Int. J. Agriworld, Vol. 7[1] January 2026

Online ISSN: 2582-7537
Journal's URL:http://www.svpss.in/ijaw
DOI: https://doi.org/10.51470/IJAW.2026.07.01.57
Received 19.12.2025 Revised 12.01.2026 Accepted 22.01.2026
ORIGINAL ARTICLE **OPEN ACCESS**

Comparative Bioinformatics and Phylogenetic Analysis of the Hemoglobin Beta (HBB) Gene Among Selected Mammalian Species

Urvashi Vashishtha, Malti* and Lavi Sharma
Department of Zoology, Janta Vedic College, Baruat (Baghpat) -250611, U.P.
***Corresponding E. mail:** maltitomar@gmail.com

Abstract

The hemoglobin beta (HBB) gene is one of the most extensively studied globin genes because of its critical role in oxygen transport and its high degree of evolutionary conservation among vertebrates. The present study investigated sequence conservation and evolutionary relationships of the HBB coding sequence among selected mammalian species using computational bioinformatics approaches. Complete coding sequences (CDS) of the HBB gene from *Homo sapiens*, *Pan troglodytes*, *Chlorocebus sabaeus*, Mus musculus, and Rattus norvegicus were retrieved from the National Center for Biotechnology Information (NCBI) database. Multiple sequence alignment was performed using Clustal Omega and MUSCLE integrated within the Phylogeny.fr platform. Comparative analysis demonstrated strong sequence conservation among primate species, particularly between *Homo sapiens* and *Pan troglodytes*, which showed 99.77% sequence identity. Rodent species exhibited comparatively lower similarity with primate sequences, ranging from approximately 82% to 83%, while *Mus musculus* and *Rattus norvegicus* displayed 92.12% identity. Phylogenetic reconstruction using Maximum Likelihood and Neighbor-Joining methods produced consistent clustering patterns that separated primates and rodents into distinct monophyletic lineages. The findings support established mammalian evolutionary relationships and demonstrate the usefulness of the HBB gene as a molecular marker in comparative genomics and evolutionary biology studies.

Keywords: *HBB gene, hemoglobin beta, comparative genomics, phylogenetics, MUSCLE alignment, BLASTn, molecular evolution, bioinformatics*.

Introduction

Bioinformatics has become an essential component of modern biological research because it enables efficient analysis of large-scale molecular datasets generated through genome sequencing and comparative genomics studies. The integration of computational methods with molecular biology has significantly improved the understanding of gene structure, sequence conservation, and evolutionary relationships among organisms. Bioinformatics integrates computational methods with biological data analysis (Lesk, 2019). Comparative genomics relies heavily on sequence analysis tools and databases (Mount, 2004). Sequence alignment tools and phylogenetic methods are now widely applied in comparative genomics, molecular evolution, medicine, and functional genomics.

Hemoglobin is an oxygen-transporting metalloprotein present in vertebrate red blood cells. Adult hemoglobin consists of two alpha-globin and two beta-globin polypeptide chains that together facilitate oxygen transport from the lungs to peripheral tissues (Perutz, 1970). The beta-globin chain is encoded by the hemoglobin beta (HBB) gene located on chromosome 11 in humans (Huisman, 1993). Because of its essential physiological function, the HBB gene has remained highly conserved throughout mammalian evolution (Hardison, 2012).

Mutations within the HBB gene are associated with several inherited hemoglobinopathies, including sickle cell disease and beta-thalassemia (Weatherall & Clegg, 2001). Consequently, the HBB gene has been extensively studied in molecular genetics, evolutionary biology, and medical research. Comparative analysis of HBB sequences among mammals provides valuable information regarding evolutionary conservation, species divergence, and molecular adaptation (Graur & Li, 2000).

Sequence similarity analysis using the Basic Local Alignment Search Tool (BLAST) allows rapid identification of homologous regions among nucleotide sequences (Altschul *et al.,* 1990). Multiple sequence alignment methods such as MUSCLE further enable identification of conserved and variable nucleotide regions across species (Edgar, 2004). Phylogenetic trees based on aligned sequences provides insight into evolutionary relationships and divergence patterns among organisms (Felsenstein, 2004).

The present study aimed to investigate the evolutionary conservation and phylogenetic relationships of the HBB coding sequence among selected mammalian species using bioinformatics tools including BLASTn, MUSCLE alignment, and phylogenetic analysis.

Materials and Methods

Sequence Retrieval

Complete coding sequences (CDS) of the HBB gene were retrieved from the National Center for Biotechnology Information (NCBI) nucleotide database. Only coding regions were selected to ensure uniform sequence comparison and minimize alignment bias caused by untranslated regions (UTRs), EST fragments, or partial transcript sequences.

The analyzed species and accession numbers included:

Table 1: Accession Numbers of HBB Coding Sequences Retrieved from NCBI for Selected Mammalian Species

Species	Accession Number
Homo sapiens	NM_000518.5
Pan troglodytes	XM_508242.5
Chlorocebus sabaeus	NM_001329918.1
Mus musculus	NM_008220.5
Rattus norvegicus	NM_033234.1

Coding sequences were downloaded in FASTA format using the "Send to → Coding sequences (CDS) → FASTA" option available through the NCBI interface.

Multiple Sequence Alignment

Sequence alignment methods are fundamental for identifying homologous regions among biological sequences (Needleman & Wunsch, 1970). Local alignment approaches further improved the identification of conserved sequence regions among related organisms (Smith & Waterman, 1981). Multiple sequence alignment of HBB coding sequences was performed using the MUSCLE algorithm integrated within the Clustal Omega and Phylogeny.fr platforms. Default alignment parameters were used during analysis. The generated alignment was examined to identify conserved and variable nucleotide regions among the selected mammalian species.

The alignment refinement step was further processed using Gblocks integrated within the Phylogeny.fr workflow to eliminate poorly aligned regions and improve phylogenetic reliability.

Percent Identity Matrix Analysis

Pairwise nucleotide sequence similarity among species was evaluated using the Clustal Omega percent identity matrix. Sequence identity values were interpreted as indicators of evolutionary conservation and divergence.

Phylogenetic Analysis

Phylogenetic reconstruction was performed using the Phylogeny.fr "One Click" workflow (Dereeper *et al.,* 2008). The pipeline integrated MUSCLE for sequence alignment, Gblocks for alignment refinement, PhyML for Maximum Likelihood tree construction, and TreeDyn for phylogenetic tree visualization (Guindon & Gascuel, 2003).

Both Maximum Likelihood and Neighbor-Joining phylograms were analyzed to evaluate evolutionary relationships among the selected mammalian species. Branch lengths represented nucleotide substitutions per site, and bootstrap or approximate likelihood-ratio test (aLRT) support values were used to assess tree reliability (Felsenstein, 1985).

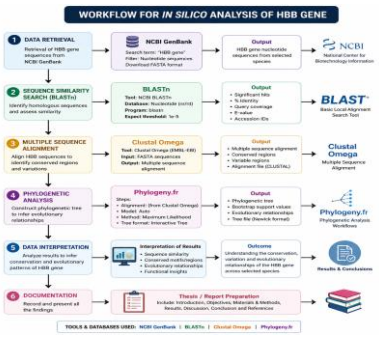

Fig. 1. Workflow diagram illustrating the in-silico bioinformatics analysis of the Hemoglobin Beta (HBB) gene using NCBI GenBank, BLASTn, Clustal Omega, and Phylogeny.fr for sequence retrieval, alignment, phylogenetic reconstruction, and evolutionary interpretation.

Results

Phenotypic evaluation of fragrance

The BLASTn searches were conducted using the NCBI BLAST web interface to compare HBB coding sequences among selected mammals (Johnson *et al.,* 2008). The BLASTn analysis demonstrated a high degree of sequence conservation among primate HBB coding sequences. *Homo sapiens* and *Pan troglodytes* showed extremely high similarity with 99.77% sequence identity and complete query coverage, indicating a close evolutionary relationship. *Chlorocebu ssabaeus* also exhibited strong sequence conservation with human HBB sequences, showing 96.85% identity.

Comparatively lower sequence similarity was observed between primates and rodent species. *Mu smusculus* and *Rattus norvegicus* demonstrated approximately 82–83% sequence identity with primate HBB sequences, reflecting greater evolutionary divergence. Overall, the results supported established evolutionary relationships among primates and rodents and demonstrated the functional conservation of the HBB gene across diverse mammalian species.

Table 2: Comparative Sequence Identity Relative to Human HBB Coding Sequence

S. No.	Species	Percent Identity (%)	Query Coverage (%)
1	*Homo sapiens*	100.00	100
2	*Pan troglodytes*	99.77	100
3	*Chlorocebus sabaeus*	97.08	100
4	*Mus musculus*	82.88	100
5	*Rattus norvegicus*	82.88	100

Multiple Sequence Alignment

Multiple sequence alignment of HBB coding sequences revealed a high degree of nucleotide conservation among primate species (Figure 2). Conserved nucleotide motifs were observed throughout the coding region, particularly within functionally important regions associated with hemoglobin structure and oxygen transport (Dickerson & Geis, 1983). The MUSCLE-based multiple sequence alignment revealed extensive nucleotide conservation among the analyzed mammalian HBB coding sequences. The alignment demonstrated particularly high sequence similarity among primate species, especially between *Homo sapiens* and *Pan troglodytes*, which showed 99.77% sequence identity. Conserved nucleotide regions were observed throughout most of the coding sequence, indicating strong functional preservation of the beta-globin gene. Comparatively greater nucleotide variation was detected within rodent sequences, including *Mus musculus* and *Rattus norvegicus*, which showed approximately 82–83% similarity with primate sequences.

Comparatively greater nucleotide variation was observed in rodent sequences, including Mus musculus and Rattus norvegicus. Despite this divergence, several conserved regions remained preserved across all analyzed mammals, indicating strong functional constraints acting on the HBB gene.

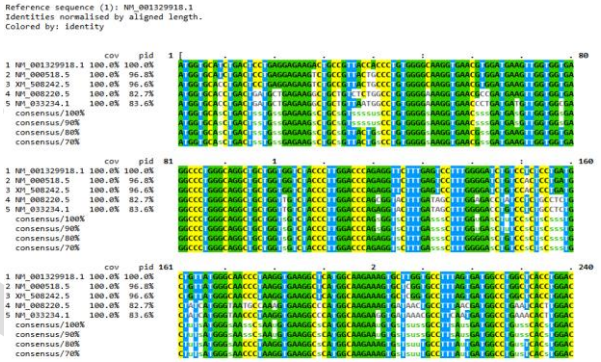

Fig. 2: Multiple sequence alignment of HBB coding sequences among selected mammalian species generated using the MUSCLE algorithm

The Gblocks refinement analysis retained 444 conserved nucleotide positions from the original multiple sequence alignment for phylogenetic reconstruction (Figure 3). Removal of poorly aligned and highly variable regions improved the overall alignment quality and reduced phylogenetic noise. The retained conserved positions represented evolutionarily informative regions suitable for accurate comparative and phylogenetic analysis of mammalian HBB coding sequences.

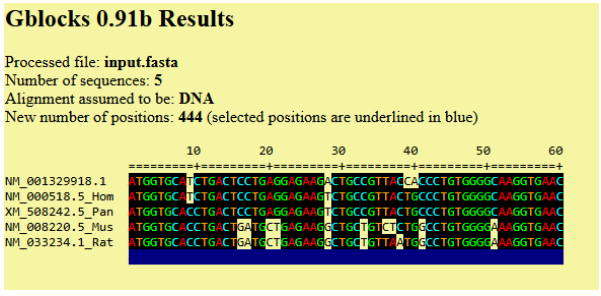

Fig. 3: Gblocks refinement of aligned HBB coding sequences showing selection of conserved nucleotide regions used for phylogenetic reconstruction

Percent Identity Matrix

The percent identity matrix demonstrated substantial sequence conservation among primate species. The highest similarity was observed between *Homo sapiens* and *Pan troglodytes* (99.77%), indicating extremely close evolutionary relatedness. *Chlorocebus sabaeus* also exhibited strong sequence similarity with both human and chimpanzee sequences, showing identity values of 96.85% and 96.62%, respectively.

Comparatively lower similarity values were observed between primates and rodents. Sequence identity values between primates and rodents ranged approximately from 82% to 83%, reflecting greater evolutionary divergence. *Mus musculus* and *Rattus norvegicus* showed 92.12% identity, supporting their close evolutionary relationship within the rodent lineage.

Table 3. Percent Identity Matrix of HBB Coding Sequences Among Selected Mammalian Species

Species	*Chlorocebus sabaeus*	*Homo sapiens*	*Pan troglodytes*	*Mus musculus*	*Rattus norvegicus*
Chlorocebus sabaeus	**100.00**	96.85	96.62	82.66	83.56
Homo sapiens	96.85	**100.00**	99.77	82.88	82.88
Pan troglodytes	96.62	99.77	**100.00**	83.11	83.11
Mus musculus	82.66	82.88	83.11	**100.00**	92.12
Rattus norvegicus	83.56	82.88	83.11	92.12	**100.00**

Interpretation of Percent Identity Matrix

The high sequence identity observed between *Homo sapiens* and *Pan troglodytes* reflected their close evolutionary relationship within the primate lineage. Similarly, the comparatively high similarity between *Mus musculus* and *Rattus norvegicus* supported their close phylogenetic association within rodents. Lower sequence identity values observed between primates and rodents indicated greater evolutionary divergence resulting from gradual accumulation of nucleotide substitutions over evolutionary time.

Phylogenetic Analysis

Phylogenetic reconstruction using Maximum Likelihood and Neighbor-Joining methods produced consistent evolutionary clustering patterns. The generated phylograms separated the analyzed mammals into two major monophyletic lineages corresponding to primates and rodents.

The primate lineage consisted of *Homo sapiens*, *Pan troglodytes*, and *Chlorocebus sabaeus*. Human and chimpanzee sequences formed the closest cluster, reflecting minimal nucleotide divergence and strong sequence conservation. *Chlorocebus sabaeus* branched within the primate lineage while remaining evolutionarily distinct from the human-chimpanzee cluster.

The rodent lineage consisted of *Mus musculus* and *Rattus norvegicus*, which formed a separate well-supported clade. Longer branch lengths separating rodents from primates indicated greater evolutionary divergence.

The scale bar within the phylograms represented nucleotide substitutions per site rather than percentage divergence. High branch support values demonstrated strong confidence in the inferred phylogenetic relationships.

The Maximum Likelihood phylogram demonstrated evolutionary relationships consistent with established mammalian taxonomy (Figure 3). Primate species, including *Homo sapiens, Pan troglodytes*, and *Chlorocebus sabaeus*, formed a distinct monophyletic clade, whereas Mus musculus and Rattus norvegicus formed a separate rodent lineage. The close clustering of human and chimpanzee sequences reflected minimal nucleotide divergence and strong evolutionary conservation of the HBB coding sequence. Branch support values approaching 1.0 indicated strong confidence in the inferred phylogenetic relationships among the analyzed species. The scale bar represented approximately 0.07 nucleotide substitutions per site, indicating evolutionary divergence among the analyzed mammalian species.

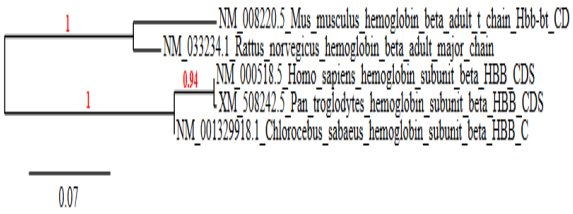

Fig. 4: Maximum Likelihood phylogram of HBB coding sequences among selected mammalian species generated using Phylogeny.fr.

The Neighbor-Joining phylogram separated the analyzed mammalian species into two major evolutionary lineages corresponding to primates and rodents (Figure 5). *Homo sapiens* and *Pan troglodytes* formed the closest cluster within the primate lineage, consistent with their high sequence identity value of 99.77%. *Chlorocebus sabaeus* clustered within the primate-associated branch while remaining evolutionarily distinct from the human–chimpanzee subgroup. *Mus musculus* and *Rattus norvegicus* formed a separate rodent clade, reflecting their close evolutionary relationship and comparatively high sequence similarity. The branch lengths represented nucleotide substitutions per site and indicated greater evolutionary divergence between primates and rodents.

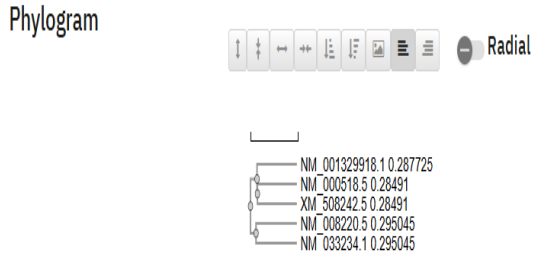

Fig. 5: Neighbor-Joining phylogram illustrating evolutionary relationships among selected mammalian HBB coding sequences

Discussion

The present investigation demonstrated strong evolutionary conservation of the HBB coding sequence among mammals (Li, 1997). Comparative sequence analysis revealed extremely high similarity among primates, particularly between *Homo sapiens* and *Pan troglodytes*, which exhibited 99.77% sequence identity. This observation is consistent with the close evolutionary relationship between humans and chimpanzees established through previous molecular and genomic studies.

The HBB gene encodes the beta-globin subunit of hemoglobin, which performs an essential role in oxygen transport and cellular respiration. Genes associated with critical physiological functions are generally subjected to strong purifying selection, thereby preserving important nucleotide regions over evolutionary time (Nei & Kumar, 2000). The high degree of sequence conservation observed among primates therefore reflects the functional importance of the beta-globin protein.

Although rodents displayed lower sequence similarity with primates, the HBB coding sequence remained substantially conserved across all analyzed mammals. Sequence identity values ranging from approximately 82% to 83% between primates and rodents suggest progressive accumulation of nucleotide substitutions during mammalian evolution while maintaining the overall functional integrity of the gene (Kimura, 1983).

The close clustering of *Mus musculus* and *Rattus norvegicus* within the phylogenetic tree was supported by their comparatively high sequence identity (92.12%). This finding is consistent with established rodent evolutionary relationships and demonstrates that HBB sequences can effectively resolve mammalian phylogeny (Avise, 2004).

The phylogenetic topology generated during the present analysis corresponded closely with accepted mammalian evolutionary relationships. Both Maximum Likelihood and Neighbor-Joining phylograms separated primates and rodents into distinct monophyletic clades, confirming the reliability of the alignment and sequence dataset used in the study.

The use of complete coding sequences rather than partial transcripts or EST fragments improved the biological consistency of the analysis and reduced alignment bias. Standardization of sequence regions was essential for generating accurate percent identity values and phylogenetic inference.

Overall, the findings highlight the usefulness of bioinformatics approaches in comparative genomics and molecular evolutionary research. Publicly accessible platforms such as NCBI, Clustal Omega, and Phylogeny.fr provide reliable tools for sequence analysis, alignment, and phylogenetic reconstruction without requiring advanced computational infrastructure.

Conclusion

The present study demonstrated substantial evolutionary conservation of the HBB coding sequence among selected mammalian species using comparative bioinformatics approaches. High sequence similarity observed among primates, particularly between *Homo sapiens* and *Pan troglodytes*, reflected their close evolutionary relationship, whereas rodent species exhibited comparatively greater divergence. Multiple sequence alignment and phylogenetic reconstruction consistently supported established mammalian evolutionary relationships and confirmed the suitability of the HBB gene as a molecular marker in comparative genomics and molecular evolution studies.

REFERENCE

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Avise, J.C. (2004). Molecular Markers, Natural History and Evolution.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36 (Web Server issue), W465–W469. https://doi.org/10.1093/nar/gkn180

Dickerson, R.E., & Geis, I. (1983). Hemoglobin: Structure, Function, Evolution, and Pathology.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*.

Felsenstein, J. (2004). Inferring phylogenies. *Sinauer Associates*.

Graur, D., & Li, W.H. (2000). Fundamentals of Molecular Evolution.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*.

Hardison, R.C. (2012). Evolution of hemoglobin and its genes. Cold Spring Harbor Perspectives in Medicine, 2(12), a011627. https://doi.org/10.1101/cshperspect.a011627

Huisman, T.H.J. (1993). The structure and function of normal and abnormal haemoglobins. *Baillière's Clinical Haematology*.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, 36(Web Server issue), W5–W9. https://doi.org/10.1093/nar/gkn201

Kimura, M. (1983). The Neutral Theory of Molecular Evolution.

Lesk, A.M. (2019). Introduction to bioinformatics (5th ed.). *Oxford University Press*.

Li, W.H. (1997). Molecular Evolution. *Sinauer Associates.*

Mount, D.W. (2004). Bioinformatics: Sequence and genome analysis (2nd ed.). *Cold Spring Harbor Laboratory Press*.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453. https://doi.org/10.1016/0022-2836(70)90057-4

Nei, M., & Kumar, S. (2000). Molecular Evolution and Phylogenetics. *Oxford University Press*.

Perutz, M.F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature*, 228, 726–739.

Smith, T.F., & Waterman, M.S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1), 195–197. https://doi.org/10.1016/0022-2836(81)90087-5

Weatherall, D.J., & Clegg, J.B. (2001). The thalassaemia syndromes (4th ed.). *Blackwell Science*.

CITATION OF THIS ARTICLE

Vashishtha, U., Malti and Sharma, L. (2026). Comparative Bioinformatics and Phylogenetic Analysis of the Hemoglobin Beta (HBB) Gene Among Selected Mammalian Species, *Int. J. Agriworld*, 7 [1]: 57-62.

