Tracking the Evolution of the Hemoglobin Beta Gene Across Species

Sami H. Saliby

Lebanese American University (LAU)

GitHub

Introduction

The Hemoglobin Beta (HBB) gene plays a central role in oxygen transport in vertebrates and has long served as a model for studying molecular evolution. In this study, I investigated the conservation of the HBB gene across selected species to infer evolutionary relationships and levels of sequence divergence.

Methodology

Sequence Retrieval

To retrieve the human *HBB* gene sequence, I utilized the Entrez Programming Utilities (E-utilities) via Biopython. The following query was submitted to the NCBI Nucleotide database through a command-line script:

Homo sapiens [Organism] AND HBB [Gene]

From the search results, I extracted the accession number corresponding to the complete DNA sequence of the human hemoglobin subunit beta gene. Using this accession number, the full nucleotide sequence was downloaded in FASTA format for subsequent analysis.

data > Humans_gene > 1394355623.fasta

```
>NG_059281.1 Homo sapiens hemoglobin subunit beta (HBB), RefSeqGene (LRG_1232) on chromosome 11
AACGAATGAGTAAATGAGTAAATGAAGGAATGATTATTCCTTGCTTTAGAACTTCTGGAA
TTAGAGGACAATATTAATAATACCATCGCACAGTGTTTCTTTGTTGTTAATGCTACAACA
TACAAAGAGGAAGCATGCAGTAAACAACCGAACAGTTATTTCCTTTCTGATCATAGGAGT
AATATTTTTTTCCTTGAGCACCATTTTTGCCATAGGTAAAATTAGAAGGATTTTTAGAAC
TTTCTCAGTTGTATACATTTTTAAAAATCTGTATTATATGCATGTTGATTAATTTTAAAC
TTACTTGAATACCTAAACAGAATCTGTTGTTTCCTTGTGTTTGAAAGTGCTTTCACAGTA
ACTCTGTCTGTACTGCCAGAATATACTGACAATGTGTTATAGTTAACTGTTTTGATCACA
ACATTTTGAATTGACTGGCAGCAGAAGCTCTTTTATATCCATGTGTTTTCCTTAAGTCAT
TATACATAGTAGGCACTGAGAACTCTTTATATCTGAATAAGATATTTAGGAACCACTGGT
TTACATATCAGAAGCAGAGCTACTCAGGGCATTTTGGGGAAGATCACTTTCACATTCCTG
AGCATAGGGAAGTTCTCATAAGAGTAAGATATTAAAAGGAGATACTTGTGTGGTATTCGA
AAGACAGTAAGAGAGATTGTAGACCTTATGATCTTGATAGGGAAAACAAACTACATTCCT
```

Figure 1: Partial FASTA sequence of the Homo sapiens HBB gene retrieved using Biopython and Entrez API.

Homology Search and Species Selection

To explore the evolutionary conservation of this gene, I employed the Basic Local Alignment Search Tool for nucleotides (BLASTn) via the NCBI web interface. The retrieved human HBB gene sequence was used as the query against the Nucleotide collection database. Sequences from five primate species were selected

The species selected were:

- *Aotus azarae* (Azara's night monkey)
- *Lagothrix lagotricha* (Brown woolly monkey)
- *Cebus apella* (Tufted capuchin)
- *Saimiri sciureus* (Common squirrel monkey)
- *Gorilla gorilla* (Western gorilla)

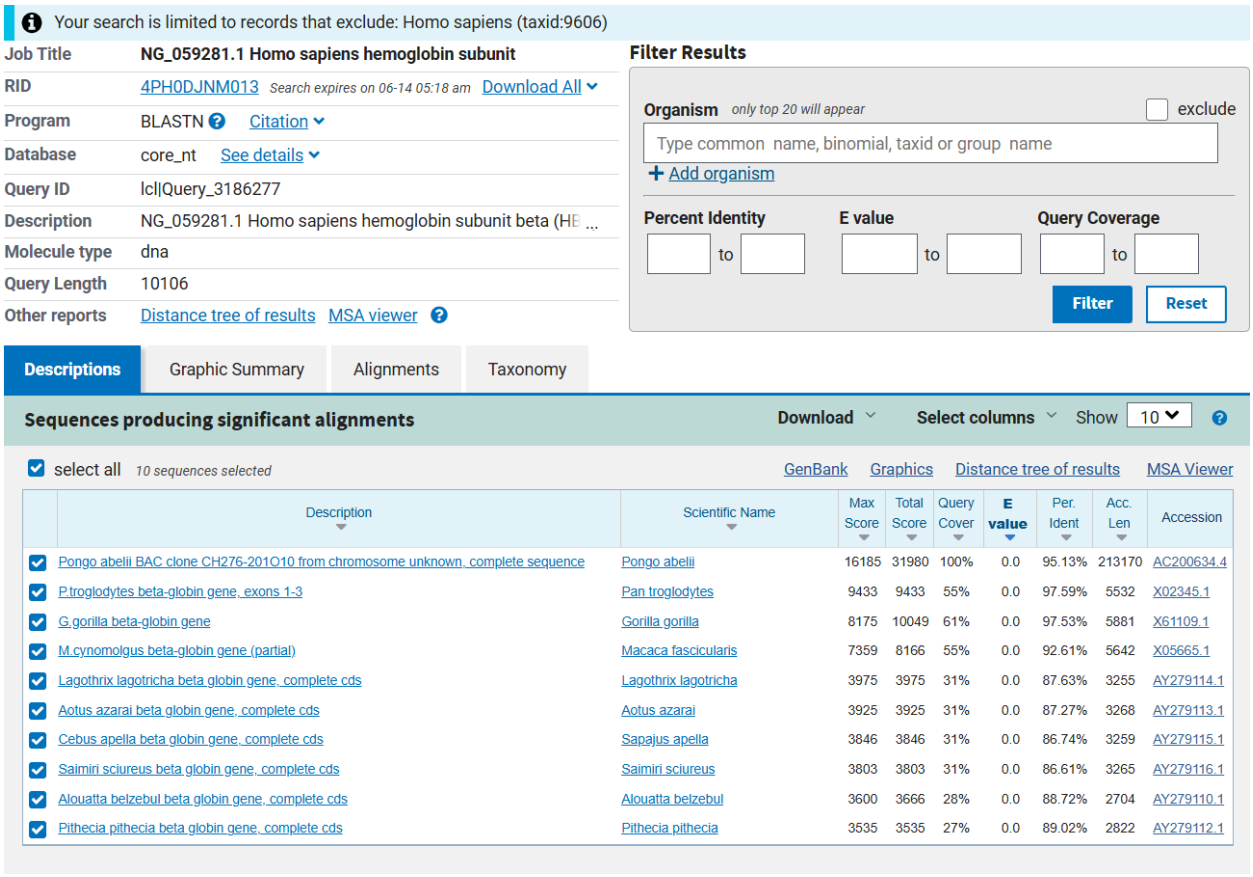

Figure 2: NCBI BLAST result using Homo Sapiens Hemoglobin Gene sequence as the query

Table 1: Table showing species name, accession ID of sequence and Identity Percentage with Human Hemoglobin Beta gene

Species Name	Accession Number	% Identity with Human HBB
A. Azarai	AY279113.1	87.27
L. lagotricha	AY279114.1	87.63
C. Apella	AY279115.1	86.74
S. sciureus	AY279116.1	86.61
G. gorilla	X61109.1	97.53

Pairwise Sequence Alignment Analysis

To quantify sequence divergence, pairwise global alignments were conducted using EMBOSS Needle between the *Homo sapiens* HBB gene and orthologous sequences from *Gorilla gorilla* and *Saimiri sciureus*. The alignment statistics are summarized below:

1. **Human vs. Gorilla (G. gorilla)**

- **Identity:** 5445 / 10447 (52.1%)
- **Similarity:** 52.1%
- **Gap Openings:** 4907 / 10447 (47.0%)
- **Alignment Score:** 26581.0

2. **Human vs. Squirrel Monkey (S. sciureus)**

- **Identity:** 2851 / 10190 (28.0%)
- **Similarity:** 28.0%
- **Gap Openings:** 7009 / 10190 (68.8%)
- **Alignment Score:** 12353.0

The alignment between *Homo sapiens* and *Gorilla gorilla* exhibits a substantially higher level of nucleotide identity (52.1%) and a correspondingly higher alignment score (26581.0), reflecting the close evolutionary relationship between the two species. The relatively lower proportion of gaps (47.0%) indicates that the gene structure has been more conserved, with fewer indels interrupting synteny. This is consistent with the known divergence time of humans and gorillas, estimated at approximately 8–10 million years ago.

In contrast, the alignment with *Saimiri sciureus*, a New World monkey, shows a considerably lower identity (28.0%) and a much higher proportion of gaps (68.8%), suggesting extensive structural divergence and sequence decay. The lower alignment score (12353.0) further supports the greater evolutionary distance, which aligns with the earlier divergence of New World monkeys from the human lineage over 35 million years ago.

Together, these comparisons reinforce the hypothesis that the HBB gene sequence has undergone progressive divergence across primate evolution, with greater conservation observed in more closely related species.

Multiple Sequence Alignment

Using MUSCLE from my Ubuntu terminal, I performed a multiple sequence alignment (MSA) of all 6 sequences.

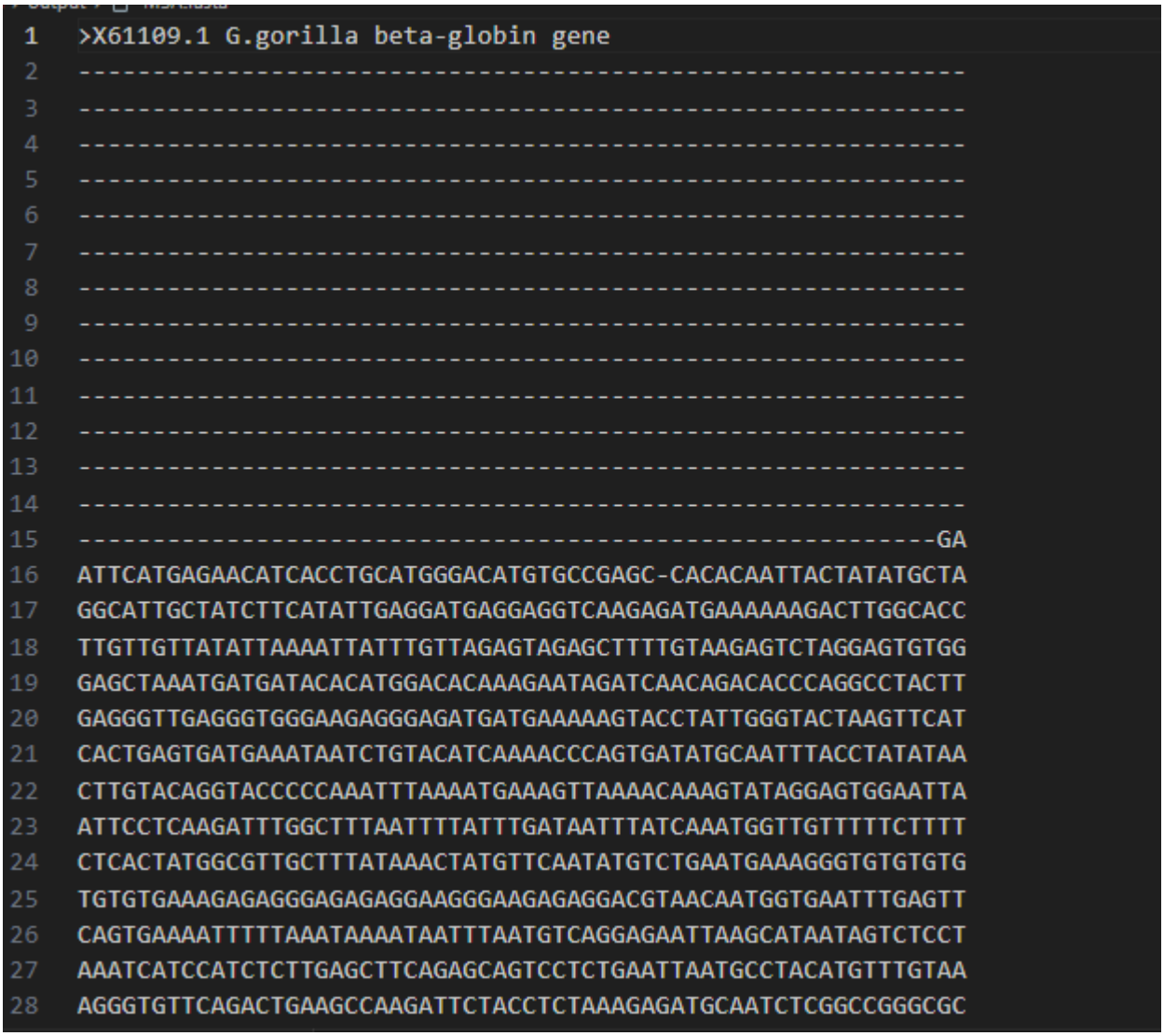

Figure 3: Snippet of the Multiple Sequence Alignment

An alignment logo was created using Skylign (default color scheme) which revealed several conserved domains. Even amongst non-conserved regions, the variability was very minimal, indicating conservation of genetic makeup of the HBB gene. This result is expected because the function of HBB protein is unchanged between organisms, its function is highly specific and shared by these organisms. Since structure determines function and function is conserved, it is expected that structure will also be conserved. Hemoglobin binds and releases oxygen in a very

precise way, even small changes to the beta-globin structure can drastically reduce oxygen delivery. This functional importance creates purifying or negative selection, which eliminates harmful mutations. Moreover, regions like the heme-binding site, subunit interaction surfaces, and structural motifs must stay the same. Lastly, the beta-globin gene belongs to a gene family that evolved from a common ancestor making them orthologs.

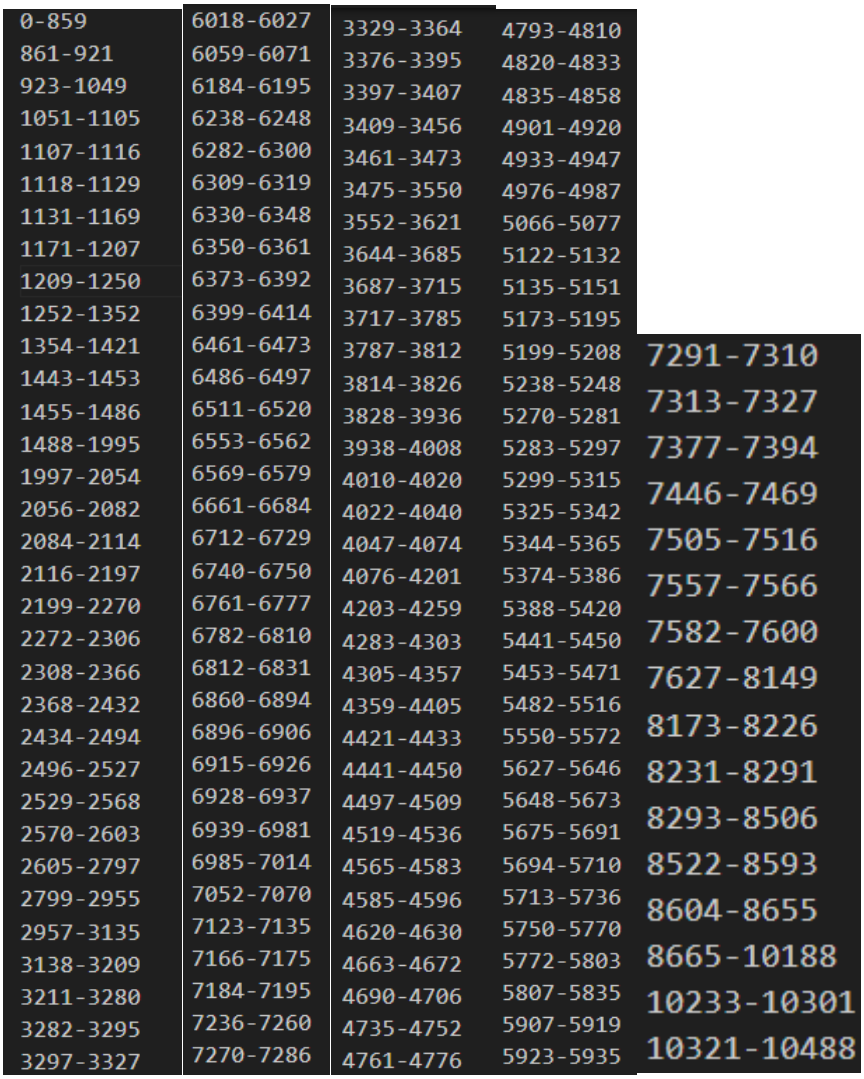

Figure 4: List of conserved domains from MSA

Phylogenetic Tree Construction

Using Mega12, a phylogenetic tree was created. “The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branch corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analytical procedure encompassed 6 coding nucleotide sequences using 1st, 2nd, 3rd, and non-coding positions. The pairwise deletion option was applied to all ambiguous positions for each sequence pair resulting in a final data set comprising 10,489 positions. Evolutionary analyses were conducted in MEGA12 utilizing up to 8 parallel computing threads.”

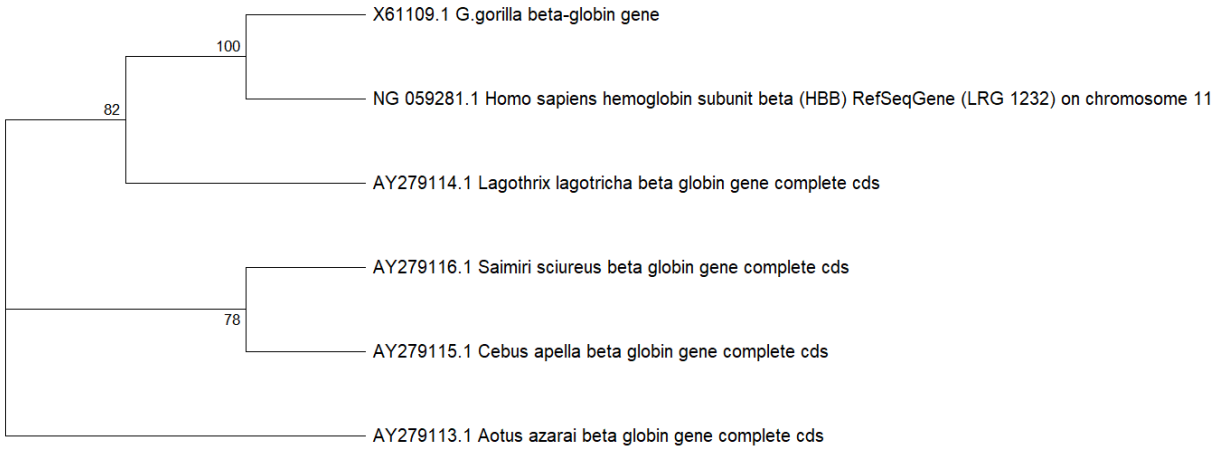

Figure 5: Phylogenetic Tree created using Mega12

This phylogenetic tree reflects expected evolutionary relationships among primates, with humans and gorillas clustering closely together, consistent with their shared recent common ancestor. New

World monkeys, including Saimiri, Cebus, and Aotus, form a distinct clade, supporting their divergence from Old World primates. Although Lagothrix (a New World monkey) appears closer to the human-gorilla group than anticipated, this may result from conserved sequences in the HBB gene due to functional constraints.

Conclusion:

The human HBB gene sequence was retrieved from NCBI's database using Entrez. Then, using NCBI online BLAST tool, similar sequences were retrieved, and a pairwise alignment was performed between the most and least similar sequences to the human sequence using EMBOSS. Next, a multiple sequence alignment was performed using MUSCLE from the terminal, and a logo was created using Skylign. Analysis of the MSA shows several conserved domains and overall similarities between the sequences reflecting the negative evolution and close evolutionary distances between the chosen species. Lastly, a phylogenetic tree was created using MEGA12 establishing primate evolutionary patterns while highlighting gene-specific variations.