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

Name: Garima Rana

Course: BioinformHer Mini Project – Module 2

Date: 11 June 2025

Table of Contents

1. Project Objective
2. Sequence Retrieval & BLAST Search
3. Pairwise Sequence Alignment
4. Multiple Sequence Alignment (MSA)
5. Sequence Logo Generation
6. Phylogenetic Tree Construction
7. Conclusion
8. References

“BioinformHer Mini Project - Module 2”

Introduction– Haemoglobin is an essential protein responsible for transporting oxygen in the blood. The beta subunit of hemoglobin, encoded by the HBB gene, plays a crucial role in this function. Studying how this gene has evolved across different species helps us understand the conservation of vital genes and molecular evolution in vertebrates.”

“In this mini project, I analyzed the HBB gene across six species—Human, Chimpanzee, Cow, Mouse, Chicken, and Zebrafish—using several bioinformatics tools such as BLAST(alignment tool), Clustal omega (multiple sequence alignment), Skylign(sequence logo generation) and MegaX(phylogenetic analysis) . The aim was to study its evolutionary conservation by retrieving sequences, performing alignments, creating sequence logos, and building a phylogenetic tree.

Task 1: Sequence Retrieval & BLAST Search

- Retrieved the human HBB protein sequence from NCBI in FASTA format.
- Used BLASTp to identify homologous sequences in other species: chimpanzee, cow, mouse, chicken, and zebrafish
- Recorded the % identity and accession numbers.
- The FASTA sequences of these proteins were downloaded and recorded the percentage identity for each species compared to the human HBB.”

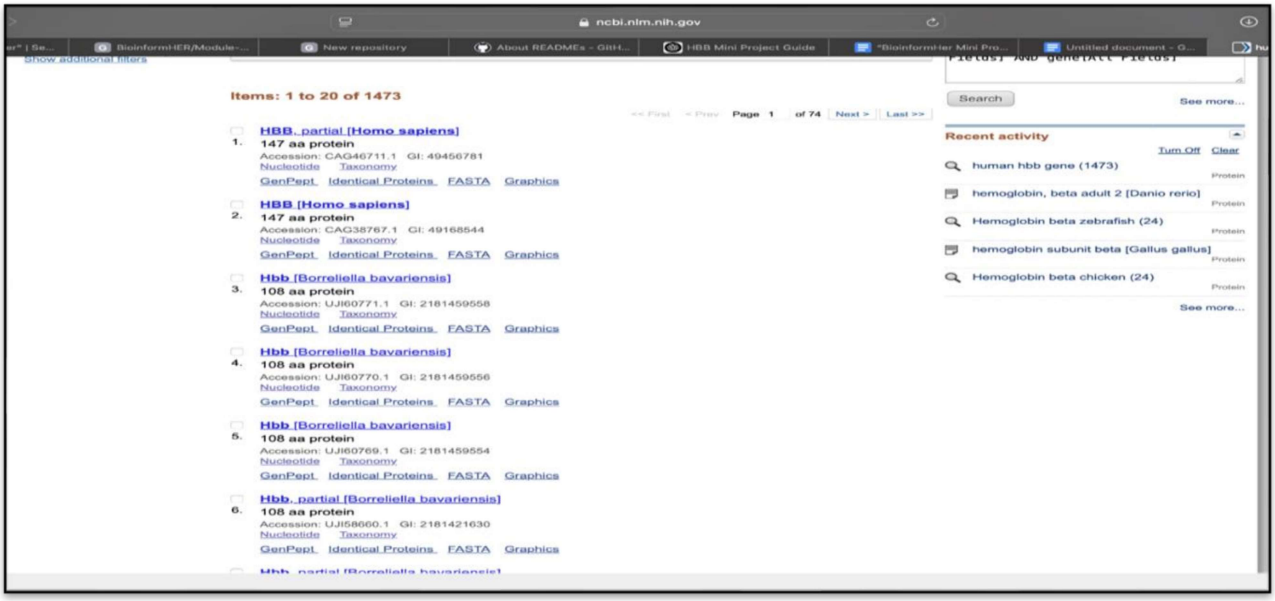
  Procedure-

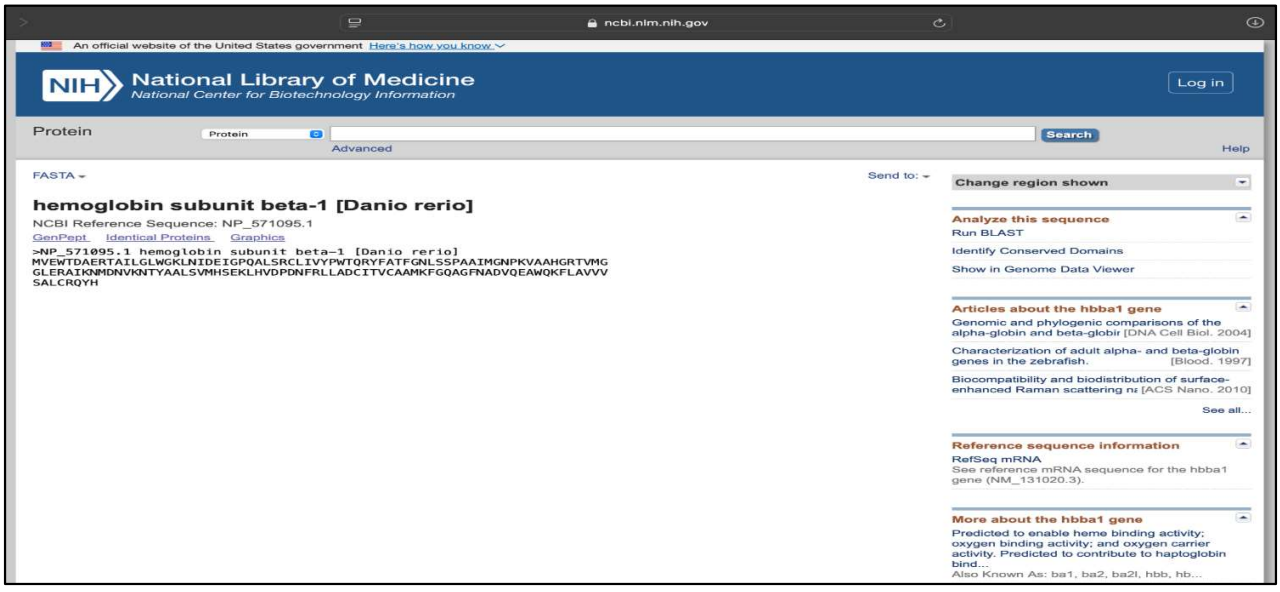
  The first step involved retrieving the human Hemoglobin Beta (HBB) protein sequence from the NCBI database in FASTA format. This sequence was then used as a query in a BLASTp search to identify homologous proteins in five other species: chimpanzee, cow, mouse, chicken, and zebrafish. Each species was selected to represent varying evolutionary distances from humans. The BLAST results included percentage identity and accession numbers, summarized in the table below.

1. Go to NCBI.
2. Search for the **Human HBB gene –** nucleotide or protein.
3. Perform **BLAST** with that sequence.
4. Select HBB sequences from **5 other species**:
   - Chimpanzee
   - Cow
   - Mouse
   - Chicken
   - Zebrafish
5. Download their **FASTA sequences**.

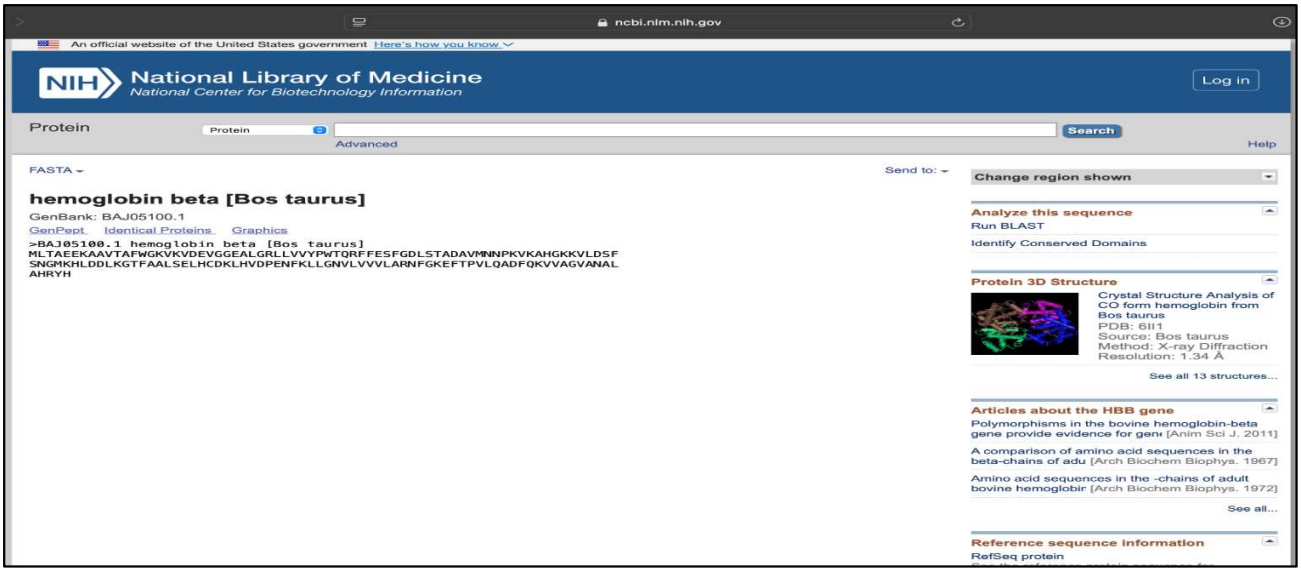
Result Table:

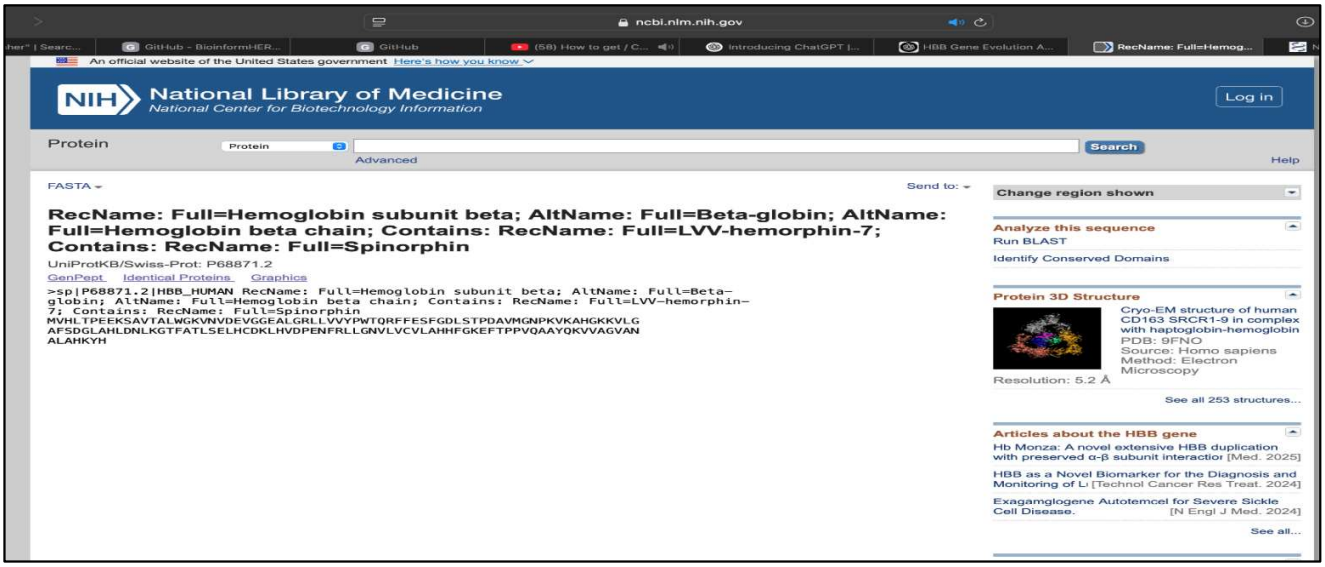
Species	Accession number	% similarity with humans
Chimpanzee	XP_508242.1	98%
cow	NP_776342.1	78%
mouse	NP_001265090.1	80%
chicken	NP_990820.1	69%
zebrafish	NP_001076303.1	49%

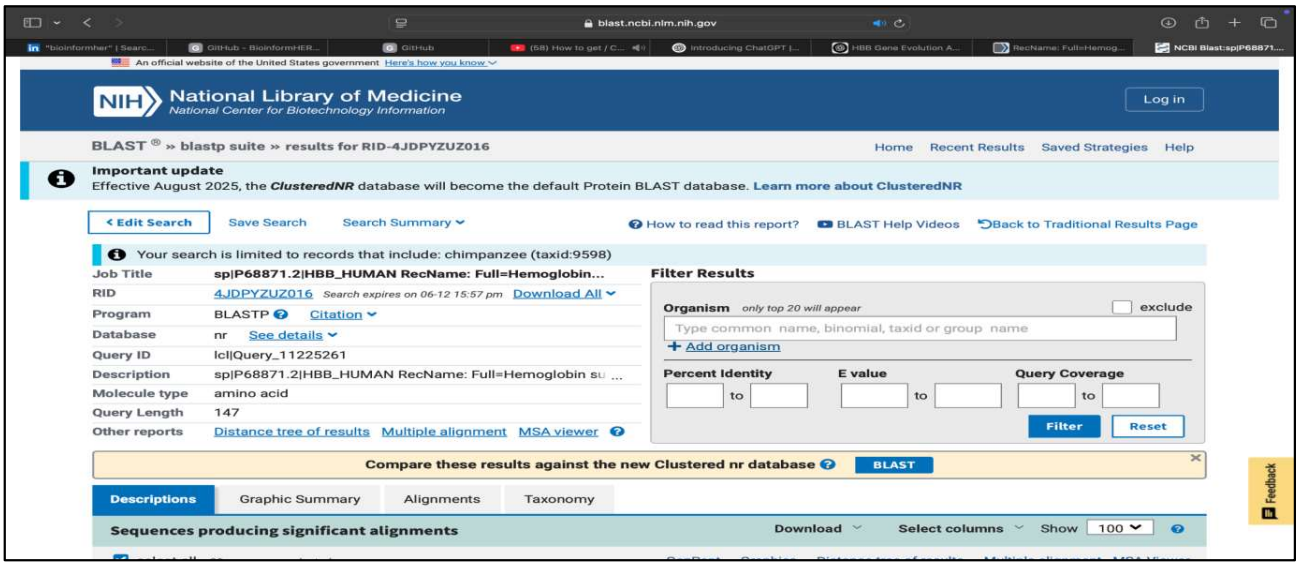


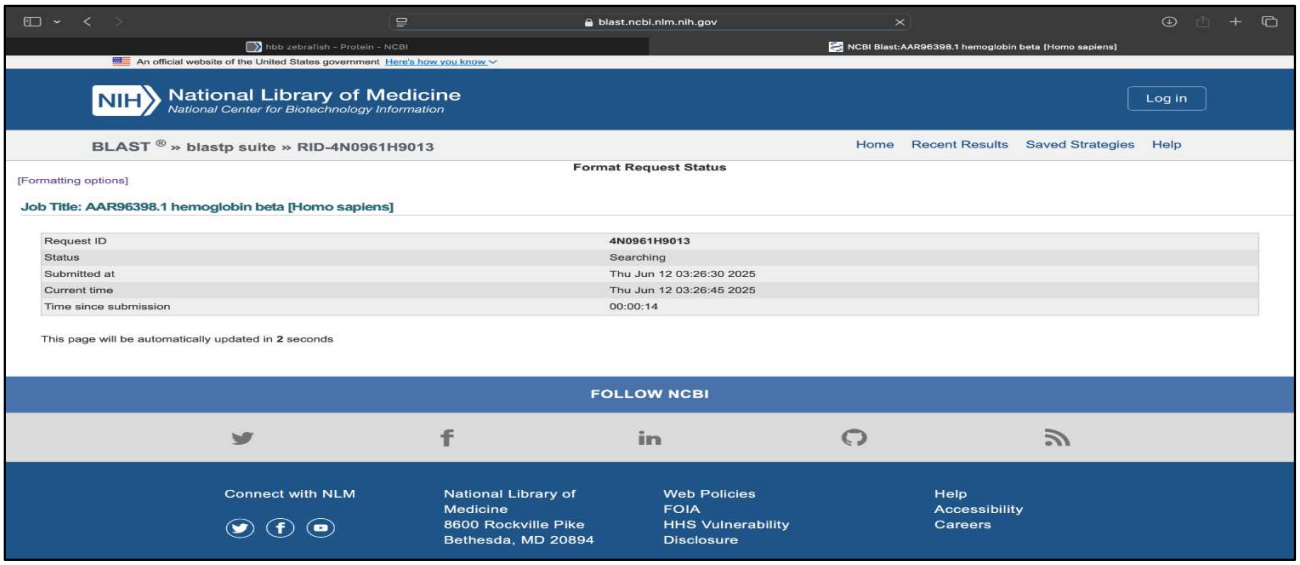


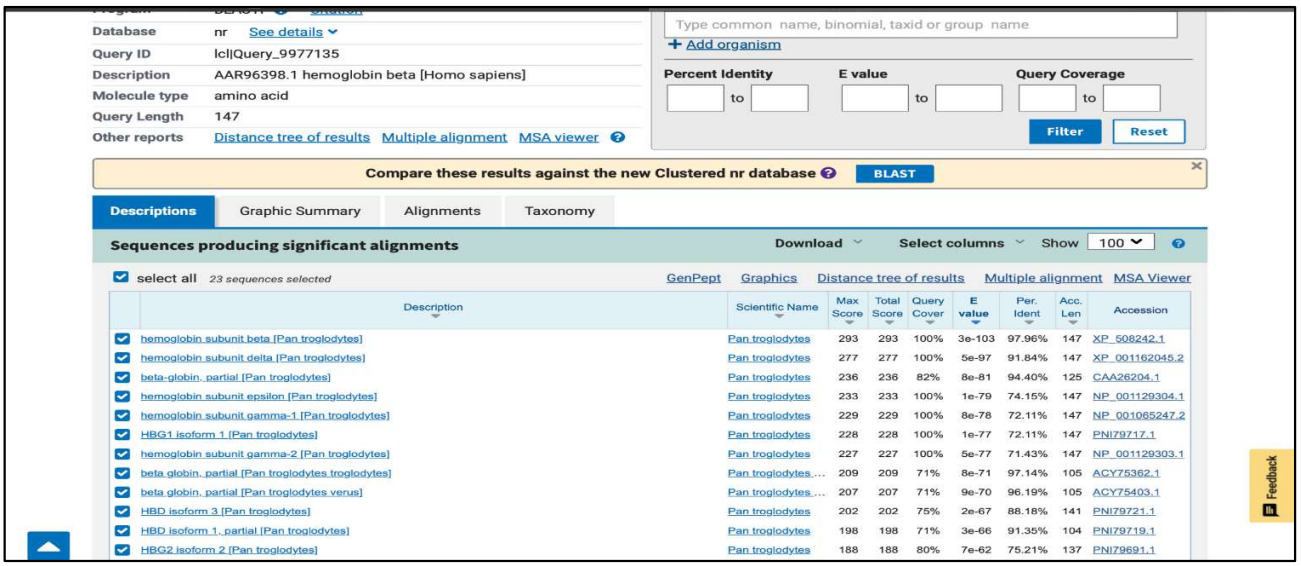
National Library of Medicine
National Center for Biotechnology Information
Log in
Protein
hbb Homo sapiens
Search
Create alert
Advanced
Help
Species
Animals (1,309)
Bacteria (8)
Viruses (158)
Customize ...
Source databases
PDB (8)
RefSeq (8)
UniProtKB / Swiss-Prot (10)
Customize ...
Sequence length
Custom range...
Molecular weight
Custom range...
Release date
Custom range...
Revision date
Custom range...
Clear all
Show additional filters
Summary
20 per page
Sort by Default order
Send to:
GENE
Was this helpful?
HBB — hemoglobin subunit beta
Homo sapiens (human)
Also known as: CD113t-C, ECYT6, beta-globin
Gene ID: 3043
RefSeq products
Orthologs
Genome Data Viewer
New - Visualize gene across multiple species
RefSeq Sequences
Items: 1 to 20 of 1475
HBB, partial [Homo sapiens]
1. 147 aa protein
Accession: CAG46711.1 GI: 49456781
Nucleotide Taxonomy
GenPept Identical Proteins FASTA Graphics
Filters: Manage Filters
Results by taxon
Top Organisms [Tree]
Homo sapiens (1304)
Hepatitis B virus (158)
Borreliella bavariensis (7)
Macaca fascicularis (1)
Papio anubis (1)
All other taxa (4)
More...
Find related data
Database: Select
Search details
hbb[All Fields] AND ("Homo sapiens"[Organism] OR Homo sapiens[All Fields])
Search
See more...
Recent activity
Turn Off Clear
hbb Homo sapiens (1475)
human hbb gene (1473)


ncbi.nlm.nih.gov
An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
Protein
Search
Advanced
Help
FASTA
Send to:
hemoglobin beta [Bos taurus]
GenBank: BAJ05100.1
GenPept Identical Proteins Graphics
>BAJ05100.1 hemoglobin beta [Bos taurus]
MLTAEEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMNNPKVKAHGKKVLDSF
SNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFGKEFTPVLQADFQKVVAGVANAL
AHRYH
Change region shown
Analyze this sequence
Run BLAST
Identify Conserved Domains
Protein 3D Structure
Crystal Structure Analysis of CO form hemoglobin from Bos taurus
PDB: 6II1
Source: Bos taurus
Method: X-ray Diffraction
Resolution: 1.34 Å
See all 13 structures...
Articles about the HBB gene
Polymorphisms in the bovine hemoglobin-beta gene provide evidence for gen [Anim Sci J. 2011]
A comparison of amino acid sequences in the beta-chains of adu [Arch Biochem Biophys. 1967]
Amino acid sequences in the -chains of adult bovine hemoglobir [Arch Biochem Biophys. 1972]
See all...
Reference sequence information
RefSeq protein

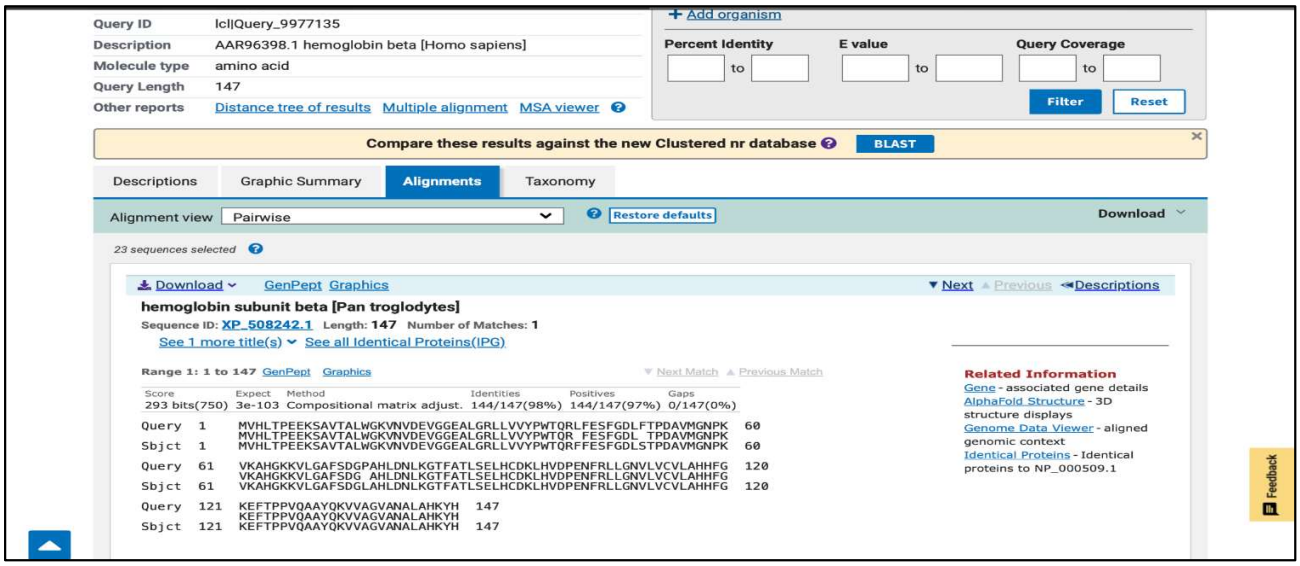

ncbi.nlm.nih.gov
An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
Protein
Search
Advanced
Help
FASTA
Send to:
RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName: Full=LVV-hemorphin-7; Contains: RecName: Full=Spinorphin
UniProtKB/Swiss-Prot: P68871.2
GenPept Identical Proteins Graphics
>sp|P68871.2|HBB_HUMAN RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName: Full=LVV-hemorphin-7; Contains: RecName: Full=Spinorphin
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
Change region shown
Analyze this sequence
Run BLAST
Identify Conserved Domains
Protein 3D Structure
Cryo-EM structure of human CD163 SRCR1-9 in complex with haptoglobin-hemoglobin
PDB: 9FNO
Source: Homo sapiens
Method: Electron Microscopy
Resolution: 5.2 Å
See all 253 structures...
Articles about the HBB gene
Hb Monza: A novel extensive HBB duplication with preserved α-β subunit interactior [Med. 2025]
HBB as a Novel Biomarker for the Diagnosis and Monitoring of L [Technol Cancer Res Treat. 2024]
Exagamglogene Autotemcel for Severe Sickle Cell Disease. [N Engl J Med. 2024]
See all...


blast.ncbi.nlm.nih.gov
An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST ® » blastp suite » results for RID-4JDPYZUZ016
Home Recent Results Saved Strategies Help
Important update
Effective August 2025, the ClusteredNR database will become the default Protein BLAST database. Learn more about ClusteredNR
Edit Search Save Search Search Summary
How to read this report? BLAST Help Videos Back to Traditional Results Page
Your search is limited to records that include: chimpanzee (taxid:9598)
Job Title sp|P68871.2|HBB_HUMAN RecName: Full=Hemoglobin...
RID 4JDPYZUZ016 Search expires on 06-12 15:57 pm Download All
Program BLASTP Citation
Database nr See details
Query ID lcl|Query_11225261
Description sp|P68871.2|HBB_HUMAN RecName: Full=Hemoglobin su ...
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Filter Results
Organism only top 20 will appear exclude
Type common name, binomial, taxid or group name
Add organism
Percent Identity to
E value to
Query Coverage to
Filter Reset
Compare these results against the new Clustered nr database BLAST
Descriptions Graphic Summary Alignments Taxonomy
Sequences producing significant alignments
Download Select columns Show 100
Feedback

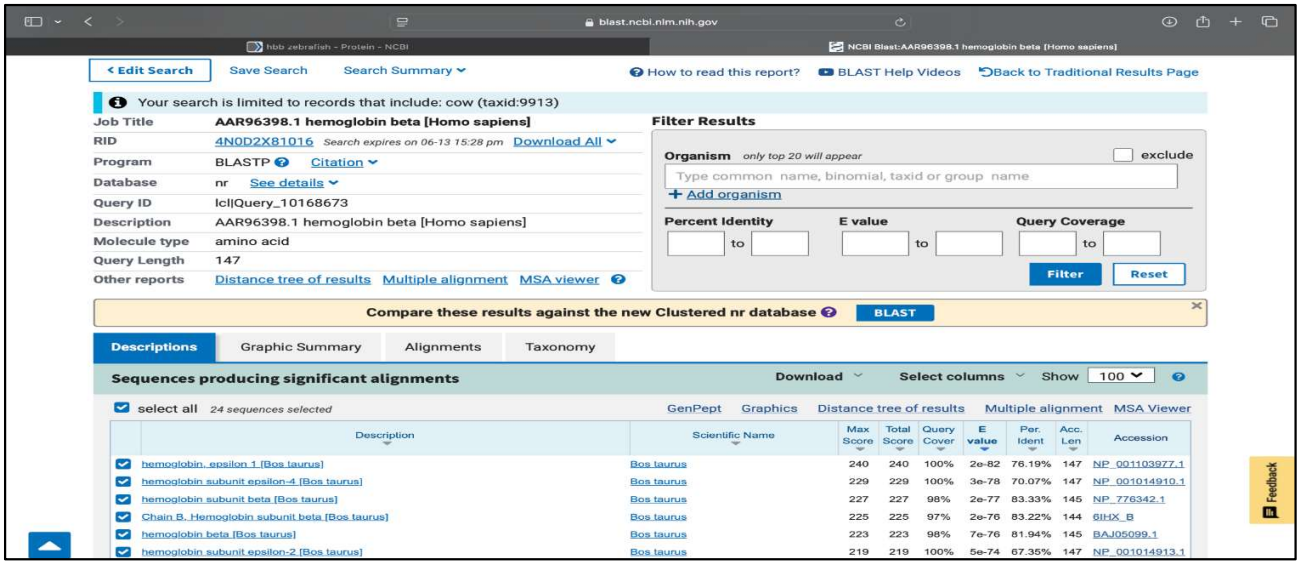

blast.ncbi.nlm.nih.gov
hbb zebrafish - Protein - NCBI
NCBI Blast:AAR96398.1 hemoglobin beta [Homo sapiens]
An official website of the United States government Here's how you know
National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST ® » blastp suite » RID-4N0961H9013
Home Recent Results Saved Strategies Help
Format Request Status
[Formatting options]
Job Title: AAR96398.1 hemoglobin beta [Homo sapiens]
Request ID 4N0961H9013
Status Searching
Submitted at Thu Jun 12 03:26:30 2025
Current time Thu Jun 12 03:26:45 2025
Time since submission 00:00:14
This page will be automatically updated in 2 seconds
FOLLOW NCBI
Connect with NLM
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894
Web Policies
FOIA
HHS Vulnerability Disclosure
Help
Accessibility
Careers

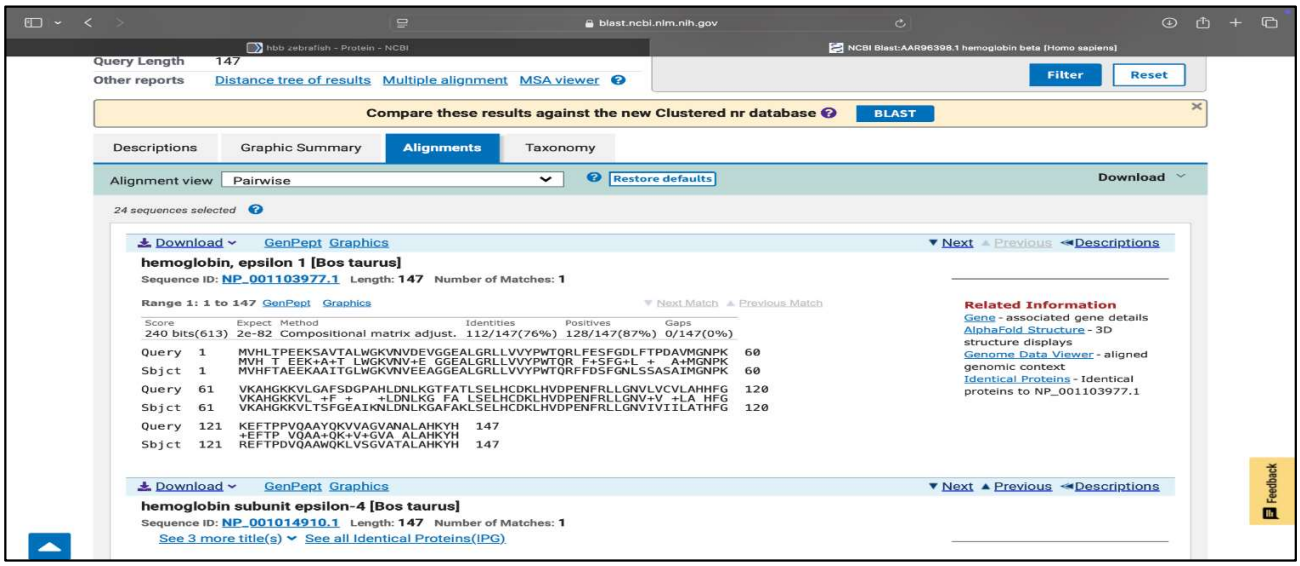

Database nr See details
Query ID lcl|Query_9977135
Description AAR96398.1 hemoglobin beta [Homo sapiens]
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Type common name, binomial, taxid or group name
Add organism
Percent Identity to
E value to
Query Coverage to
Filter Reset
Compare these results against the new Clustered nr database BLAST
Descriptions Graphic Summary Alignments Taxonomy
Sequences producing significant alignments
Download Select columns Show 100
select all 23 sequences selected
GenPept Graphics Distance tree of results Multiple alignment MSA Viewer
Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession
hemoglobin subunit beta [Pan troglodytes] | Pan troglodytes | 293 | 293 | 100% | 3e-103 | 97.96% | 147 | XP_508242.1
hemoglobin subunit delta [Pan troglodytes] | Pan troglodytes | 277 | 277 | 100% | 5e-97 | 91.84% | 147 | XP_001162045.2
beta-globin, partial [Pan troglodytes] | Pan troglodytes | 236 | 236 | 82% | 8e-81 | 94.40% | 125 | CAA26204.1
hemoglobin subunit epsilon [Pan troglodytes] | Pan troglodytes | 233 | 233 | 100% | 1e-79 | 74.15% | 147 | NP_001129304.1
hemoglobin subunit gamma-1 [Pan troglodytes] | Pan troglodytes | 229 | 229 | 100% | 8e-78 | 72.11% | 147 | NP_001065247.2
HBG1 isoform 1 [Pan troglodytes] | Pan troglodytes | 228 | 228 | 100% | 1e-77 | 72.11% | 147 | PNI79717.1
hemoglobin subunit gamma-2 [Pan troglodytes] | Pan troglodytes | 227 | 227 | 100% | 5e-77 | 71.43% | 147 | NP_001129303.1
beta globin, partial [Pan troglodytes troglodytes] | Pan troglodytes... | 209 | 209 | 71% | 8e-71 | 97.14% | 105 | ACY75362.1
beta globin, partial [Pan troglodytes verus] | Pan troglodytes... | 207 | 207 | 71% | 9e-70 | 96.19% | 105 | ACY75403.1
HBD isoform 3 [Pan troglodytes] | Pan troglodytes | 202 | 202 | 75% | 2e-67 | 88.18% | 141 | PNI79721.1
HBD isoform 1, partial [Pan troglodytes] | Pan troglodytes | 198 | 198 | 71% | 3e-66 | 91.35% | 104 | PNI79719.1
HBG2 isoform 2 [Pan troglodytes] | Pan troglodytes | 188 | 188 | 80% | 7e-62 | 75.21% | 137 | PNI79691.1
Feedback

Query ID lcl|Query_9977135
Description AAR96398.1 hemoglobin beta [Homo sapiens]
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Add organism
Percent Identity
E value
Query Coverage
Filter
Reset
Compare these results against the new Clustered nr database BLAST
Descriptions
Graphic Summary
Alignments
Taxonomy
hover to see the title
click to show alignments
Show Conserved Domains
Alignment Scores
< 40
40 - 50
50 - 80
80 - 200
>= 200
23 sequences selected
Putative conserved domains have been detected, click on the image below for detailed results.
Hb-beta-like
Globin-like superfamily
Distribution of the top 23 Blast Hits on 23 subject sequences
Query
Feedback


Query ID lcl|Query_9977135
Description AAR96398.1 hemoglobin beta [Homo sapiens]
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Add organism
Percent Identity
E value
Query Coverage
Filter
Reset
Compare these results against the new Clustered nr database BLAST
Descriptions
Graphic Summary
Alignments
Taxonomy
Alignment view Pairwise
Restore defaults
Download
23 sequences selected
Download GenPept Graphics
Next Previous Descriptions
hemoglobin subunit beta [Pan troglodytes]
Sequence ID: XP_508242.1 Length: 147 Number of Matches: 1
See 1 more title(s) See all Identical Proteins(IPG)
Range 1: 1 to 147 GenPept Graphics
Next Match Previous Match
Score Expect Method Identities Positives Gaps
293 bits(750) 3e-103 Compositional matrix adjust. 144/147(98%) 144/147(97%) 0/147(0%)
Query 1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRLFESFGDLFTPDAVMGNPK 60
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQR FESFGDL TPDAVMGNPK
Sbjct 1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK 60
Query 61 VKAHGKKVLGAFSDGPAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
VKAHGKKVLGAFSDG AHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Sbjct 61 VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
Query 121 KEFTPPVQAAYQKVVAGVANALAHKYH 147
KEFTPPVQAAYQKVVAGVANALAHKYH
Sbjct 121 KEFTPPVQAAYQKVVAGVANALAHKYH 147
Related Information
Gene - associated gene details
AlphaFold Structure - 3D structure displays
Genome Data Viewer - aligned genomic context
Identical Proteins - Identical proteins to NP_000509.1
Feedback

blast.ncbi.nlm.nih.gov
hbb zebrafish - Protein - NCBI
NCBI Blast:AAR96398.1 hemoglobin beta [Homo sapiens]
Description AAR96398.1 hemoglobin beta [Homo sapiens]
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Percent Identity
E value
Query Coverage
Filter
Reset
Compare these results against the new Clustered nr database BLAST
Descriptions
Graphic Summary
Alignments
Taxonomy
hover to see the title
click to show alignments
Show Conserved Domains
Alignment Scores
< 40
40 - 50
50 - 80
80 - 200
>= 200
24 sequences selected
Putative conserved domains have been detected, click on the image below for detailed results.
Hb-beta-like
Globin-like superfamily
Distribution of the top 24 Blast Hits on 24 subject sequences
Query
Feedback


blast.ncbi.nlm.nih.gov
hbb zebrafish - Protein - NCBI
NCBI Blast:AAR96398.1 hemoglobin beta [Homo sapiens]
< Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Back to Traditional Results Page
Your search is limited to records that include: cow (taxid:9913)
Job Title AAR96398.1 hemoglobin beta [Homo sapiens]
RID 4N0D2X81016 Search expires on 06-13 15:28 pm Download All
Program BLASTP Citation
Database nr See details
Query ID lcl|Query_10168673
Description AAR96398.1 hemoglobin beta [Homo sapiens]
Molecule type amino acid
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Filter Results
Organism only top 20 will appear
exclude
Type common name, binomial, taxid or group name
Add organism
Percent Identity
E value
Query Coverage
Filter
Reset
Compare these results against the new Clustered nr database
BLAST
Descriptions
Graphic Summary
Alignments
Taxonomy
Sequences producing significant alignments
Download
Select columns
Show 100
select all 24 sequences selected
GenPept Graphics Distance tree of results Multiple alignment MSA Viewer
Description
Scientific Name
Max Score
Total Score
Query Cover
E value
Per. Ident
Acc. Len
Accession
hemoglobin, epsilon 1 [Bos taurus] Bos taurus 240 240 100% 2e-82 76.19% 147 NP_001103977.1
hemoglobin subunit epsilon-4 [Bos taurus] Bos taurus 229 229 100% 3e-78 70.07% 147 NP_001014910.1
hemoglobin subunit beta [Bos taurus] Bos taurus 227 227 98% 2e-77 83.33% 145 NP_776342.1
Chain B, Hemoglobin subunit beta [Bos taurus] Bos taurus 225 225 97% 2e-76 83.22% 144 6IHX_B
hemoglobin beta [Bos taurus] Bos taurus 223 223 98% 7e-76 81.94% 145 BAJ05099.1
hemoglobin subunit epsilon-2 [Bos taurus] Bos taurus 219 219 100% 5e-74 67.35% 147 NP_001014913.1
Feedback


blast.ncbi.nlm.nih.gov
hbb zebrafish - Protein - NCBI
NCBI Blast:AAR96398.1 hemoglobin beta [Homo sapiens]
Query Length 147
Other reports Distance tree of results Multiple alignment MSA viewer
Filter
Reset
Compare these results against the new Clustered nr database
BLAST
Descriptions
Graphic Summary
Alignments
Taxonomy
Alignment view Pairwise
Restore defaults
Download
24 sequences selected
Download GenPept Graphics
Next Previous Descriptions
hemoglobin, epsilon 1 [Bos taurus]
Sequence ID: NP_001103977.1 Length: 147 Number of Matches: 1
Range 1: 1 to 147 GenPept Graphics
Next Match Previous Match
Score Expect Method Identities Positives Gaps
240 bits(613) 2e-82 Compositional matrix adjust. 112/147(76%) 128/147(87%) 0/147(0%)
Query 1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRLFESFGDLFTPDAVMGNPK 60
MVH T EEK+A+T LWGKVNV+E GGEALGRLLVVYPWTQR F+SFG+L + A+MGNPK
Sbjct 1 MVHFTAEEKAAITGLWGKVNVEEAGGEALGRLLVVYPWTQRFFDSFGNLSSASAIMGNPK 60
Query 61 VKAHGKKVLGAFSDGPAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
VKAHGKKVL +F + +LDNLKG FA LSELHCDKLHVDPENFRLLGNV+V +LA HFG
Sbjct 61 VKAHGKKVLTSFGEAIKNLDNLKGAFAKLSELHCDKLHVDPENFRLLGNVIVIILATHFG 120
Query 121 KEFTPPVQAAYQKVVAGVANALAHKYH 147
+EFTP VQAA+QK+V+GVA ALAHKYH
Sbjct 121 REFTPDVQAAWQKLVSGVATALAHKYH 147
Related Information
Gene - associated gene details
AlphaFold Structure - 3D structure displays
Genome Data Viewer - aligned genomic context
Identical Proteins - Identical proteins to NP_001103977.1
Download GenPept Graphics
Next Previous Descriptions
hemoglobin subunit epsilon-4 [Bos taurus]
Sequence ID: NP_001014910.1 Length: 147 Number of Matches: 1
See 3 more title(s) See all Identical Proteins(IPG)
Feedback

Task 2: Pairwise Sequence Alignment

Performed pairwise alignment using EMBOSS Needle.The FASTA sequences of these proteins were downloaded and recorded the percentage identity for each species compared to the human HBB.

Procedure-

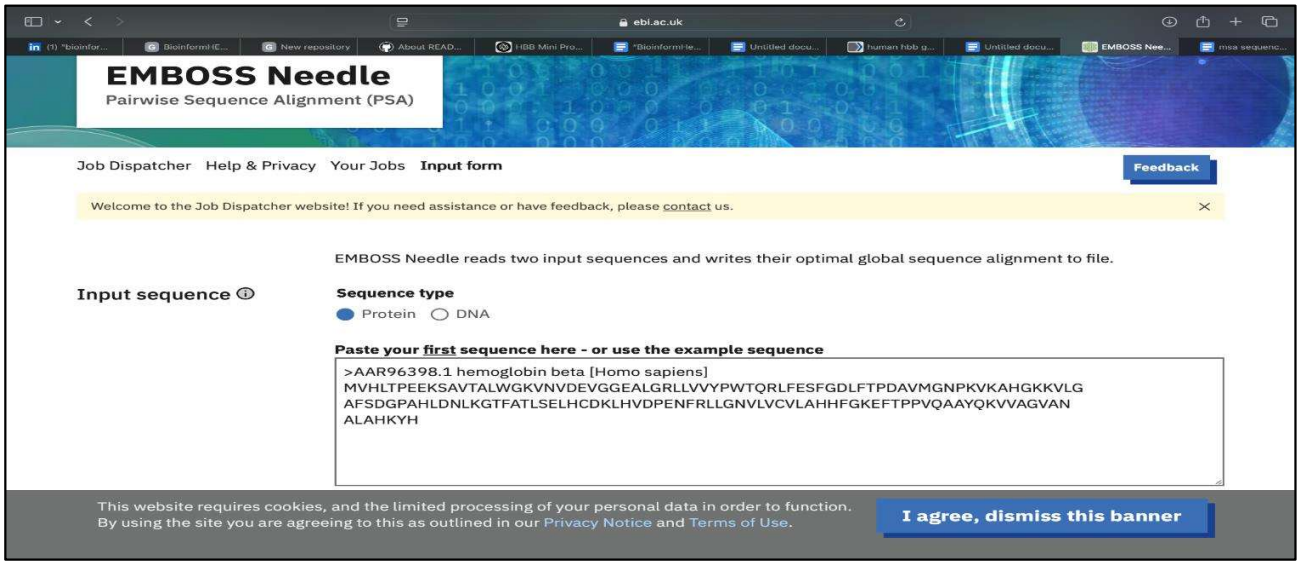
Using EMBOSS Needle, pairwise alignments were performed to compare the human HBB sequence against that of two other species — the closely related chimpanzee and the more distantly related zebrafish. These alignments helped assess the conservation of amino acids and the evolutionary divergence. As expected, the human and chimpanzee sequences were nearly identical, while the zebrafish sequence showed more divergence and reduced similarity.

Steps:

1. Use an EMBOSS Needle or NCBI Alignment tool.
2. Align:

- Human vs Chimpanzee (closely related)
- Human vs Zebrafish (distantly related)

Comparison	% Identity	% Similarity	Gaps
Human vs Chimpanzee	98%	99%	0
Human vs Zebrafish	49%	66%	0



Sequence with chimpanzee-

```
# Aligned_sequences: 2
# 1: AAR96398.1
# 2: XP_508242.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:      144/147 (98.0%)
# Similarity:    144/147 (98.0%)
# Gaps:            0/147 ( 0.0%)
# Score: 761.0
#
#
#=======================================

AAR96398.1          1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRLFESFGDLF      50
                      |||||||||||||||||||||||||||||||||||||||||.|||||||.
XP_508242.1         1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS      50

AAR96398.1         51 TPDAVMGNPKVKAHGKKVLGAFSDGPAHLDNLKGTFATLSELHCDKLHVD     100
                      ||||||||||||||||||||||||.|||||||||||||||||||||||||
XP_508242.1        51 TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD     100

AAR96398.1        101 PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
XP_508242.1       101 PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147


#---------------------------------------
#---------------------------------------
```

Sequence with zebrafish-

```
ebi.ac.uk
About READ...  HBB Mini Pro...  "BioinformHe...  Untitled docu...  human hbb g...
#
# Aligned_sequences: 2
# 1: AAR96398.1
# 2: NP_001003431.2
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:       72/147 (49.0%)
# Similarity:    102/147 (69.4%)
# Gaps:            0/147 ( 0.0%)
# Score: 395.0
#
#
#=======================================

AAR96398.1          1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRLFESFGDLF      50
                      ||.:.:..|:..:.:::|.|:|||||:|.:.|.|:|||||||||.|.:|||||.
NP_001003431.       1 MVQWSDSERKTIASVWSKINVDEIGPQTLARVLVVYPWTQRYFGAFGDLS      50

AAR96398.1         51 TPDAVMGNPKVKAHGKKVLGAFSDGPAHLDNLKGTFATLSELHCDKLHVD     100
                      ...|:|||||||..|||.||.|.....::|::|.|:|.|:||:|||:||:||
NP_001003431.      51 CASAIMGNPKVSEHGKTVLKALEKAVKNVDDIKTTYAKLSQLHCEKLNVD     100

AAR96398.1        101 PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147
                      |:|||.|.:.|..|:|.:|||..|.|.||:.:||:::.|..||..:|.
NP_001003431.     101 PDNFKLLADCLSIVIATNFGPAFNPSVQSTWQKLLSVVVAALTSRYF     147


#---------------------------------------
#---------------------------------------
```

Interpretation:

High conservation between human and chimpanzee HBB indicates evolutionary closeness. Lower similarity with zebrafish suggests divergence over time.As expected, the human and chimpanzee HBB sequences are highly conserved, indicating strong evolutionary similarity. On the other hand, the zebrafish HBB sequence shows lower identity, reflecting greater evolutionary divergence.

Task 3: <u>Multiple Sequence Alignment (MSA)</u>

Using Clustal Omega, I aligned the HBB protein sequences from all six species. The alignment showed several conserved regions, especially in the functional domains of the protein.

Description:

- Used Clustal Omega to align all six species' HBB protein sequences.
- Conserved residues indicate functionally important domains.

Procedure-

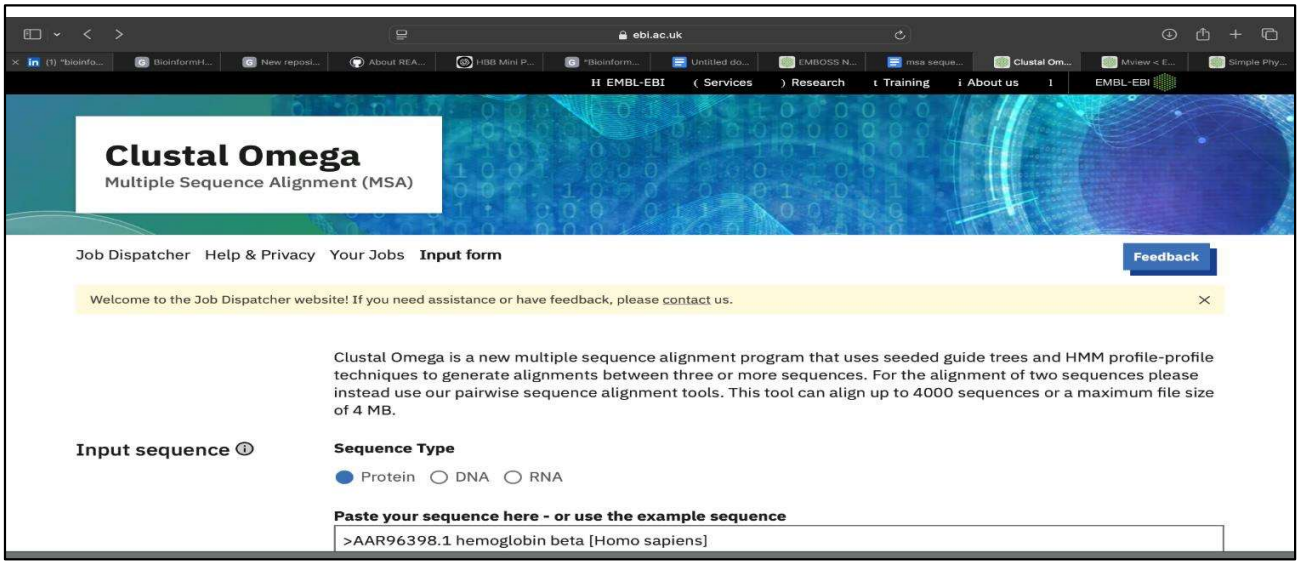
A Multiple Sequence Alignment (MSA) was performed using Clustal Omega. All six sequences were aligned to identify conserved regions and structural motifs across species. The alignment output highlighted several regions with perfect or near-perfect conservation, especially among mammalian species. These conserved regions are likely critical to the functional stability of the HBB protein

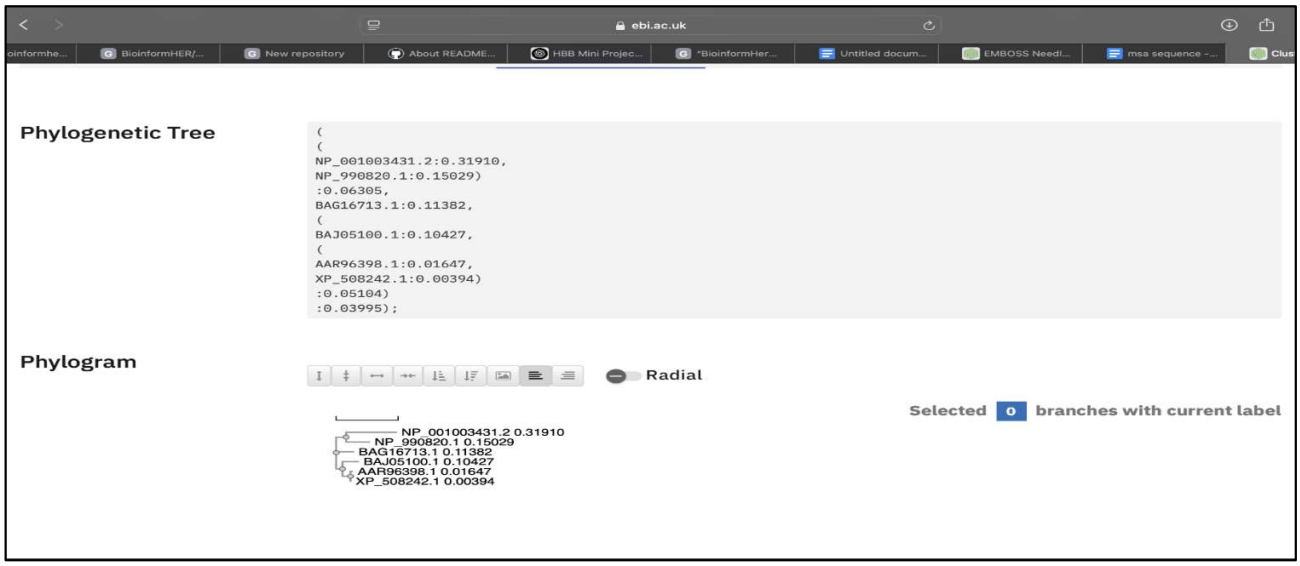
Steps:

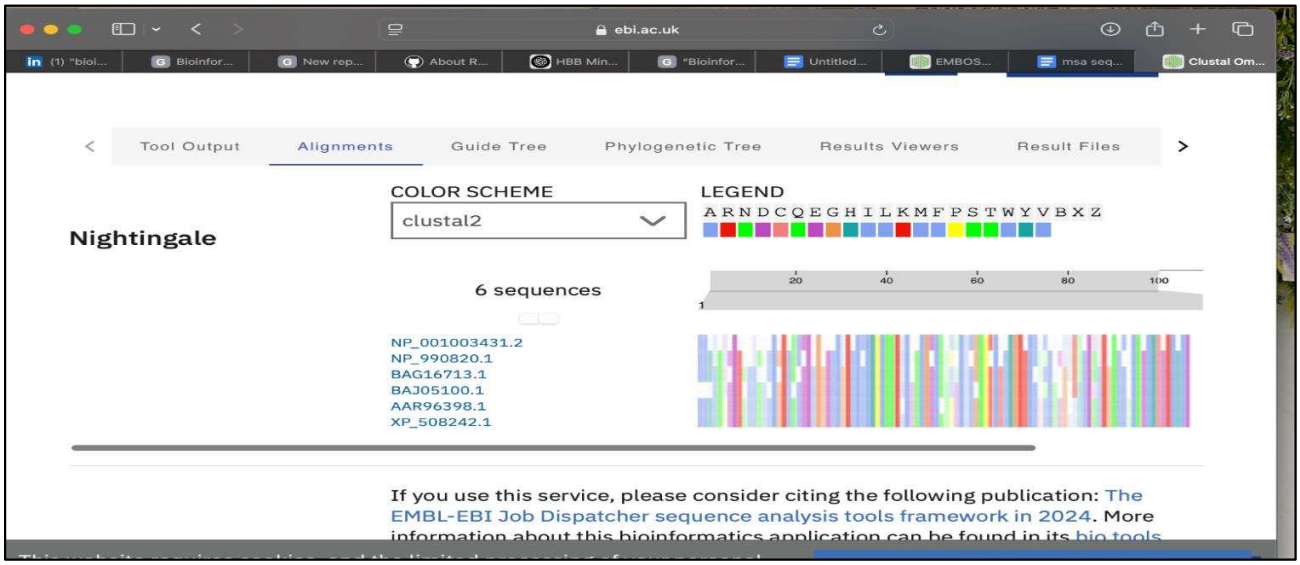
1. Use **Clustal Omega**
2. Align all 6 sequences.

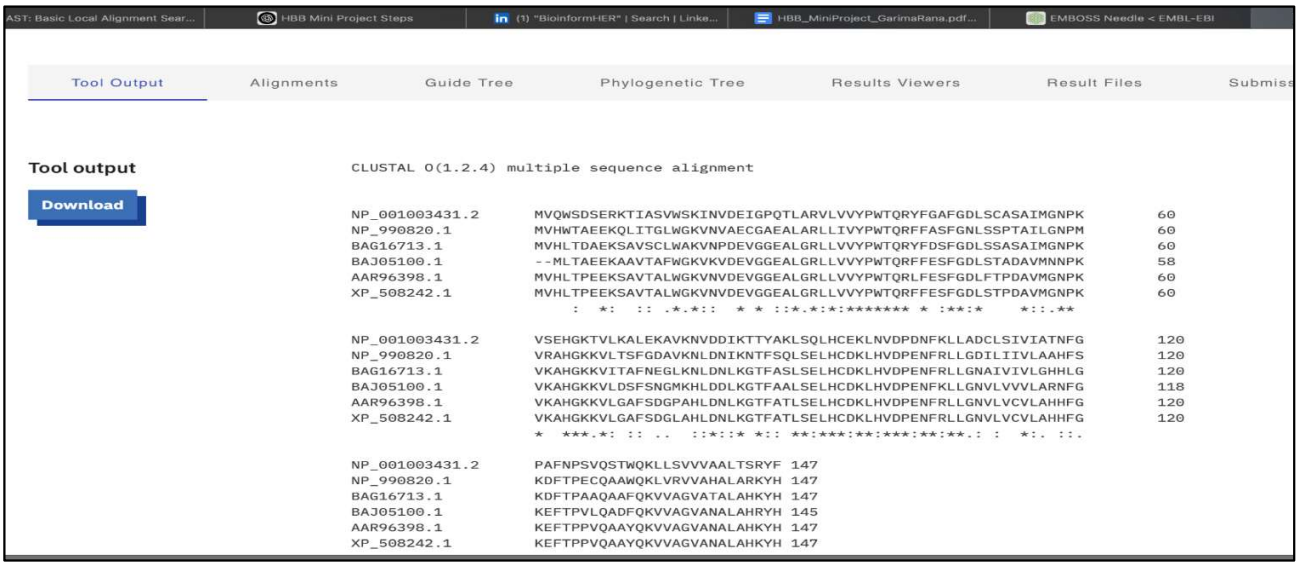
Observation:

Conserved regions suggest structurally and functionally important parts of the HBB protein.




ebi.ac.uk
Phylogenetic Tree
(
(
NP_001003431.2:0.31910,
NP_990820.1:0.15029)
:0.06305,
BAG16713.1:0.11382,
(
BAJ05100.1:0.10427,
(
AAR96398.1:0.01647,
XP_508242.1:0.00394)
:0.05104)
:0.03995);
Phylogram
Radial
Selected 0 branches with current label
NP_001003431.2 0.31910
NP_990820.1 0.15029
BAG16713.1 0.11382
BAJ05100.1 0.10427
AAR96398.1 0.01647
XP_508242.1 0.00394


ebi.ac.uk
Tool Output
Alignments
Guide Tree
Phylogenetic Tree
Results Viewers
Result Files
Nightingale
COLOR SCHEME
clustal2
LEGEND
A R N D C Q E G H I L K M F P S T W Y V B X Z
6 sequences
20
40
60
80
100
1
NP_001003431.2
NP_990820.1
BAG16713.1
BAJ05100.1
AAR96398.1
XP_508242.1
If you use this service, please consider citing the following publication: The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. More information about this bioinformatics application can be found in its bio tools


Tool Output
Alignments
Guide Tree
Phylogenetic Tree
Results Viewers
Result Files
Tool output
Download
CLUSTAL O(1.2.4) multiple sequence alignment
NP_001003431.2 MVQWSDSERKTIASVWSKINVDEIGPQTLARVLVVYPWTQRYFGAFGDLSCASAIMGNPK 60
NP_990820.1 MVHWTAEEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNPM 60
BAG16713.1 MVHLTDAEKSAVSCLWAKVNPDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNPK 60
BAJ05100.1 --MLTAEEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMNNPK 58
AAR96398.1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRLFESFGDLFTPDAVMGNPK 60
XP_508242.1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK 60
NP_001003431.2 VSEHGKTVLKALEKAVKNVDDIKTTYAKLSQLHCEKLNVDPDNFKLLADCLSIVIATNFG 120
NP_990820.1 VRAHGKKVLTSFGDAVKNLDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIVLAAHFS 120
BAG16713.1 VKAHGKKVITAFNEGLKNLDNLKGTFASLSELHCDKLHVDPENFRLLGNAIVIVLGHHLG 120
BAJ05100.1 VKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFG 118
AAR96398.1 VKAHGKKVLGAFSDGPAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
XP_508242.1 VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
NP_001003431.2 PAFNPSVQSTWQKLLSVVVAALTSRYF 147
NP_990820.1 KDFTPECQAAWQKLVRVVAHALARKYH 147
BAG16713.1 KDFTPAAQAAFQKVVAGVATALAHKYH 147
BAJ05100.1 KEFTPVLQADFQKVVAGVANALAHRYH 145
AAR96398.1 KEFTPPVQAAYQKVVAGVANALAHKYH 147
XP_508242.1 KEFTPPVQAAYQKVVAGVANALAHKYH 147

Task 4: Sequence Logo Generation

To visualize conserved amino acid positions, I uploaded the MSA file to Skylign and generated a sequence logo. The height of each letter in the logo represents how conserved that amino acid is across all six species.

Description:

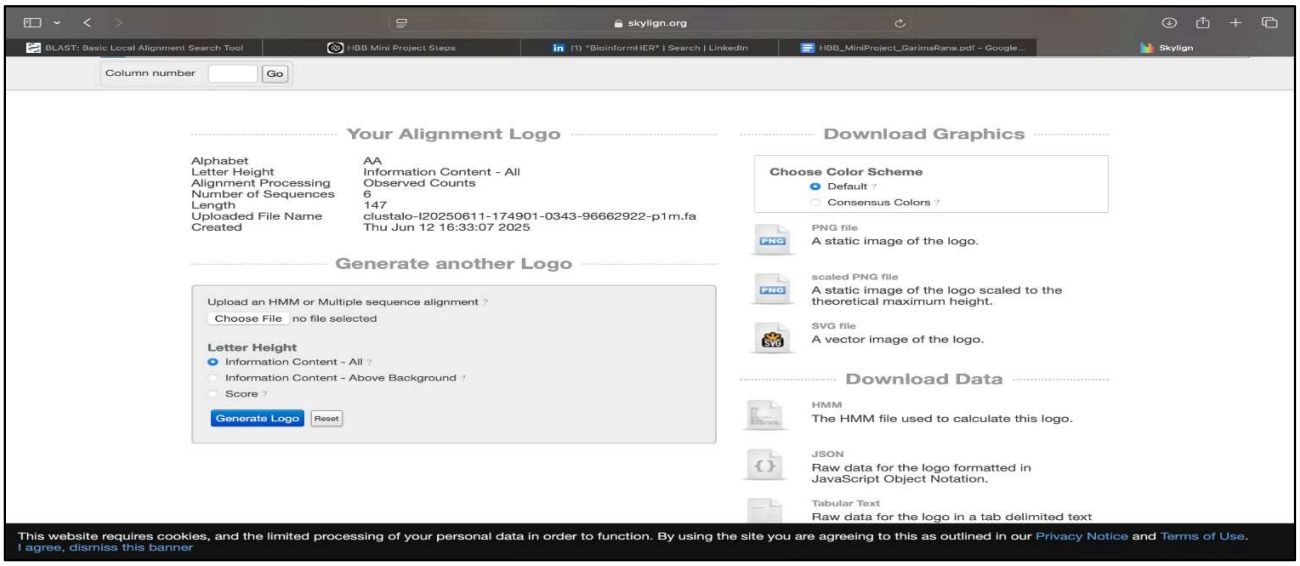
- Uploaded MSA results in Skylign to visualize conserved amino acids.

  Procedure-

  The sequence logo generated via Skylign provided a visual summary of residue conservation. Taller letters in the logo correspond to amino acids that are more conserved across the species. The conserved residues observed at specific positions suggest functional or structural importance — potentially sites involved in oxygen binding or protein folding.

  Steps:

1. Upload MSA file to **Skylign**.
2. Generate sequence logo.



Observation:

Tall letters represent highly conserved residues, indicating important structural or functional roles. As shown in the logo, some residues like Histidine and Valine are highly conserved. This suggests they may be crucial for hemoglobin's oxygen-binding function or structural integrity.

Task 5: Phylogenetic Tree Construction

I used MEGA X software to construct a phylogenetic tree based on the MSA. The resulting tree shows the evolutionary relationships among the species based on their HBB sequences.

Description:

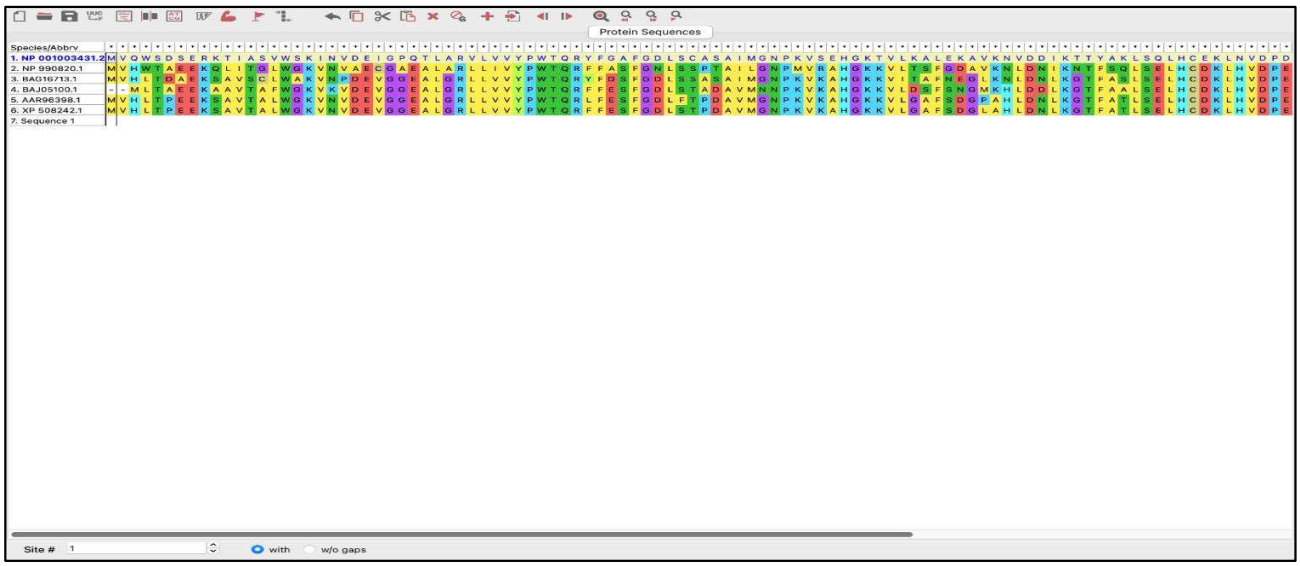
- Used MEGA X with the aligned sequences to generate a phylogenetic tree.
- Method: Neighbor-Joining

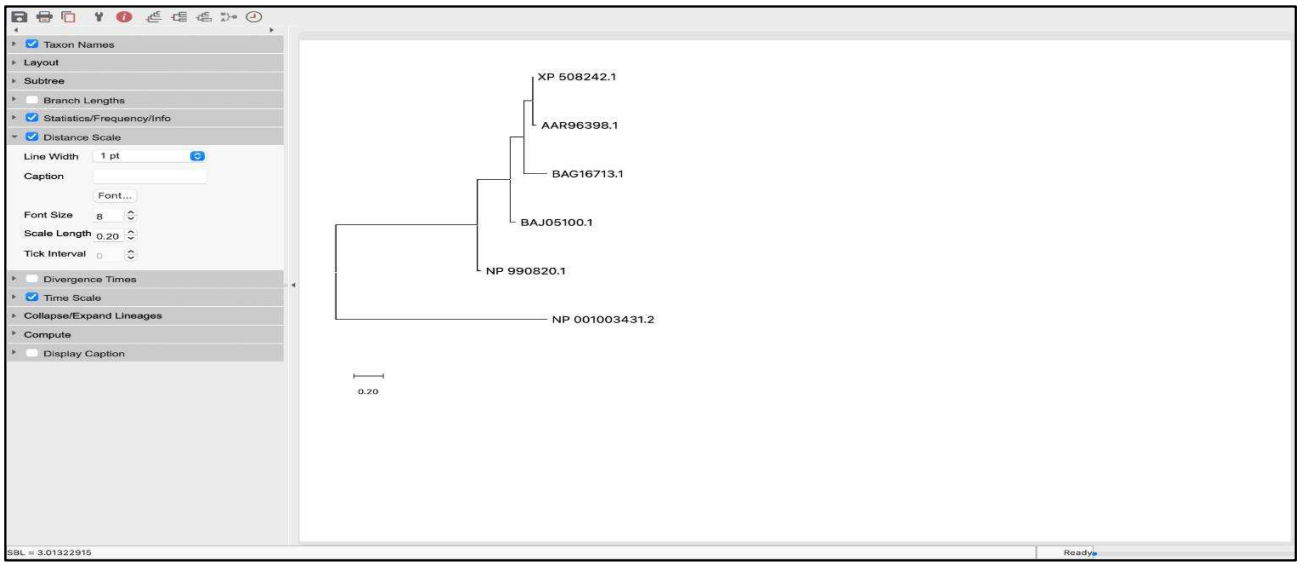
Steps:

1. Use **MEGA X** software.
2. Import the aligned file (MSA).
3. Construct a **phylogenetic tree** using the appropriate method (NJ or ML).

Interpretation:

- Chimpanzees are closest to humans.
- Zebrafish and chicken are most distant.
- Tree matches known evolutionary relationships.
- The tree clusters human and chimpanzee together, which is expected as they are closely related. Zebrafish, being the most distantly related vertebrate in this analysis, appears on a separate branch.





Conclusion

This project helped me explore how a crucial gene like HBB has evolved across species. The analysis showed that the HBB gene is highly conserved among mammals but shows significant divergence in non-mammalian species like zebrafish. These results align well with known evolutionary relationships and highlight the power of bioinformatics in evolutionary biology.

References

- NCBI GenBank and FASTA
- BLASTp – NCBI
- EMBOSS Needle
- Clustal Omega
- Skylign
- MEGA X Software