

Hemoglobin subunit beta - HBB

Laboratórios de Bioinformática

Trabalho realizado por:
Cristiana Silva – up202305464
Filipa Marinha – up202304935
Filipa Ferreira – up202305895

Motivação do Trabalho

As principais motivações para a realização deste trabalho foram:

- Aprofundar a nossa compreensão sobre conceitos biológicos e ferramentas bioinformáticas (como por exemplo BLAST, Clustal Omega, NCBI, UniProt....);
- Compreensão da homologia de proteínas;
- Análise de semelhança entre sequências e árvores filogenéticas;
- Interpretação das relações evolutivas e as similaridades funcionais entre proteínas;
- Ajuste de alinhamentos de sequências múltiplas (MSA) para identificação de regiões conservadas em sequências de proteínas.

Objetivo do Trabalho

Com a realização deste trabalho pretende-se:

- Compreender os conceitos de homologia de proteínas, semelhança de sequências e árvores filogenéticas;
- Ganhar experiência na utilização do BLAST para procurar homólogos de proteínas;
- Aprender a efetuar alinhamentos de sequências múltiplas (MSA) para identificar regiões conservadas;
- Gerar uma árvore filogenética para visualizar relações evolutivas entre sequências de proteínas.

Principais Ferramentas de Bioinformática Utilizadas



NCBI



UniProt

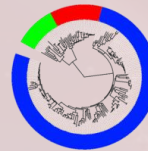


EMBL-EBI

EMBL-EBI



iTOL



Etapas da Análise

01 - Pesquisa das sequências humana e homólogas e organização num único ficheiro



02 - MSA das sequências



04 - Árvore filogenética



03 - Identificação de domínios



05 - Logo das sequências



06 - Identificação de motivos

Resultados

MSA (com cores):

CLUSTAL 0(1.2.4) multiple sequence alignment

```

XP_020754937.1      NM_000569.1      NM_018891709.1    NM_004966697.3    NM_002821273.1    NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTGDEASVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NP_000569.1      NM_018891709.1    NM_004966697.3    NM_002821273.1    NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NP_018891709.1    NM_004966697.3    NM_002821273.1    NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NM_004966697.3    NM_002821273.1    NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NM_002821273.1    NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NM_01292888.1    AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      AY09367.1         AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      AY09363.1         NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      NM_0131647.1      NM_018361446.1    NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      NM_018361446.1    NM_0131647.1      XP_033062959.1
      NM_0131647.1      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      XP_033062959.1
      MHVLTPEESAVATLTKGKVVNDEGGELGRLLVVPYQTFEFESGDLSTPDVAINPK 60
      *****

```

NP_020754937.1	VKANGVGLV, GAFSGDL, LNL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_006059.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_018189170.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_004096967.3	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_001282273.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_001292088.1	VKANGVGLV, GAFSGDL, NHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
AY009363.1	VKANGVGLV, GAFSGDL, LNL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
NP_01316467.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
XP_013164146.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12
XP_03362959.1	VKANGVGLV, GAFSGDL, AHL, DNL, KGTFAHSL, SCLHCL, HDVPE, RLL, GVL, VCV, LAHFFG	12

```

XP_002754937.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
NP_000509.1         KFTFPVQAAQYQKVVGAVNALAHKHYH 147
XP_018891709.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
XP_004069697.3      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
XP_008282173.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
NP_001292888.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
AY089367.1          KFTFPVQAAQYQKVVGAVNALAHKHYH 147
AY089363.1          KFTFPVQAAQYQKVVGAVNALAHKHYH 147
NP_001316847.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
XP_019361646.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147
XP_033062959.1      KFTFPVQAAQYQKVVGAVNALAHKHYH 147

*****

```

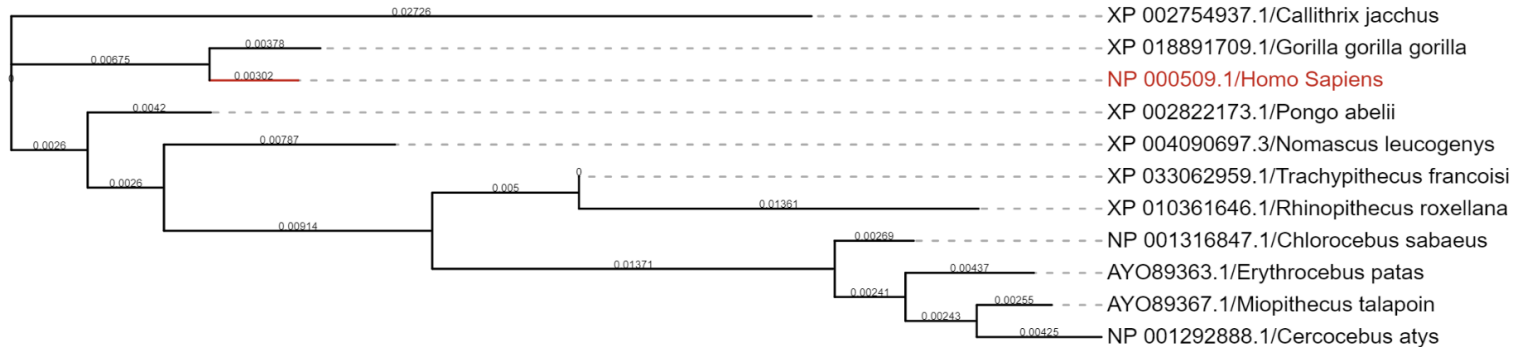
LOGO:



Resultados

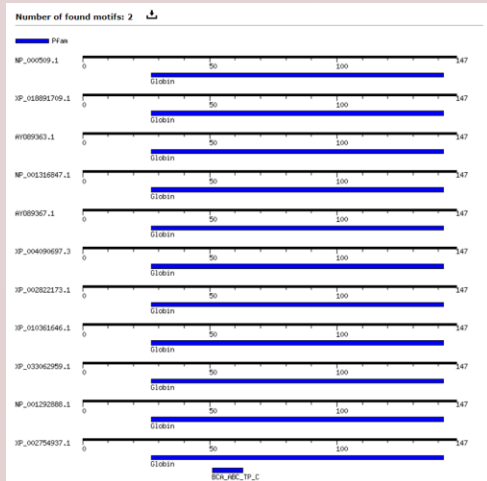
Árvore filogenética:

Tree scale: 0.01



Resultados

Motifs:



Pfam (2 motifs)

Query	Pfam	Position(Independent E-value)	Description
NP_000509.1	Globin	27..142(4.3e-33)	Detail PF00042, Globin
XP_018891709.1	Globin	27..142(2e-33)	Detail PF00042, Globin
AY089363.1	Globin	27..142(7.2e-34)	Detail PF00042, Globin
NP_001316847.1	Globin	27..142(5.6e-34)	Detail PF00042, Globin
AY089367.1	Globin	27..142(7.2e-34)	Detail PF00042, Globin
XP_004090697.3	Globin	27..142(1.2e-33)	Detail PF00042, Globin
XP_002822173.1	Globin	27..142(1.1e-33)	Detail PF00042, Globin
XP_010361646.1	Globin	27..142(2.3e-33)	Detail PF00042, Globin
XP_033062959.1	Globin	27..142(1.2e-33)	Detail PF00042, Globin
NP_001292888.1	Globin	27..142(3.9e-34)	Detail PF00042, Globin
XP_002754937.1	Globin	27..142(5.2e-33)	Detail PF00042, Globin
XP_002754937.1	BCA_ABC_TP_C	51..63(0.12)	Detail PF12399, Branched-chain amino acid ATP-binding cassette transporter

Domínios:



Conclusão

- O gene HBB é de extrema importância para o transporte de oxigênio no sangue;
- Ao comparar a sua sequência de aminoácidos com as das proteínas homólogas é possível observar que a maioria das sequências encontram-se em concordância;
- Como o gene encontra-se bem conservado (entre espécies relativamente semelhantes), podemos concluir que a maior parte da sua sequência é importante para a sua função;
- De todos os genes, o do *homo sapiens* é o segundo mais próximo do ancestral comum, sendo o do *Pongo abelii* o mais próximo;
- O gene mais parecido com o do *Homo sapiens* é o do *Gorilla gorilla gorilla*, o que apoia a hipótese de essa espécie ser das espécies mais geneticamente semelhantes ao ser humano.

Bibliografia

<https://www.youtube.com/watch?v=y5UaloYTxCk>

<https://www.ncbi.nlm.nih.gov/search/all/?term=HBB>

<https://www.uniprot.org/uniprotkb?query=HBB>

<https://itol.embl.de/>

<https://www.ebi.ac.uk/>

<https://en.vectorbuilder.com/tool/sequence-alignment/d3fb58e3-5aa0-4df3-9945-61c9dddb6ef3.html>

<https://itol.embl.de/upload.cgi>

https://www.genome.jp/tools-bin/search_motif_lib

<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>