

Project Title: Hemoglobin Beta (HBB) Gene Evolution Across Species

Author: Gargi Durbude

Email: gauridilip2001@gmail.com

Project: BioinformHER Mini-Project (Module 2)

Hemoglobin Beta (HBB) Gene Evolution Across Species

This project is part of the **BioinformHER Module 2 Mini-Project**. It explores the evolutionary conservation of the Hemoglobin Beta (HBB) gene across multiple species using various bioinformatics tools.

Objective

To investigate the evolutionary conservation of the Hemoglobin Beta (HBB) gene across species using BLAST, sequence alignment, and phylogenetic analysis.

Tasks Completed

1. Sequence Retrieval & BLAST

- Retrieved the HBB gene (protein sequence) of *Homo sapiens* from NCBI.
- Used **NCBI BLASTp** to identify homologous sequences in:
 - *Peromyscus maniculatus* (Deer mouse)
 - *Pan troglodytes verus* (Western chimpanzee)
 - *Mus musculus* (House mouse)
 - *Anas flavirostris oxyptera* (Andean yellow-billed teal)
 - *Troglodytes aedon* (House wren)
- Retrieved and stored sequences in FASTA format.

Scientific Name	Common Name	Accession Number	% Identity with Human HBB
Peromyscus maniculatus	Deer mouse	AJQ22215.1	76.87%
Myodes glareolus	Bank vole	AYL99778.1	78.23%
Mus musculus	House mouse	BAG16713.1	80.27%
Anas flavirostris oxyptera	Yellow-billed Teal (bird)	ACT81558.1	69.39%
Troglodytes aedon	House wren (bird)	ALG03129.1	68.71%

Table 1: BLAST Result

2. Pairwise Sequence Alignment

Performed using EMBOSS Needle

Human HBB vs Mus musculus

- % Identity: 118/147 (80.3%)
- % Similarity: 131/147 (89.1%)
- Gaps: 0/147 (0.0%)

Interpretation:

High conservation with no gaps indicates functional and structural importance.

Human HBB vs Troglodytes aedon (House wren)

- % Identity: 101/147 (68.7%)
- % Similarity: 122/147 (83.0%)
- Gaps: 0/147 (0.0%)

Interpretation:

Moderate identity suggests evolutionary distance, but conserved structure is still likely.

Conclusion:

- Most closely related to human HBB: Mus musculus
- More distantly related: Troglodytes aedon

3. Multiple Sequence Alignment (MSA)

Performed using Clustal Omega.

Goal: Identify conserved residues across species to understand functional and evolutionary significance.



Fig 1: MSA using Clustal Omega

4. Sequence Logo Generation

- Generated using Skylign from the MSA and helps visualize amino acid conservation.

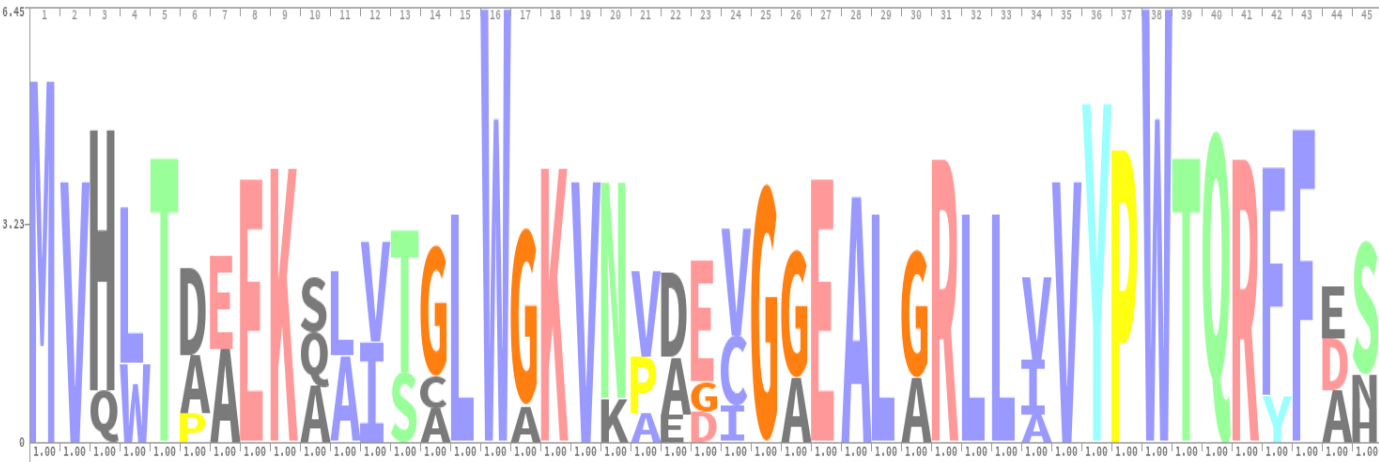


Fig 2: "Sequence Logo Representation of HBB Protein (Positions 1–45)"

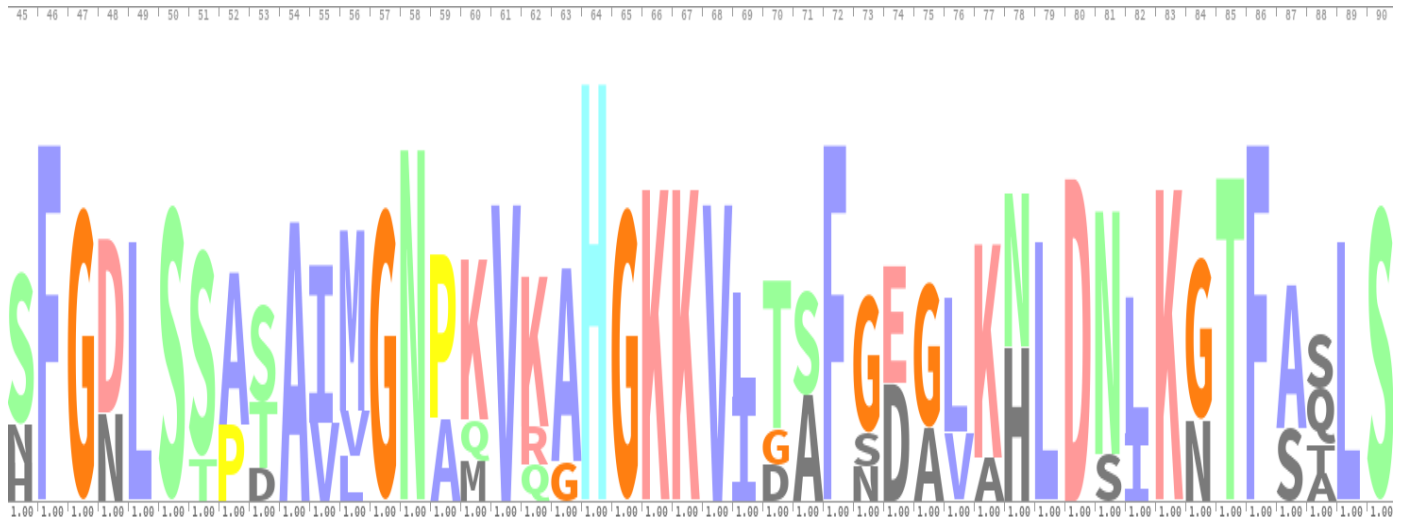


Fig 3: "Sequence Logo Representation of HBB Protein (Positions 45–90)"



Fig 4: "Sequence Logo Representation of HBB Protein (Positions 90–120)"

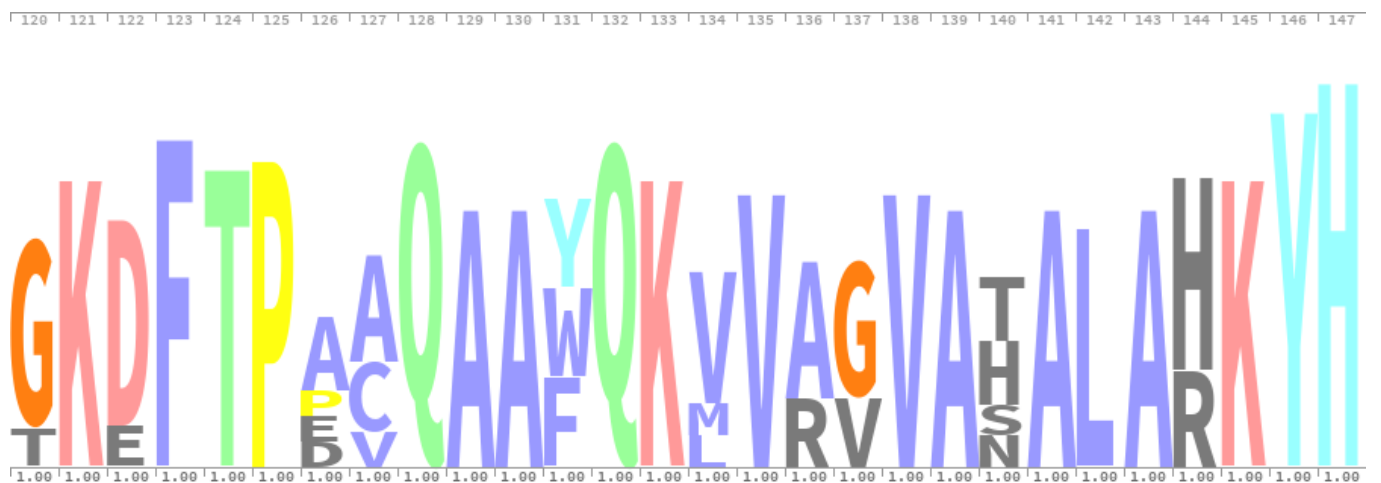


Fig 5: "Sequence Logo Representation of HBB Protein (Positions 120–147)"

Observation

Letter Height	Meaning
Tall single letter	Highly conserved amino acid at that position
Mixed tall/short	One dominant amino acid, a few variants
All short letters	High variability, no single dominant residue

Highly Conserved Positions:

Position	Dominant	Amino Acid
16	W (Tryptophan)	Very tall, almost exclusive presence → key residue
123	F (Phenylalanine)	Tall, highly conserved
139-140	V-A	Very consistent → might stabilize structure
144-147	H-K-Y-H	Highly conserved C-terminal stretch

Why Important?

These residues may indicate:

- Active or catalytic sites, Structural stability
- Evolutionary constraints, Functional conservation
- Predicting protein function through conserved domain identification etc

5. Phylogenetic Tree Construction

- Used the aligned sequences in MEGA X to build a Maximum Likelihood tree.

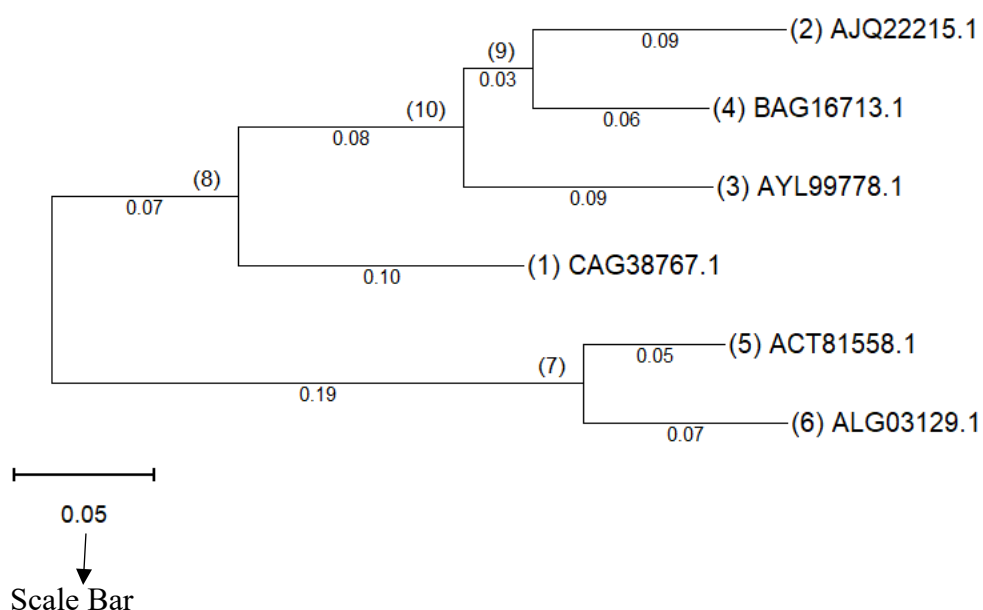


Fig 6: Phylogenetic Tree Construction using MEGA

Interpretation:

Maximum Likelihood is preferred because it is **statistically rigorous, biologically meaningful**, and produces **accurate phylogenetic trees**. In this project, a **Maximum Likelihood tree** was constructed using **MEGA X** to compare the **CAG38767.1** sequence (encoding the human β -globin gene, **HBB**) with five other related sequences. The resulting tree reflects evolutionary relationships and sequence conservation with high accuracy.

. Key Findings:

- CAG38767.1 is positioned between two main clades, indicating it shares evolutionary links with both but diverges significantly.
- Clade 1: AJQ22215.1, BAG16713.1, and AYL99778.1
 - These sequences show moderate divergence from CAG38767.1
 - Likely represent orthologs or variants from related species.
- Clade 2: ACT81558.1 and ALG03129.1
 - These sequences are more distantly related from CAG38767.1.
 - May represent more divergent homologs or possible paralogs.

Scale Interpretation:

- The scale bar (0.05) indicates the number of substitutions per site.
- Shorter branches (<0.05) imply high similarity; longer branches (>0.05) suggest greater evolutionary divergence. Conclusion:

This analysis highlights the evolutionary distance of the HBB gene from related sequences, identifying potential homologous relationships and suggesting varying degrees of sequence conservation.

Yes, this tree generally matches evolutionary expectations. The HBB gene (CAG38767.1) shows:

- Moderate divergence from AJQ22215.1, BAG16713.1, and AYL99778.1 — likely orthologs from related mammalian species.
- Greater divergence from ACT81558.1 and ALG03129.1 — possibly representing more distantly related organisms or paralogs.

Conclusion:

The tree accurately reflects expected evolutionary relationships — mammals cluster closer to humans than birds.