# A Comparative Study of Haemoglobin and Myoglobin Sequences in Selected Primates using BLAST and COBALT

**Raichel Theressa John, Risiya T R, Rosemary P Shaju, Deepthi K Dileep, Joyce Jose**
Department of Zoology, St. Thomas College (Autonomous) Thrissur.
*Corresponding author's email address: joyceofthejungle@gmail.com*

## ABSTRACT

Protein sequences of Haemoglobin (Hb) and Myoglobin (Mb) were downloaded from SWISSPROT and studied using bioinformatic tools BLAST and COBALT. It was found that Hb was much more conserved than Mb and that human sequences showed great similarity to Chimpanzees. Old World monkeys showed maximum differences. Phylogenetic trees constructed were consistent with previous studies which mapped the evolutionary route. Authors suggest a drastic overhaul of bioinformatic syllabi at practical and theory level so that students of zoology are equipped to use bioinformatic tools for their investigations in taxonomy, systematics, evolution etc.

**Keywords**: Myoglobin, Haemoglobin, BLAST, COBALT

## 1. INTRODUCTION

Globin proteins belong to the globin superfamily characterised by heme group, globular shape, the globin fold (a series of eight alpha helical segments). The globin gene family comprises of many related genes. The most well-known globins are Haemoglobin and Myoglobin, both of which can reversibly bind to oxygen via a heme prosthetic group.

There is more than one Hb gene in humans and close species, HbA, the main form of Hb, is coded for by the genes, HBA1, HBA2, and HBB. The amino acid sequences of the globin proteins in Hbs usually differ between species and these differences grow with evolutionary distance between species.

Myoglobin (Mb/MB) is found in the muscle tissue of almost all mammals. It is the primary oxygen-carrying pigment of muscle tissues. Concentration of Mb in muscle cells determine the ability of organisms to hold their breath for a longer period of time as in many aquatic mammals. Mb function relates to increased oxygen transport to muscle, oxygen storage and as a scavenger of reactive oxygen species.

Though both Hb and Mb are cytoplasm proteins that binds oxygen on a heme group, Mb is different as it has one heme group to Hb's four and this heme has higher affinity for oxygen than Hbs. Functionally the Hb transports oxygen while Mb stores oxygen.

A wide range of animals, use globins to transport and store oxygen. Hence it is natural to compare the structure and function of these proteins between species both to explore areas that vary and areas that are conserved. A comparative study can also give insights from the evolutionary analyses to better understand mechanisms of gene regulation. Since it was not financially, legally or materially possible for the investigators to isolate, purify and study globin sequences from living organisms, the protein sequences were sourced from SWISSPROT. Selection of proteins and species were constrained by the availability of these in the database.

In this context the objectives of the study was to locate and download haemoglobin and myoglobin protein sequences of selected primates from bioinformatics' databases; to compare the sequences and to construct a phylogenetic tree based on the similarities of these sequences using appropriate bioinformatics' tools

## 2. METHODOLOGY

HbA and Mb sequences of the following primates were located and downloaded from SWISSPROT and saved as '.txt' format.

| Scientific name | Common name |
|---|---|
| 1. *Homo sapiens* | Human Being |
| 2. *Pan troglodytes* | Common Chimpanzee |
| 3. *Macaca mulatta* | Rhesus Macaque |
| 4. *Gorilla gorilla gorilla* | Western lowland Gorilla |
| 5. *Semnopithecus entellus* | Northern plains Gray Langur |

The selection of species was based mainly on the availability of Hb A protein sequences and Mb sequences for all the animals. BLAST P (protein-protein BLAST) was used to obtain graphic summary, E values and S values and Distance trees showing similarities of all sequences. To construct phylogenetic trees, constrained based multiple alignment tool or COBALT was used.

## 3. RESULTS

**Amino acid sequences of Haemoglobin A for five selected primates in FASTA format**

***HOMO SAPIENS***

>sp|P69905|HBA_HUMAN Haemoglobin subunit alpha OS=*Homo sapiens* OX=9606 GN=HBA1 PE=1 SV=2

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR

***PAN TROGLODYTES***

>sp|P69907|HBA_PANTR Haemoglobin subunit alpha OS=*Pan troglodytes* OX=9598 GN=HBA1 PE=1 SV=2

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR

***MACACA MULATTA***

>sp|P63108|HBA_MACMU Haemoglobin subunit alpha OS=*Macaca mulatta* OX=9544 GN=HBA PE=1 SV=2

MVLSPADKSNVKAAWGKVGGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTLAVGHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR

***GORILLA GORILLA GORILLA***

>sp|P01923|HBA_GORGO Haemoglobin subunit alpha OS=*Gorilla gorilla gorilla* OX=9595 GN=HBA PE=1 SV=1

VLSPADKTNVKAAWGKVGAHAGDYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKG
HGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP
AEFTPAVHASLDKFLASVSTVLTSKYR

***SEMNOPITHECUS ENTELLUS***

>sp|P01924|HBA_SEMEN Haemoglobin subunit alpha OS=*Semnopithecus entellus* OX=88029 GN=HBA PE=1 SV=1

VLSPADKTNVKAAWGKVGGHGGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKG
HGKKVADALTNAVAHVDDMPHALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP
AEFTPAVHASLDKFLASVSTVLTSKYR

**Amino acid sequences of Myoglobin for five selected primates in FASTA format**

***HOMO SAPIENS***

>sp|P02144|MYG_HUMAN Myoglobin OS=*Homo sapiens* OX=9606 GN=MB PE=1 SV=2
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
SEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSK
HPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

***PAN TROGLODYTES***

>sp|P02145|MYG_PANTR Myoglobin OS=*Pan troglodytes* OX=9598 GN=MB PE=1 SV=2
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
SEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLHSK
HPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

***SEMNOPITHECUS ENTELLUS***

>sp|P68085|MYG_SEMEN Myoglobin OS=*Semnopithecus entellus*OX=88029 GN=MB PE=1 SV=2
MGLSDGEWQLVLNVWGKVEADIPSHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
SEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLELISESIIQVLQSK
HPGDFGADAQGAMNKALELFRNDMAAKYKELGFQG

***MACACA MULATTA***

>tr|F7ARV1|F7ARV1_MACMU Myoglobin OS=*Macaca mulatta* OX=9544 GN=MB PE=3 SV=1
MGLSDGEWQLVLNVWGKVEADIPSHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
SEDLKKHGVTVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLELISESIIQVLQSK
HPGDFGADAQGAMNKALELFRNDMAAKYKELGFQG

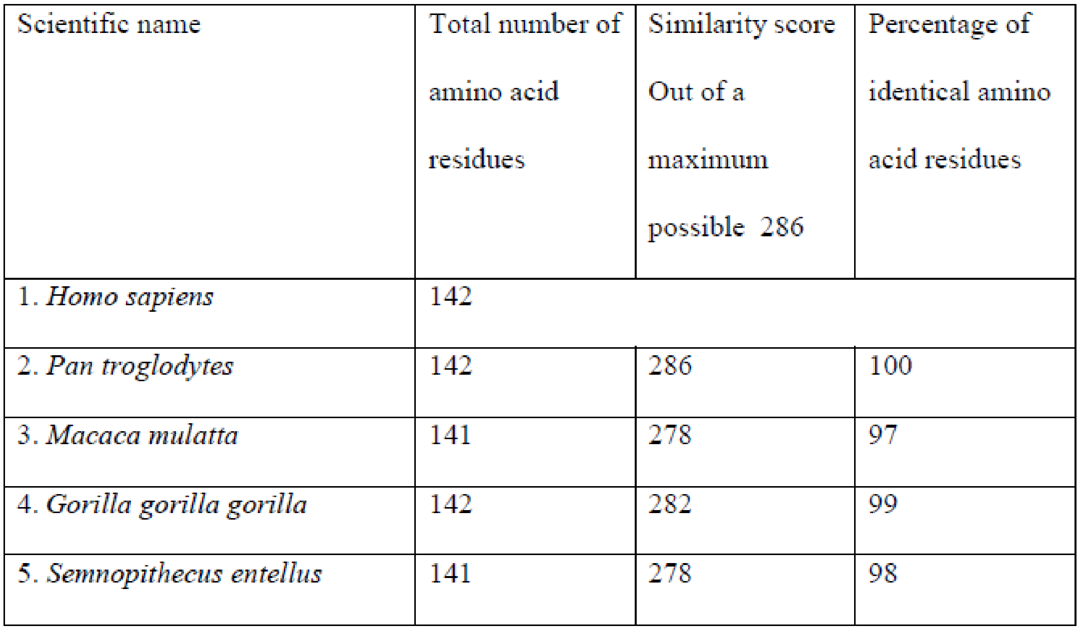
***GORILLA GORILLA GORILLA***

>tr|G3R764|G3R764_GORGO Myoglobin OS=*Gorilla gorilla gorilla* OX=9595 GN=MB PE=3 SV=1

MGLSDGEWQLVLNVWGKVEADISGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKA
SEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSK
HPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

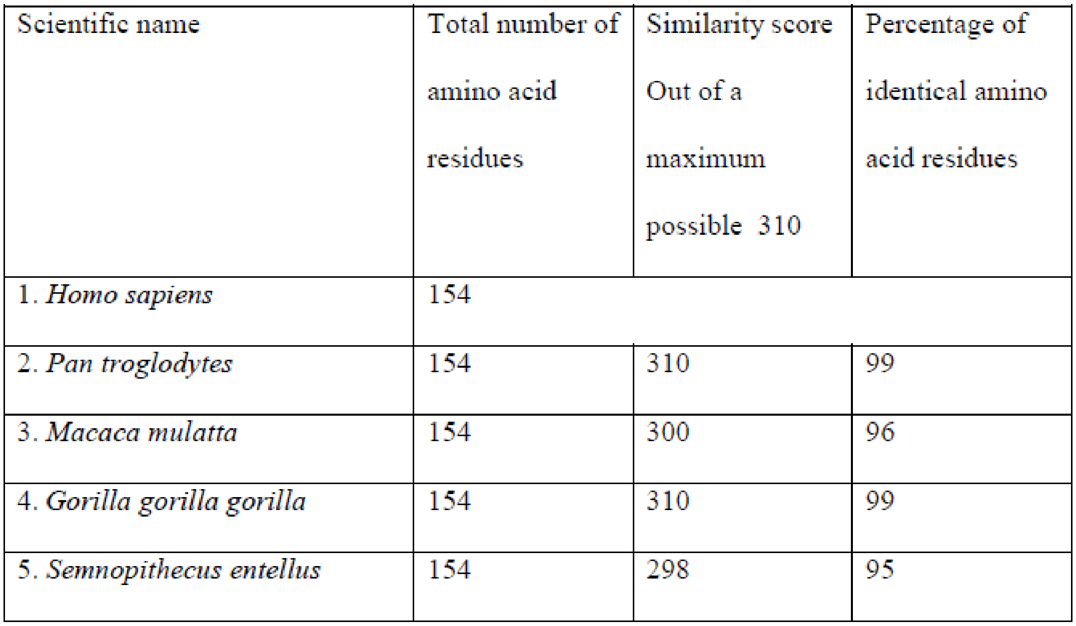
**Table 1: Explanation of the BLAST results for Hb and Mb**

**Haemoglobin**



| Scientific name | Total number of amino acid residues | Similarity score Out of a maximum possible 286 | Percentage of identical amino acid residues |
|---|---|---|---|
| 1. *Homo sapiens* | 142 | | |
| 2. *Pan troglodytes* | 142 | 286 | 100 |
| 3. *Macaca mulatta* | 141 | 278 | 97 |
| 4. *Gorilla gorilla gorilla* | 142 | 282 | 99 |
| 5. *Semnopithecus entellus* | 141 | 278 | 98 |

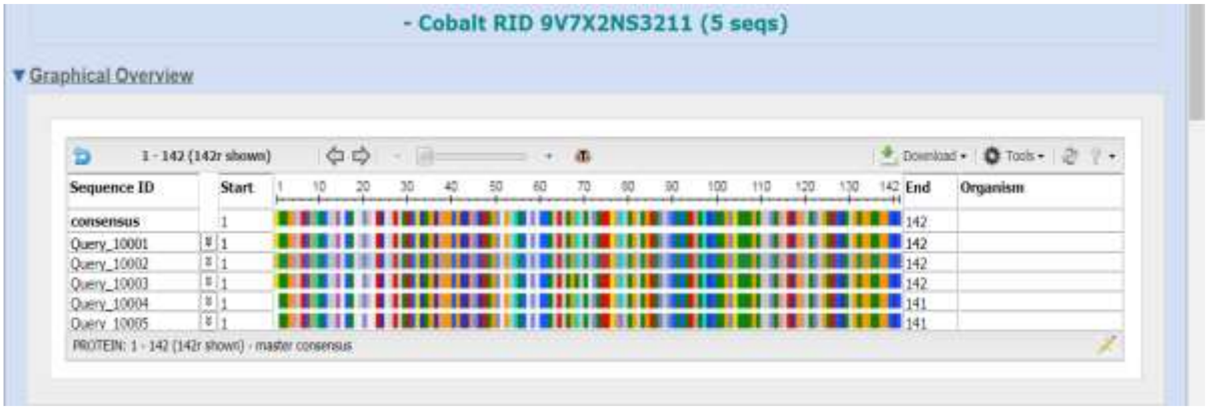
**Myoglobin**



| Scientific name | Total number of amino acid residues | Similarity score Out of a maximum possible 310 | Percentage of identical amino acid residues |
|---|---|---|---|
| 1. *Homo sapiens* | 154 | | |
| 2. *Pan troglodytes* | 154 | 310 | 99 |
| 3. *Macaca mulatta* | 154 | 300 | 96 |
| 4. *Gorilla gorilla gorilla* | 154 | 310 | 99 |
| 5. *Semnopithecus entellus* | 154 | 298 | 95 |

The Hb A sequence of *Pan troglodytes* (100%) showed maximum similarity to that of *Homosapiens* followed by *Gorilla gorilla gorilla* (99%). *Macaca mulatta* and *Semnopithecus entellus* had one residue less and showed 97% and 98% of similarity. *Pan troglodytes* had a maximum of 286out of 286 scores. All the sequences showed highly significant E scores (Table 1). Myoglobin sequences of all the five species have 154 residues. Both *Pan troglodytes* and *Gorilla gorilla gorilla* scored a maximum of 310 out of 310 while *Macaca mulatta* and *Semnopithecus entellus* had 300 and 298 respectively. Percentage of identical residues was 99% for *Pan troglodytes* and *Gorilla gorilla gorilla* and 96% and 95% for *Macaca mulatta* and *Semnopithecus entellus* respectively. All sequences showed highly significant E scores. The number of amino acid residues differed among the five species in

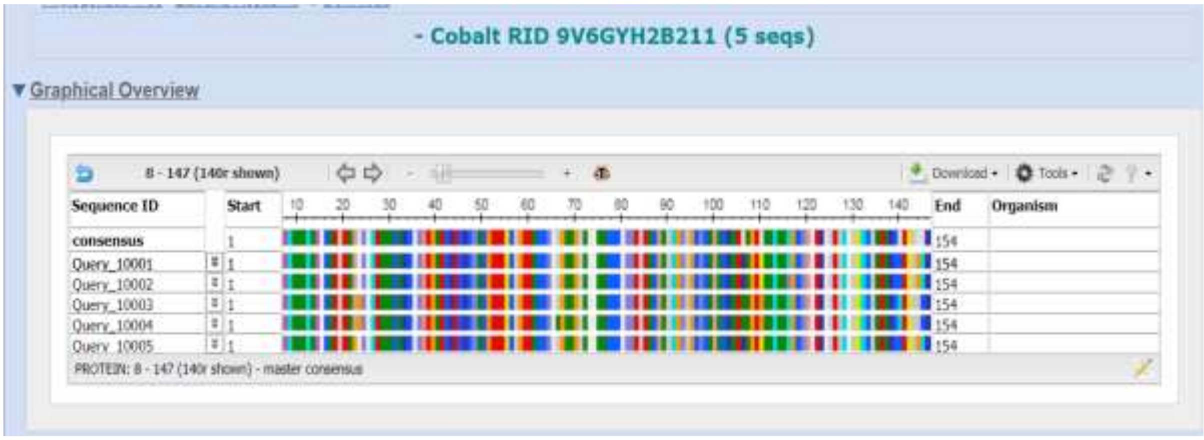
Hb A while in Mb the number of amino acid residues were the same, but overall similarity was higher in HbA when compared to Mb as is evident in Table 1.



| Sequence ID | Start | End | Organism |
|---|---|---|---|
| consensus | 1 | 142 | |
| Query_10001 | 1 | 142 | |
| Query_10002 | 1 | 142 | |
| Query_10003 | 1 | 142 | |
| Query_10004 | 1 | 141 | |
| Query_10005 | 1 | 141 | |

PROTEIN: 1 - 142 (142r shown) - master consensus

Figure 1. Colour coded map of amino acid residues obtained from COBALT multi sequencing for HbA of five selected primate species



| Sequence ID | Start | End | Organism |
|---|---|---|---|
| consensus | 1 | 154 | |
| Query_10001 | 1 | 154 | |
| Query_10002 | 1 | 154 | |
| Query_10003 | 1 | 154 | |
| Query_10004 | 1 | 154 | |
| Query_10005 | 1 | 154 | |

PROTEIN: 8 - 147 (140r shown) - master consensus

Figure 2. Colour coded map of amino acid residues obtained from COBALT multi sequencing for Mb of five selected primate species

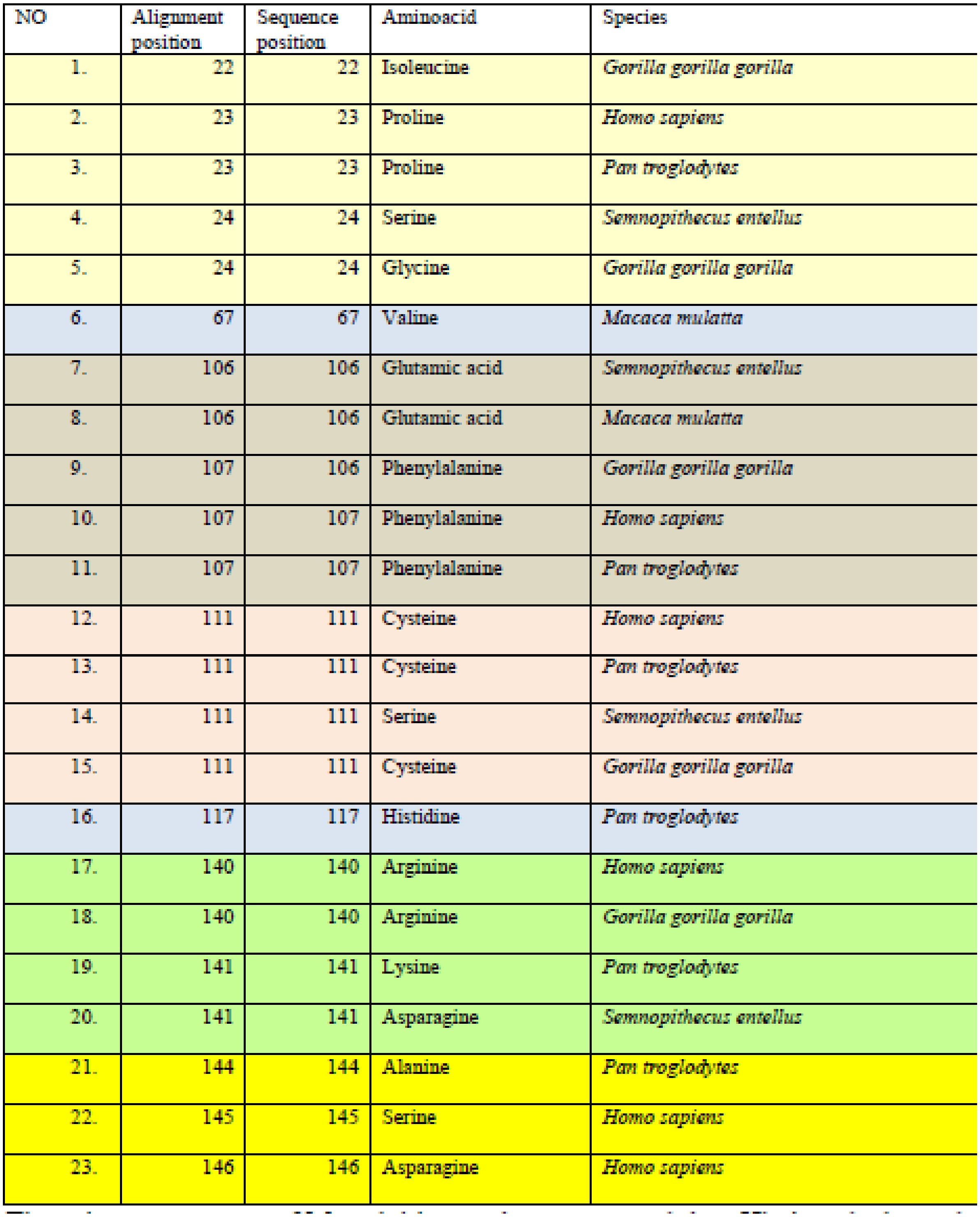
The primary structure of HbA was very conserved with only 11 residues showing differences (Fig 1 and Table 2) The primary structure of Mb was less conserved than Hb A and showed differences at 23 positions (Fig 2), the details of which are given in Table 2.

Table: 2. Summary of amino acid residue differences in HbA and Mb for five selected primate species.

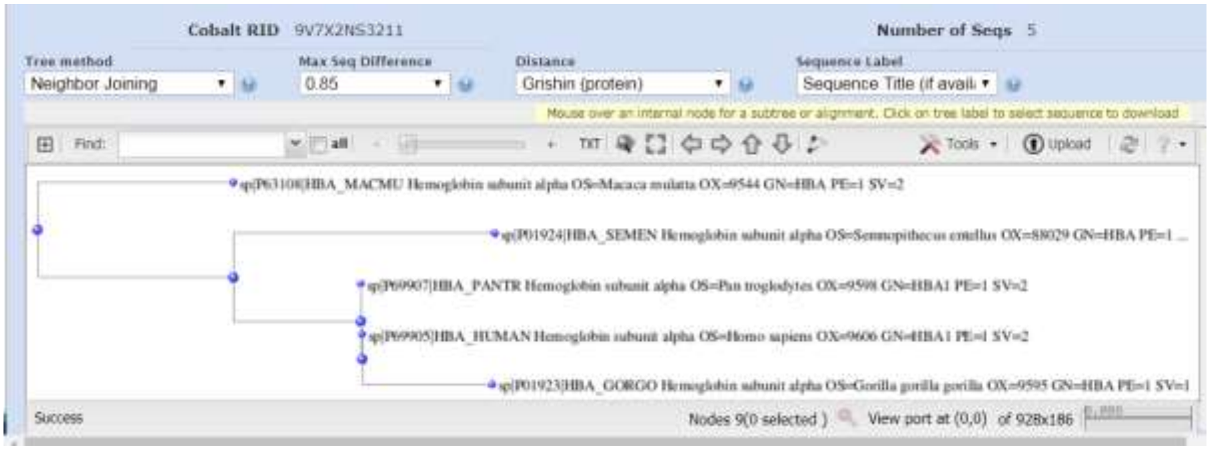
**Haemoglobin**

| No | Alignment position | Sequence position | Amino acids | Species |
|---|---|---|---|---|
| 1. | 9 | 9 | Serine | *Macaca mulatta* |
| 2. | 19 | 19 | Glycine | *Homo sapiens* |
| 3. | 20 | 19 | Alanine | *Gorilla gorilla* |
| 4. | 20 | 19 | Glycine | *Semnopithecus entellus* |
| 5. | 20 | 20 | Glycine | *Macaca mulatta* |
| 6. | 20 | 20 | Alanine | *Pan troglodytes* |
| 7. | 22 | 21 | Glycine | *Semnopithecus entellus* |
| 8. | 23 | 22 | Glycine | *Gorilla gorilla* |
| 9. | 69 | 69 | Leucine | *Macaca mulatta* |
| 10. | 72 | 72 | Glycine | *Macaca mulatta* |
| 11. | 79 | 78 | Histidine | *Semnopithecus entellus* |

## Myoglobin



| NO | Alignment position | Sequence position | Aminoacid | Species |
|---|---|---|---|---|
| 1. | 22 | 22 | Isoleucine | *Gorilla gorilla gorilla* |
| 2. | 23 | 23 | Proline | *Homo sapiens* |
| 3. | 23 | 23 | Proline | *Pan troglodytes* |
| 4. | 24 | 24 | Serine | *Semnopithecus entellus* |
| 5. | 24 | 24 | Glycine | *Gorilla gorilla gorilla* |
| 6. | 67 | 67 | Valine | *Macaca mulatta* |
| 7. | 106 | 106 | Glutamic acid | *Semnopithecus entellus* |
| 8. | 106 | 106 | Glutamic acid | *Macaca mulatta* |
| 9. | 107 | 106 | Phenylalanine | *Gorilla gorilla gorilla* |
| 10. | 107 | 107 | Phenylalanine | *Homo sapiens* |
| 11. | 107 | 107 | Phenylalanine | *Pan troglodytes* |
| 12. | 111 | 111 | Cysteine | *Homo sapiens* |
| 13. | 111 | 111 | Cysteine | *Pan troglodytes* |
| 14. | 111 | 111 | Serine | *Semnopithecus entellus* |
| 15. | 111 | 111 | Cysteine | *Gorilla gorilla gorilla* |
| 16. | 117 | 117 | Histidine | *Pan troglodytes* |
| 17. | 140 | 140 | Arginine | *Homo sapiens* |
| 18. | 140 | 140 | Arginine | *Gorilla gorilla gorilla* |
| 19. | 141 | 141 | Lysine | *Pan troglodytes* |
| 20. | 141 | 141 | Asparagine | *Semnopithecus entellus* |
| 21. | 144 | 144 | Alanine | *Pan troglodytes* |
| 22. | 145 | 145 | Serine | *Homo sapiens* |
| 23. | 146 | 146 | Asparagine | *Homo sapiens* |



Cobalt RID 9V7X2NS3211

Number of Seqs 5

Tree method: Neighbor Joining

Max Seq Difference: 0.85

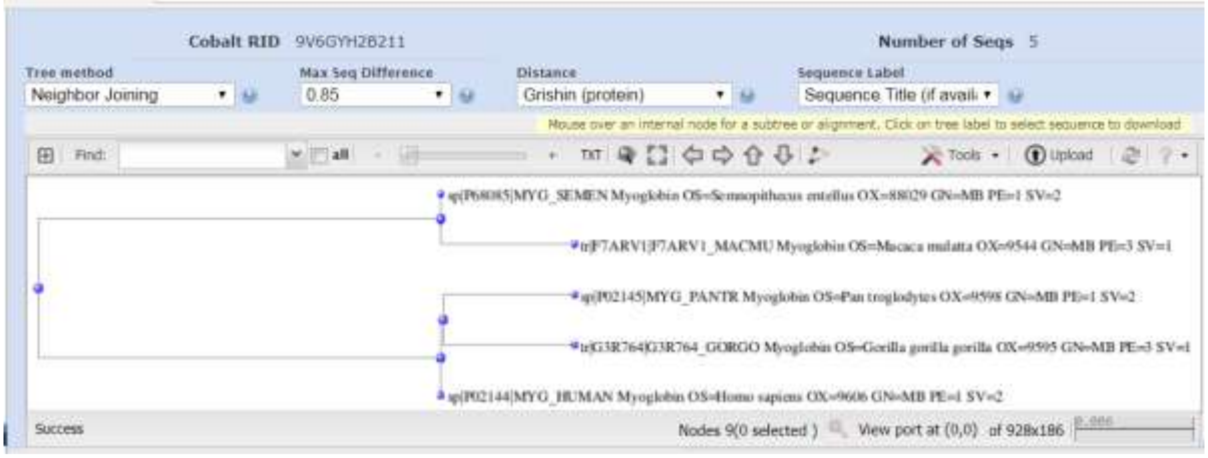
Distance: Grishin (protein)

Sequence Label: Sequence Title (if avail

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Success Nodes 9(0 selected ) View port at (0,0) of 928x186

Fig 3 Phylogenetic tree obtained from COBALT based on Hb A of five selected primates



Cobalt RID 9V6GYH2B211

Number of Seqs 5

Tree method: Neighbor Joining

Max Seq Difference: 0.85

Distance: Grishin (protein)

Sequence Label: Sequence Title (if avail

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Success Nodes 9(0 selected ) View port at (0,0) of 928x186

Fig 4. Phylogenetic tree obtained from COBALT based on Myoglobin of five selected primates

Phylogenetic trees obtained using COBALT (Figure 3) indicate that in the course of evolution of Hb A, *Macaca mulatta* was the first break off event from a common ancestor followed by *Semnopithecus entellus*.*Gorilla gorilla gorilla* broke off from the common ancestor at a much later date followed by *Homo sapiens* and *Pan troglodytes* both of which had exactly similar primary structures for the HbA protein as already vouched by pairwise alignment scores. Phylogenetic trees obtained for Mb using COBALT (Figure 4) indicate that the evolution of Mb was much different when compared to Hb A. Both *Semnopithecus entellus* and *Macaca mulatta* break off from the common ancestor at a very early date and form natural clusters. *Homo sapiens*, *Pan troglodytes* and *Gorilla gorilla gorilla* myglobins' forms the next group. Clusters formed using COBALT and CLUSTAL OMEGA suggested different results.

## 4. DISCUSSION AND CONCLUSION

The primary structure of Hb A was very conserved and showed less variations when compared to that of Myoglobin in the study organisms. It could be because the function and the degree of function may not vary much among organisms. Myoglobin primary structure was also highly conserved but showed more variations. This could be because the function of storing oxygen in muscles differ in the selected organisms based on their habit, habitat and other physiological factors which have not been examined in this study. It was observed that glycine was the amino acid which caused the differences at six instances followed by Alanine twice. Species wise it was *Macaca mulatta* (4) and *Semnopithecus entellus* (3) which showed maximum differences with HbA of *Homo sapiens*. In Myoglobin the amino acid residues which caused maximum variations were Cysteine, Phenylalanine and Serine which replaced other residues three times each. Species wise *Macaca mulatta* and *Semnopithecus entellus* were most different from the other three primates. Sequencing of HbA and Myoglobin indicate that Human beings and Chimpanzees are very similar to the extet that the primary structure of two prominent globin proteins are almost same in both the primates. Predictably Grey Langaur and Rhesus macaque, both being old world monkeys show greater similarity to each other.

Individual phylogenetic trees created using the HbA and Myoglobin sequences of the five selected species also concur with previous studies (Barnabas et al 1971), (Goodman et al 1983) and Gribaldo et al 2003, that humans are closely related to African apes especially the chimpanzee. Previous studies also concur that Gorilla due to its genetic similarity with humans and Chimpanzees had been excluded from another ape group such as gibbons and orang-utans. Humans and African apes have distinct though similar amino acid conserved and non-conserved areas with old world monkeys like Langurs and Macaques.

Based on the experience garnered in this project the authors suggest that the theory syllabus and prescribed texts on bioinformatics for final year BSc Zoology are woefully outdated and should be revised frequently to keep students abreast of the latest developments. Hands on practical sessions would also be very useful if students are to apply bioinformatic tools for the study of taxonomy, systematic, evolution and other conventional subjects.

**Acknowledgements**

The Authors thank the Head of the Department of Zoology and the College for providing necessary facilities for doing the project.

## REFERENCES

Barnabas J, Goodman M and Moore G W. (1971). Evolution of Haemoglobin in primates and other therian mammals. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*. 39 ( 3) 455-482

Gribaldo, S, Didier C, Philippe L and Herve P. (2003). Functional Divergence Prediction from Evolutionary Analysis: A Case Study of Vertebrate Haemoglobin. *Molecular Biology and Evolution* 20 (11): 1754-1759.

Goodman M, Braunitzer G, Stangl A and Schrank B. (1983). Evidence on human origins from haemoglobins of African apes. *Nature.* 303(5917):546-8.

### Web resources used

1. https://www.ncbi.nlm.nih.gov/ accessed during the month of February and March 2018
2. https://www.ncbi.nlm.nih.gov/pmc/ accessed during the month of February and March 2018
3. https://blast.ncbi.nlm.nih.gov/Blast.cgi accessed on 20/02/2018
4. http://www.clustal.org/ accessed on 21/02/2018
5. http://www.uniprot.org/ accessed on 21/02/2018