Kendra Vilfort

6/2/2025

BioinformHer Mini Project – Module 2

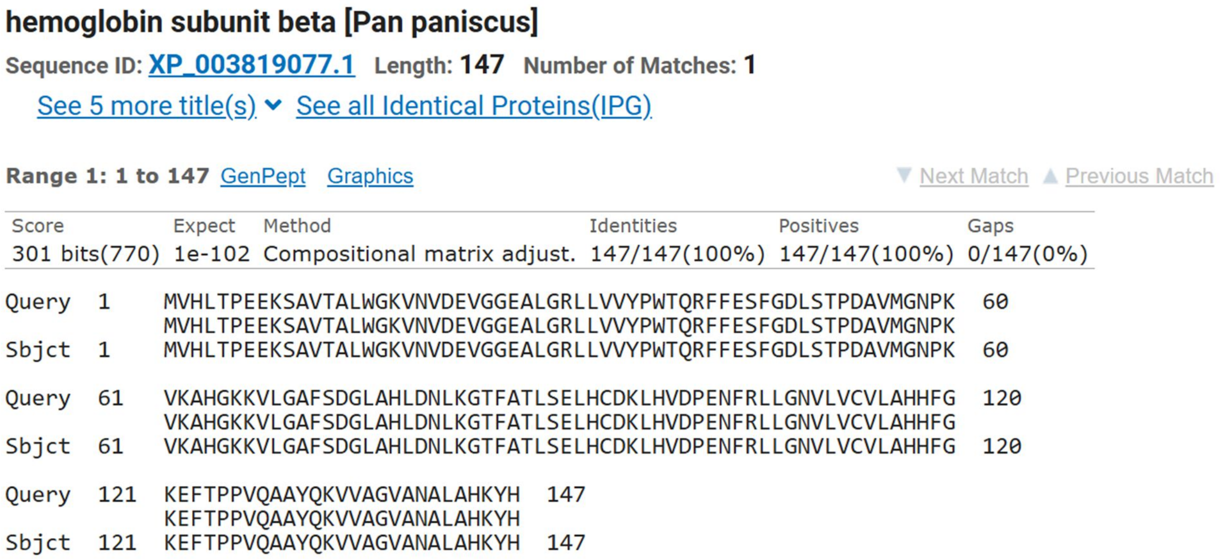
1: Sequence Retrieval & BLAST Search

To retrieve the human HBB gene from NCBI I searched for "human HBB" in the gene database, then selected the result with the description "hemoglobin subunit beta [Homo sapiens (human)]". I then went to the NCBI Reference Sequences section and selected the protein reference sequence (NP_000509.1). From the GeneBank page I selected the Run BLAST option on the sidebar, increased the max target sequence parameter to 250, excluded organism listed as "homo sapiens" and blasted the reference id.

Species name	Accession number	% identity with human HBB
Pan paniscus	XP_003819077.1	100
Gorilla gorilla gorilla	XP_018891709.1	99.32
Pongo abelii	XP_002822173.1	98.64
Nomascus leucogenys	XP_004090697.3	97.96
Vulpes vulpes	XP_025844209.1	91.16

2: Pairwise Sequence Alignment

While still in the Blast query I selected the Pan paniscus (Closely Related Species) and Vulpes vulpes (Distantly Related Species) sequences, then selected the alignments tab. I left the alignment view option as pairwise to view the alignments of each sequence with the Human HBB sequence. Pan paniscus (Pan paniscus) is more conserved with a 100% match between the query sequence and the subject organism of interest. Vulpes vulpes (red fox) has a lower match percentage, indicating it is not as conserved. This makes sense as it is expected that the chimpanzee would be more closely related to humans than the red fox. Below are my results:



hemoglobin subunit beta [Pan paniscus]

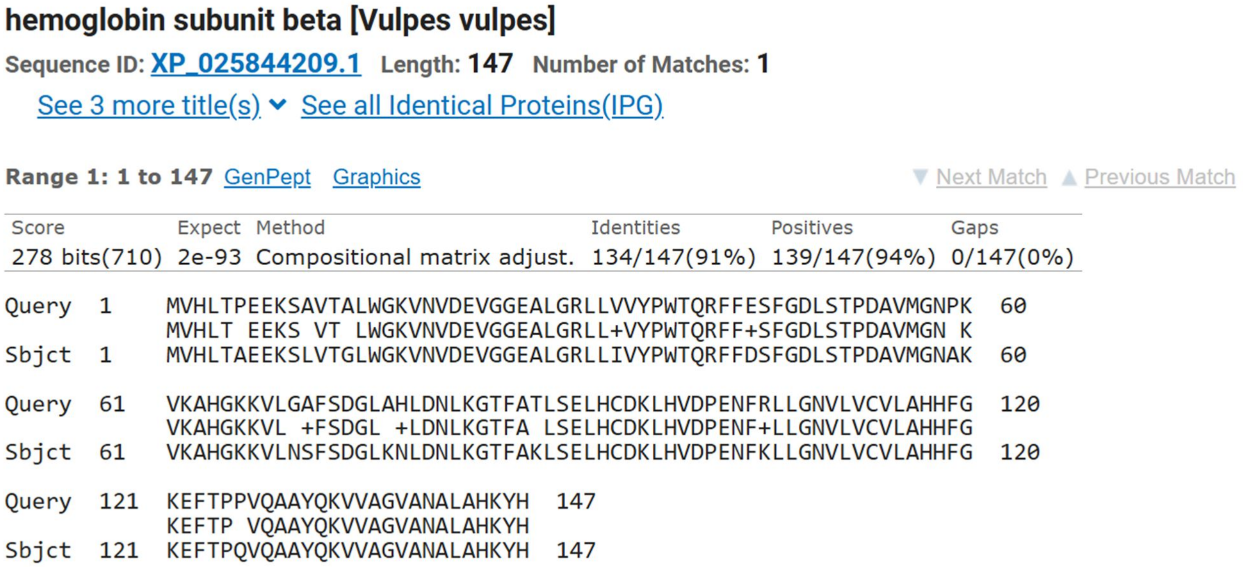
Sequence ID: XP_003819077.1 Length: 147 Number of Matches: 1

See 5 more title(s) ⌄ See all Identical Proteins(IPG)

Range 1: 1 to 147 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
301 bits(770)	1e-102	Compositional matrix adjust.	147/147(100%)	147/147(100%)	0/147(0%)

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
            MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Sbjct  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60

Query  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120
            VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Sbjct  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120

Query  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
            KEFTPPVQAAYQKVVAGVANALAHKYH
Sbjct  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
```



hemoglobin subunit beta [Vulpes vulpes]

Sequence ID: XP_025844209.1 Length: 147 Number of Matches: 1

See 3 more title(s) ⌄ See all Identical Proteins(IPG)

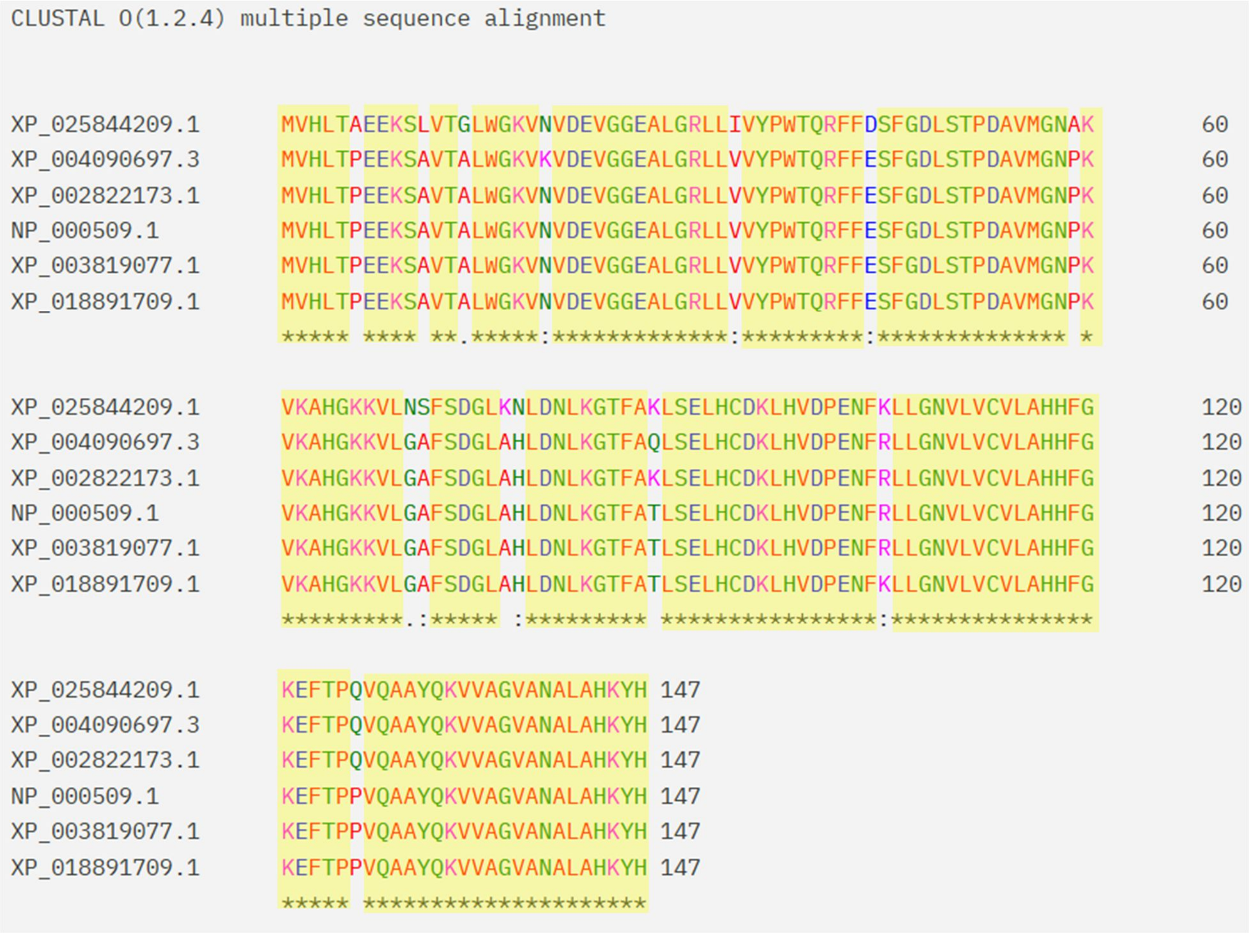
Range 1: 1 to 147 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
278 bits(710)	2e-93	Compositional matrix adjust.	134/147(91%)	139/147(94%)	0/147(0%)

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
            MVHLT EEKS VT LWGKVNVDEVGGEALGRLL+VYPWTQRFF+SFGDLSTPDAVMGN K
Sbjct  1    MVHLTAEEKSLVTGLWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLSTPDAVMGNAK  60

Query  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120
            VKAHGKKVL +FSDGL +LDNLKGTFA LSELHCDKLHVDPENF+LLGNVLVCVLAHHFG
Sbjct  61   VKAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVDPENFKLLGNVLVCVLAHHFG  120

Query  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
            KEFTP VQAAYQKVVAGVANALAHKYH
Sbjct  121  KEFTPQVQAAYQKVVAGVANALAHKYH  147
```

3: Multiple Sequence Alignment (MSA)

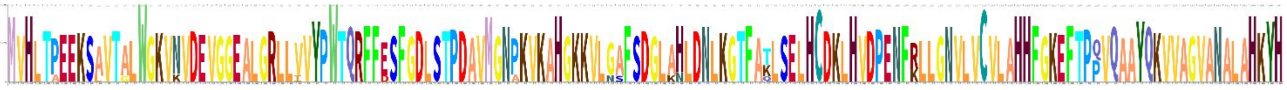
I copied FASTA sequences that were exported in task 1, along with the FASTA sequence for the human HBB protein sequence and pasted them into the Clustal Omega Input sequence box, then submitted the job. Below are the results, with the highly conserved regions highlighted in yellow.:



```
CLUSTAL O(1.2.4) multiple sequence alignment


XP_025844209.1      MVHLTAEEKSLVTGLWGKVNVDEVGGEALGRLLIVYPWTQRFFDSFGDLSTPDAVMGNAK	60
XP_004090697.3      MVHLTPEEKSAVTALWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK	60
XP_002822173.1      MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK	60
NP_000509.1         MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK	60
XP_003819077.1      MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK	60
XP_018891709.1      MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK	60
                    ***** **** **.*****:*************:*********:************** *

XP_025844209.1      VKAHGKKVLNSFSDGLKNLDNLKGTFAKLSELHCDKLHVDPENFKLLGNVLVCVLAHHFG	120
XP_004090697.3      VKAHGKKVLGAFSDGLAHLDNLKGTFAQLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG	120
XP_002822173.1      VKAHGKKVLGAFSDGLAHLDNLKGTFAKLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG	120
NP_000509.1         VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG	120
XP_003819077.1      VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG	120
XP_018891709.1      VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFKLLGNVLVCVLAHHFG	120
                    *********.:***** :********* ****************:***************

XP_025844209.1      KEFTPQVQAAYQKVVAGVANALAHKYH 147
XP_004090697.3      KEFTPQVQAAYQKVVAGVANALAHKYH 147
XP_002822173.1      KEFTPQVQAAYQKVVAGVANALAHKYH 147
NP_000509.1         KEFTPPVQAAYQKVVAGVANALAHKYH 147
XP_003819077.1      KEFTPPVQAAYQKVVAGVANALAHKYH 147
XP_018891709.1      KEFTPPVQAAYQKVVAGVANALAHKYH 147
                    ***** *********************
```

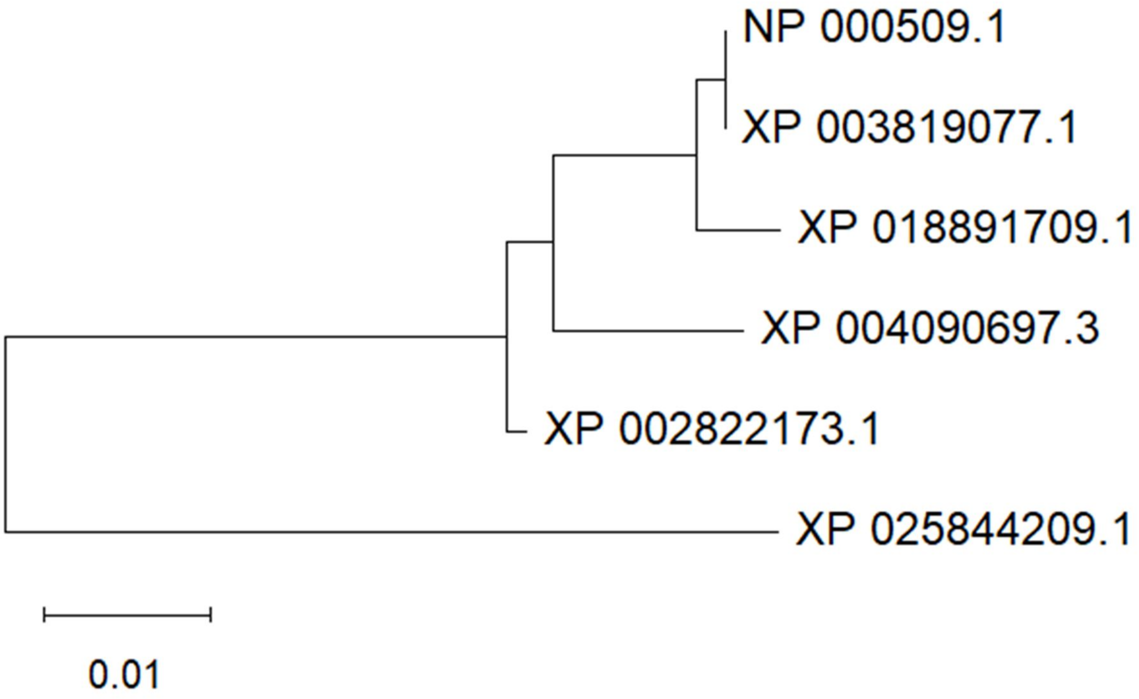
4: Sequence Logo Generation

After exporting the alignment in FASTA format converted by Seqret from Clustal Omega, I uploaded the alignment file to Skylign. The generated logo shows many highly conserved residues in the form of a single amino acid letter in the alignment position, with taller letters representing greater conservation. These region match what I have highlighted above. High conservation can indicate that the region may be biologically significant across the aligned organisms.



5: Phylogenetic Tree Construction

The phylogenetic tree shows the Pan paniscus (chimpanzee; XP_003819077.1) as the most closely related organism to the homo sapien (NP_000509.1) based on HBB. I expected all of the organisms categorized as primates to be the more closely related than Vulpes vulpes (Red fox; XP_025844209.1), and the tree matches this expectation.


NP 000509.1
XP 003819077.1
XP 018891709.1
XP 004090697.3
XP 002822173.1
XP 025844209.1
0.01