

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

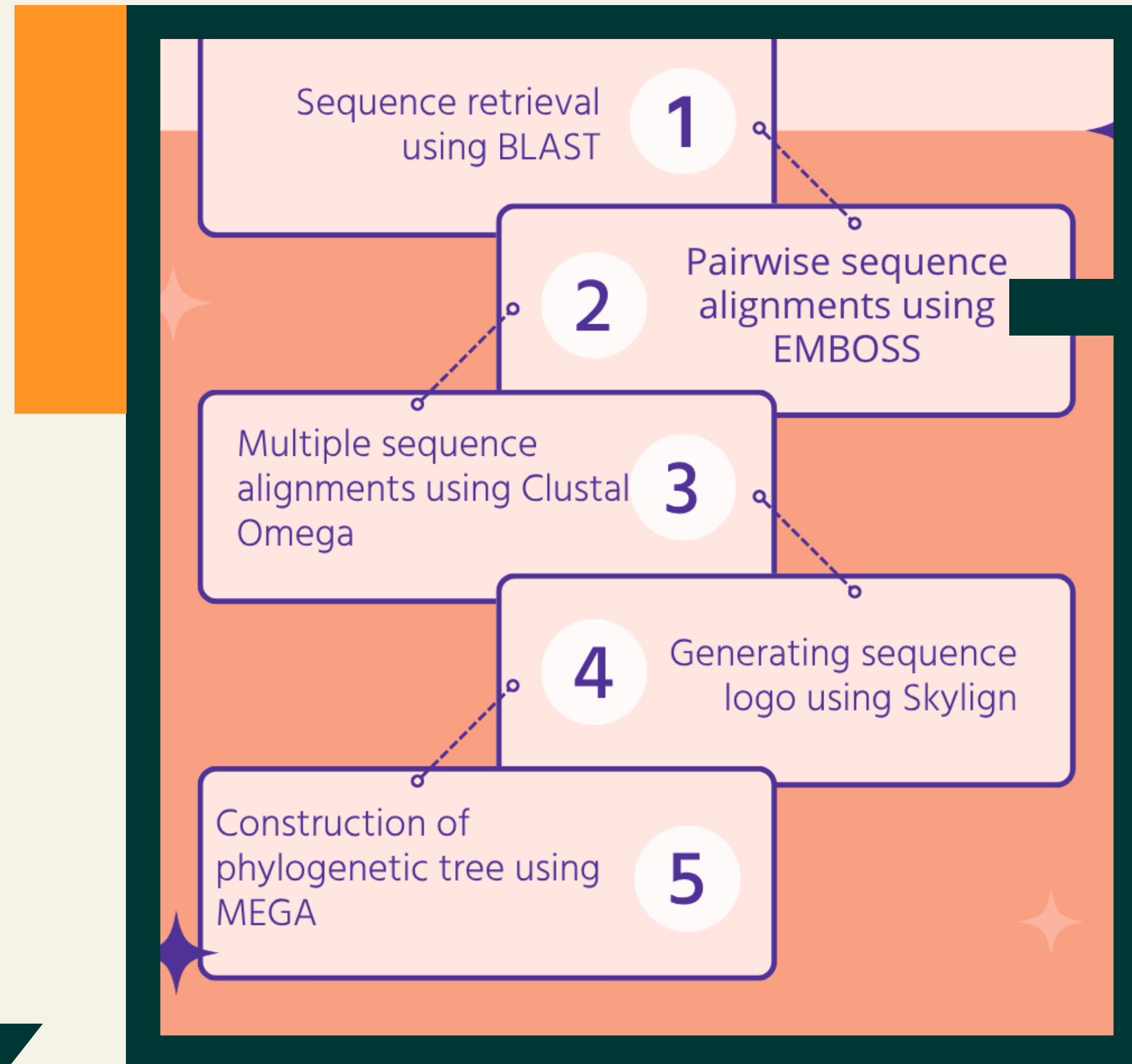
- Aim: To explore the evolutionary conservation of the HBB gene across 6 different species using various bioinformatics tools.

- 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), *Danio rerio*(zebra fish), *Lagothrix lagotricha* (woolly monkey), *Rattus norvegicus* (brown rat) and *Bos taurus* (cow)
- 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]
 - *Lagothrix lagotricha* HBB: [MH382905]
 - *Rattus norvegicus* HBB: [NM_033234]
 - *Bos taurus* HBB: [NM_173917]



PAIRWISE SEQUENCE ALIGNMENT

- **Tool Used:** EMBOSS Needle
- **Description:** The pairwise alignment was performed to identify conserved regions between two sequences: *Homo sapiens* HBB and *Mus musculus* HBB; *Danio rerio* HBB and *Lagothrix lagotricha* HBB; *Rattus norvegicus* HBB and *Bos taurus* HBB
- **Results:** The alignment output indicated conserved nucleotides across the sequences, with a significant number of matches highlighted by vertical lines in the alignment output.

```
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 648
# Identity:   474/648 (73.1%)
# Similarity: 474/648 (73.1%)
# Gaps:       43/648 ( 6.6%)
# Score: 2630.0
#
#=====

EMBOSS_001      1  -----TGCTTCTGACATAGTTGTGTTGACTCACAACTCAGAAACAGACA
                  |||...|||...|||...|||...|||...|||...|||...|||...|||
EMBOSS_001      1  ACACTTGCTTCTGACACAACCGTGTTCACTAGCAACTACACAAACAGACA

EMBOSS_001     46  CCATGGTGCACCTGACTGATGCTGAGAAGGCTGCTGTTAATGGCCTGTGG
                  |||||...|||...|||...|||...|||...|||...|||...|||
EMBOSS_001     51  CCATG-----CTGACTGCTGAGGAGAAGGCTGCCGTCACCGCCTTTTGG

EMBOSS_001     96  GGAAAGGTGAACCCTGATGATGTTGGTGGCGAGGCCCTGGGCAGGCTGCT
                  ||...|||...|||...|||...|||...|||...|||...|||...|||
EMBOSS_001     95  GGCAAGGTGAAAGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGCTGCT

EMBOSS_001    146  GGTGTCTACCCTTGGACCCAGAGGTACTTTGATAGCTTTGGGGACCTGT
                  |||...|||...|||...|||...|||...|||...|||...|||...|||
```


MULTIPLE SEQUENCE ALIGNMENT (MSA)

- **Tool Used:** ClustalW
- **Description:** MSA was conducted to align the 6 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.



```
-----AATTATGAATTAGGTACAAAGC----- 150
CAGCCATCCTTGGCCTGTGGGGAAAGCTCAATATCGATGAAATCGGACCTCAGGCCCTAT 240
CTGCTGTCTCTTGCCTGTGGGGAAAGGTGAACCTCCGATGAAGTTGGTGGTGAGGCCCTGG 88
CTGCTGTTAATGGCCTGTGGGGAAAGGTGAACCTGATGATGTTGGTGGCGAGGCCCTGG 135
CTGCCGTTACCGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGG 88
CTGCCGTACCGCCTTTTGGGGCAAGGTGAAAGTGGATGAAGTTGGTGGTGAGGCCCTGG 134
                                     **** * ** **

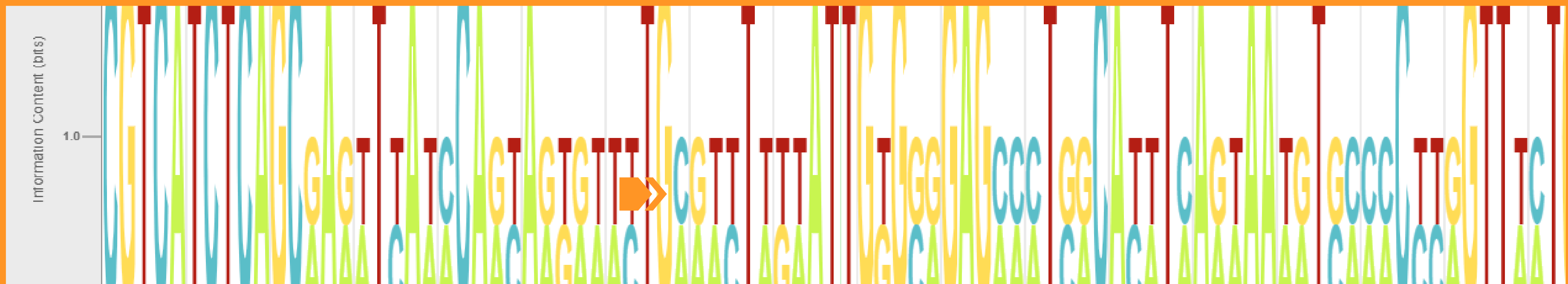
TCAAATAATTTATTTTTTC--AGGTTAGCAAGAACTTTTTTTTTT---TTTTCTGAGAT-- 203
CCAGATGTCTGATCGTGTATCCCTGGACTCAGAGATATTTGCCCACATTCGGCAACCTGT 300
GCAGGCTGCTGGTTGTCTACCCCTTGGACCCAGCGGTACTTTGATAGCTTTGGAGACCTAT 148
GCAGGCTGCTGGTTGTCTACCCCTTGGACCCAGAGGTACTTTGATAGCTTTGGGGACCTGT 195
GCAGGCTGCTGGTGGTTTACCCCTTGGACCCAGAGGTTCTTTGATTCCTTTGGGGATCTGT 148
GCAGGCTGCTGGTTGTCTACCCCTGGACTCAGAGGTTCTTTGAGTCCTTTGGGGACTTGT 194
**      * * * *      * ** * **      **      * *

---GGAGCATTGCTATGGTTGCCAGGCTGGAGTGCAATGGCATGATCCAGGCTCACTGC 260
CAAGCCCCGCTGCGATCATGGGTAAACCCAAAGTGGCAGCTCATGGGAGGACTGTGATGG 360
CCTCTGCCTCTGCTATCATGGGTAAAGGCAAGTGAAGGCCCATGGCAAGAAGGTGATAA 208
CCTCTGCCTCTGCTATCATGGGTAAACCTAAGGTGAAGGCCCATGGCAAGAAGGTGATAA 255
CAACTCCTGATGCTGTTATGAGCAACCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTAG 208
CCACTGCTGATGCTGTTATGAACAACCCTAAGGTGAAGGCCCATGGCAAGAAGGTGCTAG 254
                                     *** * *      * *      ***      ****      *

AACATCTGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTC-----CCAA 308
GAGGTCTTGAGAGAGCCATCAAGAACATGGACAACGTCAAGAACACCTATGCCGCCCTCA 420
```

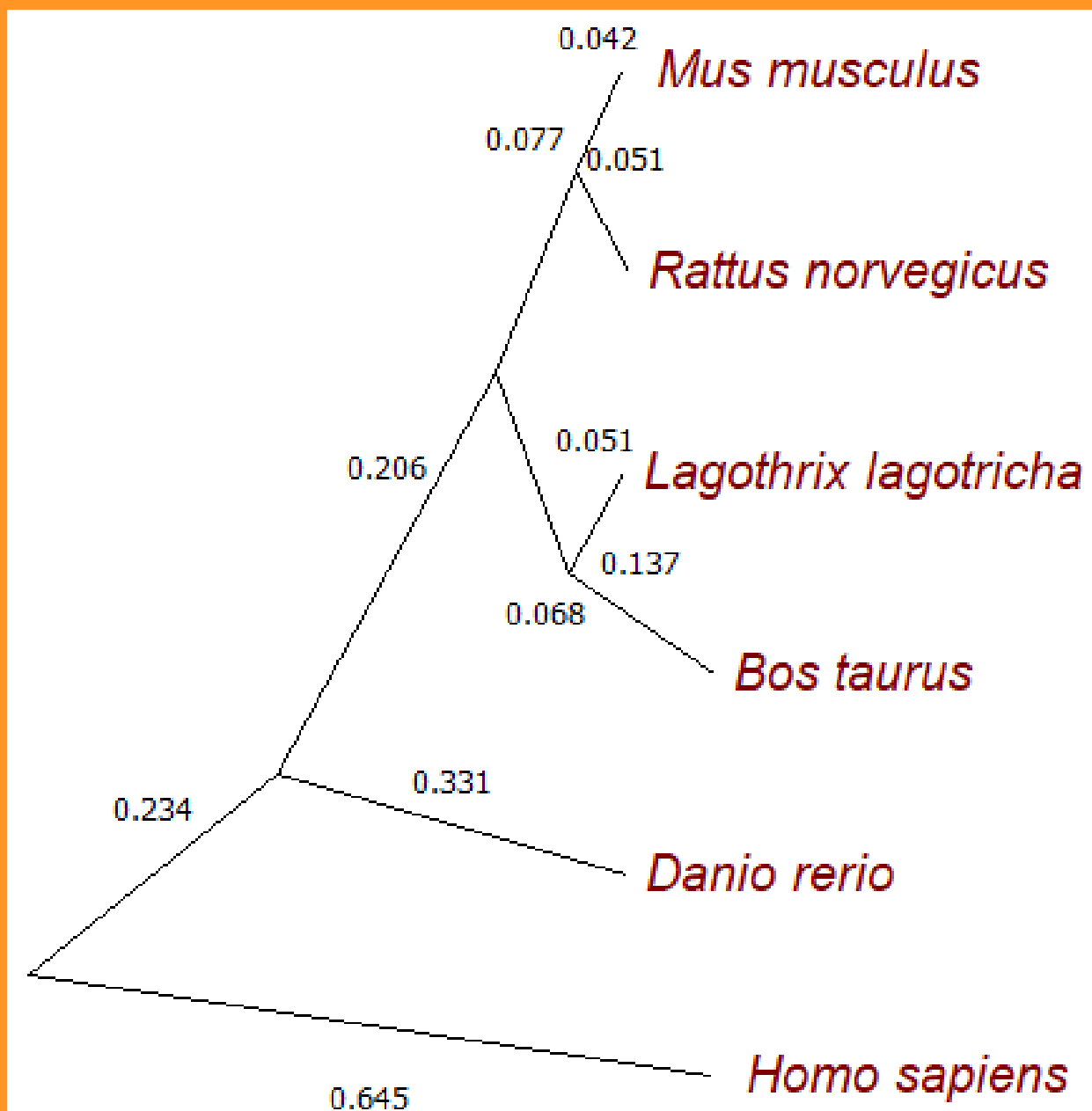
SEQUENCE LOGO GENERATION

- **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
- **Description:** The phylogenetic tree illustrates the evolutionary relationships among the species based on their HBB gene sequences.
- **Results:** The numbers along the branches represent genetic distances between the species. Longer branches indicate greater genetic divergence.



RESULT & CONCLUSION

- **Pairwise Alignment:** The alignment results showed a high identity (49.4%) and similarity (49.4%) across the sequences, indicating conserved regions; however, gaps (9.8%) were present, reflecting variability.
- **Multiple Sequence Alignment:** Highlighted critical conserved regions across species, reinforcing functional importance and evolutionary conservation.
- **Sequence Logo:** Visualized conservation levels, with taller columns indicating positions under strong evolutionary pressure, suggesting functional relevance.
- **Phylogenetic Tree:** Revealed evolutionary relationships, with closer ties between *Mus musculus* and *Rattus norvegicus*, while *Homo sapiens* and *Danio rerio* were more distantly related, supporting the use of specific models in research.
- Overall, these analyses collectively underscore the evolutionary conservation of the HBB gene, its critical functional sites, and the phylogenetic relationships among the studied species