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

Project Objective:

The objective of this mini-project is to apply the bioinformatics skills acquired in Module 2 to explore the evolutionary conservation of the Hemoglobin Beta (HBB) gene across six different species. This involves retrieving gene sequences using BLAST, performing pairwise and multiple sequence alignments to assess similarity, generating sequence logos to visualize conserved residues, and constructing a phylogenetic tree to infer evolutionary relationships. Through this analysis, we aim to understand how the HBB gene has been conserved or diversified across species and gain insights into its functional and evolutionary significance.

Project Tasks

1: Sequence Retrieval & BLAST Search

Species name	Accession number	% Identity with human HBB
Pan troglodytes	XM_508242.5	100%
Sus scrofa	X86791.1	81.94%
Bos taurus	AB512624.1	80.76%
Mus musculus	EF605506.1	79.03%

2: Pairwise Sequence Alignment

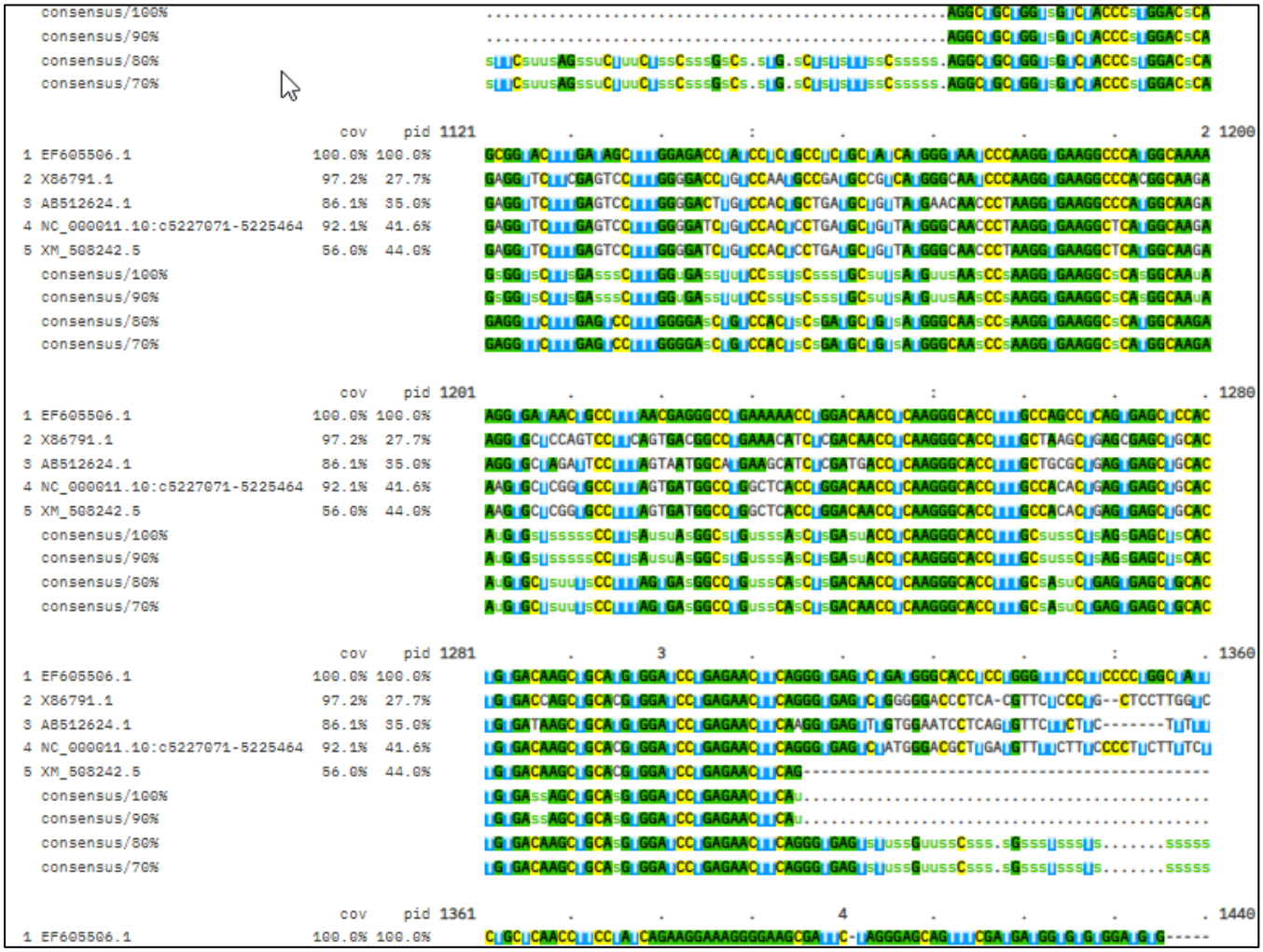
Homosapien Vs Pan troglodytes	Homosapien Vs Mus musculus
Length: 1619	**Length:** 1703
Identity: 627/1619 (38.7%)	**Identity:** 751/1703 (44.1%)
Similarity: 627/1619 (38.7%)	**Similarity:** 751/1703 (44.1%)
Gaps: 991/1619 (61.2%)	**Gaps:** 791/1703 (46.4%)
Score: 2622.0	**Score:** 2279.5

Interpretation:

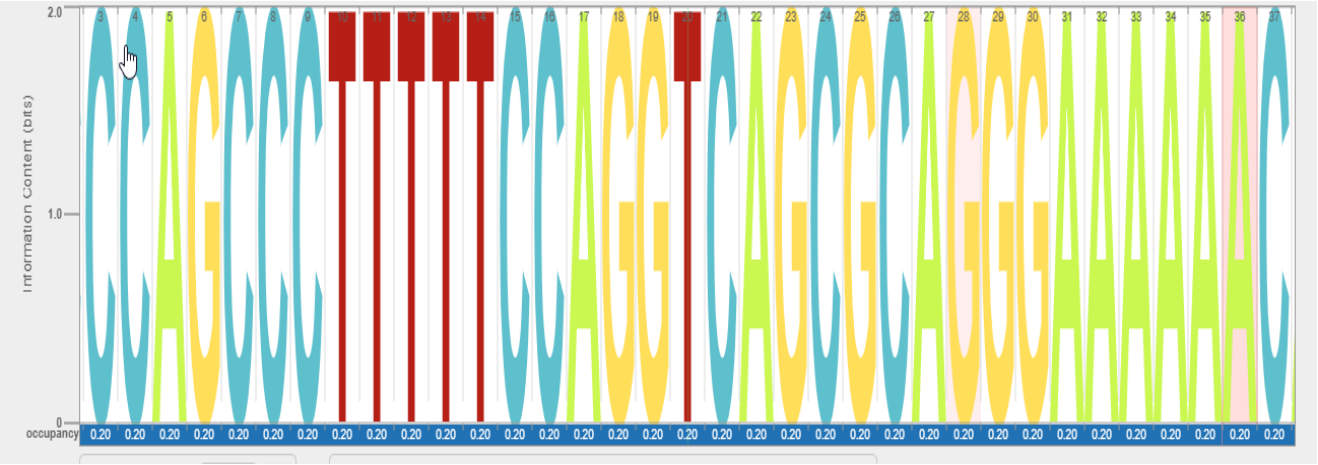
- Higher sequence identity (44.1%) was observed between Human and Mouse HBB genes compared to Human and Chimpanzee (38.7%). However, Chimpanzee alignment has a higher score (2622.0) despite lower identity, likely due to better match regions with fewer substitutions. Gap percentage is significantly higher in Human–Chimpanzee (61.2%) than in Human–Mouse (46.4%), suggesting more alignment interruptions in chimpanzee, which may be due to differences in transcript regions included.

- Human vs. Mouse HBB appears more conserved based on identity percentage and lower gap content. This result is a bit unexpected, since chimpanzees are evolutionarily closer to humans. The discrepancy might be due to differences in transcript lengths or sequence formatting in the alignment files.

3: Multiple Sequence Alignment (MSA)

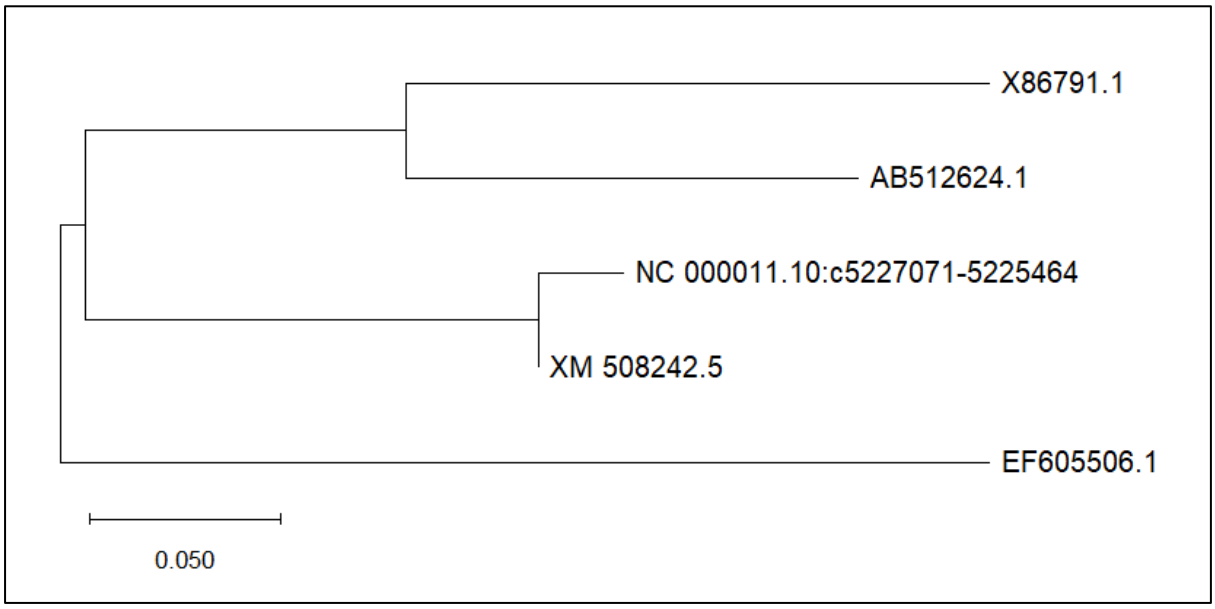


4: Sequence Logo Generation



The sequence logo generated from the multiple sequence alignment (MSA) of the HBB gene across selected species using Skylign visually represents the conservation of amino acid residues at each position in the alignment. In the logo, taller stacks of letters indicate positions with higher conservation, while shorter or more mixed stacks reflect greater variability. Several positions exhibit tall, single-letter stacks, signifying highly conserved residues. These conserved amino acids are likely critical for the structural integrity and biological function of the hemoglobin beta protein. Such regions may correspond to essential domains involved in oxygen binding, heme interaction, or subunit folding. Their evolutionary conservation suggests that mutations in these sites could disrupt protein function and are therefore likely to be deleterious, which explains their preservation across species. This highlights the biological importance of these conserved residues in maintaining the functional role of hemoglobin in oxygen transport.

5: Phylogenetic Tree Construction



Interpretation:

The phylogenetic tree constructed from the HBB gene sequences shows the evolutionary relationships among five species based on sequence similarity. In the tree, NC000011.10:c5227071-5225464 and XM 508242.5 are grouped closely together, indicating that these two species (likely human and chimpanzee or another primate) share a recent common ancestor and have highly similar HBB sequences. This suggests strong evolutionary conservation of the gene in these species. In contrast, EF605506.1 forms the most distant branch, suggesting it is the least closely related among the group—possibly representing a non-primate mammal like a rodent. The tree topology aligns with expected evolutionary relationships, where primates are more closely related to each other, and other mammals like rodents or distant species show greater divergence in their HBB gene sequences.