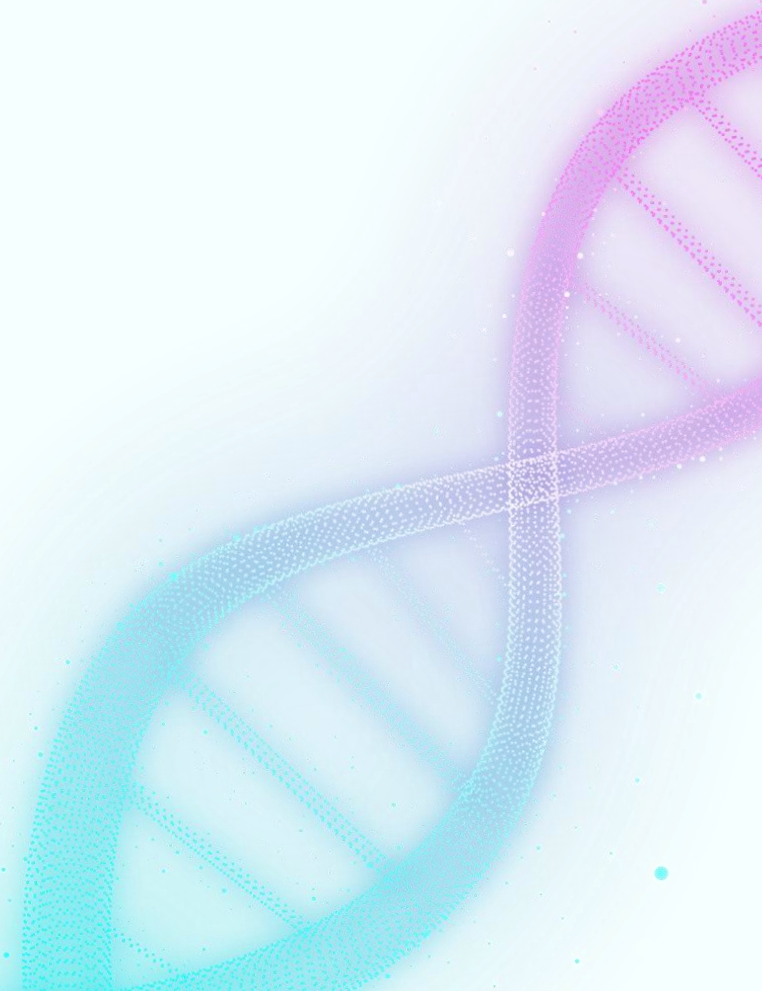


# BioinformHer Mini Project – Module 2

**Name:** Marco Hany Maher

**Date:** 11/6/2025

**Email:** marcohany1122@gmail.com



# TABLE OF CONTENTS

**01**

**Sequence Retrieval  
& BLAST Search**

**02**

**Pairwise Sequence  
Alignment**

**03**

**Multiple Sequence  
Alignment (MSA)**

**04**

**Sequence Logo  
Generation**

**05**

**Phylogenetic Tree  
Construction**

**06**

**CONCLUSION**



**01**

# **Sequence Retrieval & BLAST Search**



# Retrieve the human HBB gene (nucleotide or protein) from NCBI

I retrieved the **Hemoglobin subunit beta (HBB)** protein sequence for *Homo sapiens* from the NCBI Protein database. Represents a significant advancement in the field of medicine.

- Accession Number: NP\_000509.1
- Protein Length: 147 amino acids
- Source: Homo sapiens (human)

FASTA File   sequence.fasta



An official website of the United States government [Here's how you know](#)

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Protein   [Advanced](#) [Help](#)

GenPept

### hemoglobin subunit beta [Homo sapiens]

NCBI Reference Sequence: NP\_000509.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS	NP_000509	147 aa	linear	PRI 27-APR-2025
DEFINITION	hemoglobin subunit beta [Homo sapiens].			
ACCESSION	NP_000509			
VERSION	NP_000509.1			
DBSOURCE	REFSEQ: accession <a href="#">NM_000518.5</a>			
KEYWORDS	RefSeq; MANE Select.			
SOURCE	Homo sapiens (human)			
ORGANISM	<a href="#">Homo sapiens</a>			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

**Analyze this sequence**

- [Run BLAST](#)
- [Identify Conserved Domains](#)
- [Highlight Sequence Features](#)
- [Find in this Sequence](#)
- [Show in Genome Data Viewer](#)

**Protein 3D Structure**

[Cryo-EM structure of human](#)

# Used BLAST to find homologs in five other species

We retrieved the human Hemoglobin Beta (HBB) protein sequence from NCBI and used BLASTp to find homologs in five other species. The table below shows the most similar sequences identified, including % identity and accession numbers.

Scientific Name	Per. ident	Accession
Pan troglodytes	100%	XP_508242.1
Bos taurus	84.72%	NP_776342.1
Mus musculus	80.27%	NP_001265090.1
Gallus gallus	69.39%	NP_990820.1
Danio rerio	50%	NP_001003431.2

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

NP\_000509.1

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ ClusteredNR database **RECOMMENDED** [Learn more...](#)

Non-redundant protein sequences (nr) [?](#)

Organism Optional

☐ exclude [Add organism](#)

☐ exclude

☐ exclude

☐ exclude

☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search database nr using Blastp (protein-protein BLAST)

☒ Show results in a new window



**02**

# **Pairwise Sequence Alignment**



# Pairwise Sequence with EMBOSS Needle

## Closely Related Species (Chimpanzee)

- Identity: 147/147 (100.0%)
- Similarity: 147/147 (100.0%)
- Number of gaps: 0/147 (0.0%)

## Distantly Related Species (Zebrafish)

- Identity: 73/147 (49.7%)
- Similarity: 105/147 (71.4%)
- Number of gaps: 0/147 (0.0%)

So, The alignment between Human and Chimpanzee HBB shows high identity and similarity, reflecting their close evolutionary relationship. In contrast, the Human-Zebrafish alignment has lower identity and similarity, indicating greater divergence. Thus, the chimpanzee HBB sequence is more conserved compared to the zebrafish.

## Closely Related Species

VS

## Distantly Related Species

```
#=====
#
# Aligned_sequences: 2
# 1: NP_000509.1
# 2: XP_508242.1
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:   147/147 (100.0%)
# Similarity: 147/147 (100.0%)
# Gaps:       0/147 ( 0.0%)
# Score: 780.0
#
#
#=====
```

NP_000509.1	1	MVHLTPEEKSAVTALWGKNVDEVGGEALGRLLVVYPWTQRFFESFGDLS	50
XP_508242.1	1	MVHLTPEEKSAVTALWGKNVDEVGGEALGRLLVVYPWTQRFFESFGDLS	50
NP_000509.1	51	TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTATLSELCOKLHVD	100
XP_508242.1	51	TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTATLSELCOKLHVD	100
NP_000509.1	101	PENFRLLGNVLVCVLAHHFGKEFTPPVQAAQYQKVAVGAVANALAHKYH	147
XP_508242.1	101	PENFRLLGNVLVCVLAHHFGKEFTPPVQAAQYQKVAVGAVANALAHKYH	147

```
#=====
#
# Aligned_sequences: 2
# 1: NP_000509.1
# 2: NP_001003431.2
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:   73/147 (49.7%)
# Similarity: 105/147 (71.4%)
# Gaps:       0/147 ( 0.0%)
# Score: 408.0
#
#
#=====
```

NP_000509.1	1	MVHLTPEEKSAVTALWGKNVDEVGGEALGRLLVVYPWTQRFFESFGDLS	50
		... :.....: .:  : .:. .:     : .:	
NP_001003431.	1	MVQWSDSERKTIASVWSKINVDEIGPQTLARVLVYPWTQRYFAGFGDLS	50
NP_000509.1	51	TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTATLSELCOKLHVD	100
		...:     ..   .   .....: .:.  : : :	
NP_001003431.	51	CASAIMGNPKVSEHGKTVLKALEKAVKNVDDIKTTYAKLSQLHCEKLNVD	100
NP_000509.1	101	PENFRLLGNVLVCVLAHHFGKEFTPPVQAAQYQKVAVGAVANALAHKYH	147
		: : : .:. .:. .:. .:.  : : .:. .:. .:	
NP_001003431.	101	PDNFKLLADCLSIIVATNFGPAFNPSVQSTWQKLLSVVVAALTSRYF	147





**03**

## **Multiple Sequence Alignment (MSA)**



We performed a Multiple Sequence Alignment (MSA) using six HBB protein sequences from the following species: Human, Chimpanzee, Cow, Mouse, Chicken, and Zebrafish. The alignment was generated using Clustal Omega tool, The alignment shows several highly conserved regions across all species, conserved regions are colored in the Table below.

We performed a Multiple Sequence Alignment (MSA) using six HBB protein sequences from the following species: Human, Chimpanzee, Cow, Mouse, Chicken, and Zebrafish. The alignment was generated using Clustal Omega tool, The alignment shows several highly conserved regions across all species, conserved regions are colored in the Table below.

NP_001003431.2	PAFNPSVQSTWQKLLSVVVAALTSRYF	147
NP_990820.1	KDFTPECQAAWQKLVRVVAHALARKYH	147
NP_001265090.1	KDFTPAQAQAAFKVQVAGVATALAHKYH	147
NP_000509.1	KEFTPPVQAAYQKVQVAGVANALAHKYH	147
XP_508242.1	KEFTPPVQAAYQKVQVAGVANALAHKYH	147
NP_776342.1	KEFTPVLQADFQKVQVAGVANALAHRYH	145
	* * * * *	



## Conserved Regions



**04**

## **Sequence Logo Generation**



# Sequence Logo Generation

## 1. What do you observe?

The sequence logo shows the level of **amino acid conservation** across the aligned Hemoglobin Beta sequences from six species (human, chimpanzee, cow, mouse, chicken, zebrafish).

Each position in the alignment is represented by a **stack of letters** :

- The **height of the stack** shows how conserved that position is (measured in bits of information).
- The **size of each letter** shows the relative frequency of that amino acid at that position.
- The first few residues (e.g., **MVHLTPEEKS**) are **highly conserved**, showing tall, dominant letters.

## 2. Are there highly conserved residues?

Yes, several positions are highly conserved, especially near the beginning and middle of the sequence:

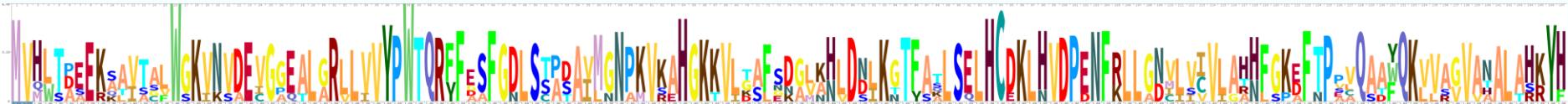
- **Positions** like M (start), V, H, L, T, P, E, K are nearly identical across all species.
- **Later**, residues such as Q, R, E, F, H, K, D also show high conservation.

## 3. Why might those regions be important?

Highly conserved regions are usually: **Functionally important** involved in oxygen binding or interaction with other subunits in the hemoglobin tetramer.

**Structurally important** necessary to maintain the correct folding and stability of the protein. For HBB, **this includes**:

- Heme-binding sites
- Alpha-beta chain interfaces
- Residues that maintain globin structure





**05**

# **Phylogenetic Tree Construction**



## 1. Which species are most closely related based on HBB?

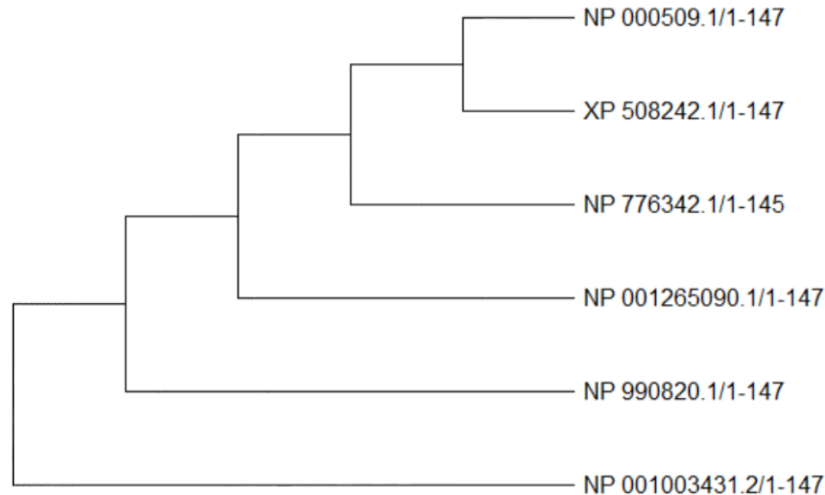
From the tree: **Human** (NP\_000509.1) is most closely related to **Pan troglodytes** (XP\_508242.1) — they are in the same and are closest on the tree.

Next closest is **Bos taurus** (NP\_776342.1), the cow.

Followed by **Mus musculus** (NP\_001265090.1), the mouse.

Then **Gallus gallus** (NP\_990820.1), the chicken.

Farthest from humans is **Danio rerio** (NP\_001003431.2), the zebrafish.



## 2. Does this tree match what you expect evolutionarily?

**Yes**, the tree is consistent with known evolutionary relationships:

- **Humans and chimpanzees** (Pan troglodytes) share a very recent common ancestor.
- **Cows** (Bos taurus) are also mammals and closer to humans
- **Mice** (Mus musculus) are still mammals but slightly more diverged from humans than cows in this case
- **Chickens** (Gallus gallus) are birds and more distantly related.
- **Zebrafish** (Danio rerio) are fish, representing an even more ancient divergence.

# CONCLUSIONS

## CONCLUSIONS

1

**HBB** is highly conserved among vertebrates, especially in functionally critical regions.

2

The highest sequence identity was observed between humans and chimpanzees, reflecting their close evolutionary relationship.

3

**Zebrafish** showed the **lowest similarity** and highlighting its early divergence from mammals.

4

**Multiple sequence alignment** revealed several **highly conserved amino acids**, likely essential for the structure and function of hemoglobin.

5

The **sequence logo** visualized these conserved sites, confirming functional importance.

6

The **phylogenetic tree** accurately reflected expected evolutionary relationships, with species grouped according to known evolutionary distances.