

BioinformHer Mini Project – Module Capstone

Title: Tracking the Evolution of the Hemoglobin Beta (HBB) Gene across Species

Project Objective: The skills learned in Module 2 was used to investigate the evolutionary conservation of the HBB Gene across six species.

The Hemoglobin Beta (HBB) gene plays a critical role in oxygen transport, making its evolutionary conservation a subject of significant interest. In this study, the HBB gene's nucleotide sequence was retrieved from humans (*Homo sapiens*) and compared with its counterparts in five other species: chimpanzee (*Pan troglodytes*), cow (*Bos taurus*), mouse (*Mus musculus*), chicken (*Gallus gallus*), and zebrafish (*Danio rerio*). The sequences were obtained from NCBI using BLAST, and their Gene IDs, accession numbers, and percent identities with the human HBB gene were documented.

S/N	Species Name	Gene ID	Accession Number	% Identity with Human HBB
1	Human [<i>Homo sapiens</i>]	3043	NC_000011.10	-
2	Chimpanzee [<i>Pan troglodytes</i>]	450978	NC_072407.2	98.2%
3	Cow [<i>Bos taurus</i>]	280813	NC_037342.1	63.5%
4	Mouse [<i>Mus musculus</i>]	15132	NC_000073.7	53.5%
5	Chicken [<i>Gallus gallus</i>]	396485	NC_052532.1	43.8%
6	Zebrafish [<i>Danio rerio</i>]	30216	NC_007114.7	27.8%

Pairwise Alignment Report

Human vs. Chimpanzee HBB Gene

```
#=====
#
# Aligned_sequences: 2
# 1: c5227071-5225464
# 2: c9358653-9357035
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1619
# Identity:   1590/1619 (98.2%)
# Similarity: 1590/1619 (98.2%)
# Gaps:       11/1619 ( 0.7%)
# Score: 7878.0
#
#
#=====

c5227071-5225      1  -----ACATTGCTTCTGACACAACTGTGTTCACTAGCAACCTC      39
                   |||
c9358653-9357      1  ATCTATTGCTTACATTGCTTCTGACACAACTGTGTTCACTAGCAACCTC      50

c5227071-5225     40  AAACAGACACCATGGTGATCTGACTCCTGAGGAGAAGTCTGCCGTTACT      89
                   |||
c9358653-9357     51  AAACAGACACCATGGTGACCTGACTCCTGAGGAGAAGTCTGCCGTTACT     100

c5227071-5225     90  GCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGG     139
                   |||
```

A pairwise alignment between the human and chimpanzee HBB genes, performed using EMBOSS Needle, revealed striking conservation. The alignment spanned 1,619 nucleotides, with 98.28% identity (1,590 identical bases) and only 0.78% gaps (11 gaps, all in the human sequence). The high alignment score of 7,878.0 underscores the close evolutionary relationship between humans and chimpanzees, reflecting strong functional constraints on the HBB gene due to its vital role in hemoglobin. The HBB gene is ****highly conserved**** between humans and chimpanzees, reflecting strong evolutionary constraint due to its critical role in hemoglobin. The minimal divergence (1.8%) aligns with the close phylogenetic relationship between the two species.

Human vs. Zebrafish HBB Gene

```
# Report file: stdout|
#####
#
#=====
#
# Aligned sequences: 2
# 1: c5227071-5225464
# 2: c55104355-55103441
# Matrix: EDNAFULL
# Gap penalty: 10.0
# Extend penalty: 0.5
#
# Length: 1830
# Identity:      508/1830 (27.8%)
# Similarity:    508/1830 (27.8%)
# Gaps:          1137/1830 (62.1%)
# Score: 1042.5
#
#=====
c5227071-5225      1  -----ACA-TTTGCTTCT-----
12
c55104355-551      1  CGTCATCTCAGCAAGATTATCCAATAAGATTTTGCATCTTTATTGGGCG
50
c5227071-5225      13  GACACAACTGTGTTCACCTAGCAACCTCA--AACAGACACCATGG---TG
56
c55104355-551      51  GAGACA-----TCAC-----ACATCAATAAAATGCACCCCGGTTCTG
88
c5227071-5225      57  CATCT-----GACT-----CCTGAGGAGA-
75
c55104355-551      89  AATTTCTGCTAGTGTGATTCTACTGGACTTTAAGCAAGCAGCTAAGGACAT
138
c5227071-5225      76  -----AGT-----CTGCCGTT
86
c55104355-551      139  CCTAAGCGTAACATGGTTGAGTGGACAGATGCCGAGCGCACAGCCATC
188
c5227071-5225      87  ACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCT
136
c55104355-551      189  CTGGCCTGTGGGGAAAGCTCAATATCGATGAAATCGGACCTCAGGCCCT
238
c5227071-5225      137  GGGCAGGTGGTATCA--AGGTTAC---AAGACAG-----GTT
169
...||  |||||  ...|||  ...|||  |||
```

In contrast, the alignment of the human HBB gene with zebrafish displayed marked divergence. The alignment length was 1,830 nucleotides, but the identity dropped to 27.8% (508 identical bases), with gaps accounting for 62.1% of the alignment (1,137 positions). The low alignment score of 1,042.5 further emphasized the weak similarity, suggesting significant evolutionary divergence or a non-orthologous relationship between the two sequences. This alignment highlights a non-conserved relationship between human HBB and the zebrafish. The low identity, high gaps, and lack of functional coherence suggest the sequences are not homologs in a classical sense.

Multiple Sequence Alignment (MSA) Report

```

CLUSTAL O(1.2.4) multiple sequence alignment

NC_000073.7:c103492369-103490845 ----- 0
NC_037342.1:48362354-48363996 ----- 0
NC_000011.10:c5227071-5225464 ----- 0
NC_072407.2:c9358653-9357035 ----- 0
NC_052532.1:c196005887-196004356 ----- 0
NC_007114.7:c55104355-55103441 CGTCATCTCAGCAGATTATCCAAGATTTCGATCTTTTATGGGCGGAGACATCAC 60

NC_000073.7:c103492369-103490845 -----ACACTTGCCTTTTGA 15
NC_037342.1:48362354-48363996 -----ACACTTGCCTTCTGAC 15
NC_000011.10:c5227071-5225464 -----ACATTGCTCTTGAC 15
NC_072407.2:c9358653-9357035 -----ATCTATTGCTTACATTTCCTTGAC 26
NC_052532.1:c196005887-196004356 -----ACCAGCGTGTATCCCCAGGGAGCAAGGCCAGACCTCTCCGG--- 45
NC_007114.7:c55104355-55103441 ACATCAATAAATGACCCGGGTTCTGAATTTCTGCTAGTGATTCTTACTGGACTT--- 117
*

NC_000073.7:c103492369-103490845 CACTTGAGATCATCTCCAAGCTTCTATACCTCACACCATGGTTCACCTTCACAGCTGAGGA 75
NC_037342.1:48362354-48363996 ACAAC--CGTGTTCACCTAGCAACTACACAAA---CAGACACCATGCTGACTGCTGAGGA 69
NC_000011.10:c5227071-5225464 ACAAC--TGTGTTCACTAGCAACTCAACAGACACCATGGTGCATCTGACTCCTGAGGA 73
NC_072407.2:c9358653-9357035 ACAAC--TGTGTTCACTAGCAACTCAACAGACACCATGGTGCATCTGACTCCTGAGGA 84
NC_052532.1:c196005887-196004356 --TAGCAGCAGCAGACATCACTCAACGGCGGCGATGGTGGACTTGGCTTCTGAGGA 103
NC_007114.7:c55104355-55103441 --TAAGCAAGCACCTAAGGACATCTAAGCGTAACATGGTTGAGTGGACAGATGCCGA 175
** ** **

NC_000073.7:c103492369-103490845 GAAGGCAGCTATCACAGCATCTGGGATAAAGTGGACTTGGAAAAAGTTGGAGGAGAAAC 135
NC_037342.1:48362354-48363996 GAAGGCTGCCGTCACCGCTTTTGGGCAAGGTGAAGTTGGATGAAGTTGGTGGTGAAGC 129
NC_000011.10:c5227071-5225464 GAAGTCTGCCGTTACTGCCCTGTGGGCAAGGTGAAGCTGGATGAAGTTGGTGGTGAAGC 133
NC_072407.2:c9358653-9357035 GAAGTCTGCCGTTACTGCCCTGTGGGCAAGGTGAAGCTGGATGAAGTTGGTGGTGAAGC 144
NC_052532.1:c196005887-196004356 GAAGCAGCTCATCAACGGCCTCTGGGCAAGGTCAATGTGGCGAAATGGGGGCTGAAGC 163
NC_007114.7:c55104355-55103441 GCGCACAGCATCCTTGGCCTGTGGGGAAGCTCAATATCGATGAATCGGACCTCAGGC 235
* * * * *

NC_000073.7:c103492369-103490845 TCTGGGAAGTGAAGGAATGGAGGGAATTGTCTCT----- 169
NC_037342.1:48362354-48363996 CCTGGGCAAGTAGGTATCCCACTACAAGCAGGTTAAGGAGAGTGAATGCACCTGGG 189
NC_000011.10:c5227071-5225464 CCTGGGCAAGTTGGTATCAAGGTTACAAGCAGGTTAAGGAGACAGTAAAGAACTGGGC 193
NC_072407.2:c9358653-9357035 CCTGGGCAAGTTGGTATCAAGGTTACAAGCAGGCTAAGGAGACAGTGAAGAACTGGGC 204
NC_052532.1:c196005887-196004356 CCTGGCCAGGTAGGTCCAGCTTCGGGCTCGGCTC----- 198
NC_007114.7:c55104355-55103441 CCTATCCAGGTATCACCTCTTCTTACTTTGTAGAAG----- 270
** ****

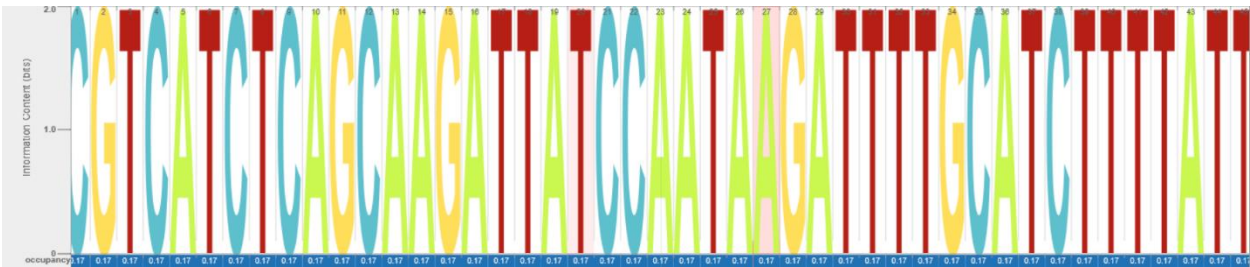
NC_000073.7:c103492369-103490845 --TGTGCATGGCAGAAATTCAGGGTTTCTATAGGGT-----TTTGGGCACACTAAT 222
NC_037342.1:48362354-48363996 CGTGTGAGCAGCAGAGCGTCTCTGAGATTTCAGAGAGCT----GCTGGCTTCTCT--- 240
NC_000011.10:c5227071-5225464 ATGTGAGACAGCAGAGCTCTGTGGTTTCTGATAGGC----ACTGACTCTCT---G 248
NC_072407.2:c9358653-9357035 ATGTGAGACAGCAGAGCTCTGTGGTTTCTGATAGGC----ACTGACTCTCT---G 256
NC_052532.1:c196005887-196004356 -----AGCTGCTCCTCGGTA----GGGAACGGCTGCAG-TGGAGGTGT 239

```

A multiple sequence alignment (MSA) of all six species was conducted using CLUSTAL OMEGA. The results highlighted highly conserved functional domains, particularly in the 5' coding region, which includes the initiation codon and early exonic sequences. Human and chimpanzee sequences were nearly identical (>98% identity), while cow and mouse showed high similarity (~80–90%) with minor substitutions. Chicken and zebrafish exhibited lower but detectable homology (~60–70%), indicating deep evolutionary conservation of certain functional regions.

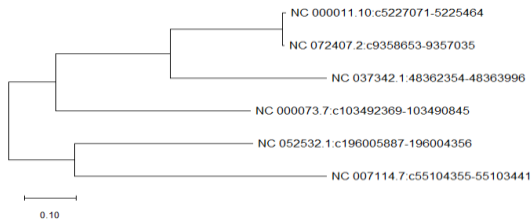
The MSA also provided evolutionary insights. The near-identical sequences of humans and chimpanzees reflect their recent common ancestry. In contrast, the higher divergence observed in cow and mouse sequences, along with the limited conservation in chicken and zebrafish outside core functional domains, illustrates the varying evolutionary pressures and timelines across these species.

Sequence Logo Generation



To further analyze conservation patterns, a sequence logo was generated from the MSA using Skylign. The logo visually emphasized conserved amino acids, reinforcing the findings of the MSA. There are conserved areas in the logo which indicates how HBB has evolved over the years

Phylogenetic Tree Construction



Additionally, a phylogenetic tree constructed with MEGA X revealed the expected evolutionary relationships. Humans and chimpanzees clustered closely, while chicken and zebrafish formed another related group. Cow and mouse, however, did not show close proximity to the human-chimpanzee branch, aligning with their intermediate divergence levels.

Conclusion

This study demonstrates the evolutionary conservation and divergence of the HBB gene across six species. The high conservation between humans and chimpanzees contrasts sharply with the limited similarity to zebrafish, highlighting the gene's functional importance and the impact of evolutionary distance. The MSA and phylogenetic analysis provide a comprehensive view of the HBB gene's evolutionary history, offering valuable insights into its role and conservation across diverse species.