

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
<i>Peromyscus maniculatus</i>	Deer mouse	AJQ22215.1	76.87%
<i>Myodes glareolus</i>	Bank vole	AYL99778.1	78.23%
<i>Mus musculus</i>	House mouse	BAG16713.1	80.27%
<i>Anas flavirostris oxyptera</i>	Yellow-billed Teal (bird)	ACT81558.1	69.39%
<i>Troglodytes aedon</i>	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*

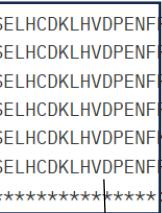
3. Multiple Sequence Alignment (MSA)

Performed using Clustal Omega.

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

CLUSTAL O(1.2.4) multiple sequence alignment

ACT81558.1	MVHWTAEEKQLITGLWGKVNADCGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNPM
ALG03129.1	MVQWTAEEKQLITGLWGKVNVAECGGEALARLLIVYPWTQRFFANFGNLSSATAIVGNPK
CAG38767.1	MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPAVMGNPK
AYL99778.1	MVHLTDAEKAISGLWGKVNADGVGAEALGRLLVVYPWTQRFFEHFGLSSASAVMGNAQ
AJQ22215.1	MVHLTDAEKAALVTGLWGKVKEEIGGEALGRLLAVVYPWTQRFFDSFGDLSSASAIMGNAK
BAG16713.1	MVHLTDAEKSACWAKVNPDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNPK
	: * ** :: ***.: *.*.*. *****: * *:***: *::**
ACT81558.1	VRAHGKKVLTSFGDAVKNLDNIKNTFAQLSELHCDKLHVDPENFRLLGDILIIVLASHFT
ALG03129.1	VQAHGKKVLTSFGEAVKNLDSIKNTFSQISELHCDKLHVDPENFRLLGDILVVLAAHFG
CAG38767.1	VKAHGKKVLGAFTDGLAHLNDNLKGTFATISELHCDKLHVDPENFRLLGNVLVCVLAHHFG
AYL99778.1	VKAHGKKVITAFGDGLKHLNDNLKGTFSAISELHCDKLHVDPENFRLLGNMIVIVLSHDLG
AJQ22215.1	VKGHGKKVIDSFEGGLKHLNDNLKGTFASISELHCDKLHVDPENFKLLGNMIVIVMAHHLG
BAG16713.1	VKAHGKKVITAFNEGKLNDNLKGTFASISELHCDKLHVDPENFRLLGNAIVIVLGHHLG
	*:*****: *:.:. :*.:.***: *****:*****: ***: :: *:. .:
ACT81558.1	KDFTPECQAAWQKMVRVVAHALARKYH
ALG03129.1	KDFTPDCQAAWQKLVRVVAHALARKYH
CAG38767.1	KEFTPVQAAQKVVAGVANALAHKYH
AYL99778.1	KDFTPAAQAAFQKVVAGVASALAHKYH
AJQ22215.1	KDFTPAAQAAFQKVVAGVATALAHKYH
BAG16713.1	KDFTPAAQAAFQKVVAGVATALAHKYH
	*:*** ***:***: * ** ***:***



Conserve region

Fig 1: MSA using Clustal Omega

4. Sequence Logo Generation

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

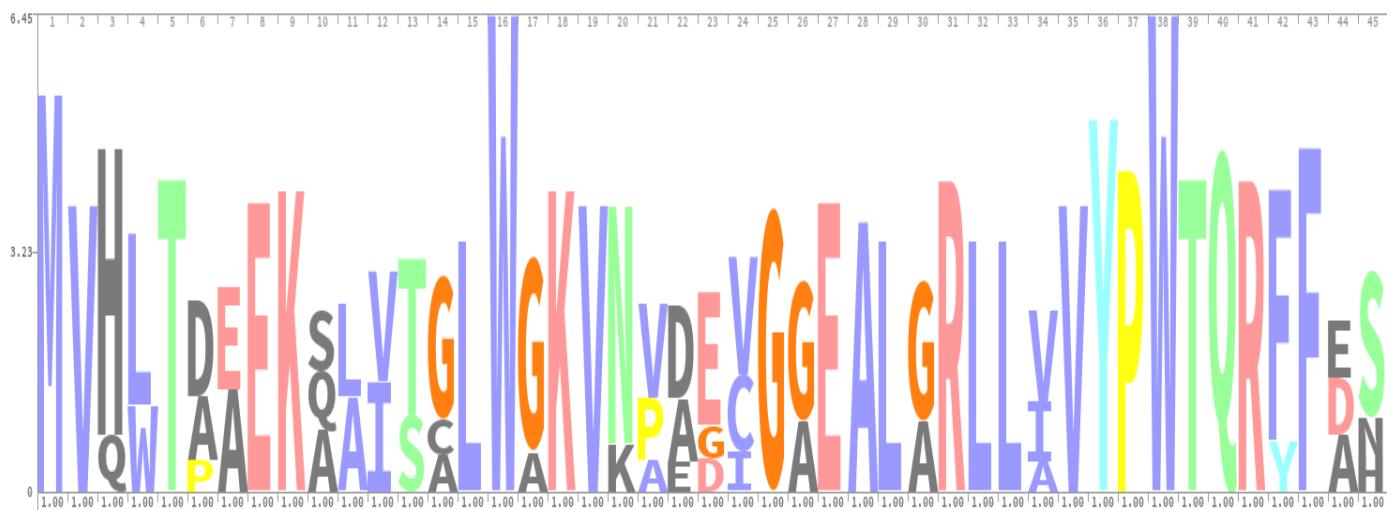


Fig 2: "Sequence Logo Representation of HBB Protein (Positions 1–45)"

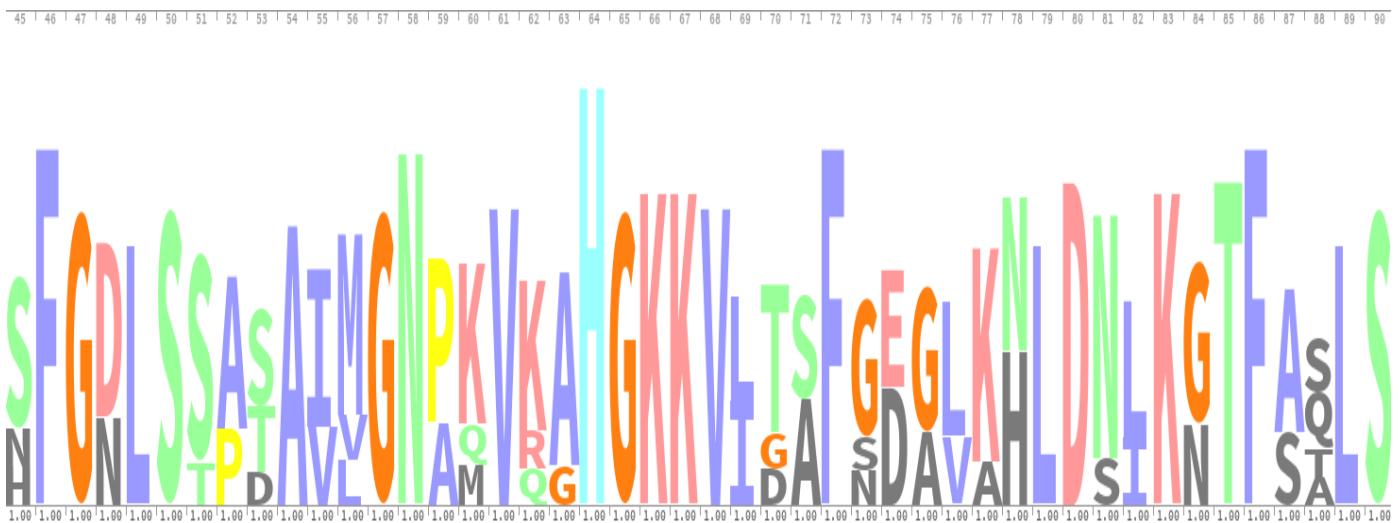


Fig 3: "Sequence Logo Representation of HBB Protein (Positions 45–90)"



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

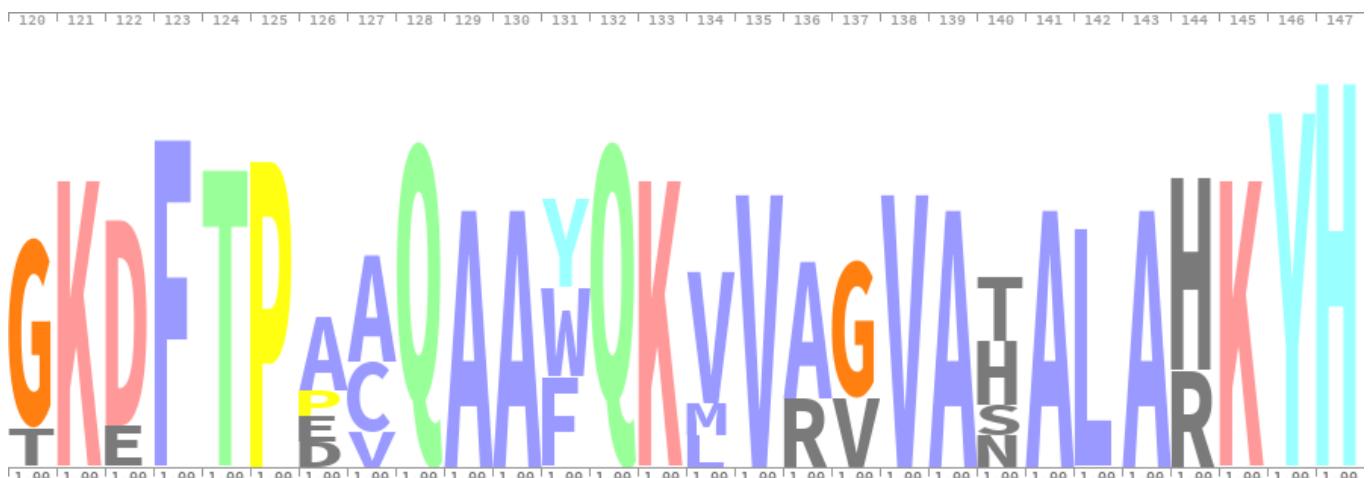


Fig 5: "Sequence Logo Representation of HBB Protein (Positions 120–147)"

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.

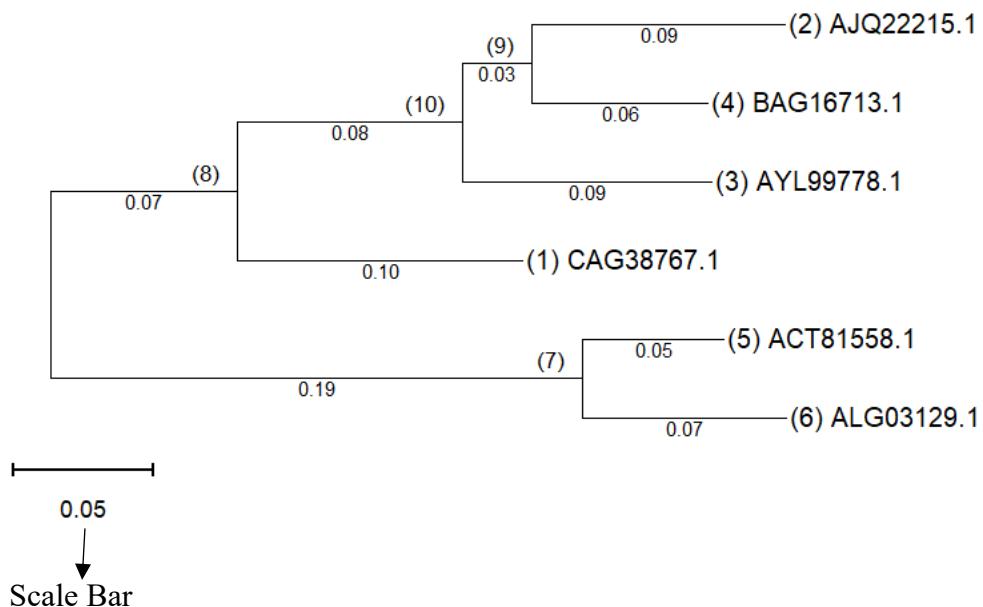


Fig 6: Phylogenetic Tree Construction using MEGA

Interpretation:

1. Most Closely Related Species Based on HBB

From the tree, the most closely related sequences are:

- **AJQ22215.1 and BAG16713.1**, which cluster together with a very **short branch length (~0.03)**, suggesting **high sequence similarity** and **recent common ancestry**.
- These two are also closely related to **AYL99778.1**, forming a tight clade.

2. Evolutionary Consistency

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.

Interpretation:

- **Shorter branches (<0.05)** indicate closer evolutionary relationships (fewer mutations).
- **Longer branches (>0.05)** represent more mutations and thus greater divergence.

This tree supports known patterns of **β -globin gene conservation** among closely related species and divergence in distant lineages.

Conclusion:

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

Author

Gargi Durbude

 gauridilip2001@gmail.com

 Project under BioinformHER Mini-Project Module 2