

HBB Evolution

## **Tracking the Evolution of the Hemoglobin Beta Gene Across Species**

Sami H. Saliby

Lebanese American University (LAU)

[GitHub](#)

# HBB Evolution

## 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.

## HBB Evolution

```
data > Humans_gene > 1394355623.fasta
1  >NG_059281.1 Homo sapiens hemoglobin subunit beta (HBB), RefSeqGene (LRG_1232) on chromosome 11
2  AACGAATGAGTAAATGAGTAAATGAAGGAATGATTATTCCTTGCTTTAGAACTTCTGGAA
3  TTAGAGGACAATATTAATAATACCATCGCACAGTGTTCCTTGTTGTTAATGCTACAACA
4  TACAAAGAGGAAGCATGCAGTAAACAACCGAACAGTTATTCCTTCTGATCATAGGAGT
5  AATATTTTTTTCCTTGAGCACCATTTTGCCATAGGTAAAATTAGAAGGATTTTAGAAC
6  TTTCTCAGTTGTATACATTTTAAAAATCTGTATTATATGCATGTTGATTAATTTAAAC
7  TTACTTGAATACCTAAACAGAATCTGTTGTTTCCTTGTTGTTGAAAGTGCTTTCACAGTA
8  ACTCTGCTGTACTGCCAGAATATACTGACAATGTGTTATAGTTAACTGTTTTGATCACA
9  ACATTTTGAATTGACTGGCAGCAGAAGCTCTTTTATATCCATGTGTTTCCTTAAGTCAT
10 TATACATAGTAGGCACTGAGAACTTTTATATCTGAATAAGATATTAGGAACCACTGGT
11 TTACATATCAGAAGCAGAGCTACTCAGGCATTTTGGGGAAGATCACTTTCACATTCCTG
12 AGCATAGGGAAGTTCTCATAAGAGTAAGATATTAAGGAGATACTTGTGTGGTATTCGA
13 AAGACAGTAAGAGAGATTGTAGACCTTATGATCTTGATAGGGAACAACTACATTCCT
```

*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)

## HBB Evolution

**i** Your search is limited to records that exclude: Homo sapiens (taxid:9606)

Job Title	NG_059281.1 Homo sapiens hemoglobin subunit
RID	4PHODJNM013 <small>Search expires on 06-14 05:18 am</small> <a href="#">Download All</a> ▼
Program	BLASTN <a href="#">Citation</a> ▼
Database	core_nt <a href="#">See details</a> ▼
Query ID	lcl Query_3186277
Description	NG_059281.1 Homo sapiens hemoglobin subunit beta (HE ...
Molecule type	dna
Query Length	10106
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> <a href="#">?</a>

**Filter Results**

**Organism** only top 20 will appear ☐ exclude  
  
[+ Add organism](#)

**Percent Identity**  to   
**E value**  to   
**Query Coverage**  to

**Descriptions** | Graphic Summary | Alignments | Taxonomy

**Sequences producing significant alignments** [Download](#) ▼ [Select columns](#) ▼ Show  [?](#)

☒ select all 10 sequences selected
[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-201O10 from chromosome unknown, complete sequence</a>	<a href="#">Pongo abelii</a>	16185	31980	100%	0.0	95.13%	213170	<a href="#">AC200634.4</a>
<input checked="" type="checkbox"/>	<a href="#">P.troglodytes beta-globin gene, exons 1-3</a>	<a href="#">Pan troglodytes</a>	9433	9433	55%	0.0	97.59%	5532	<a href="#">X02345.1</a>
<input checked="" type="checkbox"/>	<a href="#">G.gorilla beta-globin gene</a>	<a href="#">Gorilla gorilla</a>	8175	10049	61%	0.0	97.53%	5881	<a href="#">X61109.1</a>
<input checked="" type="checkbox"/>	<a href="#">M.cynomolgus beta-globin gene (partial)</a>	<a href="#">Macaca fascicularis</a>	7359	8166	55%	0.0	92.61%	5642	<a href="#">X05665.1</a>
<input checked="" type="checkbox"/>	<a href="#">Lagothrix lagotricha beta globin gene, complete cds</a>	<a href="#">Lagothrix lagotricha</a>	3975	3975	31%	0.0	87.63%	3255	<a href="#">AY279114.1</a>
<input checked="" type="checkbox"/>	<a href="#">Aotus azarai beta globin gene, complete cds</a>	<a href="#">Aotus azarai</a>	3925	3925	31%	0.0	87.27%	3268	<a href="#">AY279113.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cebus apella beta globin gene, complete cds</a>	<a href="#">Sapajus apella</a>	3846	3846	31%	0.0	86.74%	3259	<a href="#">AY279115.1</a>
<input checked="" type="checkbox"/>	<a href="#">Saimiri sciureus beta globin gene, complete cds</a>	<a href="#">Saimiri sciureus</a>	3803	3803	31%	0.0	86.61%	3265	<a href="#">AY279116.1</a>
<input checked="" type="checkbox"/>	<a href="#">Alouatta belzebul beta globin gene, complete cds</a>	<a href="#">Alouatta belzebul</a>	3600	3666	28%	0.0	88.72%	2704	<a href="#">AY279110.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pithecia pithecia beta globin gene, complete cds</a>	<a href="#">Pithecia pithecia</a>	3535	3535	27%	0.0	89.02%	2822	<a href="#">AY279112.1</a>

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.

## HBB Evolution

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.

## HBB Evolution

```
1 >X61109.1 G.gorilla beta-globin gene
2 -----
3 -----
4 -----
5 -----
6 -----
7 -----
8 -----
9 -----
10 -----
11 -----
12 -----
13 -----
14 -----
15 -----GA
16 ATTCATGAGAACATCACCTGCATGGGACATGTGCCGAGC-CACACAATTACTATATGCTA
17 GGCATTGCTATCTTCATATTGAGGATGAGGAGGTCAAGAGATGAAAAAGACTTGGCACC
18 TTGTTGTTATATTTAAATTTTGTAGAGTAGAGCTTTGTAAGAGTCTAGGAGTGTGG
19 GAGCTAAATGATGATACACATGGACACAAAGAATAGATCAACAGACACCCAGGCCTACTT
20 GAGGGTTGAGGGTGGGAAGAGGGAGATGATGAAAAAGTACCTATTGGGTACTAAGTTCAT
21 CACTGAGTGATGAAATAATCTGTACATCAAACCCAGTGATATGCAATTTACCTATATAA
22 CTTGTACAGGTACCCCAAATTTAAATGAAAGTTAAACAAAGTATAGGAGTGGAATTA
23 ATTCCTCAAGATTTGGCTTTAATTTTATTTGATAATTTATCAAATGGTTGTTTTCTTTT
24 CTCCTATGGCGTTGCTTTATAAACTATGTTCAATATGTCTGAATGAAAGGGTGTGTGTG
25 TGTGTGAAAGAGAGGGGAGAGAGGAAGGGAAGAGAGGACGTAACAATGGTGAATTTGAGTT
26 CAGTGAAATTTTTAAATAAAATAATTTAATGTCAGGAGAATTAAGCATAATAGTCTCCT
27 AAATCATCCATCTCTTGAGCTTCAGAGCAGTCCTCTGAATTAATGCCTACATGTTTGTA
28 AGGGTGTTTCAGACTGAAGCCAAGATTCTACCTCTAAAGAGATGCAATCTCGGCCGGGCGC
```

*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

## HBB Evolution

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.

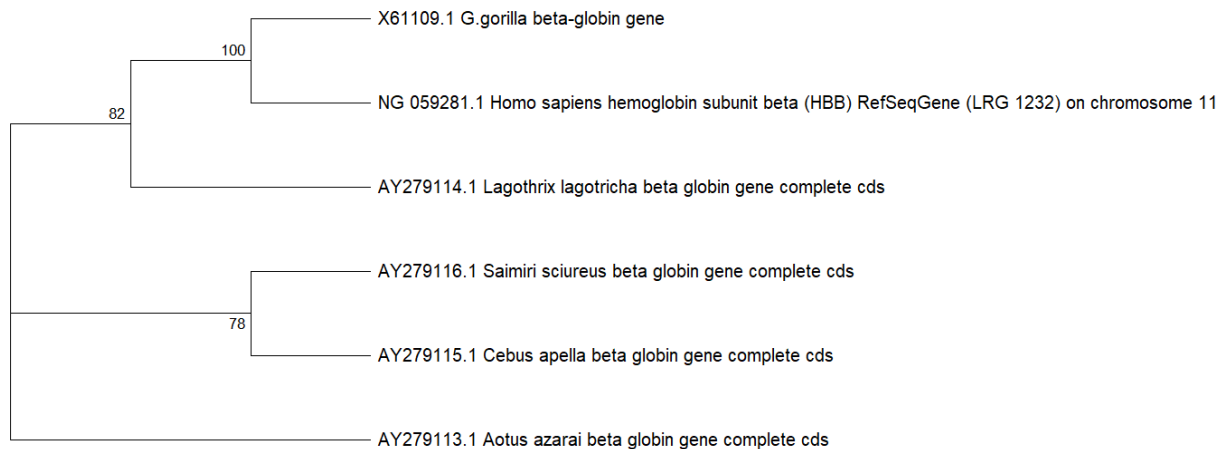
0-859	6018-6027	3329-3364	4793-4810	
861-921	6059-6071	3376-3395	4820-4833	
923-1049	6184-6195	3397-3407	4835-4858	
1051-1105	6238-6248	3409-3456	4901-4920	
1107-1116	6282-6300	3461-3473	4933-4947	
1118-1129	6309-6319	3475-3550	4976-4987	
1131-1169	6330-6348	3552-3621	5066-5077	
1171-1207	6350-6361	3644-3685	5122-5132	
1209-1250	6373-6392	3687-3715	5135-5151	
1252-1352	6399-6414	3717-3785	5173-5195	
1354-1421	6461-6473	3787-3812	5199-5208	7291-7310
1443-1453	6486-6497	3814-3826	5238-5248	7313-7327
1455-1486	6511-6520	3828-3936	5270-5281	7377-7394
1488-1995	6553-6562	3938-4008	5283-5297	7446-7469
1997-2054	6569-6579	4010-4020	5299-5315	7505-7516
2056-2082	6661-6684	4022-4040	5325-5342	7557-7566
2084-2114	6712-6729	4047-4074	5344-5365	7582-7600
2116-2197	6740-6750	4076-4201	5374-5386	7627-8149
2199-2270	6761-6777	4203-4259	5388-5420	8173-8226
2272-2306	6782-6810	4283-4303	5441-5450	8231-8291
2308-2366	6812-6831	4305-4357	5453-5471	8293-8506
2368-2432	6860-6894	4359-4405	5482-5516	8522-8593
2434-2494	6896-6906	4421-4433	5550-5572	8604-8655
2496-2527	6915-6926	4441-4450	5627-5646	8665-10188
2529-2568	6928-6937	4497-4509	5648-5673	10233-10301
2570-2603	6939-6981	4519-4536	5675-5691	10321-10488
2605-2797	6985-7014	4565-4583	5694-5710	
2799-2955	7052-7070	4585-4596	5713-5736	
2957-3135	7123-7135	4620-4630	5750-5770	
3138-3209	7166-7175	4663-4672	5772-5803	
3211-3280	7184-7195	4690-4706	5807-5835	
3282-3295	7236-7260	4735-4752	5907-5919	
3297-3327	7270-7286	4761-4776	5923-5935	

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

## HBB Evolution

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.