

Tracking the Evolution of the Hemoglobin Beta (HBB) Gene Across Species

This project highlights the importance of the HBB gene in understanding its evolutionary conservation and function. Comparative analysis reveals the preservation of critical regions across species, offering insights into gene evolution, protein function, and their relevance to health and disease.

1. Sequence Retrieval & BLAST Search

Methodology

- The human HBB protein sequence (Ref Seq: **NP_000509.1**) was retrieved from the NCBI database.
- A **BLASTP** search was performed to find homologous HBB protein sequences in Six other species: chimpanzee, cow, Horse, Mouse, Chicken, and Zebrafish.
- The top hits were selected based on high sequence identity and annotated as hemoglobin subunit beta in each species.
- Sequences were downloaded in FASTA format for downstream analysis.

Results

Species	Scientific Name	Taxonomy ID	Accession Number	% Identity with Human HBB
Cow	<i>Bos taurus</i>	9913	NP_776342.1	84.72%
Horse	<i>Equus caballus</i>	9796	NP_001157490.1	83.67%
Mouse	<i>Mus musculus</i>	10090	NP_001265090.1	80.27%
Chimpanzee	<i>Pan troglodytes</i>	9598	XP_508242.1	100%
Zebrafish	<i>Danio rerio</i>	7955	NP_001003431.2	50.00%
Chicken	<i>Gallus gallus</i>	9031	NP_990820.1	69.39%

BLASTP results show high HBB similarity in mammals, with chimpanzee closest to human, followed by horse, cow, and mouse. Lower identity in chicken and zebrafish reflects greater evolutionary distance.

2. Pairwise Sequence Alignment

Methodology

Pairwise alignments were performed using the EMBOSS Needle tool:

- **Human vs Chimpanzee** (closely related)
- **Human vs Zebrafish** (distantly related)

Sequences used:

- Hemoglobin subunit beta [Homo Sapiens] - NCBI Reference Sequence: NP_000509.1
- Hemoglobin, beta adult 2 [Danio rerio] - NCBI Reference Sequence: NP_001003431.2
- Hemoglobin subunit beta [Pan troglodytes] - NCBI Reference Sequence: XP_508242.1

Results

A. Human vs Chimpanzee

- **% Identity:** 100%
- **% Similarity:** 100%
- **Gaps:** 0
- **Interpretation:** The HBB gene in chimpanzees is identical to that in humans, which aligns with their close evolutionary relationship.

B. Human vs Zebrafish

- **% Identity:** 49.7%
 - **% Similarity:** 71.4%
 - **Gaps:** 0
 - **Interpretation:** Significant divergence is observed, indicating zebrafish are distantly related to humans. However, some conservation is retained, suggesting functionally important regions.
-

3. Multiple Sequence Alignment (MSA)

Tool Used: Clustal Omega

- All six HBB protein sequences were aligned.

BioinformHer Mini Project – Module 2 Capstone

Cow	Horse	Mouse	Chimpanzee	Zebrafish	Chicken
NP_776342.1	NP_001157490.1	NP_001265090.1	XP_508242.1	NP_001003431.2	NP_990820.1

Results

CLUSTAL O(1.2.4) multiple sequence alignment

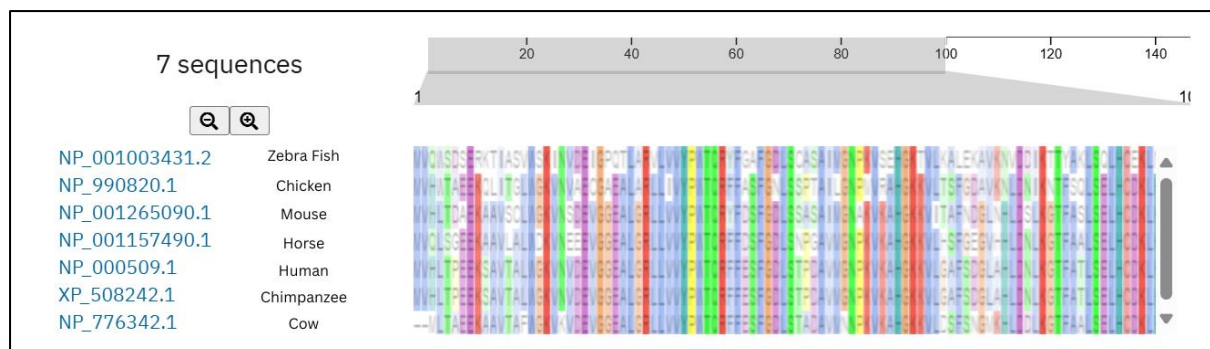
```

NP_001003431.2  MVQWSDSERKTIASVWSKINVDEIGPQTLARVLVVPWTQRYFGAFGDLSCASAIMGNPK 60
NP_990820.1     MVHWTAEKQLITGLWGKVNVAECGAELARLLIVYPWTQRFASFGLSSPTAILGNPM 60
NP_001265090.1 MVHLTDAEKAASVCLWGKVNSEVGGGALGRLLVVPWTQRYFDSFGDLSSASAIMGNAK 60
NP_001157490.1 MVQLSGEEKAAVLALWDKVNNEEVGGGALGRLLVVPWTQRFDSFGDLSPGAVMGNPK 60
NP_000509.1     MVHLTPEEKSAVTALWGKVNDEVGGGALGRLLVVPWTQRFESFGDLSTPDAMGNPK 60
XP_508242.1     MVHLTPEEKSAVTALWGKVNDEVGGGALGRLLVVPWTQRFESFGDLSTPDAMGNPK 60
NP_776342.1     --MLTAEKAAVTAFWGKVKVDEVGGGALGRLLVVPWTQRFESFGDLSTADAMNNPK 58
                  :  *:  :  .*.*:  *  *  :*.*:*****:  :*:**  *::.*

NP_001003431.2  VSEHGKTVLKALEKAVKNVDDIKTTYAKLSQLHCEKLNVDPNFKLLADCLSIIVATNFG 120
NP_990820.1     VRAHGKKVLTSTFGDAVKNLNLIKNTFSQLSELHCDKLHVDPENFRLLGDILIIIVLAHFS 120
NP_001265090.1 VKAHGKKVITAFNDGLNHLDSLKGTFASELHCDKLHVDPENFRLLGNMIVIVLGHHLG 120
NP_001157490.1 VKAHGKKVLHSFGEVHHLNLIKGTFAALSELHCDKLHVDPENFRLLGNVLVVVLARHFG 120
NP_000509.1     VKAHGKKVLGAFSDGLAHLNLIKGTFAATSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
XP_508242.1     VKAHGKKVLGAFSDGLAHLNLIKGTFAATSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
NP_776342.1     VKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFG 118
                  *  ***.*:  ::  ..  :*:.*  *::  **:*:*:*:*:*:*:*:*:*:  :  *:.  :.

NP_001003431.2  PAFNPSVQSTWQKLLSVVVAALTSRYF 147
NP_990820.1     KDFTPQCQAAWQKLVRVVAHALARKYH 147
NP_001265090.1 KDFTPAAQAAFQKVAVGATALAHKYH 147
NP_001157490.1 KDFTPQLQASYQKVAVGVANALAHKYH 147
NP_000509.1     KEFTPPVQAAQKVAVGVANALAHKYH 147
XP_508242.1     KEFTPPVQAAQKVAVGVANALAHKYH 147
NP_776342.1     KEFTPVLQADFQKVAVGVANALAHRYH 145

```



A multiple sequence alignment of HBB proteins across six species shows high conservation among mammals, especially in the N-terminal region. Key motifs related to heme-binding and oxygen transport remain conserved even in distant species like chicken and zebrafish, highlighting evolutionary pressure to preserve these functional regions.

4. Sequence Logo Generation

Tool used: Skylign

To **visualize conserved residues** in the HBB protein across six species using a sequence logo. The height of each letter indicates how conserved an amino acid is at that position.

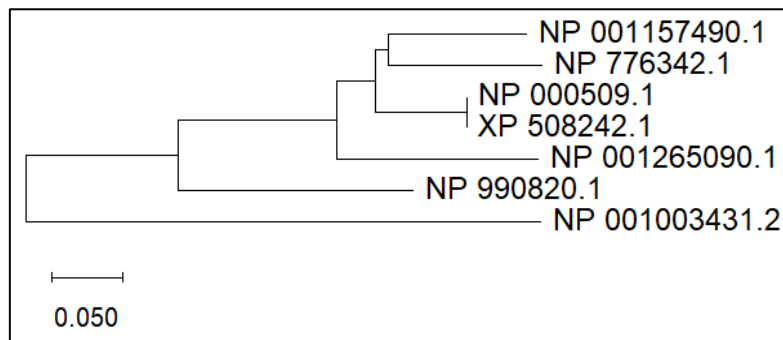


Highly conserved amino acids, such as **Valine (V)**, **Histidine (H)**, **Lysine (K)**, and **Glutamic acid (E)**, are consistently present in the N-terminal region across all species, reflecting their essential role in **heme binding** and **oxygen transport**. Variations observed in species like chicken and zebrafish involve substitutions in less critical regions, often with **similar or functionally tolerable residues**, suggesting adaptive changes without compromising protein function.

5. Phylogenetic Tree Construction

A phylogenetic tree was constructed using the aligned HBB protein sequences of human, chimpanzee, cow, mouse, chicken, and zebrafish in **MEGA X software** (Neighbor-Joining method, Poisson model).

Results



The HBB-based phylogenetic tree shows human and chimpanzee as closest relatives, followed by cow and mouse. Chicken and zebrafish diverge earlier, reflecting their distant relation to mammals. The tree aligns with known evolutionary relationships, highlighting HBB as a reliable marker for phylogenetic analysis.
