

BioinformHer Mini Project – Module 2 Capstone

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

Project Objective:

The objective of this project is to investigate the evolutionary conservation of the Hemoglobin Subunit Beta (HBB) gene across six species—Human, Mouse, Chimpanzee, Chicken, Cow, and Zebrafish—using bioinformatics techniques learned in Module 2. The study involves retrieving HBB protein sequences from public databases, performing multiple sequence alignment (MSA) to compare sequence similarities, generating a sequence logo to visualize conserved amino acid residues, and constructing a phylogenetic tree using MEGA X to infer evolutionary relationships. This project aims to explore how conserved the HBB gene is across species and what that implies about functional constraints and evolutionary divergence.

Task 1: Sequence Retrieval & BLAST Search

Table 1: Percentage Identity of HBB Gene Sequences in Different Species Compared to Human

Species	Accession Number	% identity with human HBB
Pan troglodytes (chimpanzee)	NP_001129304.1	75.51%
Bos taurus (cow)	NP_001103977.1	77.55%
Mus musculus (mouse)	NP_001265090.1	80.27%
Gallus gallus (chicken)	NP_990820.1	69.39%
Danio rerio (zebrafish)	NP_001003431.2	50.00%

Task 2: Pairwise Sequence Alignment

Species Chosen:2

Closely related → Mouse

Distantly related → Zebrafish

Tool Used: EMBOSS Needle

Table 2: Pairwise Sequence Comparison of Human HBB Gene with Mouse and Zebrafish Using EMBOSS Needle

Comparison	%Identity	%Similarity	Gaps	Interpretation
Human vs Mouse	80.3%	89.8%	0%	Highly conserved; close evolutionary link
Human vs Zebrafish	49.7%	71.4%	0%	Moderately conserved; distant relation

The human and mouse HBB protein sequences are highly conserved, with 89.8% similarity and no gaps. This reflects their close evolutionary relationship as both are primates. The differences are minimal, mostly conservative substitutions that retain functional similarity.

The alignment shows moderate conservation between human and zebrafish HBB proteins. While there's no gap, the % identity is only 49.7%, indicating greater evolutionary divergence. Still, some functionally important regions remain conserved, suggesting that the haemoglobin function is preserved across vertebrates, despite evolutionary distance.

Task 3: Multiple Sequence Alignment (MSA)

Tool Used: MUSCLE



Figure 1: CLUSTAL multiple sequence alignment of beta-globin protein sequences from six vertebrates using MUSCLE (v3.8). Asterisks (*) denote conserved residues, colons (:) denote residues with strongly similar properties, and periods (.) indicate weak similarity

NP_000509.1 (Human)
NP_001003431.2 (zebrafish)
NP_990820.1 (Chicken)
NP_001129304.1 (Chimpanzee)
NP_001103977.1 (cow)
NP_001265090.1 (Mouse)

Highly Conserved Segments:

Table 3: Highly Conserved Amino Acid Regions in HBB Across Species

Region 1:	VVYPWTQRFFDSFGNLSS
Region 2:	HCDKLHVDPENFRLLGN
Region 3:	QAAWQKLIV

Task 4: Sequence Logo Generation

Tools Used: Skylign



Figure 2: Sequence logo generated from beta-globin protein multiple sequence alignment using Skylign. The height of each letter corresponds to the conservation level of that amino acid at each position.

What do you observe?

The sequence logo above, generated from a multiple sequence alignment using MUSCLE and visualised via Skylign, shows varying levels of amino acid conservation across the aligned protein sequences. Certain positions have tall, dominant letters indicating high conservation, while others show a mixture of residues, indicating variability.

Are there highly conserved residues?

Yes, several residues are highly conserved across the sequences. These include:

V (Valine), L (Leucine), F (Phenylalanine), G (Glycine), K (Lysine), W (Tryptophan), H (Histidine), C (Cysteine).

These residues are shown as large, single letters, indicating minimal variation at those sites among all sequences.

Why might those regions be important?

These conserved residues are likely crucial for the structural and functional integrity of the protein. In the context of haemoglobin or globin-like proteins:

- Histidine (H) is often involved in binding the heme group.
- Cysteine (C) may be important in disulfide bond formation or structural stability.
- Tryptophan (W) and Phenylalanine (F) are bulky hydrophobic residues often found in the core of the protein, stabilizing the fold.
- Other residues like Valine, Leucine, Glycine, and Lysine contribute to hydrophobic interactions, flexibility, or charge balance, which are essential for correct protein function.

These conserved sites have likely been preserved throughout evolution due to their essential biological roles.

Task 5: Phylogenetic Tree Construction

Tool Used: MEGA X

Which species are most closely related based on HBB?

The tree shows that Chimpanzee (NP_001129304.1) and Cow (NP_001103977.1) cluster together closely, indicating a high degree of similarity in their HBB protein sequences.

Mouse (NP_4001265090.1) and Human (NP_000509.1) also appear closely related and somewhat to the above group.

Zebrafish (NP_001003431.2) is most distantly related to all, indicating it is the outgroup or evolutionarily farthest.

Does this tree match what you expect evolutionarily?

Yes, the tree largely matches expected evolutionary relationships:

Humans and chimpanzees, both primates, are expected to cluster closely. Mice and cows are mammals and grouped somewhat near the primates. Chicken (NP_990820.1), a bird, branches off earlier. Zebrafish, a distant vertebrate species, is most distantly related, consistent with evolutionary timelines.

Despite Mouse and Human being more distant in general evolutionary terms compared to Human and Chimpanzee, the functional constraints on the HBB gene may have caused the sequences in these two species to remain more similar due to similar oxygen-transport roles and physiological requirements.