TRACKING THE EVOLUTION OF THE HEMOGLOBIN BETA (HBB) GENE ACROSS SPECIES

1. Sequence Retrieval and BLAST Search

hemoglobin subunit beta [Homo sapiens]

NCBI Reference Sequence: NP_000509.1

GenPept   Identical Proteins   Graphics

>NP_000509.1 hemoglobin subunit beta [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH

Species Name	Accession Number	% Identity with Human HBB
Chimpanzee (*Pan troglodytes)*	XP_508242.1	100.00%
Cow (*Bos taurus*)	NP_776342.1	84.72 %
Mouse (*Mus musculus*)	NP_001265090.1	80.27%
Chicken (*Gallus gallus*)	NP_990820.1	69.39 %
Zebrafish (*Danio rerio*)	NP_001013045.1	51.35 %

2. Pairwise Sequence Alignment

(a) Human vs Chimpanzee

```
NP_000509.1      1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS    50
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
XP_508242.1      1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS    50

NP_000509.1     51 TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD   100
                   |||||||||||||||||||||||||||||||||||||||||||||||||
XP_508242.1     51 TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD   100

NP_000509.1    101 PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147
                   |||||||||||||||||||||||||||||||||||||||||||||||
XP_508242.1    101 PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147


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

Identity: 100.0%

Similarity: 100.0%

Gaps: 0

Percentage identity and similarity is highest between humans and chimpanzees and lowest between humans and zebrafish. This indicates that humans and chimpanzees have had a closer evolutionary relationship compared to zebrafish.

(b) Human vs Zebrafish

```
NP_000509.1      1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS    50
                   ||..|..|::|:..||||:|:||:|.:||.|.|:|||||||:|.:||:||
NP_001013045.    1 MVEWTDAERTAILGLWGKLNIDEIGPQALSRCLIVYPWTQRYFATFGNLS    50

NP_000509.1     51 TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD   100
                   :|.|:|||||.|||:.|:|.....:.::||:|.|:|.||.:|.:|||||
NP_001013045.   51 SPAAIMGNPKVAAHGRTVMGGLERAIKNMDNVKNTYAALSVMHSEKLHVD   100

NP_000509.1    101 PENFRLLGNVLVCVLAHHFGKE-FTPPVQAAYQKVVAGVANALAHKYH    147
                   |:|||||.:.:....|..||:. |...||.|:||.:|.|.:||..:||
NP_001013045.  101 PDNFRLLADCITVCAAMKFGQAGFNADVQEAWQKFLAVVVSALCRQYH    148
```

Identity: 76/148 (51.4%)

Similarity: 106/148 (71.6%)

Gaps: 1/148 (0.7%)

3. Multiple Sequence Alignment

```
CLUSTAL O(1.2.4) multiple sequence alignment


NP_001013045.1      MVEWTDAERTAILGLWGKLNIDEIGPQALSRCLIVYPWTQRYFATFGNLSSPAAIMGNPK      60
NP_990820.1         MVHWTAEEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNPM      60
NP_001265090.1      MVHLTDAEKAAVSCLWGKVNSDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNAK      60
NP_000509.1         MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK      60
XP_508242.1         MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK      60
NP_776342.1         --MLTAEEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMNNPK      58
                      *  *:  :  :***::  * * :**.* *:*******:* :**:**:  *::.*
```

```
NP_001013045.1      VAAHGRTVMGGLERAIKNMDNVKNTYAALSVMHSEKLHVDPDNFRLLADCITVCAAMKFG     120
NP_990820.1         VRAHGKKVLTSFGDAVKNLDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIVLAAHFS     120
NP_001265090.1      VKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPENFRLLGNMIVIVLGHHLG     120
NP_000509.1         VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG     120
XP_508242.1         VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG     120
NP_776342.1         VKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFG     118
                     * ***:.*: .:  .: ::*.:*.*:: ** :*.:******:**:**.: :   . ::.
```

```
NP_001013045.1      QAGFNADVQEAWQKFLAVVVSALCRQYH         148
NP_990820.1         K-DFTPECQAAWQKLVRVVAHALARKYH         147
NP_001265090.1      K-DFTPAAQAAFQKVVAGVATALAHKYH         147
NP_000509.1         K-EFTPPVQAAYQKVVAGVANALAHKYH         147
XP_508242.1         K-EFTPPVQAAYQKVVAGVANALAHKYH         147
NP_776342.1         K-EFTPVLQADFQKVVAGVANALAHRYH         145
                    : *.   *  :**.: *. **.::**
```

4. Sequence Logo Generation
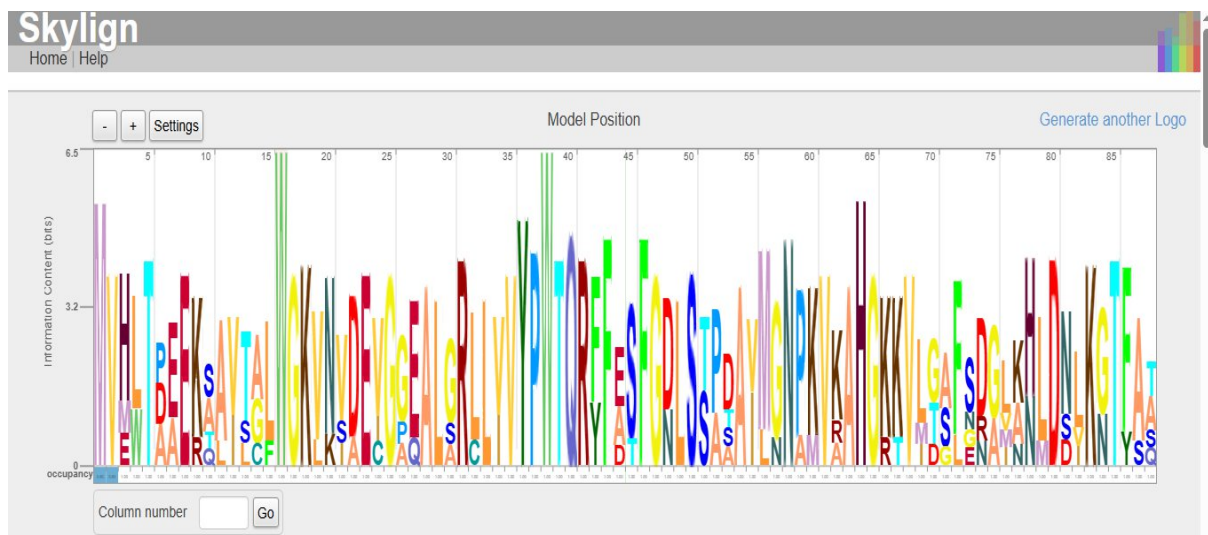
There are a number of highly conserved residues in certain regions This could indicate that these regions are of great importance either structurally or functionally.

5. Phylogenetic Tree Construction
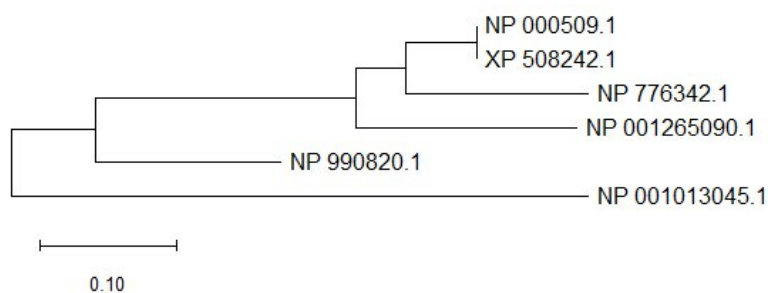
The species that are most closely related are human and chimpanzee. This indicates that they had a common ancestor in the course of evolution.