem a “e-values” baixos
- Dependem da porção alinhada com a sequência utilizada na procura (cover)

- Dependem da homologia
- Pequenas porções podem apresentar homologia alta mesmo que a restante proteína não apresente homologia significativa (motivos conservados?)
- Difícil estabelecer a partir de que “score” ou “e-value” o resultado é significativo



BLAST - Alinhamento

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

hemoglobin beta chain [Homo sapiens]
Sequence ID: [gb|AAD19696.1|](#) Length: 147 Number of Matches: 1

[Range 1: 1 to 147](#) [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

| | Score | Expect | Method | Identities | Positives | Gaps |
|-------|---------------|--------|---|--------------|--------------|-----------|
| | 299 bits(766) | 1e-101 | Compositional matrix adjust. | 146/147(99%) | 146/147(99%) | 0/147(0%) |
| Query | 1 | | MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK | | | 60 |
| | | | MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRF ESFGDLSTPDAVMGNPK | | | |
| Sbjct | 1 | | MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFLESFGDLSTPDAVMGNPK | | | 60 |
| Query | 61 | | VKAHGKKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG | | | 120 |
| | | | VKAHGKKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG | | | |
| Sbjct | 61 | | VKAHGKKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG | | | 120 |
| Query | 121 | | KEFTPPVQAAYQKVVAGVANALAHKYH | 147 | | |
| | | | KEFTPPVQAAYQKVVAGVANALAHKYH | | | |
| Sbjct | 121 | | KEFTPPVQAAYQKVVAGVANALAHKYH | 147 | | |

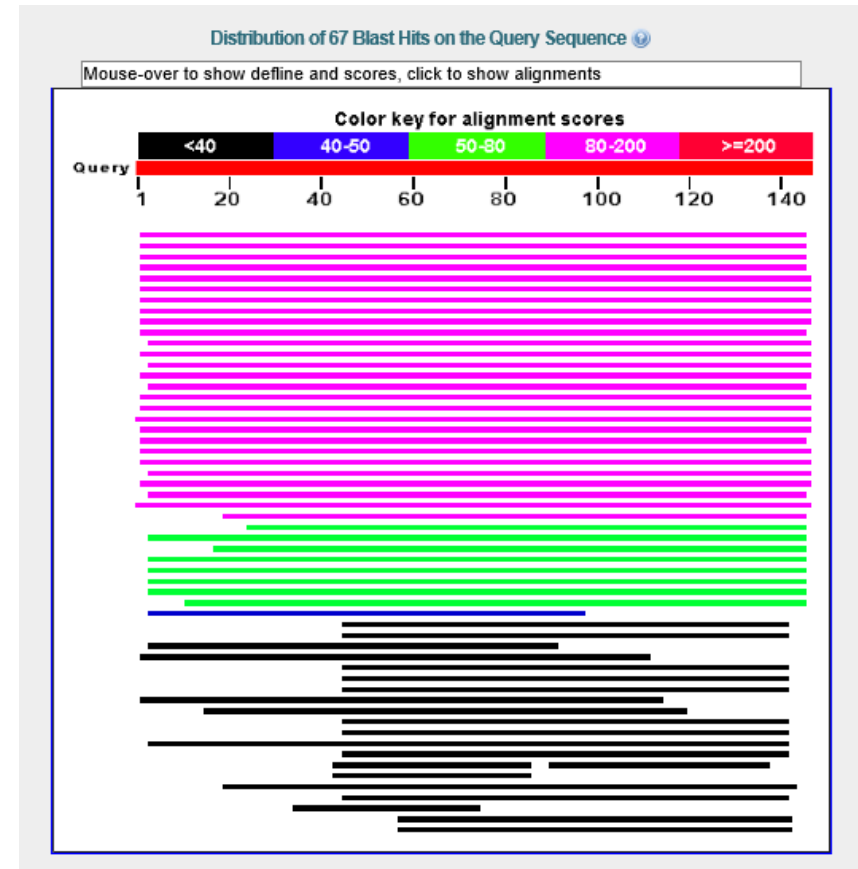
Related Information
[Gene](#) - associated gene details

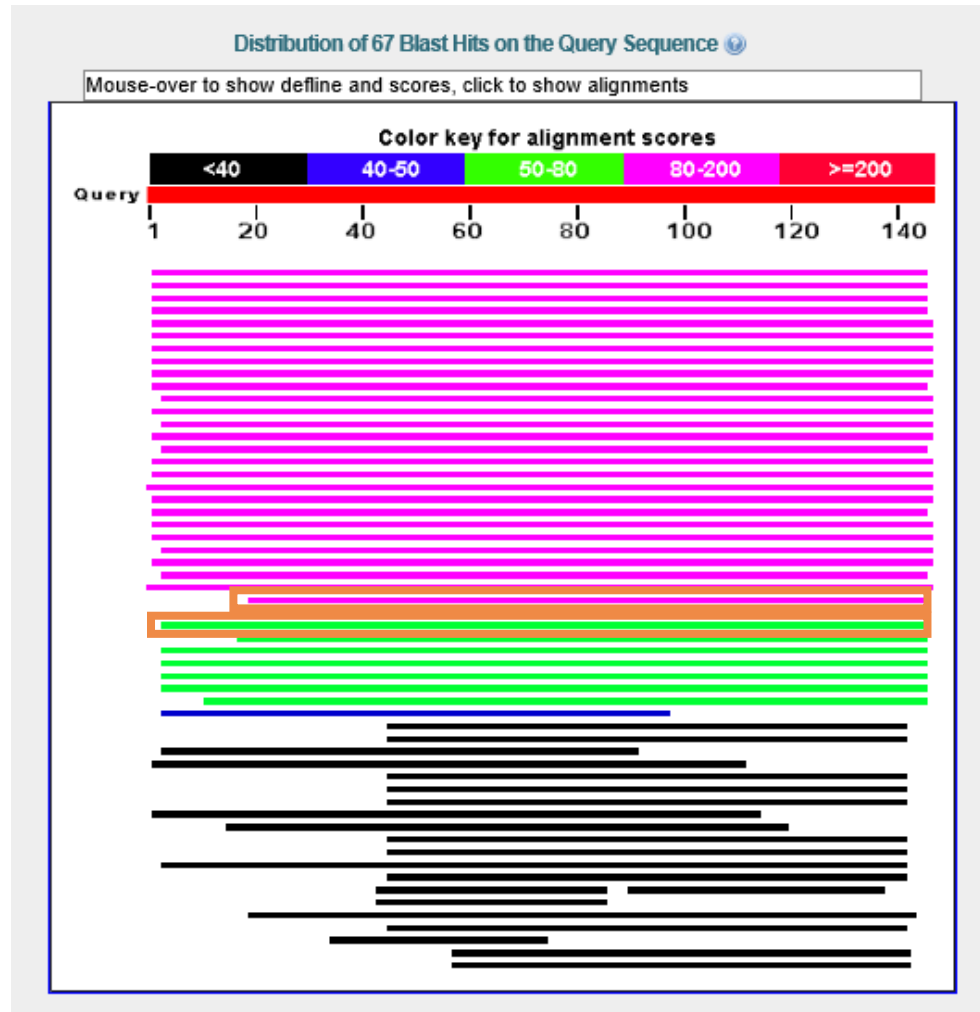
BLAST

■ Exemplo:

- Fazer a mesma busca mas verificar a existência de homólogos fora do grupo taxonómico tetrapoda
- Os scores vão ser mais baixos (menor homologia)
- Em muitos casos o alinhamento não se estenderá por toda a proteína da busca



BLAST



- Uma sequência que não alinhou em toda a extensão da sequência de procura tem um score maior do que uma sequência que alinhou em praticamente toda a extensão
 - score e e-value dependem da extensão do alinhamento e da homologia encontrada

[Download](#) [GenPept](#) [Graphics](#)

RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain [Heterodontus portusjacksoni]

Sequence ID: [sp|P02021.1|HBA_HETPO](#) Length: 148 Number of Matches: 1

Range 1: 27 to 147 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

| Score | Expect | Method | Identities | Positives | Gaps |
|----------------|--|------------------------------|-------------|-------------|-----------|
| 83.6 bits(205) | 2e-19 | Compositional matrix adjust. | 45/127(35%) | 64/127(50%) | 6/127(4%) |
| Query 20 | NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL | 79 | | | |
| | N + G EAL R+ VY T+ +F+ + D + P +KAHG KV+ A + HL | | | | |
| Sbjct 27 | NAEAFGAEALARMFTVYAATKSYFKDYKDFTA-----AAPSIIKAGAKVVTALAKACDHL | 81 | | | |
| Query 80 | DNLKGTFTLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAYQKVVAGVA | 139 | | | |
| | D+LK L+ H +L VDP NF+ L L LA H EF+P A K + V | | | | |
| Sbjct 82 | DDLKTHLHKLATFHGSELKVDPANFYLSYCLEVALAVHLT-EFSPETHCALDKFLTNVC | 140 | | | |
| Query 140 | NALAHKY | 146 | | | |
| | + L+ +Y | | | | |
| Sbjct 141 | HELSSRY | 147 | | | |

[Download](#) [GenPept](#) [Graphics](#)

RecName: Full=Globin D, coelomic [Caudina arenicola]

Sequence ID: [sp|P80017.3|GLBD CAUAR](#) Length: 159 Number of Matches: 1

Range 1: 13 to 157 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

| Score | Expect | Method | Identities | Positives | Gaps |
|----------------|--|------------------------------|-------------|-------------|-----------|
| 53.9 bits(128) | 1e-08 | Compositional matrix adjust. | 43/147(29%) | 64/147(43%) | 6/147(4%) |
| Query 4 | LIPEEKSAVTALWGKVNVDEVG--GEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV | 61 | | | |
| | LTP EK + + W ++ G + R+ P QR F LS P + + ++ | | | | |
| Sbjct 13 | LTPAEKDLIRSTWDQLMTHRTGFVADVIRIFHNDPTAQRKFPQMAGLS-PAELRTSRQM | 71 | | | |
| Query 62 | KAHGKKVLGAFSDGLAHLN--LKGTFATLSELHCDKLHVDPENFRLLGNVLCVLAHHF | 119 | | | |
| | AH +V + + +D L ATL+ H DK HV +N+ L G VL+ + | | | | |
| Sbjct 72 | HAHAIRVSALMTTYIDEMDTEVLPELLATLIRTH-DKNHVGKKNYDLFGKVLMEAIKAEI | 130 | | | |
| Query 120 | GKEFTPPVQAAYQKVVAGVANALAHKY | 146 | | | |
| | G FT V A+ K A V L K+ | | | | |
| Sbjct 131 | GVGFTKQVHDAWAKTFAIVQGVLIITKH | 157 | | | |

Linha de comando

BLAST no Bioinformatics Linux

- O Bioinformatics Linux contem o docker do **blast+**.
- Sintaxe para executar a shell do docker:
 - *docker run --rm -it *
*-v ~/dockermounts/blastdb:/blast/blastdb:rw *
*-v ~/dockermounts/queries:/blast/queries:rw *
*-v ~/dockermounts/results:/blast/results:rw *
*ncbi/blast *
/bin/bash

BLAST no Bioinformatics Linux

- Parâmetros passados usados dentro do container:
 - identificação do programa: blastp, blastn, blastx and tblastx
 - **-db** identificação da Base de Dados
 - **-query** localização do ficheiro de *input*
 - **-out** especificação do ficheiro de *output*
- Exemplo :

```
$ blastp -db swissprot -query mypeptides.tfa -out mypeptides.out
```