



Instituto Politécnico Nacional  
Escuela Superior de Cómputo

Bioinformatics

Practice 6 - UniProt

Group: 3CV9

*Student:*  
Ramos Diaz Enrique

*Professor:*  
Rosas Trigueros Jorge Luis

*Development Date:*  
November 18th 2020

*Due Date:*  
November 25th 2020

# 1 Theoretical Framework

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc). The UniProt consortium and host institutions EMBL-EBI, SIB and PIR are committed to the long-term preservation of the UniProt databases [1]. Figure 1 shows how UnitProt is conformed.

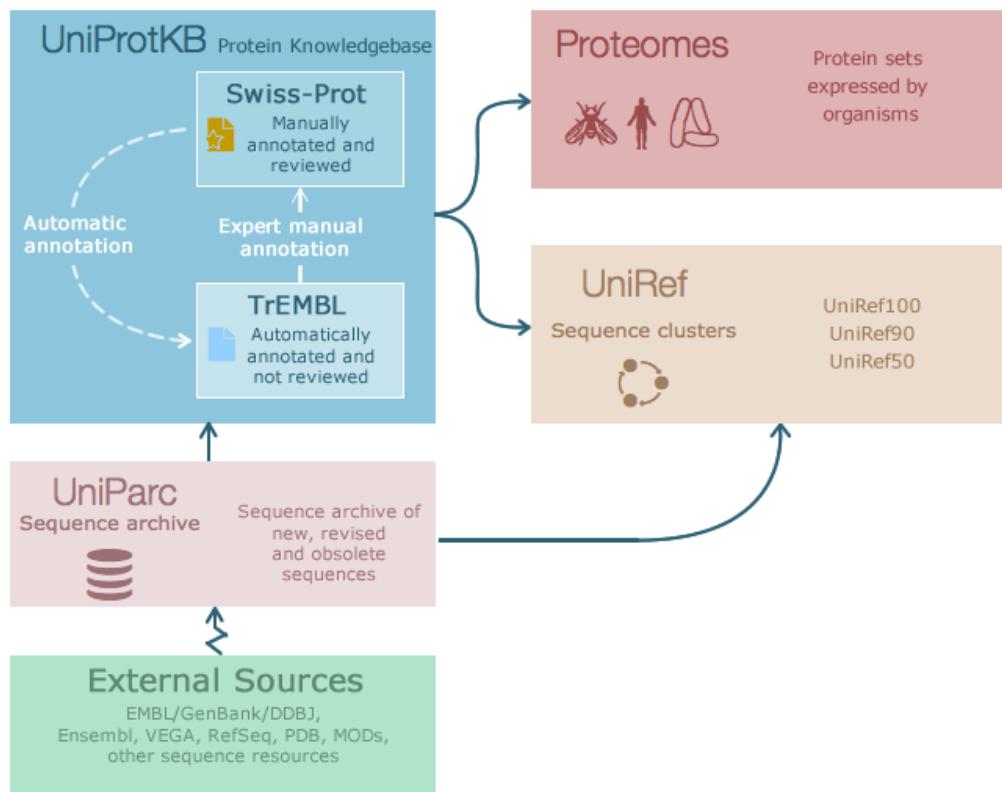


Figure 1: UnitProt.

UniProt is a collaboration between the European Bioinformatics Institute (EMBL-EBI), the SIB Swiss Institute of Bioinformatics and the Protein Information Resource (PIR). EMBL-EBI and SIB together used to produce Swiss-Prot and TrEMBL, while PIR produced the Protein Sequence Database (PIR-PSD).

These two data sets coexisted with different protein sequence coverage and annotation priorities. TrEMBL (Translated EMBL Nucleotide Sequence Data Library) was originally created because sequence data was being generated at a pace that exceeded Swiss-Prot's ability to keep up.

Meanwhile, PIR maintained the PIR-PSD and related databases, including iProClass, a database of protein sequences and curated families. In 2002 the three institutes decided to pool their resources and expertise and formed the UniProt consortium [1].

## 2 Material and Equipment

- UniProt web page [2].
- Text editor.
- Google [3].

## 3 Practice Development

The objective of this practice is to use the UnitProt web page [2] alongside two of its tools: Align [4] and BLAST [5]. The first one make a comparison between the proteins entered of different species of living beings (animals, plants, insects, etc.), and tells how related, similar or close are each of them according to their similitude of its protein sequence, or how well it matches between them. On the second one just one protein of a living being is entered and it searches for proteins from other living beings that are similar, indicating which other species could be related to this one entered.

### 3.1 Searching animal species

In this practice there will be performed two aligns and two blasts with different examples of animal species, using the **hemoglobin** protein in all of them. It's important that all the selected animals have their hemoglobin sequence available on the UnitProt database [2].

#### 3.1.1 Animal set 1

For the first analysis, the selected species are the following (scientific names obtained from Google [3]):

1. Human - *Homo sapiens*
2. Dog - *Canis lupus familiaris*
3. Kramer Parrot - *Psittacula krameri*
4. Lion - *Panthera leo*
5. Giant Turtle - *Chelonoidis niger*
6. European Otter - *Lutra lutra*
7. Nile Crocodile - *Crocodylus niloticus*
8. Indian Cobra - *Naja naja*
9. Black Fruit Bat - *Pteropus alecto*
10. Ostrich - *Struthio camelus*
11. Sea cow - *Hydrodamalis gigas*
12. Cat - *Felis catus*
13. Polar Bear - *Ursus maritimus*
14. Black Bear - *Ursus thibetanus*
15. Malay Bear - *Helarctos malayanus*
16. Chimpanzee - *Pan troglodytes*
17. Grizzly Bear - *Ursus arctos horribilis*
18. Raccoon - *Procyon lotor*
19. Badger - *Meles meles*
20. Panda Bear - *Ailuropoda melanoleuca*

### 3.1.2 Animal set 2

For the second analysis, the selected species are the following:

- |   |   |
|---|---|
| 1. Platypus - <i>Ornithorhynchus anatinus</i>   | 11. Dugong - <i>Dugong dugon</i>                  |
| 2. Dolphin - <i>Delphinus delphis</i>           | 12. Sperm Whale - <i>Physeter macrocephalus</i>   |
| 3. Wild Pig - <i>Sus scrofa</i>                 | 13. Beluga - <i>Delphinapterus leucas</i>         |
| 4. Seal - <i>Phoca vitulina</i>                 | 14. Hedgehog - <i>Erinaceus europaeus</i>         |
| 5. Real Duck - <i>Anas platyrhynchos</i>        | 15. Opossum - <i>Didelphis virginiana</i>         |
| 6. White Swan - <i>Cygnus olor</i>              | 16. Turkey - <i>Meleagris gallopavo</i>           |
| 7. Canadian Beaver - <i>Castor canadensis</i>   | 17. Rabbit - <i>Oryctolagus cuniculus</i>         |
| 8. Walrus - <i>Odobenus rosmarus</i>            | 18. Emperor Penguin - <i>Aptenodytes forsteri</i> |
| 9. Muskrat - <i>Ondatra zibethicus</i>          | 19. Guinea Pig - <i>Cavia porcellus</i>           |
| 10. Amazonian Tapir - <i>Tapirus terrestris</i> | 20. Pigeon - <i>Columba livia</i>                 |

## 3.2 Building the sets of hemoglobin sequences

Now that the animal species are defined for both analyses, their hemoglobin sequences have to be searched in UnitProt, typing the animal's name on the search bar. Figure 2 shows the results for the "dog hemoglobin" search.

| Entry  | Entry name | Protein names            | Gene names | Organism                            | Length |
|--------|------------|--------------------------|------------|-------------------------------------|--------|
| P60529 | HBA_CANLF  | Hemoglobin subunit alpha | HBA        | <i>Canis lupus familiaris</i> (Dog) | 141    |
| P68871 | HBB_HUMAN  | Hemoglobin subunit beta  | HBB        | <i>Homo sapiens</i> (Human)         | 147    |
| P60524 | HBB_CANLF  | Hemoglobin subunit beta  | HBB        | <i>Canis lupus familiaris</i> (Dog) | 146    |
| O97492 | CATA_CANLF | Catalase                 | CAT        | <i>Canis lupus familiaris</i> (Dog) | 527    |
| Q9TUS3 | CUBN_CANLF | Cubilin                  | CUBN       | <i>Canis lupus familiaris</i> (Dog) | 3,620  |

Figure 2: UnitProt search results for dog hemoglobin.

Hemoglobin protein has two sub-units: alpha and beta. It is recommended to pick the same

sub-unit for all the cases in an analysis. For this practice, the alpha sub-unit of the protein will be chosen for all the animals selected.

Figure 3a shows the dog's hemoglobin sub-unit alpha entry in UniProt. There are several formats available for the protein sequence, but for the analyses that will be performed on this practice, the **FASTA (canonical)** is the necessary one (see Figure 3b).

The figure consists of two screenshots of the UniProtKB interface. Both screenshots show the same protein entry: P60529 (HBA\_CANLF), which is Hemoglobin subunit alpha from Canis lupus familiaris (Dog).  
 (a) Entry: This screenshot shows the main entry page. At the top, there are tabs for Display, BLAST, Align, Format, Add to basket, and History. Below these are sections for Entry (highlighted in blue), Publications, Feature viewer, Feature table, Function (with 'Function' checked), Names & Taxonomy (with 'HBA' checked), Subcellular location (with 'None' checked), and Miscellaneous. The Function section contains the text: "Involved in oxygen transport from the lung to the various peripheral tissues." The Miscellaneous section contains the text: "The sequence shown is one of two non-allelic alpha chains from dog."  
 (b) FASTA Format: This screenshot shows the same entry page, but with a dropdown menu open under the 'Format' tab. The 'View format' dropdown has 'FASTA (canonical)' selected. Other options in the dropdown include Text, XML, RDF/XML, and GFF. The rest of the page content is identical to (a), including the detailed description of the protein's function and its role in oxygen transport.

Figure 3: Dog Hemoglobin sub-unit alpha.

The Dog Hemoglobin subunit alpha sequence in FASTA format looks similar to the next block code:

```

1 >sp|P60529|HBA_CANLF Hemoglobin subunit alpha OS=Canis lupus familiaris
   → OX=9615 GN=HBA PE=1 SV=1
2 VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPFSDLSPGSAQVKAHGK
3 KVADALTTAVAHLDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTP
4 VHASLDKFFAAVSTVLT SKYR
  
```

In each of the sequences from the selected animal species, their respective scientific name has to be put between the first '>' character and the "sp" string, as shown below:

```

1 >Canis_lupus_familiaris sp|P60529|HBA_CANLF Hemoglobin subunit alpha
   → OS=Canis lupus familiaris OX=9615 GN=HBA PE=1 SV=1
2 VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPFSDLSPGSAQVKAHGK
3 KVADALTTAVAHLDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTP
4 VHASLDKFFAAVSTVLT SKYR
  
```

In this way, the hemoglobin sequences for the [Animal set 1](#) with the previous fix are:

```

1 >Homo_sapiens sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens
   → OX=9606 GN=HBA1 PE=1 SV=2
2 MVLS PADK TNVKA WGKV GAH AGEY GAE ALER MF LS FPTT KTY FP FH FD LS HG SAQ VKG HG
3 KK VAD ALTNA VAH VDD MPN A L S A L S D L H AH K L R V DP V N F K L L S H C L L V T L A A H L P A E F T P
4 AV HAS LDK F L A S V S T V L T SKYR
5
  
```

6 >Canis\_lupus\_familiaris sp|P60529|HBA\_CANLF Hemoglobin subunit alpha  
  ↳ OS=Canis lupus familiaris OX=9615 GN=HBA PE=1 SV=1  
7 VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHFDLSPGSAQVKAHGK  
8 KVADALTTAVAHLDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTP  
9 VHASLDKFFAAVSTVLT SKYR  
10  
11 >Psittacula\_krameri sp|P19831|HBA\_PSIKR Hemoglobin subunit alpha  
  ↳ OS=Psittacula krameri OX=9228 GN=HBA PE=1 SV=1  
12 VLSGTDKTNVKSIFSKIGGQADDYGAEARLMFVTYPQTCKTYFPHFDVSPGSAQVKAHGK  
13 KVAGGLSEAANHIDDIATSLSKLSDLHAQKLRVDPVNFKLLGQCFLVVVAIHNPALTPE  
14 AHASLDKFCAVGLVLTAKYR  
15  
16 >Panthera\_leo sp|P18975|HBA\_PANLE Hemoglobin subunit alpha OS=Panthera leo  
  ↳ OX=9689 GN=HBA PE=1 SV=2  
17 MVLSSADKNNVKACWGKIGSHAGEYGAEARLTFCSFPTTKTYFPHFDLHSQVQAHG  
18 QKVADALTKAVVHINDLPNALSDLSDLHAYKLRVDPVNFKFLSHCLLVTLACHHPEEFTP  
19 AVHASLDKFFSAVSTVLT SKYR  
20  
21 >Chelonoidis\_niger sp|P83135|HBAA\_CHENI Hemoglobin subunit alpha-A  
  ↳ OS=Chelonoidis niger OX=66189 GN=HBAA PE=1 SV=2  
22 MVLTAGDKANVKTVWSKGSHLEYGSETLERLFIVYPSTKTYFPHFDLHHDSAQVRAHG  
23 RKVLSALGEAVNHIDDIPIGALSKLSDLHAQTLRVDPVNFKLLNLCFVVVGRHHPTILTP  
24 EVHVS LDKFLSAVATALT SKYR  
25  
26 >Lutra\_lutra sp|P10892|HBA\_LUTLU Hemoglobin subunit alpha OS=Lutra lutra  
  ↳ OX=9657 GN=HBA PE=1 SV=1  
27 VLSPADKTNVKSTWDKIGGHAGDYGGEALRTFVSFPTTKTYFPHFDLHSQVKAHGK  
28 KVADALTNAVAHMDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPAEFTP  
29 VHASLDKFFSAVSTVLT SKYR  
30  
31 >Crocodylus\_niloticus sp|P01998|HBA\_CRONI Hemoglobin subunit alpha  
  ↳ OS=Crocodylus niloticus OX=8501 GN=HBA PE=1 SV=1  
32 VLSSDDKCNVKAVWSKGSHLEYGAEALERMFCAYPQTCKTYFPHFDLHSQVRAHGK  
33 KVFAALHEAVNHIDDLPGALCRSELHSLRVDPVNFKFLAQCVLVVVAIHHPGSLTPE  
34 VHASLDKFCAVSSVLT SKYR  
35  
36 >Naja\_naja sp|Q9PRL9|HBA1\_NAJNA Hemoglobin subunit alpha-1 OS=Naja naja  
  ↳ OX=35670 PE=1 SV=1  
37 VLTDEDKARVRASWVPVGKNAELYGSETLTRMFAAHPTTKTYFPHFDLSPGSNNLRAHGK  
38 KVIDAITEAVNNLDDVAGTLSKLSDLHAQKLRVDPVNFKLLAHCLLVTIAAHNGGVLKPE  
39 VIVSLDKFLGDLSKDLVSKYR  
40  
41 >Pteropus\_alecto sp|P14389|HBA\_PTEAL Hemoglobin subunit alpha OS=Pteropus  
  ↳ alecto OX=9402 GN=HBA PE=1 SV=1

42 VLSSTDKSNVKAADKVGGHVGEYGAEARMFLSFPTTKTYFPHFDLAHGSSQVKAHKG  
43 KVGDALTNAGHIDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTASHLPSDFTPA  
44 VHASLDKFLASVSTVLTSKYR  
45  
46 >Struthio\_camelus sp|P01981|HBA\_STRCA Hemoglobin subunit alpha-A OS=Struthio  
  → camelus OX=8801 GN=HBAA PE=1 SV=1  
47 VLSGTDKTNVKGIFSKISSHAEYGAETLERMFITYPQTCKTYFPHFDLHHGSAQIKAHKG  
48 KVANALIEAVNHIDDISGALKSDLHAQKLRVDPVNFKLLGQCFLVVVAIHPSALTPE  
49 VHASLDKFLCAVGAVLTAKYR  
50  
51 >Hydrodamalis\_gigas tr|AOA481WNT7|AOA481WNT7\_HYDGI Hemoglobin subunit alpha  
  → OS=Hydrodamalis gigas OX=63631 GN=HBA PE=3 SV=1  
52 MVLSAEDKTNVKTFWGKLGHTAEYGGEALERMFLSFPTTKTYFPHFDMKHDSDQIKAHG  
53 KKVALDALTRAVGHLDLPGTLSESDLHAHKLRVDPINFKLLSHCLLVTLSGHPEDFTP  
54 PVHASLDKFLSNVSTVLTSKYR  
55  
56 >Felis\_catus sp|P07405|HBA\_FELCA Hemoglobin subunit alpha OS=Felis catus  
  → OX=9685 GN=HBA PE=1 SV=1  
57 VLSAADKSNVKAACWGKIGSHAGEYGAEARLERTFCSFPTTKTYFPHFDLHSQVKAHQ  
58 KVADALTQAVAHMDDLPTAMSALSDLHAYKLRVDPVNFKFLSHCLLVTLACHHPAEFTP  
59 VHASLDKFFSAVSTVLTSKYR  
60  
61 >Ursus\_maritimus sp|P68235|HBA\_URSMA Hemoglobin subunit alpha OS=Ursus  
  → maritimus OX=29073 GN=HBA PE=1 SV=2  
62 MVLSPADKSNVKATWDKIGSHAGEYGGGEALERTFASFPTTKTYFPHFDLSPGSQVKAHG  
63 KKVALDALTTAACGHLDLPGALSALSDLHAHKLRVDPVNFKFLSHCLLVTLASHHPAEFTP  
64 AVHASLDKFFSAVSTVLTSKYR  
65  
66 >Ursus\_thibetanus sp|P68236|HBA\_URSTH Hemoglobin subunit alpha OS=Ursus  
  → thibetanus OX=9642 GN=HBA PE=1 SV=2  
67 MVLSPADKSNVKATWDKIGSHAGEYGGGEALERTFASFPTTKTYFPHFDLSPGSQVKAHG  
68 KKVALDALTTAACGHLDLPGALSALSDLHAHKLRVDPVNFKFLSHCLLVTLASHHPAEFTP  
69 AVHASLDKFFSAVSTVLTSKYR  
70  
71 >Helarctos\_malayanus sp|P68237|HBA\_HELMA Hemoglobin subunit alpha  
  → OS=Helarctos malayanus OX=9634 GN=HBA PE=1 SV=2  
72 MVLSPADKSNVKATWDKIGSHAGEYGGGEALERTFASFPTTKTYFPHFDLSPGSQVKAHG  
73 KKVALDALTTAACGHLDLPGALSALSDLHAHKLRVDPVNFKFLSHCLLVTLASHHPAEFTP  
74 AVHASLDKFFSAVSTVLTSKYR  
75  
76 >Pan\_troglodytes sp|P69907|HBA\_PANTR Hemoglobin subunit alpha OS=Pan  
  → troglodytes OX=9598 GN=HBA1 PE=1 SV=2  
77 MVLSPADKTNVKAAGKVGGAHAGEYGAEARMFLSFPTTKTYFPHFDLHSQVKGHG  
78 KKVALDNLNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAHLPAEFTP

```

79 AVHASLDKFLASVSTVLTSKYR
80
81 >Ursus_arctos_horribilis tr|A0A3Q7UWX5|A0A3Q7UWX5_URSAR hemoglobin subunit
  → alpha OS=Ursus arctos horribilis OX=116960 GN=HBA1 PE=3 SV=1
82 MVLS PADKSNVKATWDKIGSHAGEYGEALERTFASFPTT KTYFPHFDLSPGSAQVKAHG
83 KK VAD ALTTAAGH LDDLP GALSALSDL HAHKL RVDPVNFKFLSHC LLVT LASHHPAEFTP
84 AVHASLDKFFSAVSTVLTSKYR
85
86 >Procyon_lotor sp|P18977|HBA_PROLO Hemoglobin subunit alpha OS=Procyon lotor
  → OX=9654 GN=HBA PE=1 SV=1
87 VLSPADKANI KATWDKIGGHAGEYGEALERTFASFPTT KTYFPHFDLSPGSAQVKAHGK
88 KVAD ALTLAVGHL DDLPGAL SALSALSDL HAYKL RVDPVNFKLLSHC LLVT LACHHPAEFTP
89 VH ASLDKFFT SVSTVLTSKYR
90
91 >Meles_meles sp|P01953|HBA_MELME Hemoglobin subunit alpha OS=Meles meles
  → OX=9662 GN=HBA PE=1 SV=1
92 VLSPADKANI KATWDKIGGHAGEYGEALERTFASFPTT KTYFPHFDLSPGSAQVKG HGK
93 KVAD ALTNAVAH LDDLP GALSALSDL HAYKL RVDPVNFKLLSHC LLVT LACHHPAEFTP
94 VH ASLDKFLSSVSTVLTSKYR
95
96 >Ailuropoda_melanoleuca sp|P18970|HBA_AILME Hemoglobin subunit alpha
  → OS=Ailuropoda melanoleuca OX=9646 GN=HBA PE=1 SV=2
97 MVLS PADKTNVKATWDKIGGHAGEYGEALERTFASFPTT KTYFPHFDLSPGSAQVKAHG
98 KK VAD ALTTAVGHL DDLPGAL SALSALSDL HAHKL RVDPVNFKLLSHC LLVT LASHHPAEFTP
99 AVHASLDKFFSAVSTVLTSKYR

```

By contrast, the hemoglobin sequences for the [Animal set 2](#) with the fix are:

```

1 >Ornithorhynchus_anatinus sp|P01979|HBA_ORNAN Hemoglobin subunit alpha
  → OS=Ornithorhynchus anatinus OX=9258 GN=HBA PE=1 SV=1
2 ML TDAEKKEVTALWGKAAGHGE EYGAEALERLFQAFPTT KTYFSHFDLSPGSAQIKAHGK
3 KVAD ALSTAAGHFDDMDSAL SALSALSDL HAHKL RVDPVNFKLLAH CILVV LARHCPGEFTPS
4 AHAAMD KFLSKV ATVLTSKYR
5
6 >Delphinus_delphis tr|D3TTJ8|D3TTJ8_DELDE Hemoglobin alpha2 chain
  → OS=Delphinus delphis OX=9728 PE=2 SV=1
7 MVLS PADKTNVKGTWAKIGNHS AEGAEALERMFINF P STKTYFSHFDLGHG SAQIK GHG
8 KK VAD ALTKAVGHI DNLPD ALSELSDL HAHKL RVDPVNFKLLSHC LLVT LALHLPADFTP
9 SV HASLDKFLASVSTVLTSKYR
10
11 >Sus_scrofa sp|P01965|HBA_PIG Hemoglobin subunit alpha OS=Sus scrofa OX=9823
  → GN=HBA PE=1 SV=1
12 VLSAADKANVKA AWGKVGGQAGA HGA EALER MFILGFPTT KTYFPHFNLSHGSDQVKAHGQ
13 KVAD ALTKAVGHL DDLPGAL SALSALSDL HAHKL RVDPVNFKLLSHC LLVT LAAHHPDDFNPS
14 VH ASLDKFLANVSTVLTSKYR

```

15  
16 >Phoca\_vitulina sp|P09908|HBA\_PHOVI Hemoglobin subunit alpha OS=Phoca  
17    ↳ vitulina OX=9720 GN=HBA PE=1 SV=1  
18 VLSPADKTNVKATWDKIGGHAGEYGGAEALERTFTAFPTTKTYFPHFDSLHGSQVKAHGK  
19 KVADALTTAVAHMDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHPADFTPA  
20 VHASLDKFFSAVSTVLTSKYR  
21  
22 >Anas\_platyrhynchos sp|P01988|HBA\_ANAPP Hemoglobin subunit alpha-A OS=Anas  
23    ↳ platyrhynchos platyrhynchos OX=8840 GN=HBAA PE=1 SV=2  
24 MVLSAADKTNVKGVFSKIGGHAEYGAETLERMFIAYPQTCKTYFPHFDSLHGSQVKAHGK  
25 KKVAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPAALTP  
26 EVHASLDKFMCAGVAVLTAKYR  
27  
28 >Cygnus\_olor sp|P01992|HBA\_CYGOL Hemoglobin subunit alpha-A OS=Cygnus olor  
29    ↳ OX=8869 GN=HBAA PE=1 SV=1  
30 VLSSADDKANIKAWSKVGGHGAEYGAEALERMFHSFPTTCKTYFPHFDSLHGSQVKAHGK  
31 KVAAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPAALTP  
32 VHASLDKFLCAVGAVLTAKYR  
33  
34 >Castor\_canadensis tr|A0A250Y9Z1|A0A250Y9Z1\_CASCN Hemoglobin alpha, adult  
35    ↳ chain 2 OS=Castor canadensis OX=51338 GN=HBA-A2 PE=3 SV=1  
36 MVLSSDDKANIKAWSKVGGHGAEYGAEALERMFHSFPTTCKTYFPHFDSLHGSQVKAHGK  
37 KVAAEALTNASNHLDDLYSALSDLSDLHAHKLRVDPVNFKLLSHSLLVTLANHHPNEFTP  
38 SIHASLDKFFAAVGTVLTSKYR  
39  
40 >Odobenus\_rosmarus sp|P10778|HBA\_ODORO Hemoglobin subunit alpha OS=Odobenus  
41    ↳ rosmarus divergens OX=9708 GN=HBA PE=1 SV=1  
42 VLSPADKTNVKTWDKLGGHAGEYGGAEALERTFMSFPTTCKTYFPHFDSLPGSAQVKAHGK  
43 KVADALTTAVAHIDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPAEFTPA  
44 VHASLDKFFSTVSTVLTSKYR  
45  
46 >Ondatra\_zibethicus sp|P01944|HBA\_ONDZI Hemoglobin subunit alpha OS=Ondatra  
47    ↳ zibethicus OX=10060 GN=HBA PE=1 SV=1  
48 VLSGEDKNNIKATAWKIGGHAAEYGAEALERMFVVYPTTCKTYFPHFDSLPGSAQVKAHGK  
49 KVADALTTAVGHLDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLANHIPADFTPA  
50 VHASLDKFLASVSTVLTSKYR  
51  
52 >Tapirus\_terrestris sp|P01962|HBA\_TAPTE Hemoglobin subunit alpha-1/2  
53    ↳ OS=Tapirus terrestris OX=9801 PE=1 SV=1  
54 VLSPTDKTNVKAWSKVGSHAGEYGAEALERMFGLGPFTTCKTYFPHFDSLHGSQVQAHGK  
55 KVGDALTQAVGHLDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLALHHPPDDFTPA  
56 IHASLDKFLSNVSTVLTSKYR  
57

51 >Dugong\_dugon tr|AOA481WNF4|AOA481WNF4\_DUGDU Hemoglobin subunit alpha  
  ↳ OS=Dugong dugon OX=29137 GN=HBA PE=3 SV=1  
52 MVLSAEDKTNVKTFWGKLGHTAEYGSEALERMFNAFPATKTYFPHFDMSHGSDQIKAHG  
53 KKVAEALTRAVGHLDLPGTLDSDLHAHRLRVDPINFKLLSHCLLVTLSNHLPPDFTP  
54 PVHASLDKFLSNVSTVLTSKYR  
55  
56 >Physeter\_macrocephalus tr|AOA2Y9FTRO|AOA2Y9FTRO\_PHYMC hemoglobin subunit  
  ↳ alpha OS=Physeter macrocephalus OX=9755 GN=LOC102988256 PE=3 SV=1  
57 MVLSPADKTNIAAWAKVGNHAADFGAEALERMFMSFPSTKTYFSHFDLGHNSTQVKGHG  
58 KKVADALTКАVGHLDLPGTLDSDLHAHKLKVDPVNFKLLSHCLLVTLAHLPGDFTP  
59 PVHASLDKFASVSTVLTSKYR  
60  
61 >Delphinapterus\_leucas tr|AOA2Y9PXW7|AOA2Y9PXW7\_DELLE hemoglobin subunit  
  ↳ alpha OS=Delphinapterus leucas OX=9749 GN=LOC111185089 PE=3 SV=1  
62 MVLSPADKTNVKGTWAKIGNHSAEYGAEALERMFISFPSTKTYFSHFDLGHGSAQIKGHG  
63 KKVADALTКАVGHIDNLPDALSELSDLHAHKLKVDPVNFKLLSHCLLVTLAHLPADFTP  
64 SVHASLDKFASVSTVLTSKYR  
65  
66 >Erinaceus\_europaeus sp|P01949|HBA\_ERIEU Hemoglobin subunit alpha  
  ↳ OS=Erinaceus europaeus OX=9365 GN=HBA PE=1 SV=1  
67 VLSATDKANVKTFWGKLGHHGGEYGGAEALDRMFQAHPPTKTYFPHFDLNPAGSAQVKGHGK  
68 KVADALTTAVNNLDDVPGALSALSDLHAHKLKVDPVNFKLLSHCLLVTLAHHPADFTP  
69 VHASLDKFATVATVLTSKYR  
70  
71 >Didelphis\_virginiana sp|P01976|HBA\_DIDVI Hemoglobin subunit alpha  
  ↳ OS=Didelphis virginiana OX=9267 GN=HBA PE=1 SV=1  
72 VLSANDKTNVKGAWSKGVGGNSGAYMGEALYRTFLSFPTTKTYFPFNYDFSAGSAQIKTQGQ  
73 KIADAVGLAVAHLDDMPMTALSSSDLHAHELKVDPVNFKFLCHNVLTMAAHLGKDFTP  
74 IHASMDKFASVSTVLTSKYR  
75  
76 >Meleagris\_gallopavo sp|P81023|HBA\_MELGA Hemoglobin subunit alpha-A  
  ↳ OS=Meleagris gallopavo OX=9103 GN=HBAA PE=1 SV=2  
77 MVLSAADKNNVKGIFTKIAGHAEYGAETLERMFITYPPTKTYFPHFDFTHGSEQIKAHG  
78 KKVVAALEAANHIDDIAGTLSKLSDLHAHKLKVDPVNFKLLGQCFLVVVAIHHPAALTP  
79 EVHASLDKFCAVGTVLTAKYR  
80  
81 >Oryctolagus\_cuniculus sp|P01948|HBA\_RABIT Hemoglobin subunit alpha-1/2  
  ↳ OS=Oryctolagus cuniculus OX=9986 PE=1 SV=2  
82 MVLSPADKTNIKTAWEKIGSHGGEYGAEEVERMFLGFPTTKTYFPHFDFTHGSEQIKAHG  
83 KKVSEALTКАVGHLDLPGALSTSDLHAHKLKVDPVNFKLLSHCLLVTLANHPSEFTP  
84 AVHASLDKFLANVSTVLTSKYR  
85  
86 >Aptenodytes\_forsteri sp|P01980|HBA\_APTFO Hemoglobin subunit alpha  
  ↳ OS=Aptenodytes forsteri OX=9233 GN=HBA PE=1 SV=3

```

87 MVLSANDKSNVKSIFS KISSHAEYGAETLERMFTTYPQT KTYFPHFDLHHGSAQVKAHG
88 KKVAALIEAANHIDDIAGALS KSDLHAEKLRVDPVNFKLLGQCFMVVVAIHHPSALTP
89 EIHASLDKFLCAVGNVLTSKYR

90
91 >Cavia_porcellus sp|P01947|HBA_CAVPO Hemoglobin subunit alpha OS=Cavia
   → porcellus OX=10141 GN=HBA PE=1 SV=2
92 MVLSAADKNNVKTTWDKIGGHAAEYVAEGLTRMFTSFPTT KTYFHHIDVSPGSGDIKAHG
93 KKVA DALTTAVGHLDLPTALSTLSDVHAHKL RVDPVNFKFLNHCLLVT LA AHLGADFTP
94 SIHASLDKFFASVSTVLTSKYR

95
96 >Columba_livia sp|P21871|HBA_COLLI Hemoglobin subunit alpha-A OS=Columba
   → livia OX=8932 GN=HBAA PE=1 SV=2
97 MVLSANDKSNVKA VFAKIGGQAGDLGGEALERLFIT YPQT KTYFPHFDLSHGSAQIKGHG
98 KKVAE ALVEAANHIDDIAGALS KSDLHAQKL RVDPVNFKLLGHCFLVVAVHFPSLLTP
99 EVHASLDKFVLA VGTVL TAKYR

```

### 3.3 Performing an Align

In the UnitProt homepage, select *Align* on the tools bar (see Figure 2). Here all the species entered will be analyzed to determine which one are related and to view their characteristics alongside each other [4]. Figure 4 shows the Align interface.

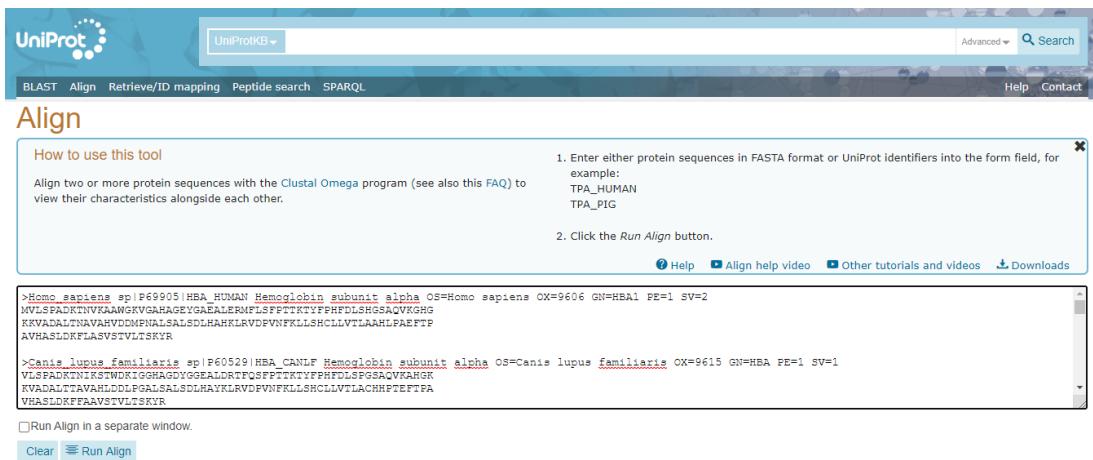


Figure 4: Align tool on UnitProt [4].

Copy and paste the animal's hemoglobin sequences from the sets that were collected on the [Building the sets of hemoglobin sequences](#) subsection on the text area, then run the process clicking the *Run Align* button and wait until the process ends.

### 3.3.1 Align for animal set 1

Hypothesis for the [Animal set 1](#) is that the human and the chimpanzee will be strongly related; also all the bears will be despite of being from different ecosystems and regions. Similar case will happen for the raccoon and the badger (lion and cat too), due the simply fact that at first glance they look similar. Maybe the crocodile and the cobra will have something in common for being reptiles, and the otter and sea cow may have associations because they are aquatic mammals.

Job status: COMPLETED

|   |    |  |                      |     |
|---|----|--|----------------------|-----|
| Homo_sapiens  | 1  | MVLSPADKTNVAAWGVKGVAHAGEYGAELERMFLSFTETKT    | YFPFDLSHGSQVKGHH     | 59  |
| Canis_lupus_fam..   | 1  | -VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFTTKT    | YFPFDLSPGSQVKKAHG    | 59  |
| Psittacula_kram..   | 1  | -VLSGTDKTNVKSIFSKIGQQADDYGAELERMFTVTFQTKT    | YFPFDLSPGSQVKKAHG    | 59  |
| Panthera_leo  | 1  | MVLSADKNNVACWGKIGSHAGEYGAELERTFCSFETKT       | YFPFDLSHGSQVQAHG     | 60  |
| Chelonoidis_niger   | 1  | MVLTAGDKANVTKTVWSKGVSHEEYGSSETLERLFIVV       | YFPFDLHHGSAQVRAHG    | 60  |
| Lutra_lutra_  | 1  | -VLSPADKTNVMSWDKIGGHAGEYGGAEALERTFVSEFTKT    | YFPFDLSHGSQVKKAHG    | 59  |
| Crocodylus_nilo..   | 1  | -VLSDDDKCNVCAWWSKGVSQHAGLEEYGAELERMFCAYE     | YFPFDLSHGSQVIRAHG    | 59  |
| Naja_naja   | 1  | -VLTDEDKARVFAWVWPVGNAELYGSETTRMFAAHFTKT      | YFPFDLSPGSNNLRAHG    | 59  |
| Pteropus_alecto   | 1  | -VLSSTDKNVAAWDKVGHVGEYGAELERMFLSFFETKT       | YFPFDLALHGGSQVKAHG   | 59  |
| Struthio_camelus  | 1  | -VLSGTDKTNVGIFSKISSHAEEYGAETLERMFITYPE       | YFPFDLHHGSAQIKAHG    | 59  |
| Hydrodamalis_gi..   | 1  | MVLSAEDKTNVTFWKGCLAHAEYGGAEALERMFLSFFETKT    | YFPFDLDMKHSDQIKAHG   | 60  |
| Felis_catus   | 1  | -VLSAADKSNVAAWGVKGISHAGEYGGAEALERTFCSFETKT   | YFPFDLSHGSQVKAHG     | 59  |
| Ursus_maritimus   | 1  | MVLSPADKSNVATWDKICSHAGEYGGAEALERTFASFFETKT   | YFPFDLSPGSQVKAHG     | 60  |
| Ursus_thibetanus  | 1  | MVLSPADKSNVATWDKIGSHAGEYGGAEALERTFASTFETKT   | YFPFDLSPGSQVKAHG     | 60  |
| Helarctos_malay..   | 1  | MVLSPADKSNVATWDKIGSHAGEYGGAEALERTFASFFETKT   | YFPFDLSPGSQVKAHG     | 60  |
| Pan_troglodytes   | 1  | MVLSPADKTNVAAWGVKGASHAGEYGGAEALERTFASFFETKT  | YFPFDLSPGSQVKAHG     | 60  |
| Ursus_arctos_ho..   | 1  | MVLSPADKSNVATWDKIGSHAGEYGGAEALERTFASFFETKT   | YFPFDLSPGSQVKAHG     | 60  |
| Procyon_lotor   | 1  | -VLSPADKANIKATWDKIGGHAGEYGGAEALERTFASFFETKT  | YFPFDLSPGSQVKAHG     | 59  |
| Meles_meles   | 1  | -VLSPADKANIKATWDKIGGHAGEYGGAEALERTFASFFETKT  | YFPFDLSPGSQVKAHG     | 59  |
| Ailuropoda_mela..   | 1  | MVLSPADKTNVATWDKIGGHAGEYGGAEALERTFASFFETKT   | YFPFDLSPGSQVKAHG     | 60  |
| *** : * * . . : : . . : **.*: * * . * ** *****: . * : : *** |    |  |                      |     |
| Homo_sapiens  | 61 | KKVADALTNAVAHVDMPNAAISALSDLHAKLRLDPVNFKLL    | SHCLLVTIAAHLPAEFTP   | 120 |
| Canis_lupus_fam..   | 60 | KKVADALTAAVHLDLPGAAISALSDLHAYKLRLDPVNFKLL    | SHCLLVTIACHHPTEFTP   | 119 |
| Psittacula_kram..   | 60 | KKVAGGLSEPAHNDIATSIKSLDLHAKQLRLDPVNFKLL      | GQCFLVVVVAIHNPNSALTP | 119 |
| Panthera_leo  | 61 | QKVADALTAKVWHINDLPNAISLSDLHAYKLRLDPVNFKLL    | SHCLLVTIACHPPEEFTP   | 120 |
| Chelonoidis_niger   | 61 | RKVLSAALGEAVNHIDDIPGAAISKSLDLHQTLRVDPVNFKLL  | NICFLVVVVGRLHPTILTP  | 120 |
| Lutra_lutra_  | 60 | KKVADALTNAVAHMDDIPGAAISALSDLHAYKLRLDPVNFKLL  | SHCLLVTIACHPAEFTP    | 119 |
| Crocodylus_nilo..   | 60 | KKVFAALHEAVNHIDDIPGAAICRLSELHAAHSLRVDPVNFKLL | AQCVIIVVVVAIHPGSLTP  | 119 |
| Naja_naja   | 60 | KKVIDATEAVVNLLDDVAGTISKSLDLHQAQKLRVDPVNFKLL  | LAHCLLVTIAAHNGGVLKP  | 119 |
| Pteropus_alecto   | 60 | KKVGDALETAVGHDDLPGAAISALSDLHAYKLRLDPVNFKLL   | SHCLLVTIACHLPSDFTP   | 119 |
| Struthio_camelus  | 60 | KKVANALIEAVNHIDDISGAISKSLDLHQAQKLRVDPVNFKLL  | GQCFLVVVVAIHNPNSALTP | 119 |
| Hydrodamalis_gi..   | 61 | KKVADALTTRAVGHLDDLPGTISELSDLHAKLRLDPVNFKLL   | SHCLLVTLSGLHPEDFTP   | 120 |
| Felis_catus   | 60 | QKVADALTQAVAHMDMIPTAMALSALSDLHAYKLRLDPVNFKLL | SHCLLVTIACHHPAEFTP   | 119 |
| Ursus_maritimus   | 61 | KKVADALTAAAGHLDDLPGAAISALSDLHAKLRLDPVNFKLL   | SHCLLVTIACHHPAEFTP   | 120 |
| Ursus_thibetanus  | 61 | KKVADALTAAAGHLDDLPGAAISALSDLHAKLRLDPVNFKLL   | SHCLLVTIACHHPAEFTP   | 120 |
| Helarctos_malay..   | 61 | KKVADALTAAAGHLDDLPGAAISALSDLHAKLRLDPVNFKLL   | SHCLLVTIACHHPAEFTP   | 120 |
| Pan_troglodytes   | 61 | KKVADALTNAVAHVDMPNAAISALSDLHAKLRLDPVNFKLL    | SHCLLVTIACHHPAEFTP   | 120 |
| Ursus_arctos_ho..   | 61 | KKVADALTAAAGHLDDLPGAAISALSDLHAYKLRLDPVNFKLL  | SHCLLVTIACHHPAEFTP   | 120 |
| Procyon_lotor   | 60 | KKVADALTAAVGHLDDLPGAAISALSDLHAYKLRLDPVNFKLL  | SHCLLVTIACHHPAEFTP   | 119 |
| Meles_meles   | 60 | KKVADALTNAVAHLDLPGAAISALSDLHAYKLRLDPVNFKLL   | SHCLLVTIACHHPAEFTP   | 119 |
| Ailuropoda_mela..   | 61 | KKVADALTAAVGHLDDLPGAAISALSDLHAYKLRLDPVNFKLL  | SHCLLVTIACHHPAEFTP   | 120 |
| *** : * * . . : : . . : **.*: * * . * ** *****: . * : : *** |    |  |                      |     |

Figure 5: Align results for [Animal set 1](#). It shows the first 120 amino acids of the sequences compared that are align to indicate which ones are located on the same position (matching residues are remarked in gray).

## Tree

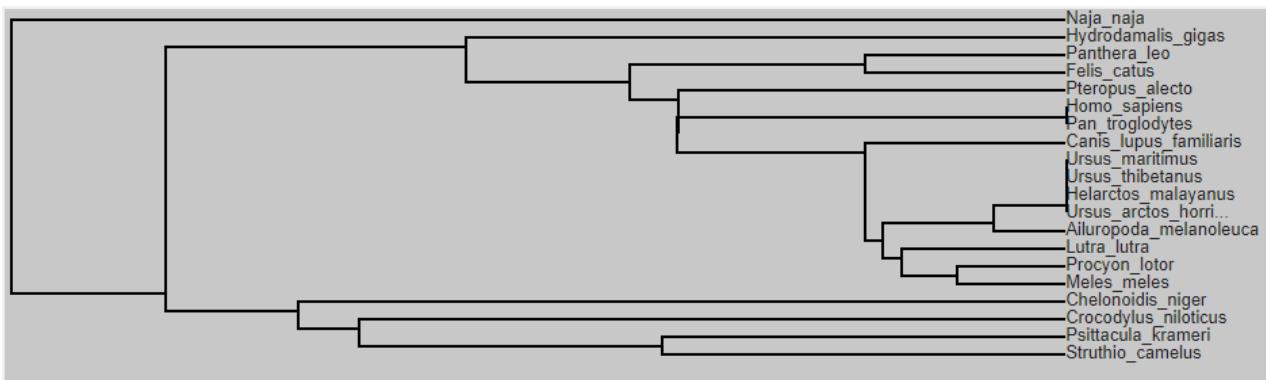


Figure 6: Phylogenetic Tree for the [Animal set 1](#), resulted by the align process.

When the align analyses is done, there are two sections of interest. The first one shown in Figure 5 indicate all the sequences compared by sections, so the matched or mismatched amino acids can be easily seen. The second one shown in Figure 6 illustrates a phylogenetic tree that tells the distance of similarity between the species analyzed; while this is higher, more different is one species from another, and on the contrary if this is closer, a species is more identical to each other.

Figure 6 confirm partially the hypothesis established for the [Animal set 1](#), specifically the relation between the human and the chimpanzee, and the bears one with the exception of the Panda Bear (see Figure 7). The cat and the lion seems to be correlated too but not too much, even though they are felines. The same happens for the aquatic mammals and the rodents.



Figure 7: More related species from the [Animal set 1](#) Align analysis.

### 3.3.2 Align for animal set 2

Hypothesis for the [Animal set 2](#) is the following: one particular species in this set is the platypus, that has physical characteristics that resembles to a duck and a beaver mixed. It's expected that this animal has strong relationship with others from the set, such as the real duck, the beaver, and even the rodents like the muskrat. The swan may be related too from the duck side. Besides that, here are more aquatic mammals than in the [Animal set 1](#): the dolphin, the

seal, the walrus, the whale, etc. must be related too. Finally, little mammals like the tapir and guinea pig for example looks similar, so they may have more things in common.

Job status: COMPLETED

Figure 8: Align results for [Animal set 2](#).

## Tree

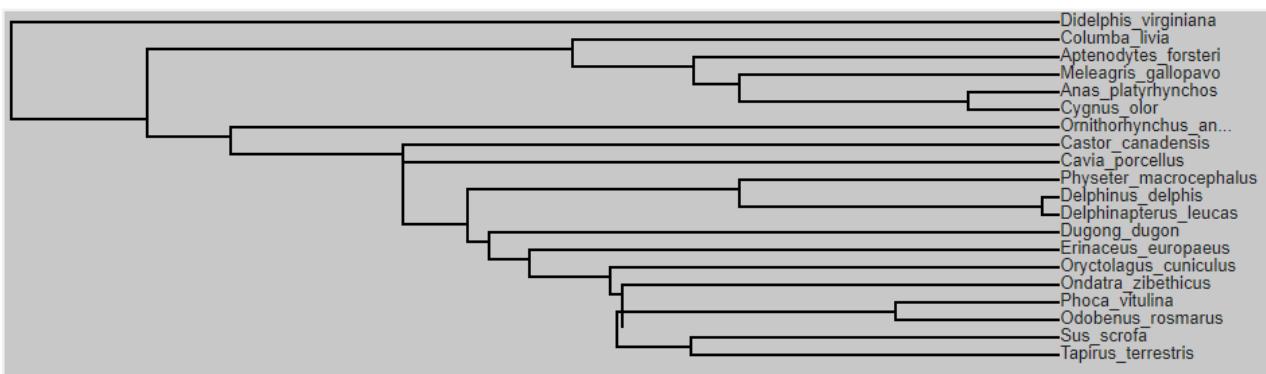


Figure 9: Phylogenetic Tree for the [Animal set 2](#), resulted by the align process.

Figure 8 corresponds to the comparison between the hemoglobin sequences of this set of animals, but the important one for the analysis is Figure 9, that shows the phylogenetic tree of this species, where it can be seen that from all the cases established on the hypothesis, only the one that implies the aquatic mammals is accomplished, but not totally because just the cases of the dolphin and the beluga, and the seal and walrus seems to be moreover related. The duck and the swan case also is satisfied, but not too strong as expected (see Figure 10).

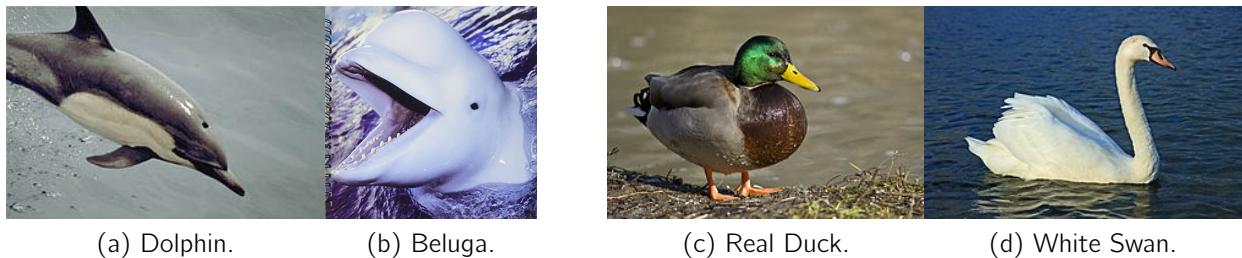


Figure 10: More related species from the [Animal set 2 Align](#) analysis.

### 3.4 Performing a BLAST

In the UnitProt homepage, select *BLAST* on the tools bar (see Figure 2). 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 [5]. Figure 11 shows the BLAST interface [5].

The screenshot shows the UniProt homepage with the BLAST tool selected. A search query for "Ornithorhynchus anatinus sp|P01979|HBA ORNAN Hemoglobin subunit alpha OS=Ornithorhynchus anatinus OX=9258 GN=HBA PE=1 SV=1" has been entered. The results panel displays the sequence: >Ornithorhynchus anatinus sp|P01979|HBA ORNAN Hemoglobin subunit alpha OS=Ornithorhynchus anatinus OX=9258 GN=HBA PE=1 SV=1 MLDIAKKEVTAIWGKAAAGHGEELYGAALERLRLQAFPTTKYSHFDLSHGSAQIKAHGK KVADALSTAAAGHFDIMOSALSALSLLAHKLRLRVDFVNFFLLAHCLILVVLARHCOPGEFTPS AHAAMDKFILSKVATVLTSKY|. Below the sequence, search parameters are shown: Target database: UniProtKB reference proteomes plus Swiss-Prot, E-Threshold: 10, Matrix: Auto, Filtering: None, Gapped: yes, Hits: 250. Buttons for "Run BLAST" and "Run BLAST in a separate window" are visible.

Figure 11: BLAST tool on UnitProt [5].

Here only the sequence from just one species has to be entered to search for others that could be related to it. Two analyses will be performed, being the selected species the **hedgehog** and the **human**.

### 3.5 BLAST for the Human

Enter the human hemoglobin sequence (collected at the [Animal set 1](#)) on the text area from the BLAST interface (see Figure 11) and run the process by clicking the *Run BLAST* button and wait until the process ends.

Figure 12 shows the result of the BLAST analysis, where the 3 principal species that could be related to the human (see Figure 13) based on their matching hemoglobin sequences are:

- Chimpanzee - *Pan troglodytes* **100.0%**
- Western lowland gorilla - *Gorilla gorilla gorilla* **99.3%**
- Sumatran orangutan - *Pongo abelii* **98.6%**

#### Overview

| <a href="#">Show all 250</a> |  |
|------------------------------|--|
| P69907                       | Hemoglobin subunit alpha ( <i>Pan troglodytes</i> ) 100.0%         |
| P69906                       | Hemoglobin subunit alpha ( <i>Pan paniscus</i> ) 100.0%            |
| P69905                       | Hemoglobin subunit alpha ( <i>Homo sapiens</i> ) 100.0%            |
| A0A341AFD2                   | hemoglobin subunit alpha ( <i>Neophocaena asiaeorientalis</i> ...) |
| A0A2J8R2J5                   | Uncharacterized protein ( <i>Pongo abelii</i> ) 98.6%              |
| P01923                       | Hemoglobin subunit alpha ( <i>Gorilla gorilla gorilla</i> ) 99.3%  |

Figure 12: BLAST results for the Human.



(a) Human. (b) Chimpanzee. (c) Western lowland gorilla. (d) Sumatran orangutan.

Figure 13: Human's more related species from the BLAST analysis.

### 3.6 BLAST for the Hedgehog

Enter the hedgehog hemoglobin sequence (collected at the [Animal set 2](#)) on the text area from the BLAST interface and run the process by clicking the *Run BLAST* button and wait until the process ends.

Figure 14 shows the result of the BLAST analysis, where the 3 principal species that could be related to the hedgehog (see Figure 15) based on their matching hemoglobin sequences are:

- Meadow Vole - *Microtus pennsylvanicus* **85.8%**
- California Mouse - *Peromyscus californicus* **85.1%**
- Weddell Seal - *Leptonychotes weddellii* **83.7%**

### Overview



Figure 14: BLAST results for the Hedgehog.



(a) Hedgehog. (b) Meadow Vole. (c) California Mouse. (d) Weddell Seal.

Figure 15: Hedgehog's more related species from the BLAST analysis.

## 4 Conclusions and recommendations

The results that provide both Align and BLAST analyses of species proteins are strong evidence of the evolution process. Sometimes it seems that an animal species could be related to another one based on their physical characteristics, and vice versa, but these kind of analyses might refute that premises, and that is perfectly logic because it intensifies the idea of a common ancestor, and give us an idea from what could have been the variations or mutations between species through time that made them start to distance.

## 5 References

- [1] T. U. Consortium, "About UniProt," <https://www.uniprot.org/help/about>, [Online; last access November 21, 2020].
- [2] ——, "UniProt: a worldwide hub of protein knowledge," <https://www.uniprot.org/>, [Online; last access November 21, 2020].
- [3] Google, "Google," <https://www.google.com.mx/>, [Online; last access November 21, 2020].
- [4] T. U. Consortium, "Align," <https://www.uniprot.org/align/>, [Online; last access November 21, 2020].
- [5] ——, "BLAST," <https://www.uniprot.org/blast/>, [Online; last access November 21, 2020].