

# **TITLE: TRACKING THE EVOLUTION OF THE HEMOGLOBIN BETA (HBB) GENE ACROSS SPECIES EIGHT SELECTED VERTEBRATE SPECIES**

## **Introduction**

Hemoglobin (Hb) is widely known as the iron-containing protein in blood that is essential for O<sub>2</sub> transport in mammals (Gell, 2018). The Hemoglobin Beta (HBB) gene encodes the beta chain of hemoglobin, a protein critical for oxygen transport in vertebrates. Due to its essential physiological function, the HBB gene is highly conserved across species, making it a valuable molecular marker for comparative genomics.

This project aims to investigate the functional conservation, divergence patterns, and evolutionary relationship of the HBB gene in eight species; *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Bos taurus* (cattle), *Mus musculus* (mouse), *Gallus gallus* (chicken), *Xenopus laevis* (Frog), *Rattus norvegicus* (Rat), and *Danio rerio* (zebrafish). We will achieve this by retrieving and analyzing protein sequences through a range of methods, including sequence retrieval and BLAST, pairwise and multiple sequence alignment, sequence logo generation, and phylogenetic tree construction.

This project integrates skills acquired from Module 2 of the BioinformHer training program, which aims to investigate molecular evolution and demonstrate practical bioinformatics competency. The findings could potentially contribute to our understanding of molecular evolution and have valuable applications in various fields.

## **Methods**

### **1. Sequence Retrieval and BLAST**

The human Hemoglobin Beta (HBB) protein sequence was retrieved by searching the NCBI Protein database, which is shown in Figure 1 and Figure 2.

**Table 1: Percentage Identity of HBB Protein Sequences Between Human and Selected Vertebrate Species**

S/N	Specie name	Accession number	% identity with HBB
1	<i>Homo Sapiens</i> (Human)	NP_000509.1	100%

2	<i>Pan troglodytes</i> (Chimpanzee)	<a href="#">XP_508242.1</a>	100%
3	<i>Bos taurus</i> (Cattle)	<a href="#">NP_776342.1</a>	84.72%
4	<i>Mus musculus</i> (House mouse)	<a href="#">NP_001265090.1</a>	80.27%
5	<i>Gallus gallus</i> (chicken)	<a href="#">NP_990820.1</a>	69.39%
6	<i>Danio rerio</i> (Zebrafish)	<a href="#">Q90485.3</a>	50.00
7	<i>Xenopus laevis</i> (Frog)	<a href="#">NP_001091375.1</a>	57.82%
8	<i>Rattus norvegicus</i> (Rat)	<a href="#">NP_942071.2</a>	73.47%

**National Library of Medicine**  
National Center for Biotechnology Information

Protein:

Species: Animals (7,077), Plants (12), Fungi (7), Protists (16), Bacteria (73), Archaea (3), Customize...

Source databases: PDB (941), RefSeq (2,026), UniProtKB / Swiss-Prot (878), Customize...

Sequence length: Custom range...

Molecular weight: Custom range...

Release date: Custom range...

Revision date: Custom range...

[Clear all](#) [Show additional filters](#)

Summary: 20 per page | Sort by Default order | Send to: [Manage Filters](#)

See [HB2 \(HEMOGLOBIN\) hemoglobin 2](#) in the Gene database  
hemoglobin reference sequences [Transcript \(1\)](#) [Protein \(1\)](#)

See the results of this search (1179 items) in our new [Identical Protein Groups](#) database.

Items: 1 to 20 of 7209

Selected: 1

☐ 1. [RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName: Full=LVV-hemorphin-7; Contains: RecName: Full=Spinorphin](#)  
147 aa protein  
Accession: P68871.2 GI: 56749856  
[PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

☒ 2. [hemoglobin subunit beta \[Homo sapiens\]](#)  
147 aa protein  
Accession: NP\_000509.1 GI: 4504349  
[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)  
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

☐ 3. [hemoglobin subunit beta \[Gallus gallus\]](#)  
147 aa protein

Results by taxon  
Top Organisms [Tree](#)  
Homo sapiens (1877)  
Anas georgica (694)  
Anser cygnoides (129)  
Oxyura jamaicensis (117)  
Anas flavirostris flavirostris (70)  
All other taxa (4322)  
[More...](#)

Find related data  
Database:

Search details  
Hemoglobin subunit beta[Protein Name]  
OR (Hemoglobin[All Fields] AND  
subunit[All Fields] AND beta[All  
Fields])  
 [See more...](#)

Recent activity

Figure 1: Screenshot of the NCBI Protein database page showing the list of search results.

blastn **blastp** blastx tblastn tblastx

BLAST programs search protein databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)  Query subrange

Or, upload file  No file chosen

Job Title

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (nr, etc.) ☐ Experimental databases

Non-redundant protein sequences (nr)

Organism Optional

☐ Mus musculus (taxid 10090) ☐ exclude

☐ Gallus gallus (taxid 9031) ☐ exclude

☐ Bos taurus (taxid 9913) ☐ exclude

☐ Pan troglodytes (taxid 9598) ☐ exclude

☐ Danio rerio (taxid 7955) ☐ exclude

☐ Xenopus laevis (taxid 8355) ☐ exclude

☐ Rattus norvegicus (taxid 10116) ☐ exclude

Exclude Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WPI) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☒ Quick BLASTP (Accelerated protein-protein BLAST)

☐ 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

Search database nr using blastp (protein-protein BLAST)

☒ Show results in a new window

Figure 2: Screenshot of the BLASTp page to identify the HBB sequence for seven other species

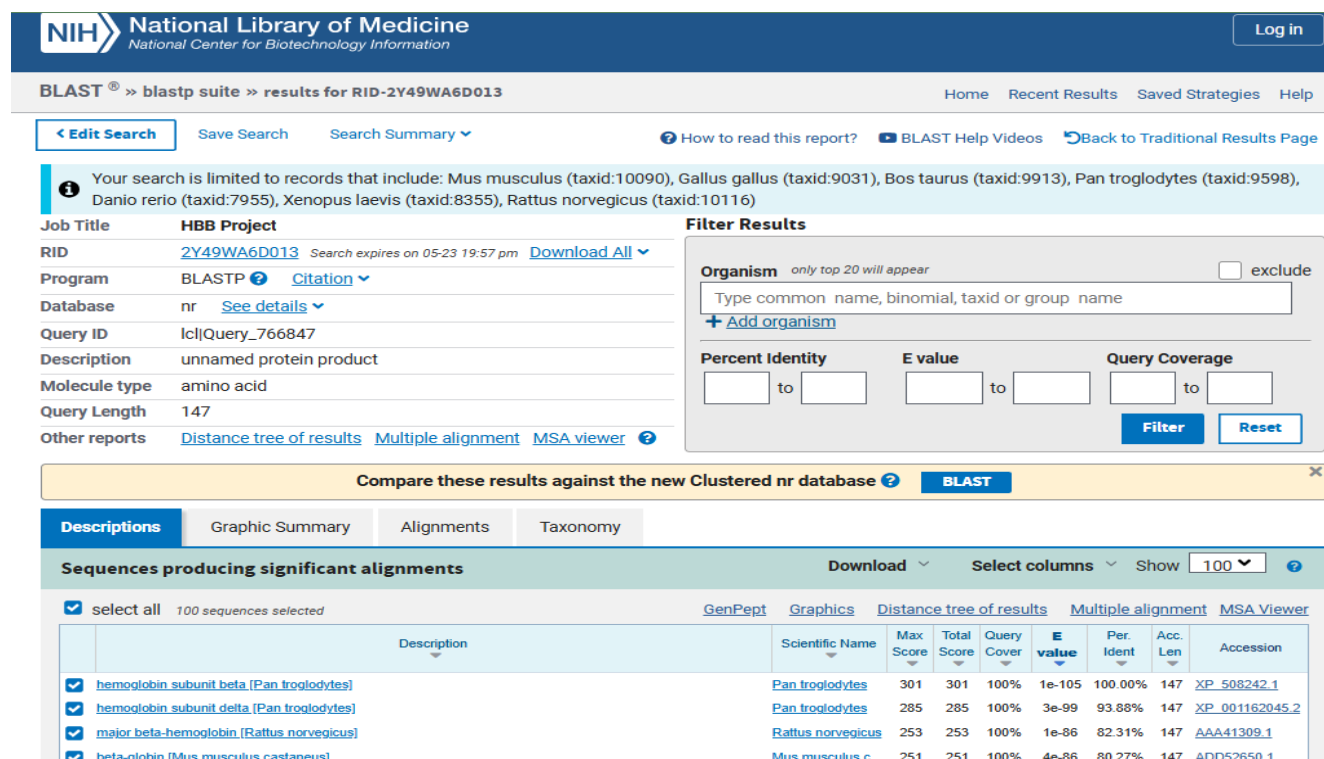


Figure 3: BLASTp results showing homologous protein sequences to the human hemoglobin beta (HBB) protein in selected vertebrate species.

<input checked="" type="checkbox"/> haemoglobin beta-h0 chain [Mus musculus]	<a href="#">Mus musculus</a>	221	221	100%	7e-74	67.35%	147	<a href="#">CAA32220.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit epsilon-2 [Bos taurus]	<a href="#">Bos taurus</a>	221	221	100%	9e-74	68.03%	147	<a href="#">NP_001014913.1</a>
<input checked="" type="checkbox"/> hemoglobin fetal subunit beta-like [Bos taurus]	<a href="#">Bos taurus</a>	220	220	98%	1e-73	78.47%	145	<a href="#">XP_024831416.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit rho [Gallus gallus]	<a href="#">Gallus gallus</a>	220	220	100%	1e-73	65.99%	147	<a href="#">NP_001004390.1</a>
<input checked="" type="checkbox"/> RecName: Full=Hemoglobin subunit beta-H0; AltName: Full=Beta-H0-globin; AltName: Full=Hemoglobin beta...	<a href="#">Mus musculus</a>	219	219	100%	3e-73	66.67%	147	<a href="#">P04443.5</a>
<input checked="" type="checkbox"/> beta globin, partial [Pan troglodytes troglodytes]	<a href="#">Pan troglodytes...</a>	218	218	71%	3e-73	100.00%	105	<a href="#">ACY75362.1</a>
<input checked="" type="checkbox"/> Chain B, PROTEIN (HEMOGLOBIN D) [Gallus gallus]	<a href="#">Gallus gallus</a>	219	219	99%	4e-73	69.18%	146	<a href="#">1HBR_B</a>
<input checked="" type="checkbox"/> hemoglobin subunit epsilon [Gallus gallus]	<a href="#">Gallus gallus</a>	219	219	100%	4e-73	65.99%	147	<a href="#">NP_001075173.3</a>
<input checked="" type="checkbox"/> hemoglobin subunit epsilon [Rattus norvegicus]	<a href="#">Rattus norvegicus</a>	218	218	100%	8e-73	67.35%	147	<a href="#">NP_001019976.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit epsilon 1 [Gallus gallus]	<a href="#">Gallus gallus</a>	217	217	100%	2e-72	67.35%	147	<a href="#">NP_001026660.1</a>
<input checked="" type="checkbox"/> minor beta-globin [Mus musculus domesticus]	<a href="#">Mus musculus d...</a>	216	216	100%	5e-72	65.99%	147	<a href="#">AAB59637.1</a>
<input checked="" type="checkbox"/> beta globin, partial [Pan troglodytes verus]	<a href="#">Pan troglodytes...</a>	214	214	71%	6e-72	99.05%	105	<a href="#">ACY75403.1</a>
<input checked="" type="checkbox"/> unnamed protein product, partial [Mus musculus]	<a href="#">Mus musculus</a>	215	215	91%	6e-72	81.34%	134	<a href="#">BAB27380.1</a>
<input checked="" type="checkbox"/> beta-H1-globin, partial [Mus musculus]	<a href="#">Mus musculus</a>	214	214	95%	4e-71	68.57%	140	<a href="#">AAA37695.1</a>
<input checked="" type="checkbox"/> beta-globin chain [Rattus norvegicus]	<a href="#">Rattus norvegicus</a>	213	213	100%	8e-71	66.67%	147	<a href="#">CAA43137.1</a>
<input checked="" type="checkbox"/> HBD isoform 3 [Pan troglodytes]	<a href="#">Pan troglodytes</a>	210	210	75%	9e-70	90.91%	141	<a href="#">PNI79721.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit beta [Bos taurus]	<a href="#">Bos taurus</a>	207	207	100%	1e-68	65.99%	147	<a href="#">XP_002707739.2</a>
<input checked="" type="checkbox"/> HBD isoform 1, partial [Pan troglodytes]	<a href="#">Pan troglodytes</a>	206	206	71%	2e-68	94.23%	104	<a href="#">PNI79719.1</a>
<input checked="" type="checkbox"/> LOW QUALITY PROTEIN, hemoglobin subunit beta-like [Bos taurus]	<a href="#">Bos taurus</a>	206	206	97%	5e-68	71.83%	143	<a href="#">XP_015330224.2</a>
<input checked="" type="checkbox"/> TPA, delta globin-like [Bos taurus]	<a href="#">Bos taurus</a>	204	204	100%	2e-67	65.31%	147	<a href="#">DAA22104.1</a>
<input checked="" type="checkbox"/> beta-globin chain, partial [Rattus norvegicus]	<a href="#">Rattus norvegicus</a>	204	204	92%	2e-67	65.93%	135	<a href="#">CAA43138.1</a>
<input checked="" type="checkbox"/> hemoglobin, epsilon 1 isoform X1 [Bos taurus]	<a href="#">Bos taurus</a>	196	196	80%	4e-64	77.78%	150	<a href="#">XP_059730899.1</a>
<input checked="" type="checkbox"/> HBG2 isoform 2 [Pan troglodytes]	<a href="#">Pan troglodytes</a>	194	194	80%	1e-63	76.92%	137	<a href="#">PNI79691.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit epsilon 1 isoform X1 [Rattus norvegicus]	<a href="#">Rattus norvegicus</a>	191	191	80%	2e-62	76.07%	124	<a href="#">XP_038962576.1</a>
<input checked="" type="checkbox"/> rCG39881, isoform CBA, a [Rattus norvegicus]	<a href="#">Rattus norvegicus</a>	186	186	71%	2e-60	82.86%	122	<a href="#">EDM18112.1</a>
<input checked="" type="checkbox"/> y2-globin, partial [Mus musculus]	<a href="#">Mus musculus</a>	185	185	85%	3e-60	73.60%	124	<a href="#">CAA24225.1</a>
<input checked="" type="checkbox"/> hypothetical protein XELAEV_18045083mg [Xenopus laevis]	<a href="#">Xenopus laevis</a>	178	178	100%	6e-57	58.50%	147	<a href="#">OCT63985.1</a>
<input checked="" type="checkbox"/> hemoglobin subunit beta-2-like [Xenopus laevis]	<a href="#">Xenopus laevis</a>	177	177	100%	2e-56	56.46%	147	<a href="#">XP_018094550.1</a>
<input checked="" type="checkbox"/> Hemoglobin subunit beta-2-like [Xenopus laevis]	<a href="#">Xenopus laevis</a>	176	176	100%	6e-56	57.82%	147	<a href="#">NP_001091375.1</a>
<input checked="" type="checkbox"/> hemoglobin larval subunit beta-1 [Xenopus laevis]	<a href="#">Xenopus laevis</a>	173	173	99%	4e-55	56.16%	147	<a href="#">NP_001079742.1</a>

Figure 4: BLASTp results of the human HBB protein against selected vertebrate species

## 2. Pairwise sequence alignment

To assess the sequence similarity of the HBB gene across species, I picked two organisms from my BLAST results: chimpanzee, which is closely related to humans, and zebrafish, which is more distantly related.

- **Chimpanzee (*Pan troglodytes*)**, a *closely related species*
- **Zebrafish (*Danio rerio*)**: a *distantly related species*

The EMBOSS Needle tool was used to perform pairwise sequence alignments between the human HBB protein sequence and that of the two selected species, which is shown in figures 5 and 6. Figure 5 reveals 100% identity and similarity, with no gaps, indicating strong evolutionary conservation between human and chimpanzee Hemoglobin Beta (HBB) protein and figure 6 shows 50% identity and 71.6% similarity, with one gap, between human and zebrafish HBB protein indicating moderate evolutionary conservation of core regions but significant sequence divergence due to evolutionary distance.

```
#
# 2: XP_508242.1
# Matrix: EBLSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:   147/147 (100.0%)
# Similarity: 147/147 (100.0%)
# Gaps:      0/147 ( 0.0%)
# Score: 780.0
#
#=====
NP_000509.1      1  MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS      50
XP_508242.1      1  MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS      50

NP_000509.1     51  TPDAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD     100
XP_508242.1     51  TPDAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD     100

NP_000509.1    101  PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
XP_508242.1    101  PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147

#-----
#-----
```

Figure 5: Pairwise Alignment of Human and Chimpanzee HBB Protein Sequences

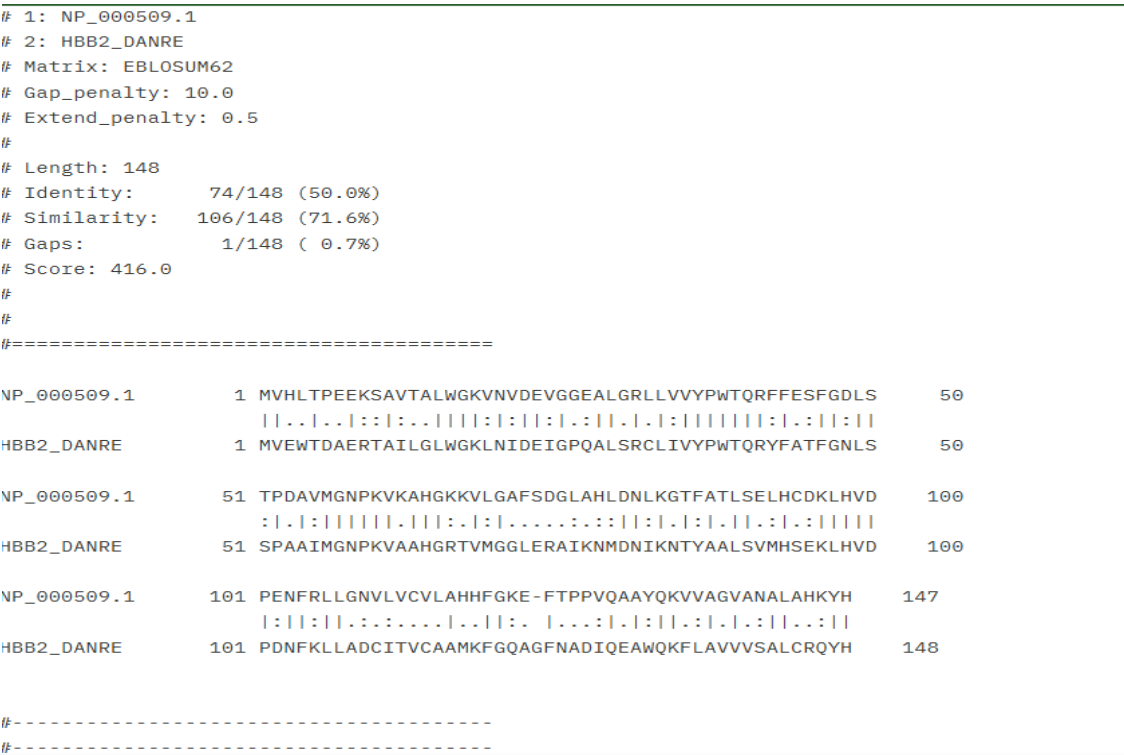


Figure 6: Pairwise Alignment of Human and Zebrafish HBB Protein Sequences

Table 2: Pairwise Alignment Statistics of HBB Protein Sequences Between Species

Comparison	% identity	% similarity	Number of gaps
Human vs Chimpanzee	100%	100%	0
Human vs Zebrafish	50%	71.6%	1

Result Interpretation

The human HBB gene is more conserved in chimpanzees than in zebrafish. The pairwise alignment between human and chimpanzee HBB proteins reveals a perfect match, with 100% identity and no gaps or substitutions. This high level of conservation reflects the close evolutionary relationship between the two species, which share a recent common ancestor. It underscores the essential and conserved role of the HBB gene in oxygen transport. In contrast, the alignment with zebrafish shows only 50% identity and approximately 72% similarity when accounting for conservative amino acid substitutions, along with a small gap. This reduced similarity highlights the greater evolutionary distance between humans and zebrafish. However, the preservation of some key regions suggests that critical functional elements of the hemoglobin protein remain conserved across vertebrates despite species-specific divergence.

### 3. Multiple sequence alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools. This tool can align up to 4000 sequences or a maximum file size of 4 MB.

Input sequence ⓘ

Sequence Type  
☒ Protein ☐ DNA ☐ RNA

Paste your sequence here - or use the example sequence  
>NP\_000509.1 hemoglobin subunit beta [Homo sapiens]  
MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTATLSELHCDKLHVDPENFRLLGNLVLCVLAHHFGKEFTPPVQAAAYQKVVAGVANALAHKYH  
>NP\_942071.2 hemoglobin, beta adult major chain [Rattus norvegicus]  
MVHLTDAEKATVNLWGKVNPEIGAESLASLLIVYPWTQRYFSKFGDLSSVSAIMGNPQVKAHGKKVINAFDDGLKHLNKLKGTGSLSELHCDKLHVDPENFRLLGNMIVIMMGHHLGKEFTPSAQAAFQKVVAGVAS

Choose File No file chosen

Use the example

Clear sequence

More example inputs

Parameters

OUTPUT FORMAT ⓘ

**Figure 7:** Multiple Sequence Alignment of Protein Sequences from Homo sapiens and Seven Other Species Using Clustal Omega

Alignment with colours		CLUSTAL O(1.2.4) multiple sequence alignment	
<div>Hide</div>			
sp Q90485.3 HBB2_DANRE		MVEWTDARTAILGLWGKLNIDEIGPQALSRLIVYPWTQRYFATFGNLSSPAAIMGNPK	60
NP_001091375.1		MVHWTAEEKAAITSVWQEVNQEQDGHDLTRLLVVYPWTQRYFSFGNLGNATAIAGNVK	60
NP_990820.1		MVHWTAEEKQLITGLWGKVNVAECGAELARLLIVYPWTQRYFSFGNLSSPTAILGNPM	60
NP_000509.1		MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRYFSFGDLSTPDAMGNPK	60
XP_508242.1		MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRYFSFGDLSTPDAMGNPK	60
NP_776342.1		--MLTAEKKAATVAFWGKVKVDEVGGEALGRLLVVYPWTQRYFSFGDLSTADAMNPNK	58
NP_942071.2		MVHLTDAEKATVNLWGKVNPEIGAESLASLLIVYPWTQRYFSKFGDLSSVSAIMGNPK	60
NP_001265090.1		MVHLTDAEKAAVSCLWGKVNDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNAK	60
		* *: : . * : : : * : : * * : * : * : * : * : * : *	
sp Q90485.3 HBB2_DANRE		VAAHGRTVMGGLERAIKNMDNIKNTYAALSMHSEKLVHVDPNFKLLADCTVCAAMKFG	120
NP_001091375.1		VRAHGKVLKSAVGDAIAHLDNVKGTLHDLVVHAFKLYVDPENFKRLGEVLVIVLASKLG	120
NP_990820.1		VRAHGKVLTSFGDAVKNLNLIKNTFSQSELHCDKLHVDPENFRLLGDILIVLAHFHS	120
NP_000509.1		VKAHGKVLGAFSGLAHLNKLKGTATLSELHCDKLHVDPENFRLLGNLVLCVLAHHFG	120
XP_508242.1		VKAHGKVLGAFSGLAHLNKLKGTATLSELHCDKLHVDPENFRLLGNLVLCVLAHHFG	120
NP_776342.1		VKAHGKVLDSFSGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNLVVVLARNFG	118
NP_942071.2		VKAHGKVINAFDDGLKHLNKLKGTGSLSELHCDKLHVDPENFRLLGNMIVIMMGHHLG	120
NP_001265090.1		VKAHGKVIITAFNDGLNHLDSLKGTFASELHCDKLHVDPENFRLLGNMIVIVLGHHLG	120
		* * : : * : . . : : : * : * : * : * : * : * : * : . : .	
sp Q90485.3 HBB2_DANRE		QAGFNADIQEAQKFLAVVVSALCRQYH	148
NP_001091375.1		S-AFTPQVQGAWEKFAVLVDALSQGYN	147
NP_990820.1		K-DFTPECQAQWKLVVVVAHALARKYH	147
NP_000509.1		K-EFTPPVQAAQKQVAVGVANALAHKYH	147
XP_508242.1		K-EFTPPVQAAQKQVAVGVANALAHKYH	147
NP_776342.1		K-EFTPVLAQDFKQVAVGVANALAHRYH	145
NP_942071.2		K-EFTPSAQAAFQKQVAVGVASALAHKYH	147
NP_001265090.1		K-DFTPAQAQAFQKQVAVGVATALAHKYH	147
		. * . * : : * : . : * : * : *	

**Figure 8:** Multiple sequence alignment (MSA) results showing all eight sequences using clustal omega

Figure 8 displays the MSA of HBB protein sequences from eight different species, including *Homo sapiens*, *Pan troglodytes*, *Bos taurus*, *Mus musculus*, *Gallus gallus*, *Danio rerio*, *Xenopus laevis*, and *Rattus norvegicus* all represented by their accession number. The alignment was performed using Clustal Omega. In this alignment, each sequence is presented along with markers indicating conserved regions:

- An asterisk (\*) denotes positions where the amino acid is identical in all sequences.
- A colon (:) indicates positions with strongly similar substitutions.
- A period (.) marks positions with weakly similar substitutions.
- A space ( ) indicates no significant conservation.

The alignment revealed several highly conserved regions, as indicated by asterisks (\*) in the consensus line. For example, the sequence motif “**VVYPWTPQR**” found between residues 35–43 is completely conserved across all species analyzed, and it had the longest stretch. This suggests that these residues are likely critical for the structural or functional integrity of the hemoglobin protein, possibly playing a role in heme binding and maintaining protein stability.

Another highly conserved stretch is observed around residues 63–65 and 98–104, which includes sequences such as “**AHG**” and “**VDP**”, emphasizing the importance of this region in the protein’s core function. In contrast, the C-terminal region (residues 120–147) displays greater sequence variability, indicating possible species-specific adaptations or reduced functional constraints in this portion of the protein.

Overall, the alignment underscores the evolutionary conservation of key functional domains within the HBB protein while also revealing areas of divergence that may reflect physiological differences across species. In contrast, regions with fewer asterisks and more gaps (or weaker similarity markers) reflect areas that have diverged during evolution, especially when comparing more distantly related species like zebrafish.

#### 4. Sequence logo generation

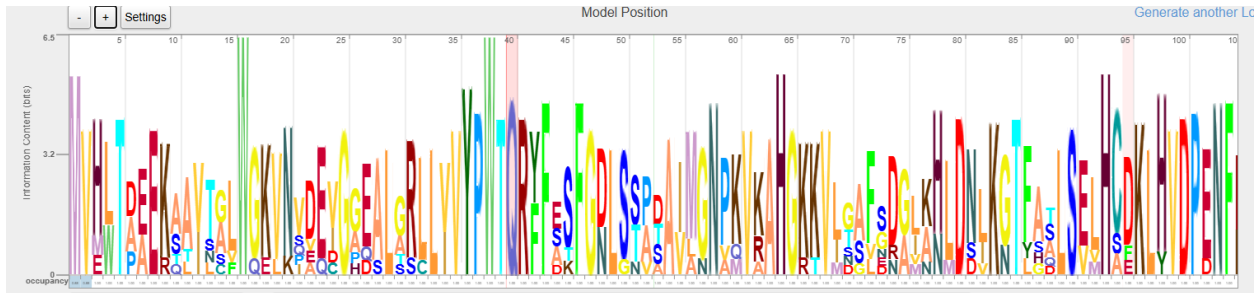


Figure 9: Half view of the Skylign-Generated Sequence Logo Displaying Conserved Regions of MSA of hemoglobin gene

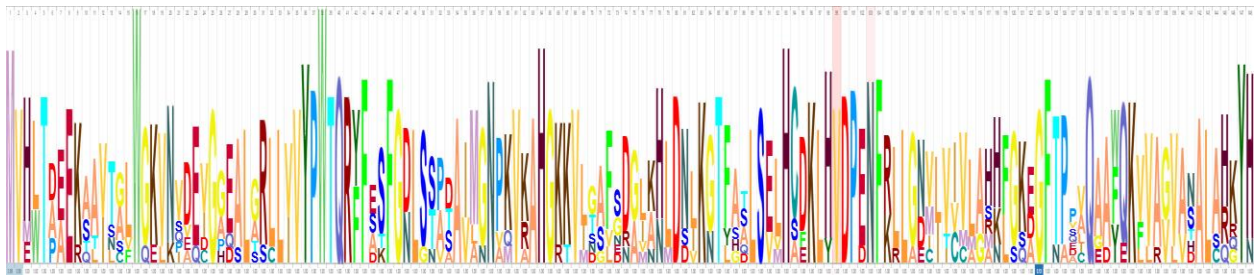


Figure 10: Full view of the Skylign-Generated Sequence Logo Displaying Conserved Regions of MSA of hemoglobin gene

**a) What do you observe**

The sequence logo displays the amino acid conservation at each position across multiple aligned sequences. Taller stacks of letters indicate higher conservation, while shorter stacks with many different letters suggest variability. Some positions show a clear preference for specific amino acids.

**b) Are there highly conserved residues?**

Yes. Highly conserved residues appear at positions such as **1 (M), 16 (W), 36 (Y), 38 (Y), 64 (H), 93 (H), 103-104 (N, F), 124 (F), 129 (Q), 147 (Y)**. These positions show dominant single-letter amino acids with high information content (measured in bits), indicating strong evolutionary conservation.

**c) Why might those regions be important?**

Conserved regions often correspond to functionally or structurally critical parts of the protein. For example:



- i. Methionine (M) at position 1 – This is often the start codon in protein translation, essential for initiating synthesis.
- ii. Tryptophan (W) and Phenylalanine (F) – These are bulky, aromatic residues that often help stabilize the protein’s core through hydrophobic interactions.
- iii. Tyrosine (Y) – An aromatic residue that can also serve as a phosphorylation site, which is important for protein regulation.
- iv. Histidine (H) – Frequently found in active or binding sites due to its ability to donate and accept protons; crucial in enzyme catalysis or metal ion binding.
- v. Asparagine (N) and Glutamine (Q) – Polar residues involved in hydrogen bonding often stabilize protein secondary structures.

## 5. Phylogenetic tree construction

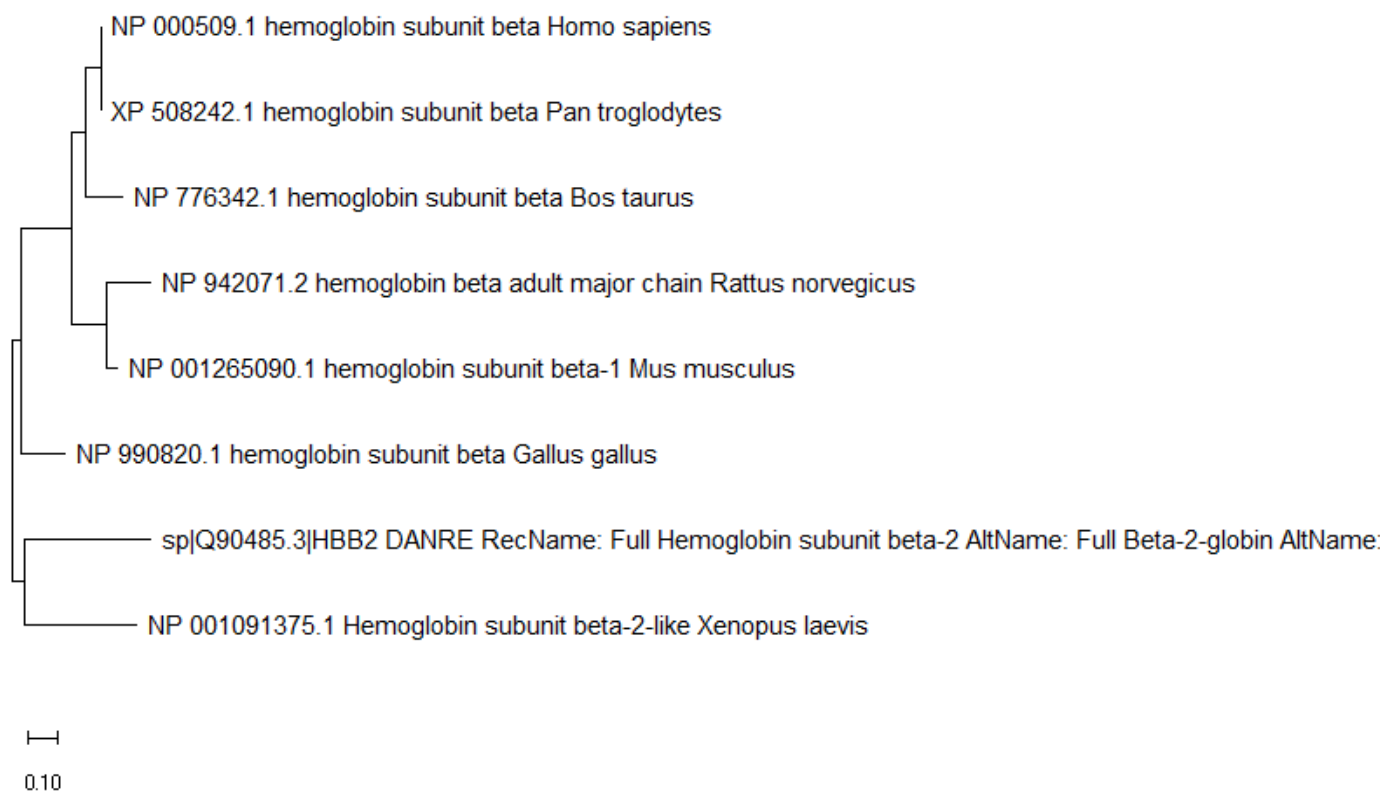


Figure 11: Phylogenetic Tree of HBB Proteins from Homo sapiens and Seven Other Vertebrates using MEGAX

**a) Which species are most closely related based on HBB?**

According to the phylogenetic tree of the hemoglobin subunit beta (HBB) sequence in Figure 11, *Homo sapiens* (human) and *Pan troglodytes* (chimpanzee) are the most closely related, as they share a common branch point and cluster, while more distantly related species (e.g., zebrafish and frog) are placed further apart. This suggests a high level of sequence similarity in their HBB genes, reflecting their recent shared ancestry.

**b) Does this tree match what you expect evolutionarily?**

Yes, the tree matches evolutionary expectations. Evolutionarily close species (like humans and chimpanzees or rats and mice) are grouped. Mammals cluster separately from non-mammals like *Gallus gallus* (chicken), *Xenopus laevis* (frog), and *Danio rerio* (zebrafish), with zebrafish being the most distantly related.

**Conclusion**

The comparative analysis of the HBB gene across selected vertebrates reveals a high degree of conservation, underscoring its critical role in hemoglobin function. Overall, the study demonstrates how bioinformatics tools can reveal evolutionary relationships and functional conservation across species.

## References

Gell D. A. (2018). Structure and function of haemoglobins. *Blood cells, molecules & diseases*, 70, 13–42. <https://doi.org/10.1016/j.bcmd.2017.10.0>