

Project Title: Bioinformatics

Task 1: DNA Sequence Analysis of the Homo sapiens HBB Gene

Domain: Bioinformatics

Intern Name: Muhammad Daniyal Hameed **Date:** December 9, 2025

1. Introduction

Bioinformatics combines biology, computer science, and information engineering to analyze and interpret biological data. A fundamental tool in this field is the Basic Local Alignment Search Tool (BLAST), which detects regions of local similarity between sequences. This program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

For this project, the gene selected for analysis is **Hemoglobin subunit beta (HBB)**. This gene provides instructions for making a protein called beta-globin, a subunit of hemoglobin, which is the protein in red blood cells that carries oxygen. Mutations in the HBB gene are clinically significant, as they are responsible for genetic disorders such as sickle cell anemia and beta-thalassemia.

The objective of this analysis is to retrieve the DNA sequence for the human HBB gene, perform a BLAST analysis to identify homologous sequences in other species, and interpret the evolutionary relationships based on sequence identity and alignment scores.

2. Methodology

The analysis was conducted using the following workflow:

1. **Sequence Retrieval:** The reference mRNA sequence for *Homo sapiens* hemoglobin subunit beta (HBB) was retrieved from the National Center for Biotechnology Information (NCBI) database. The accession number used was **NM_000518.5**.
2. **BLAST Configuration:** The NCBI Nucleotide BLAST (blastn) tool was utilized. The standard nucleotide collection (nr/nt) database was selected to ensure a comprehensive search across all available organisms.

3. **Algorithm Selection:** The "Somewhat similar sequences" (blastn) or "Highly similar sequences" (megablast) algorithm was used to optimize for finding closely related homologs.
4. **Analysis:** The resulting alignments were analyzed for Percent Identity, Query Coverage, and E-values to determine statistical significance.

3. Results

The BLAST search generated a list of sequences producing significant alignments. The results indicate a high degree of conservation for the HBB gene among primates.

3.1. Overview of Alignments The search returned multiple hits with significant alignment scores. The top hit was the query sequence itself (*Homo sapiens* HBB mRNA), serving as a positive control for the search accuracy.

National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite » results for RID-KGE92749014

HomeRecent ResultsSaved StrategiesHelp

Edit SearchSave SearchSearch Summary▼

How to read this report?BLAST Help VideosBack to Traditional Results Page

Job TitleNM_000518.5 Homo sapiens hemoglobin subunit...
RIDKGE92749014Search expires on 12-10 21:58 pmDownload All▼
ProgramBLASTN Citation▼
Database nt See details▼
Query IDIcIQuery_2478559
DescriptionNM_000518.5 Homo sapiens hemoglobin subunit beta (H...
Molecule typedna
Query Length628
Other reportsDistance tree of resultsMSA viewer?

Filter Results
Organism only top 20 will appear☐ exclude
Type common name, binomial, taxid or group name
+ Add organism
Percent IdentityE valueQuery Coverage
to to to
FilterReset

DescriptionsGraphic SummaryAlignmentsTaxonomy

Sequences producing significant alignmentsDownload▼Select columns▼Show100▼

☒ select all 100 sequences selected

GenBankGraphicsDistance tree of resultsMSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens hemoglobin subunit beta (HBB) mRNA	Homo sapiens	1160	1160	100%	0.0	100.00%	628	NM_000518.5
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes hemoglobin subunit beta (HBB) mRNA	Pan troglodytes	1155	1155	100%	0.0	99.84%	639	XM_508242.5
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus hemoglobin subunit beta (LOC100976465) mRNA	Pan paniscus	1149	1149	100%	0.0	99.68%	643	XM_003819029.5
<input checked="" type="checkbox"/> Homo sapiens hemoglobin_beta_mRNA (cDNA clone MGC:14540 IMAGE:4292125) complete cds	Homo sapiens	1142	1142	100%	0.0	99.52%	658	BC007075.1
<input checked="" type="checkbox"/> Human messenger RNA for beta-globin	Homo sapiens	1140	1140	100%	0.0	99.52%	626	V00497.1
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla gorilla hemoglobin subunit beta (LOC101126932) mRNA	Gorilla gorilla go...	1133	1133	100%	0.0	99.20%	638	XM_019036164.3
<input checked="" type="checkbox"/> Homo sapiens hemoglobin beta mRNA, complete cds	Homo sapiens	1133	1133	100%	0.0	99.36%	647	AY509193.1
<input checked="" type="checkbox"/> PREDICTED: Pongo pygmaeus hemoglobin subunit beta (HBB) mRNA	Pongo pygmaeus	1105	1105	100%	0.0	98.41%	639	XM_054440830.2
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii hemoglobin subunit beta (HBB) mRNA	Pongo abelii	1105	1105	100%	0.0	98.41%	658	XM_002822127.6
<input checked="" type="checkbox"/> PREDICTED: Symphalangus syndactylus hemoglobin subunit beta (HBB) mRNA	Symphalangus...	1099	1099	100%	0.0	98.25%	638	XM_055282707.1
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys hemoglobin subunit beta (HBB) mRNA	Nomascus leuc...	1088	1088	100%	0.0	97.93%	753	XM_004090649.3
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch hemoglobin subunit beta (HBB) mRNA	Hylobates moloch	1077	1077	100%	0.0	97.61%	638	XM_032166808.2

Figure 1: Table of sequences producing significant alignments from NCBI BLAST.

3.2. Quantitative Analysis Based on the data shown in Figure 1, the following observations were made:

- **Top Hit (Human):** The *Homo sapiens* entry (NM_000518.5) showed a **Max Score** of approximately 1155, **100% Query Coverage**, and **100% Identity**. This confirms the input sequence was correctly identified.
- **Primate Homologs:** The search identified homologous sequences in several non-human primates:
 - *Pan troglodytes* (Chimpanzee): 99-100% Identity with an E-value of 0.0.
 - *Pan paniscus* (Bonobo): High identity (>99%) and E-value of 0.0.
 - *Gorilla gorilla* (Western Gorilla): High identity (>99%) and E-value of 0.0.
 - *Pongo abelii* (Sumatran Orangutan): Slightly lower but still highly significant identity scores.
- **E-Value Significance:** All top results displayed an **E-value of 0.0**. In bioinformatics, an E-value (Expect value) represents the number of hits one can expect to see by chance. An E-value of 0.0 indicates that the similarity is statistically significant and not due to random chance.

4. Discussion and Interpretation

The results of this BLAST analysis demonstrate strong evolutionary conservation of the beta-globin gene.

4.1. Evolutionary Proximity The extremely high percent identity (99%–100%) between the human HBB sequence and those of the Chimpanzee (*Pan troglodytes*) and Bonobo (*Pan paniscus*) supports the scientific consensus that these great apes are our closest living relatives. As we move further away taxonomically for example, to Old World monkeys like *Macaca mulatta* (Rhesus macaque) we observe a slight decrease in the "Max Score" and "Percent Identity," reflecting the accumulation of genetic divergence over millions of years.

4.2. Functional Conservation The HBB gene is critical for survival, as hemoglobin function is essential for oxygen transport. Genes that perform such vital "housekeeping" functions often show high conservation because mutations are frequently deleterious and selected against. The fact that this gene remains so similar across different primate species suggests that the protein structure and function have been maintained by strong evolutionary pressure.

5. Conclusion

This project successfully utilized NCBI BLAST to analyze the *Homo sapiens* HBB gene. The results confirmed the identity of the input sequence and revealed significant homology with other primate species. The analysis illustrates the utility of bioinformatics tools in comparative genomics, allowing researchers to trace evolutionary history and identify conserved biological functions through sequence alignment.

6. References

1. National Center for Biotechnology Information (NCBI). (n.d.). *Basic Local Alignment Search Tool (BLAST)*. Retrieved from <https://blast.ncbi.nlm.nih.gov/>
2. NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 46(D1), D8–D13.