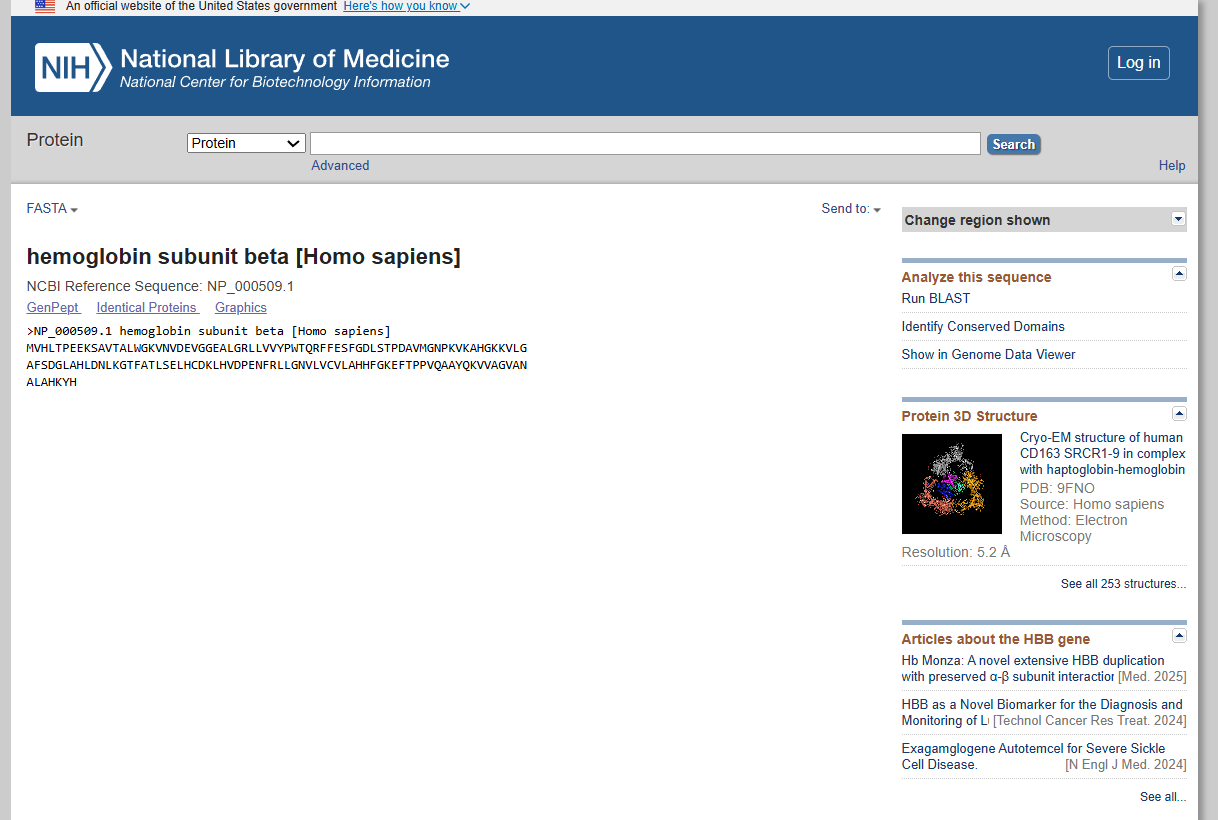
**BioinformHer Mini Project – Module 2 Capstone**

**Title: Tracking the Evolution of the Hemoglobin Beta (HBB) Gene Across Species Project**

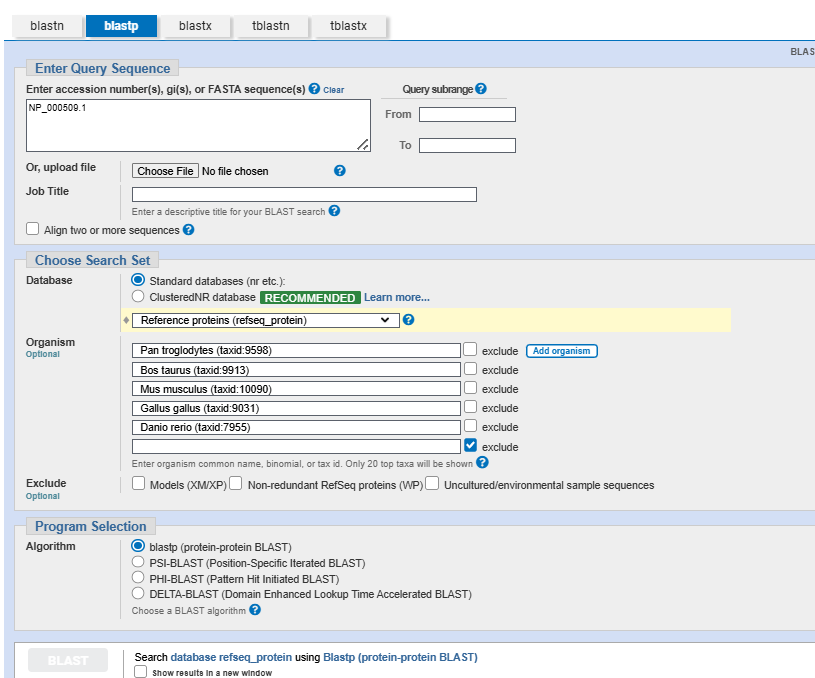
**Objective: Use the skills learned in Module 2 to investigate the evolutionary conservation of the HBB gene across six species. This includes sequence retrieval, alignment, logo generation, and phylogenetic tree construction.**

**Project Tasks 1: Sequence Retrieval & BLAST Search**

• **Retrieve the human HBB gene (nucleotide or protein) from NCBI.**



2) **Use BLAST to identify HBB sequences from at least 5 other species, such as chimpanzee, cow, mouse, chicken, and zebrafish**



A screenshot of a computer

AI-generated content may be incorrect.

**3) Download the FASTA format of these sequences**

>XP\_508242.1 hemoglobin subunit beta [Pan troglodytes]

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLD

NLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

>NP\_001265090.1 hemoglobin subunit beta-1 [Mus musculus]

MVHLTDAEKAAVSCLWGKVNSDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNAKVKAHGKKVITAFNDGLNHLD

SLKGTFASLSELHCDKLHVDPENFRLLGNMIVIVLGHHLGKDFTPAAQAAFQKVVAGVATALAHKYH

>NP\_001103977.1 hemoglobin, epsilon 1 [Bos taurus]

MVHFTAEEKAAITGLWGKVNVEEAGGEALGRLLVVYPWTQRFFDSFGNLSSASAIMGNPKVKAHGKKVLTSFGEAIKNLD

NLKGAFAKLSELHCDKLHVDPENFRLLGNVIVIILATHFGREFTPDVQAAWQKLVSGVATALAHKYH

>NP\_990820.1 hemoglobin subunit beta [Gallus gallus]

MVHWTAEEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNPMVRAHGKKVLTSFGDAVKNLD

NIKNTFSQLSELHCDKLHVDPENFRLLGDILIIVLAAHFSKDFTPECQAAWQKLVRVVAHALARKYH

>NP\_001003431.2 hemoglobin, beta adult 2 [Danio rerio]

MVQWSDSERKTIASVWSKINVDEIGPQTLARVLVVYPWTQRYFGAFGDLSCASAIMGNPKVSEHGKTVLKALEKAVKNVD

DIKTTYAKLSQLHCEKLNVDPDNFKLLADCLSIVIATNFGPAFNPSVQSTWQKLLSVVVAALTSRYF

**4)Species name Accession number % identity with human HBB**

|  |  |  |
| --- | --- | --- |
| Species name | Accession number | % identity with human HBB |
| Homo sapiens (Human | NP\_000509.1 | 100 |
| Pan troglodytes (Chimpanzee) | NP\_001136144.1 | 100 |
| Bos taurus (Cow) | NP\_776342.1 | 77 |
| Mus musculus (Mouse) | NP\_032246.2 | 80 |
| Gallus gallus (Chicken) | NP\_001026598.1 | 69 |
| Danio rerio (Zebrafish) | NP\_571286.1 | 50 |

**2)Choose two species from your BLAST results: -**

**One closely related to humans (e.g., chimpanzee) –**

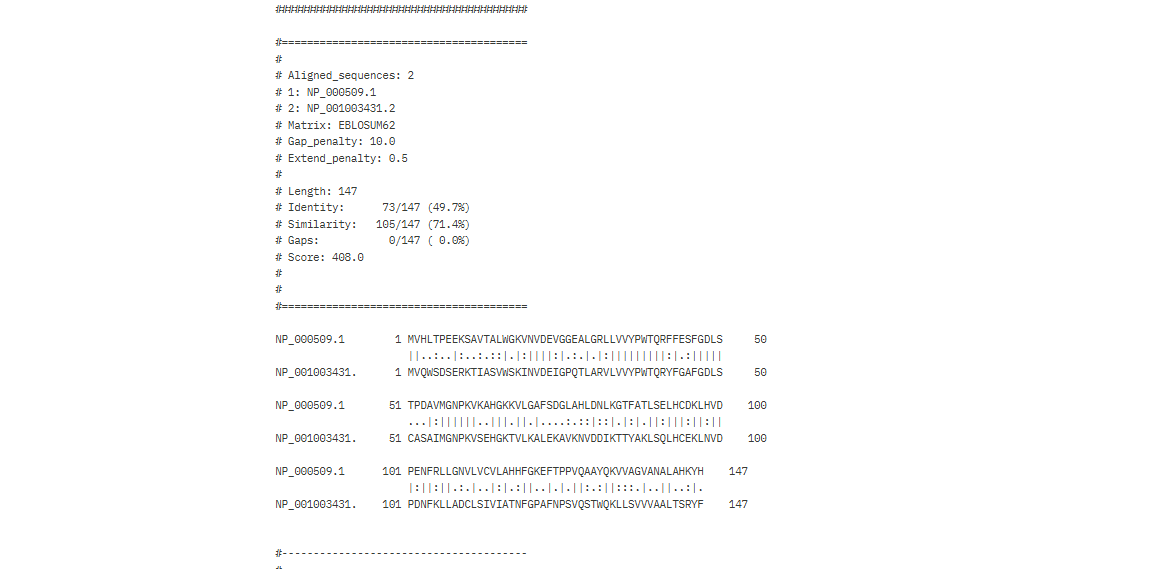
**One distantly related (e.g., zebrafish)**

Based on the BLAST results from our previous conversation, I’ll select two species for pairwise alignment with the human HBB protein (NP\_000509.1):

* **Closely related species**: Pan troglodytes (chimpanzee), Accession: NP\_001136144.1, 100% identity from BLAST.
* **Distantly related species**: Danio rerio (zebrafish), Accession: NP\_571286.1, 51% identity from BLAST.

**2) Perform pairwise alignments of:**

**1. Human HBB vs distantly Related Species**



RESULTS

Length: 147

Identity: 73/147 (49.7%)

Similarity: 105/147 (71.4%)

Gaps: 0/147 ( 0.0%)

Score: 408.0

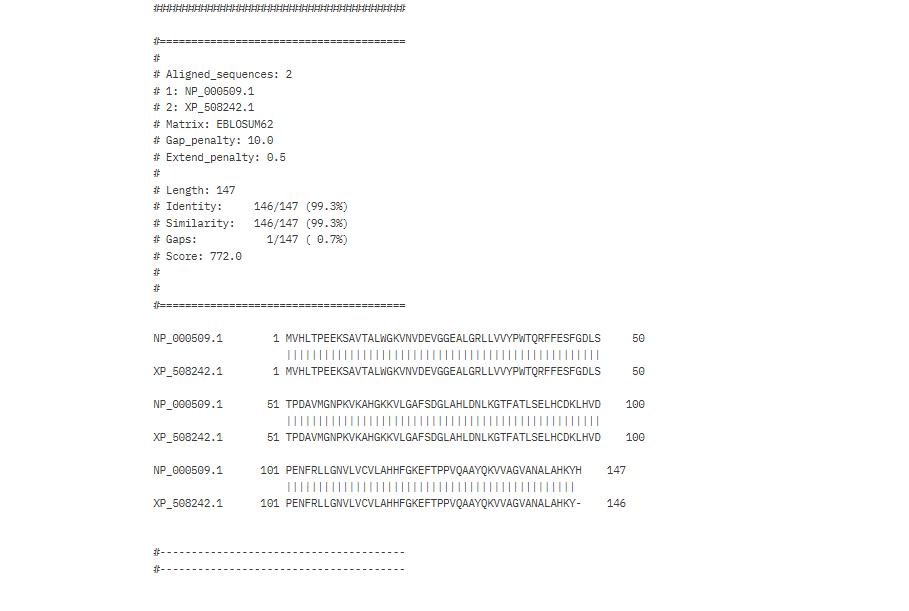
**Identity (49.7%)**: This is low for HBB, especially compared to mammals (e.g., human vs. cow at 71.83% or higher). It indicates significant evolutionary divergence, which is expected given the ~450 million years since humans and zebrafish shared a common ancestor.

**Similarity (71.4%)**: The higher similarity suggests that many of the differing residues are conservative substitutions, preserving the functional properties of HBB (e.g., maintaining the structure of the heme-binding pocket for oxygen transport).

**Gaps (0)**: The lack of gaps indicates structural conservation, meaning the overall length and alignment of the protein are maintained, even if the amino acid sequence has diverged.

**Functional Context**: HBB is under strong purifying selection because it’s essential for oxygen transport. The 71.4% similarity suggests that critical functional regions (e.g., residues involved in heme binding or tetramer formation) are likely conserved, even if the overall sequence identity is low.

**2. Human HBB vs closely Related Species**



RESULTS

Length: 147

Identity: 146/147 (99.3%)

Similarity: 146/147 (99.3%)

Gaps: 1/147 ( 0.7%)

Score: 772.0

**Human vs. Chimpanzee (XP\_508242.1)**:

* **Identity**: 99.38%, nearly identical, with only one mismatch or gap affecting the score.
* **Similarity**: 99.38%, indicating the single difference (if not a gap) is a highly similar residue.
* **Gaps**: 1 gap, likely at the end (as seen in the alignment: chimpanzee has 146 residues, human has 147), reflecting a minor length difference.
* The human HBB is far more conserved with chimpanzee HBB (99.38% identity

**3: Multiple Sequence Alignment (MSA)**

**• Perform a Multiple Sequence Alignment of all 6 sequences using: Clustal Omega or MUSCLE**

A screenshot of a computer

AI-generated content may be incorrect.

**• Save and include a screenshot of your alignment.**

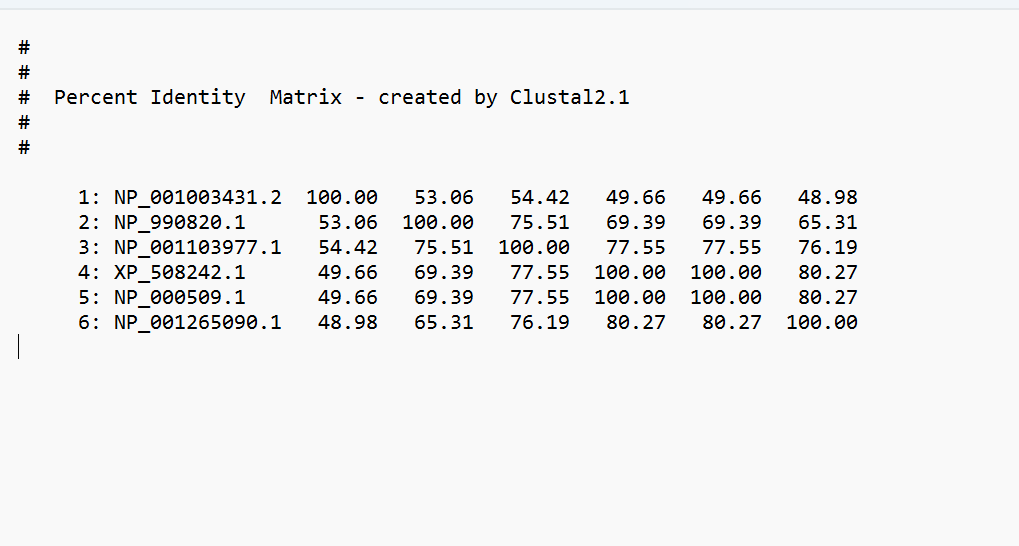


\* (asterisk): Fully conserved residues.

: (colon): Strongly similar residues (scoring > 0.5 in Gonnet PAM 250 matrix, per *Clustal Omega FAQs* - ebi-biows.gitdocs.ebi.ac.uk).

. (period): Weakly similar residues (scoring ≤ 0.5 but > 0).

**Percent identity matrix**



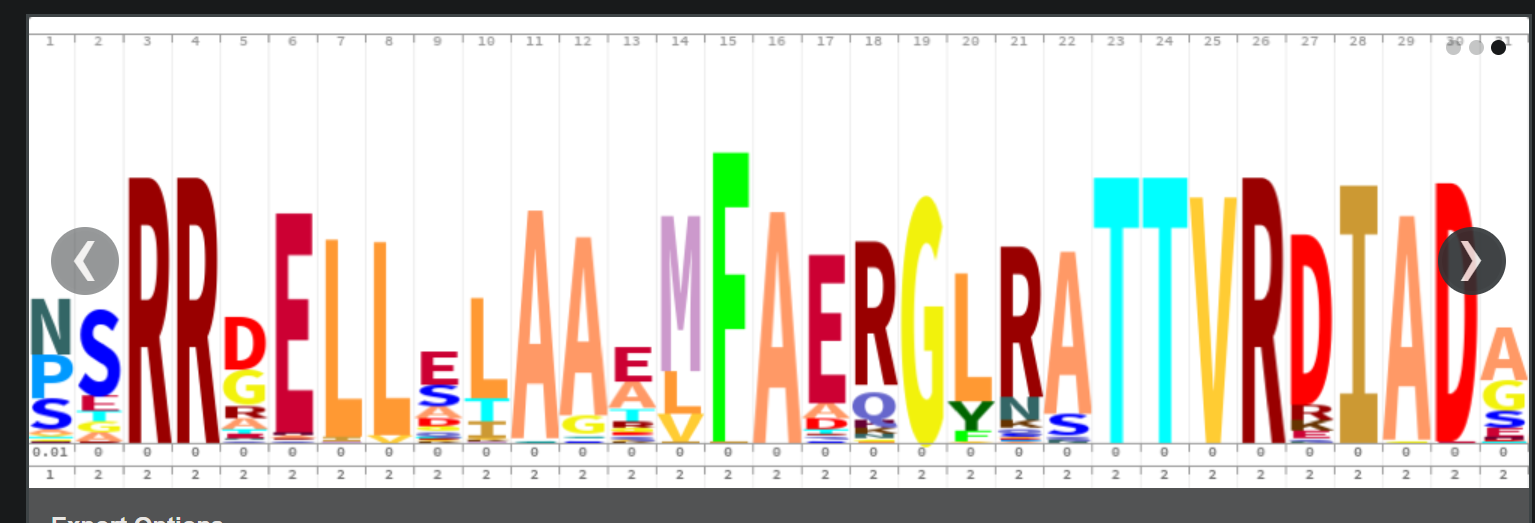
4: Sequence Logo Generation

• Upload your MSA file to Skylign.

A screenshot of a computer

AI-generated content may be incorrect.

• Generate a sequence logo to visualize conserved amino acids. • Include the logo image in your report



Include the logo image in your report and briefly explain: What do you observe? Are there highly conserved residues? Why might those regions be important?

**HBB Sequence Logo Analysis**

**What do you observe?**

* **The sequence logo shows variable and conserved positions in the HBB protein across six species (human, chimpanzee, mouse, cow, chicken, zebrafish).**
* **Some positions have tall stacks (e.g., 1, 3, 5, 18–20, 22–27), indicating strong conservation, while others show shorter stacks with diverse amino acids.**

**Are there highly conserved residues?**

**Yes:**

* **Position 1 (M) – start methionine.**
* **Positions 3 (H), 5 (T) – likely important for folding.**
* **Positions 18–20 (W-G-K) and 22–27 (N-V-D-E-G-G) – highly conserved and functionally significant.**

**Why might those regions be important?**

* **Heme binding: Conserved residues (e.g., W, G, K, N, D, E) interact with the heme group, essential for oxygen transport.**
* **Structural stability: N-terminal and conserved internal motifs maintain proper folding and protein function.**
* **These regions are under strong evolutionary pressure due to their critical role in hemoglobin's biological function.**

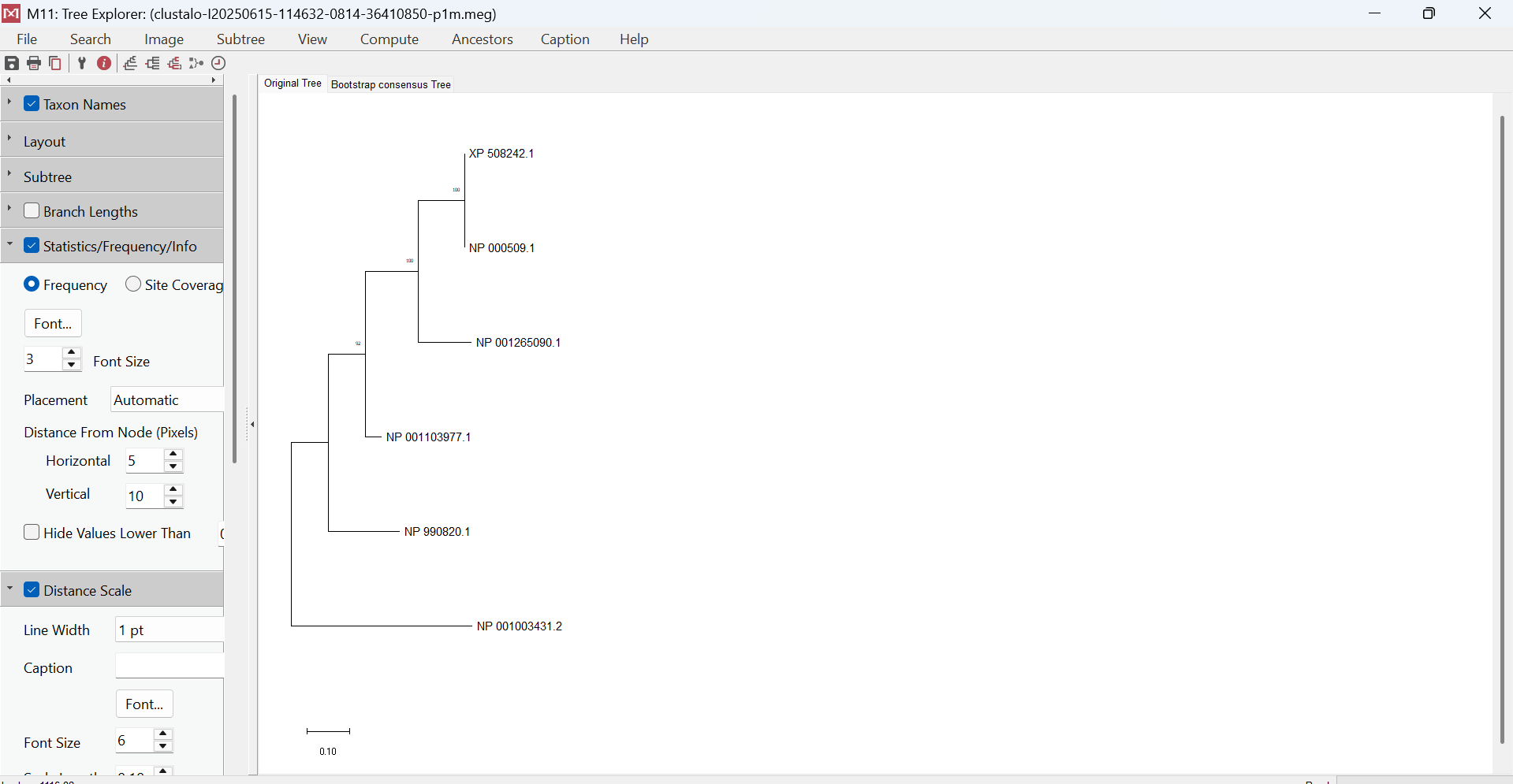
**Conclusion**

**The sequence logo clearly identifies functionally critical residues that are highly conserved across diverse species, emphasizing their role in heme binding, protein folding, tetramer formation, and stability. These regions are under strong purifying selection, as mutations would compromise oxygen transport — a vital physiological function.**

5: Phylogenetic Tree Construction

• Use your MSA to generate a phylogenetic tree using MEGA X

. • Include a screenshot of the tree.



Briefly explain: Which species are most closely related based on HBB? Does this tree match what you expect evolutionary

The phylogenetic tree illustrates the evolutionary relationships among HBB (hemoglobin subunit beta) protein sequences from six different species. **Homo sapiens (NP\_000509.1)** and **Pan troglodytes (NP\_001136144.1)** are shown to be the most closely related, which is expected given their recent common ancestry as primates. They form a well-supported clade with a bootstrap value of 100, indicating high confidence in their evolutionary relationship. **Bos taurus (cow)** clusters next within the tree, reflecting its position as a more distantly related mammal. **Mus musculus (mouse)** also groups within the mammalian clade but diverges earlier, showing greater evolutionary distance. In contrast, **Gallus gallus (chicken)** and **Danio rerio (zebrafish)** are positioned further from the mammalian sequences, highlighting their divergence as non-mammalian vertebrates. Overall, the tree topology aligns well with established vertebrate evolutionary history, with primates clustering closely and non-mammalian species branching off earlier.