

# Tracking the Evolution of the Hemoglobin Beta (HBB) Gene Across Species

## SEQUENCE RETRIEVAL AND BLAST SEARCH

1. Retrieving the FASTA sequence of the human ( *Homo sapiens* ) haemoglobin beta (HBB) gene.

Once in the NCBI platform, the search engine was adjusted to select a Protein. In the search engine, 'HBB Homosapiens' was inputted and the first result displaying the HBB for *Homo sapiens* was chosen. The FASTA sequence of the gene was then highlighted and copied from the website.

ORGANISM	SCIENTIFIC NAME	ACCESSION NUMBER OF SELECTED HUMAN HBB GENE
HUMAN	<i>Homo sapiens</i>	CAG38767

2. Retrieving the FASTA sequence of the HBB gene of other species. The species of choice for this report are: Chicken (*Gallus gallus* ), Zebra fish ( *Danio rerio* ), House mouse ( *Mus musculus* ), Pigeon ( *Columba livia* ) and Chimpanzee ( *Pan troglodytes* )

ORGANISM	SCIENTIFIC NAME	ACCESSION NUMBER OF SELECTED GENE	% IDENTITY WITH HUMAN HBB
CHICKEN	<i>Gallus gallus</i>	NP_990820.1	102/147(69%)
ZEBRA FISH	<i>Danio rerio</i>	NP_001003431	73/146(50%)
HOUSE MOUSE	<i>Mus musculus</i>	NP_001265090	118/147(80%)
PIGEON	<i>Columba livia</i>	XP_005507431	102/147(69%)
CHIMPANZEE	<i>Pan troglodytes</i>	XP_508242	147/147(100%)

## PAIRWISE SEQUENCE ALIGNMENT

Chosen species from BLAST results : **Chicken (distantly related)** and **House mouse (closely related)**. Note: Both NCBI and EMBOSS Needle were used for the Pairwise Sequence Alignment, though the final screenshots were all taken from NCBI.

**METHOD:** For each of the animal species whose HBB genes were sequenced against the human HBB gene, the human gene was inputted in the Subject Sequence segment, while the animal HBB gene was inputted in the Query Sequence segment.

**KEY:** In NCBI: (+) indicates conserved regions, ( ) blank spaces indicate regions that are neither conserved nor identical.

The plus ( + ) sign indicates regions in the gene where there is a substitution of the amino acid in that region of the gene that is different from the amino acid in the query sequence yet with strongly similar properties.

The spaces indicates mismatches in that region implying evolutionary divergence ie the amino acids in these 22 regions are neither identical nor conserved and signifies regions within the HBB gene of chickens and humans that have evolved differently either due to function or evolutionary pressures.

ORGANISM	% IDENTITY	% SIMILARITY	NUMBER OF GAPS
<i>Gallus gallus</i>	69% (102/147)	121/147(82.3%)	0/147
<i>Mus musculus</i>	80%(118/147)	132/147 (89.3%)	0/147

1. *Gallus gallus* - There was a total of 22 spaces ( ‘ ’ ) observed in the pairwise alignment between HBB gene of humans vs the HBB gene of chickens. 19 conserved regions were also observed between the sequences.(pictured below)

Descriptions

Graphic Summary

Alignments

Dot Plot

Alignment view

Pairwise

Restore defaults

Download

1 sequences selected

Download

Graphics

Next

Previous

Descriptions

CAG38767.1 HBB [Homo sapiens]

Sequence ID: Query\_7009849 Length: 147 Number of Matches: 1

Range 1: 1 to 147

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
221 bits(564)	1e-80	Compositional matrix adjust.	102/147(69%)	121/147(82%)	0/147(0%)

Query 1

MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFASFGLSSPTAILG

MVH T EEK +T LWGKVVN E G EAL RLL+VYPWTQRF SFG+LS+P A++G

Sbjct 1

MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVPWTQRFESFGDLSTPDVVMG

Query 61

VRAHGKKVLTSGDAVNLDNIKNTFSQLSELHCDKLHVDPENFRLGDIILVLA

V+AHGKKVL +F D + +LDN+K TF+ LSELHCDKLHVDPENFRLG++L+ VLA

Sbjct 61

VKAHGKKVLGAFSDGLAHLNLIKGTATLSELHCDKLHVDPENFRLGNVLCVLAH

Query 121

KDFTPECQAAWQKLVRVVAHALARKYH 147

K+FTP QAA+QK+V VA+ALA KYH

Sbjct 121

KEFTPPVQAAYQKVVAGVANALAHKYH 147

2. *Mus musculus* - There was a total of 13 spaces and 14 conserved regions observed in the pairwise alignment. This implies higher similarity with the HBB gene of human beings. (pictured below)

Descriptions Graphic Summary **Alignments** Dot Plot

Alignment view Pairwise [Restore defaults](#) Download

1 sequences selected

[Download](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**CAG38767.1 HBB [Homo sapiens]**  
Sequence ID: Query\_6448085 Length: 147 Number of Matches: 1

Range 1: 1 to 147 [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
251 bits(641)	2e-92	Compositional matrix adjust.	118/147(80%)	132/147(89%)	0/147(0%)

Query	1	MVHLTDAEKA AVSCLWGK VNSDEVGGEALGRLLVVYPWTQRYFDSFGDLSASAIMG	
Sbjct	1	MVHLT EK+AV+ LWGKVN DEVGGEALGRLLVVYPWTQR+F+SFGDLS+ A+MG	
Query	61	VKAHGKKVITAFNDGLNHLDSLKGTFA SLSELHCDKLHVDPENFRLLGNMIVIVLGH	
Sbjct	61	VKAHGKKV+ AF+DGL HLD+LKGTF A+LSELHCDKLHVDPENFRLLGN++V VL H	
Query	121	KDFTPAAQAA FQKV VAGVATALAHKYH	147
Sbjct	121	KEFTPPVQAA YQKV VAGVANALAHKYH	147

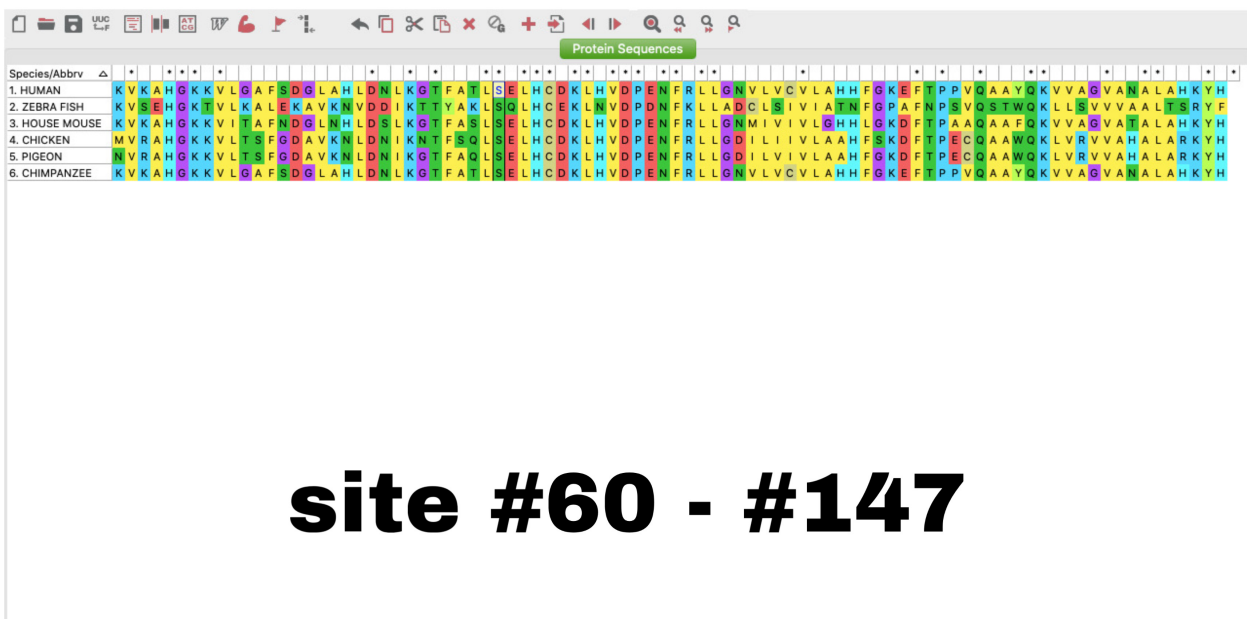
Based on the information stated above, it can be accurately concluded that the HBB gene of *Mus musculus* is more conserved with the HBB gene of *Homo sapiens* than the HBB gene of *Gallus gallus*.

## MULTIPLE SEQUENCE ALIGNMENT (MSA)

The Multiple Sequence Alignment was done using MUSCLE within the MEGA 11 Software. The Sequences were arranged in this order : Human, Zebrafish, House Mouse, Chicken, Pigeon and Chimpanzee.

Highly conserved regions observed among the six species are as follows: Sites (1, 2, 8, 16, 18, 20, 25, 29, 31, 33, 35, 36, 37, 38, 39, 40, 41, 43, 46, 47, 49, 50, 54, 57, 58, 61, 64, 65, 66, 68, 80, 83, 85, 89, 90, 92, 93, 94, 96, 97, 99, 100, 101, 103, 104, 106, 107, 114, 123, 125, 128, 132, 133, 138, 141, 142, 146) all have the same amino acid in the same position in all the HBB genes of the different species.

(Pictured below are screenshots showcasing the MSA from site #1 to site #89 and then site #60 to #site 147).

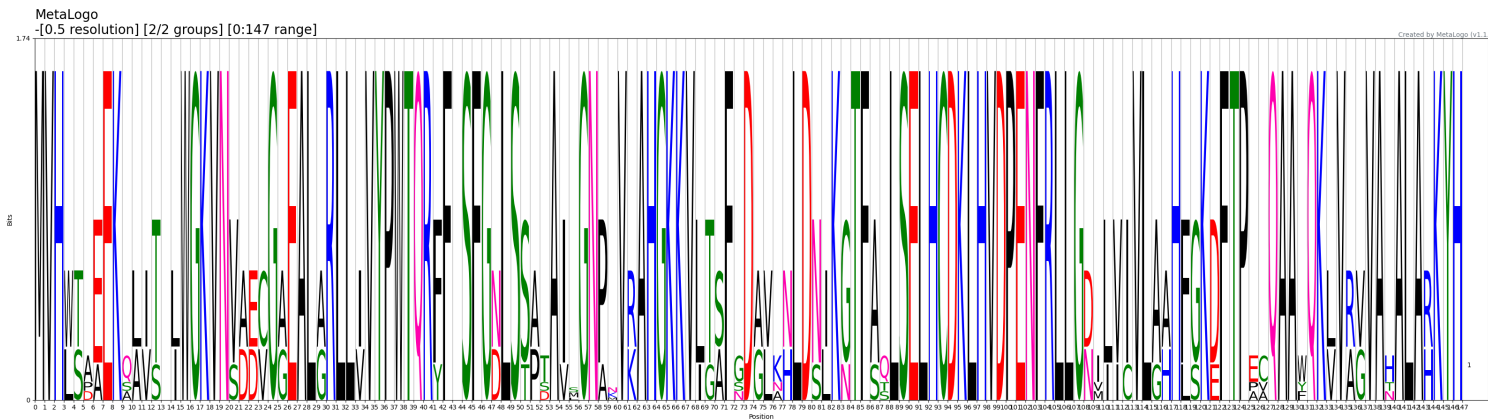


## SEQUENCE LOGO GENERATION

In order to generate the sequence logo, the web application 'MetaLogo' was utilised. This was used in stead of Skylign due to the fact that I kept encountering a proxy error message on both my phone and laptop.

**METHOD:** To use Skylign, the MSA results from MUSCLE were converted to FASTA format and then uploaded to the Skylign website. The minimum and maximum length were adjusted to 140 and

150 approximately in order to accommodate the length of all the sequences which is 147. (Pictured below is the image of the generated sequence logo).



## OBSERVATIONS

1. Presence of amino acid letters that are extremely long/ taller at specific positions.eg 1, 73-74 etc
2. Presence of amino acid letters that are shorter at certain positions.
3. Presence of amino acids that are stacked on top each other at certain positions eg 25-26, 109-110 etc

There are also highly conserved observed ie regions where a single amino acid letter stands out much taller than the rest at a particular position. eg Position 1- Methionine (M), Position 117/118 - Histidine (H), Position 99- Aspartic acid (D) etc

These regions may be important for:

- a. Structure of the gene
- b. Function of the gene eg Histidine is associated with the heme binding properties of haemoglobin which is crucial for oxygen transport .

## PHYLOGENETIC TREE CONSTRUCTION

MEGA 11 was used to generate the phylogenetic tree from the MSA.

- a. Based on the tree, the HBB gene of Chimpanzees and House Mouse share the closest evolutionary relationship, with the Human HBB gene and Chimpanzee HBB gene sharing the Most Recent Common Ancestor (MRCA).
- b. Yes and no. On one hand, I am unsurprised that the Human and Chimpanzee gene are highly similar (identical even), but on the other hand, I didn't expect there to be this much disparity in the HBB gene amongst species as I had always assumed it was the same in all blood.

(Pictured below is the phylogenetic tree)

