

EVOLUTIONARY CONSERVATION OF THE HEMOGLOBIN BETA (HBB) GENE

BIOINFORMHER MODULE 2 MINI PROJECT

INTRODUCTION

- The hemoglobin beta gene is the gene that codes for beta-globin, a vital part of hemoglobin, the protein that carries oxygen in red blood cells.
- To comprehend hemoglobin function and disease, trace the history of hemoglobinopathies, clarify evolutionary adaptations, comprehend regulatory mechanisms, and identify potential therapeutic targets, it is essential to comprehend the evolutionary conservation of the HBB gene.

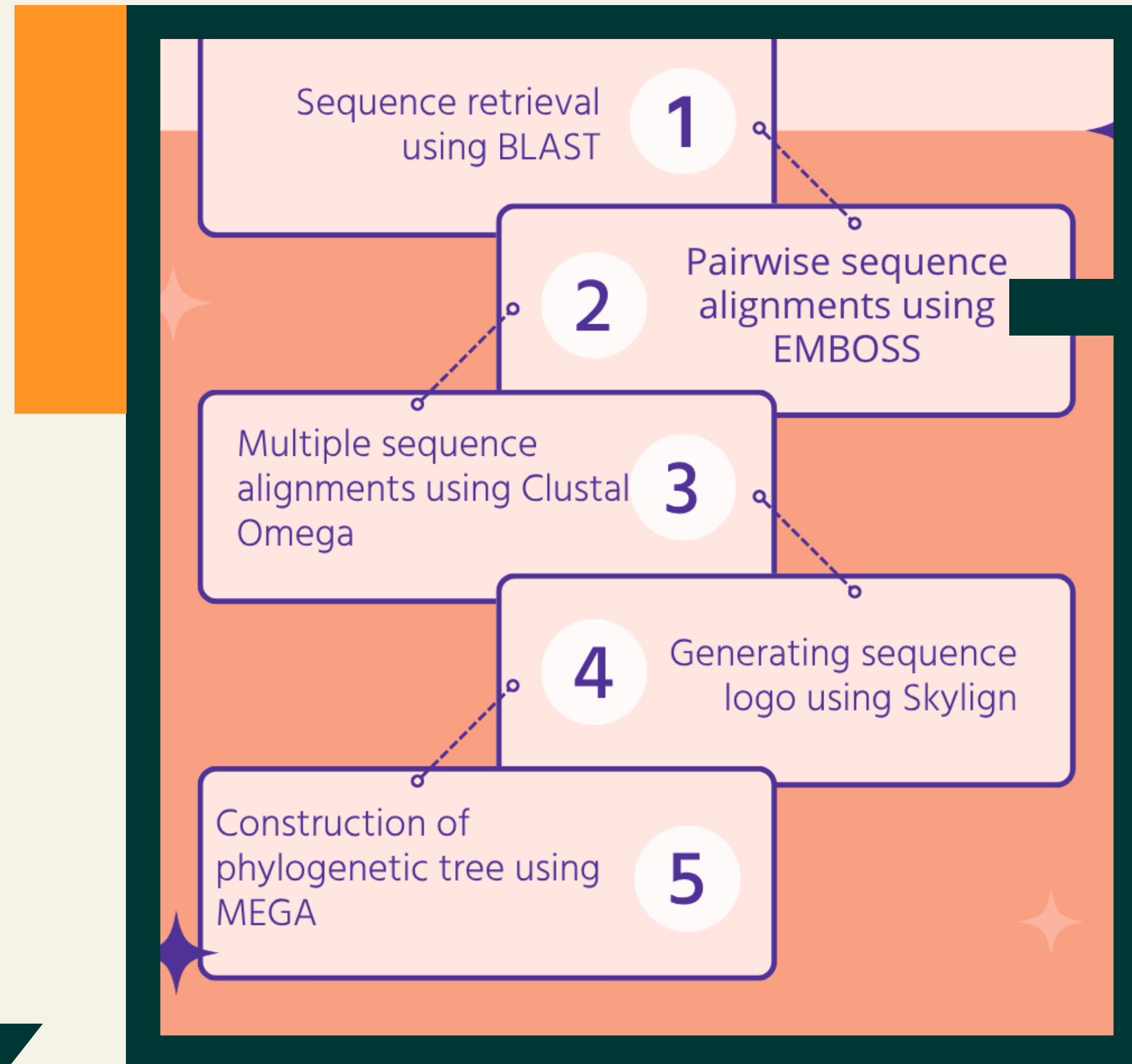


AIMS & OBJECTIVES

- Aim: To explore the evolutionary conservation of the HBB gene across different species using various bioinformatics tools.
- Objectives:
 - Perform sequence retrieval using BLAST
 - Conduct pairwise and multiple sequence alignments
 - Generate sequence logos to visualize conservation
 - Construct phylogenetic trees to infer evolutionary relationships
 - Practice documenting and presenting scientific findings



METHODOLOGY



DATA RETRIVAL

- The search was conducted using the keyword "Hemoglobin Beta" along with the species names of interest, including *Homo sapiens* (human), *Mus musculus*(mouse), and *Danio rerio*(zebra fish).
- The search results were filtered to select only the relevant entries for the HBB.
- Selected sequences were downloaded in FASTA format to ensure compatibility with the alignment tools used in subsequent analyses.
- The following sequences were acquired for the study:
- Homo sapiens HBB: [J00180]
- Mus musculus HBB: [AY410048]
- Danio rerio HBB: [NM_131020]



PAIRWISE SEQUENCE ALIGNMENT

•Tool Used: EMBOSS Needle

Description: The pairwise alignment was performed to identify conserved regions between two sequences.

Results: The alignment output indicated conserved nucleotides across the sequences, with a significant number of matches highlighted by vertical lines in the alignment output.



```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 643
# Identity:      239/643 (37.2%)
# Similarity:    239/643 (37.2%)
# Gaps:          342/643 (53.2%)
# Score: 375.0
#
#=====
EMBOSS_001      1 -----G
EMBOSS_001      1 ATGGTGCACCTGACTGATGCTGAGAAGGCTGCTGTCTCTTGCCTGTGGGG
EMBOSS_001      2 AATTCAAACAGCAGTGAAGTGAAGATTAGAA--TTGTGGAGA-GCACTGGC
EMBOSS_001      51 AA-----AG--GTGAACTCCGAT--GAAGTTGGTGGTGAGGCCCTGG-
EMBOSS_001      49 ATTTAAGAA-----TGTC-ACACTTAGAATGTGTCTCTA--GGCA
EMBOSS_001      89 -----GCAGGCTGCTGGTGTCTACCTTGGG-----CCCAGCGGTA
```


MULTIPLE SEQUENCE ALIGNMENT (MSA)

• **Tool Used:** ClustalW

Description: MSA was conducted to align multiple HBB sequences from different species.

Results: The alignment showed conserved regions across all sequences. Notably, several residues were highly conserved (indicated by asterisks), suggesting functional importance.



Multiple sequence alignment

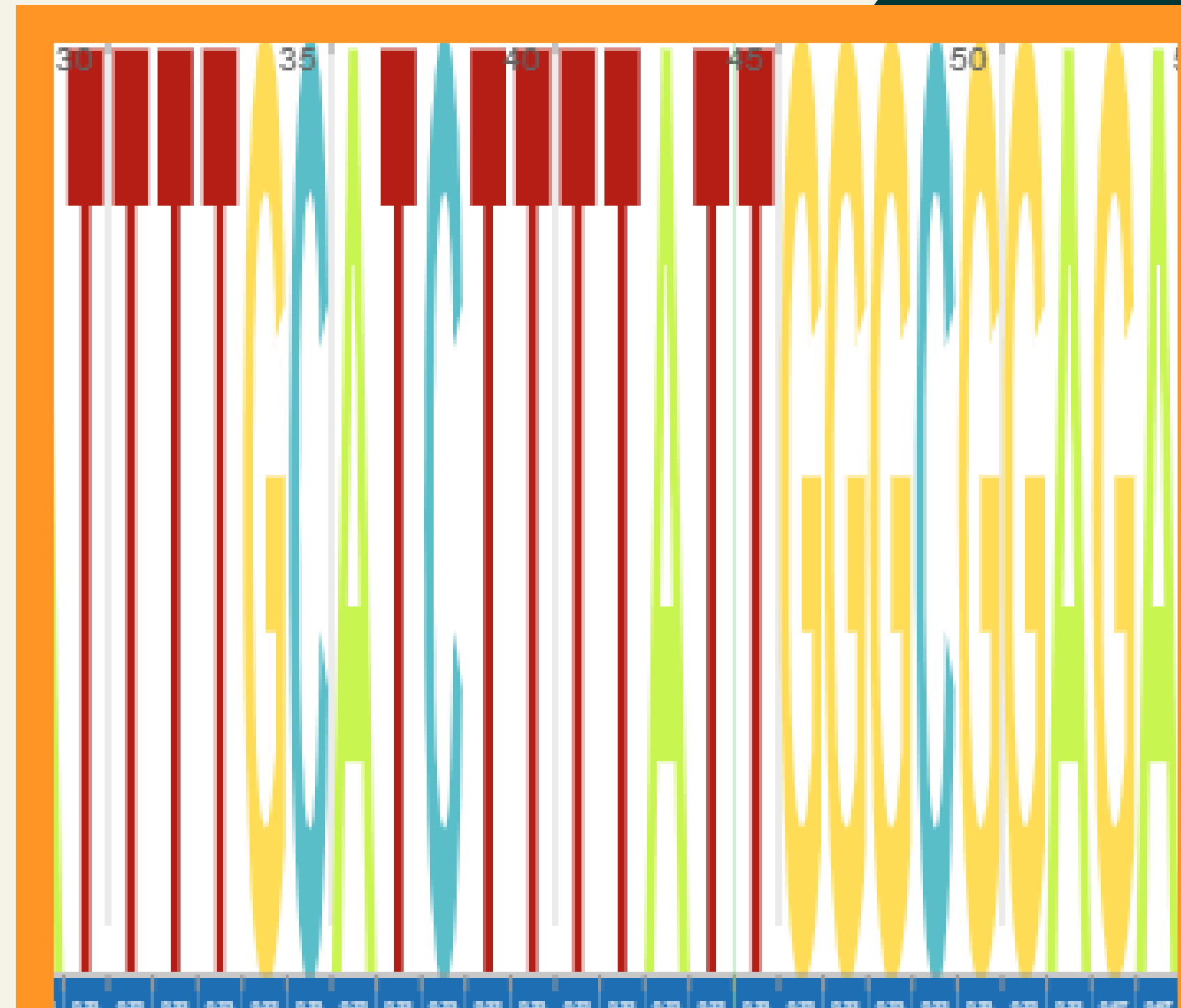
-----GAATTCAA	8
-----	0
CGTCATCTCAGCAAGATTATCCAATAAGATTTTGCATCTTTATTGGGCGGAGACATCAC	60
ACAGCAGTGAACTGAG-----ATTAGAATTGTGGAGAGCACTGGCATTAA	54
-----	0
ACATCAATAAAATGCACCCCGTTTCTGAATTTCTGCTAGTGATTCCCTACTGGACTTTAA	120
GAATGTCACACTTAGAATGTGTCTCTAGGCATTGTTGTTCTGTGCATATATC-----	106
-----ATGGTGACCTGACTGATGCTGAGAAG	27
GCAAGCACCTAAGGACAT-CCTAAAGCGTAAACATGGTTGAGTGGACAGATGCCGAGCGC	179
* ** * **	
--ATCTCAATATTCATTATCTGAAA-----ATTATGAATTAGGTACAAAGCT----	151
GCTGCTGTCTCTTGCCGTGTTGGGAAAGGTGAACTCCGATGAAGTTGGTGGTGAGGCCCTG	87
ACAGCCATCCTTGGCCTGTGTTGGGAAAGCTCAATATCGATGAAATCGGACCTCAGGCCCTA	239
* * * * * * * ***** * ** **	
--CAAATAATTTATTTTTTCAGGTTAG--CAAGAACTTTTTT-TTTTTTTC-----TG	199
GGCAGGCTGCTGGTTGTCTACCCCTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTA	147

SEQUENCE LOGO GENERATION

•**Tool Used:** Skylign

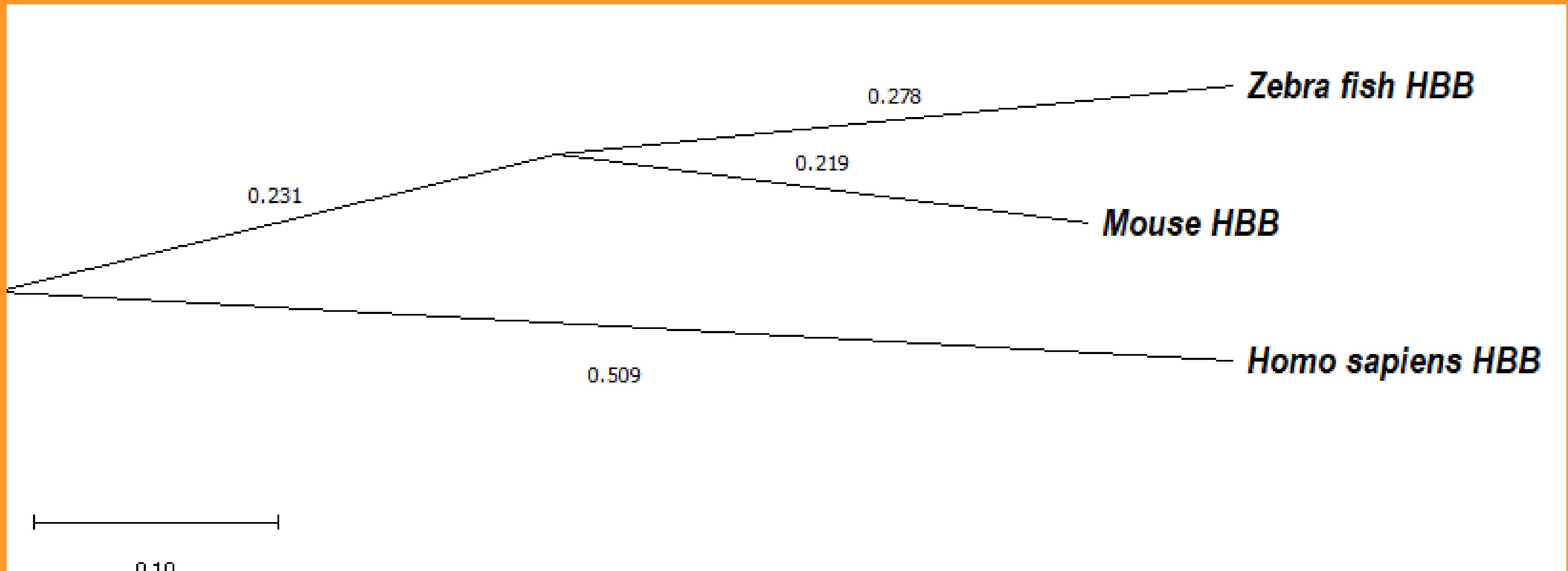
Description: A sequence logo was generated to visualize the conservation of nucleotides across the aligned sequences.

Results: The logo highlighted positions with high conservation, indicating evolutionary pressure on these sites.



PHYLOGENETIC TREE CONSTRUCTION

• Tool Used: MEGA X



RESULT & CONCLUSION

- The results of the pairwise alignment highlighted critical conserved regions in the HBB gene, suggesting these areas are essential for the gene's function across species.
- The multiple sequence alignment reinforced this observation by showing strong conservation patterns, particularly in key regions that correlate with known functional sites of the HBB protein.
- The sequence logo provided a visual representation of conservation, emphasizing positions under strong evolutionary pressure. This is significant, as it suggests that these regions are crucial for the protein's function.
- Finally, the phylogenetic tree indicated the evolutionary relationships among the species studied. The close relationship between humans and mice supports the use of mouse models in biomedical research.