doi:10.20944/preprints201807.0330.v1

PHYLOGENETIC ANALYSIS OF GLOBIN LIKE SUPERFAMILY AT MOLECULAR LEVEL

Himani malhotra & Harmanpreet Kaur Sanghu

Department of Bioinformatics and Biotechnology, Guru Nanak Girls College, Model Town, Ludhiana, Punjab

Email:himanimalhotra89@gmail.com, preetharman140394@gmail.com

ABSTRACT

Globin-like superfamily constitutes a superfamily of both extracellular and intracellular proteins that are found in 3 kingdoms of life. Globins are respiratory proteins that usually bind an oxygen molecule .Earlier, phylogenetic study of globin like superfamily was based on morphological character, however, molecular basis of their functional diversity is not completely understood. The present study focuses on the phylogenetic analysis of globin at molecular level. Firstly, proteins of SCOP families of globin like superfamily were retrieved from NCBI and then similar proteins are retrieved from PSI BLAST results and then pfam was used to predict the domains and two domains: Globin and Phycobilisome were analysed. Other analysis predicted the localization in cytoplasm of most of the proteins. The nucleotide sequences of same species for protein analysis were retrieved from NCBI, and phylogenetic analysis of aligned fasta files of proteins and nucleotides by using MEGA at molecular level was performed and analysis shows the difference on taxonomic tree and phylogenetic tree.

Keywords: Globin-like superfamily, Phylogenetic analysis, Taxonomic tree

1 INTRODUCTION

The proteins belong to globin like superfamily are present in three kingdoms of life. The globin like superfamily belongs to class of all alpha proteins and Fold Globin-like which is having 6 helices, folded leaf and partly open structure. The globin like superfamily has 4 families: Truncated Hemoglobin, Neural globins, Globins, Phycocyanin like phycobilisome proteins. Hemoglobin-like proteins found in eubacteria, cyanobacteria, protozoa, algae and plants, but not in animals or yeast. These proteins have a truncated 2-over-2 rather than the conical 3-over-3 alpha-helical sandwich fold. They lack first helix so, different from other families and 20-40 amino acids shorter than other globins. Neural globins are vertebrate globins present mainly in neural tissues. This also lack first helix but are more similar to other globins than the truncated ones and belongs to class all alpha proteins *(Giangiacome Laura et al 2005).*

Globins are Heme binding proteins. These are rich in alpha helix and surrounded by heme group. The globin fold typically consists of eight alpha helicas only. So belongs to class all alpha proteins. As these are respiratory proteins so, bind an oxygen molecule between iron of porphyrin ring and histidine of polypeptide chain(Xie LK *et al* 2016). The main subfamilies of globin family are cytoglobin, myoglobin, leghemoglobin, erythrocruorin. Myoglobin consists of 154 amino acid and contain single chain alpha helical content and also contain the heme prosthetic group. Cytoglobin is hexa-coordinated globin but hemoglobin and myoglobin are penta coordinated globins and is 25% identical to Myoglobin in aspect of amino acid similarity. Leghemoglobin is Oxygen binding heme protein, present in plants for nitrogen fixation and having molecular weight of 15,400 and 16800.Leghemoglobin (Lb) is a myoglobin-like protein of about 16 kDa, which occurs in legume root nodules at very high concentrations. These found in root nodules of plants and are alpha proteins.

Allophycocyanin belongs to phycocyanin like phycobilisome proteins which have oligomers of two different types of globin like subunits containing two extra helices at the N- terminus bind a bilin chromophore. Allophycocyanin have alpha and beta subunits, which has a single non-cyclic tetrapyrrole chromophore called phycocyanin. Functional annotation of Globin-like superfamily is as they contain heme prosthetic group so mainly help in respiration of living organism. Leghemoglobin helps in nitrogen fixation with help of root nodules in plants. Molecular data that are in the form of DNA or protein sequences can provide very useful evolutionary perspectives of existing organisms because, as organisms evolve, the genetic materials accumulate mutations over time causing phenotypic changes. Phylogenetic analysis of proteins give evolutionary information on the basis of proteins but taxonomical tree is based on the evolution on the basis of morphological characters (Balasubramanian L *et al* 2015).

Here, we carried out an extensive analysis to explain the phylogenetic relationships of proteins and nucleotides and comparing the phylogenetic tree with taxonomical tree. Firstly, we take the families under globin-like superfamily from SCOP and extracted five proteins sequence under globin super family from NCBI and PSI BLAST is done of every protein. Then we got eleven proteins in different species from PSI BLASTresult of each protein and saved their nucleotide sequences also.We made phylogenetic tree of them using software MEGA7 .Then we predicted domains of these 55 proteins of globin superfamily by using Pfam tool. Later, we analyzed the physiochemical properties using Protparam tool. The physiochemical properties analysed are number of amino acids, theoretical Pi, aliphatic index, instability index, molecular weight, GRAVY. Finally PSORT II was used to find out the location of hypothetical proteins for complete functional annotation of proteins of Globin-like superfamily.Here most of the proteins are located in cytoplasm and 1a few are in mitochondria. Then comparisons of the phylogenetic trees are done with taxonomical tree.

2 MATERIALS AND METHODS:

2.1 Search for Globin-like family and superfamily in SCOP:

Open SCOP homepage www.scop.mrc-lmb.cam.ac.uk. The database is accessible on World Wide Web (WWW) with an entry point to URL http: parallel scop.mrc-lmb.cam.ac.uk magnitude of SCOP. This database provides a detailed and comprehensive description of the structural and evolutionary relationships of the proteins of known structure (*Murzin AG et al,1995).* It also provides for each entry links to co-ordinates, images of the structure, interactive viewers, sequence data and literature references. The classification is on hierarchical levels: the first two levels, family and superfamily, describe near and distant evolutionary relationships; the third, fold, describes geometrical relationships. The key word search finds, for a word entered by the user, matches from both the text of the scop database and the headers of Brookhaven Protein Databank structure files. Search for the Globin in scop database and then click on the Globin-Like superfamily. Result shows the information about families and sub families of Globin-Like superfamily. 5 subfamilies i.e.Truncated Hemoglobin, Myoglobin, Cytoglobin, Leghemoglobin, Allophycocyanin have been identified. The protein fasta sequence for the subfamilies of families of Globin-Like superfamily is retrieved from NCBI.

2.2 SEARCH FOR SEQUENCES AND RUN PSI-BLAST: Open NCBI homepage www.ncbi.nlm.nih.gov. The national centre for biotechnology information (NCBI) www.ncbi.nlm.nih.gov serves as an international resource for scientific research community providing access to public databases and software tools for analyzing biological data, as well as performing research in computational biology. Genbank coordinates with individual laboratories and other sequence databases such as those of the European molecular biology laboratory (EMBL) and the DNA databank of Japan (Lindberg 2000).Search for the protein sequence of 5 subfamilies searched from SCOP database i.e. Truncated Hemoglobin, Myoglobin, Cytoglobin, Leghemoglobin, Allophycocyanin. Run PSI (Poition specific iteration) blast for every protein sequence of each subfamily.Results shows similar protein sequence of different species of each subfamily. Save fasta sequence for 11 proteins of each subfamily. The fasta sequence for 55 proteins are saved as separate fasta file for each each subfamily. Open NCBI homepage www.ncbi.nlm.nih.gov. Search for nucleotide sequence for the same species as proteins sequences. Save fasta sequence for 5 species from every subfamily as seprate fasta file for each subfamily.

2.3 IDENTIFICATION OF DOMAIN: Open Pfam tool homepage from www.pfam.xfam.org/ which is a comprehensive collection of protein domains and families, represented as multiple sequence alignments and as profile hidden Markov models. Pfam is a manually curated database. The current release of Pfam contains 9318 protein families. Pfam is available on the web from the consortium members using a new, consistent and improved website design in the UK (http://pfam.sanger.ac.uk/), the USA (http://pfam.janelia.org/) and Sweden (http://pfam.sbc.su.se/), as well as from mirror sites in France (http://pfam.jouy.inra.fr/) and South Korea (http://pfam.ccbb.re.kr/) .pfam is comprehensive database of conserved protein families. Click on option search by sequence, then paste the protein sequence in given box to find the matching pfam entries. Result of domain search for each sequence will appear.

2.4 PREDICTION OF PHYSIOCHEMICAL PROPERTIES:

ProtParamtool (http://web.expasy.org/protparam/) computes various physico-chemical properties that can be deduced from a protein sequence. ProtParam computes various physico-chemical properties that can be deduced from a protein sequence. (Wilkins MR *et a,1999)l* No additional information is required about the protein under consideration. Open the protparam homepage www.expasy.org/protparam. Paste the protein sequence in given box.The result showed physiochemical properties of protein.

2.5 PREDICTION OF SUBCELLULAR LOCALIZATION : Open the homepage of PSORTII www.psort.hgc.jp/form2.html.Paste the each protein sequence in search box and Click on submit option. Result page will be displayed. Results predict the protein subcellular localization of each protein sequence of each subfamily.

2.6.1 PHYLOGENETIC ANALYSIS: Open offline software of MEGA. Click on file and open saved fasta file of protein sequences. Click on the align option. Save the alignment file as .mas. Click on phylogeny and select option construct neighbor joining tree. Select the saved alignment file and Click on compute. Result shows the phylogenetic tree of query protein. Click on file and open saved fasta file of protein sequences.

2.6.2 Open offline software of MEGA. Click on file and open saved fasta file of nucleotide sequences.Click on the align option and Save the alignment file as mas. Click on phylogeny and select option construct neighbor joining tree.Select the saved alignment file. Click on compute and Result shows the phylogenetic tree of query nucleotide.

3 RESULTS:

Structural Classification of Proteins

Superfamily: Globin-like

Superfamily

Lineage:

1. Root: scop
2. Class: All alpha proteins [46456]
3. Fold: Globin-like [46457]
   core: 6 helices; folded leaf, partly opened
4. Superfamily: Globin-like [46458]
   Superfamily

Families:

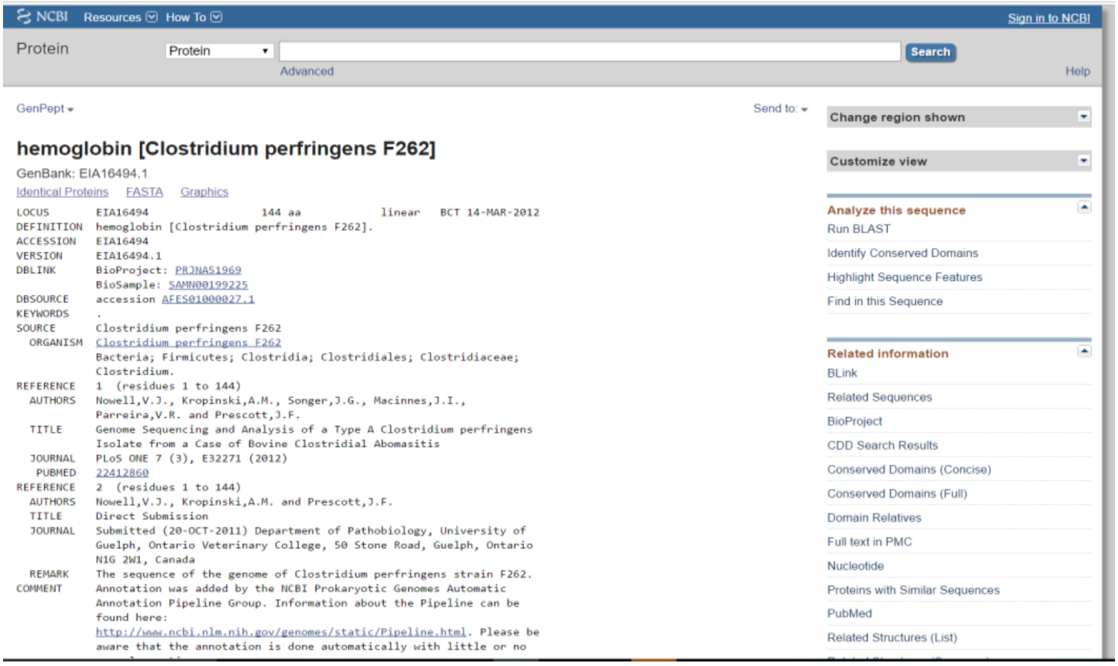
1. Truncated hemoglobin [46459] (6)
   lack the first helix (A)
2. Nerve tissue mini-hemoglobin (neural globin) [74660] (1)
   lack the first helix but otherwise is more similar to conventional globins than the truncated ones
3. Globins [46463] (81)
   Heme-binding protein
4. Phycocyanin-like phycobilisome proteins [46532] (26)
   oligomers of two different types of globin-like subunits containing two extra helices at the N-terminus
   binds a bilin chromophore

Enter search key: Search

Generated from scop database 1.75 with scopm 1.101 on Wed Jun 3 10:42:06 2009
Copyright © 1994-2009 The scop authors / scop@mrc-lmb.cam.ac.uk

FIGURE 1: Output of SCOP showing the families under the Globin-like superfamily

From SCOP we retrieved the information about the families and subfamilies belongs to Globin-Like superfamily .



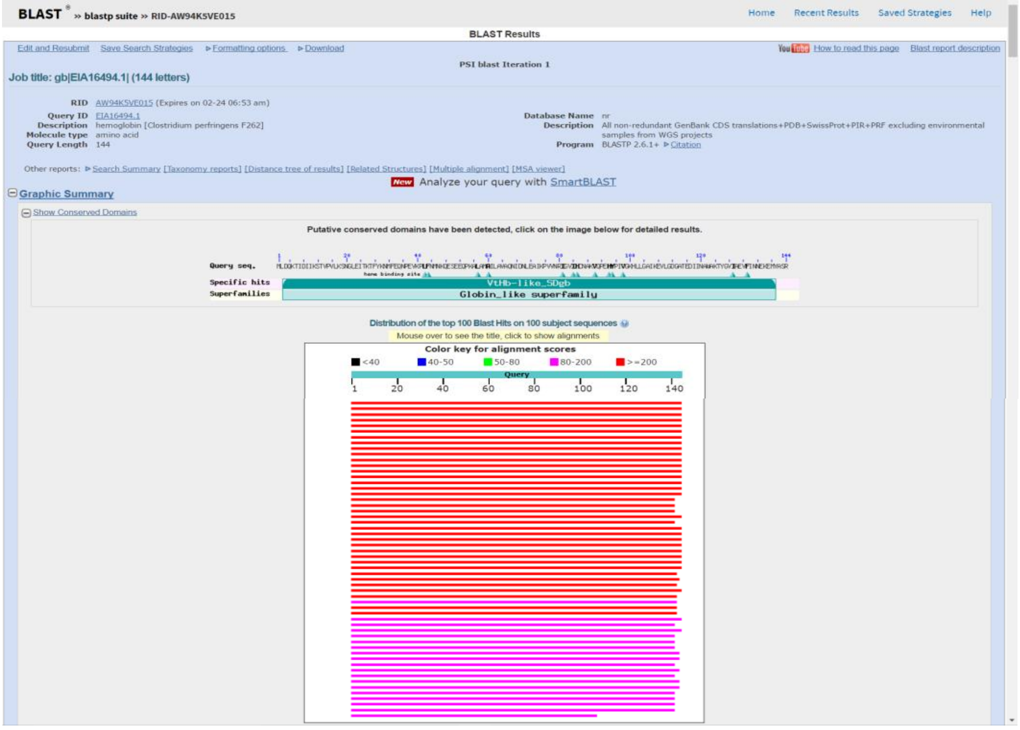
hemoglobin [Clostridium perfringens F262]

GenBank: EIA16494.1

Identical Proteins FASTA Graphics

```
LOCUS       EIA16494                 144 aa            linear   BCT 14-MAR-2012
DEFINITION  hemoglobin [Clostridium perfringens F262].
ACCESSION   EIA16494
VERSION     EIA16494.1
DBLINK      BioProject: PRJNA51969
            BioSample: SAMN00199225
DBSOURCE    accession AFES01000027.1
KEYWORDS    .
SOURCE      Clostridium perfringens F262
  ORGANISM  Clostridium perfringens F262
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (residues 1 to 144)
  AUTHORS   Nowell,V.J., Kropinski,A.M., Songer,J.G., Macinnes,J.I.,
            Parreira,V.R. and Prescott,J.F.
  TITLE     Genome Sequencing and Analysis of a Type A Clostridium perfringens
            Isolate from a Case of Bovine Clostridial Abomasitis
  JOURNAL   PLoS ONE 7 (3), E32271 (2012)
   PUBMED   22412860
REFERENCE   2  (residues 1 to 144)
  AUTHORS   Nowell,V.J., Kropinski,A.M. and Prescott,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2011) Department of Pathobiology, University of
            Guelph, Ontario Veterinary College, 50 Stone Road, Guelph, Ontario
            N1G 2W1, Canada
  REMARK    The sequence of the genome of Clostridium perfringens strain F262.
COMMENT     Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group. Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
```

FIGURE 2: Showing the NCBI result of qyery protein hemoglobin in *Clostridium perfringes*



PSI blast Iteration 1

Job title: gb|EIA16494.1| (144 letters)

RID AW94K5VE015 (Expires on 02-24 06:53 am)
Query ID EIA16494.1
Description hemoglobin [Clostridium perfringens F262]
Molecule type amino acid
Query Length 144

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.6.1+ Citation

Putative conserved domains have been detected, click on the image below for detailed results.

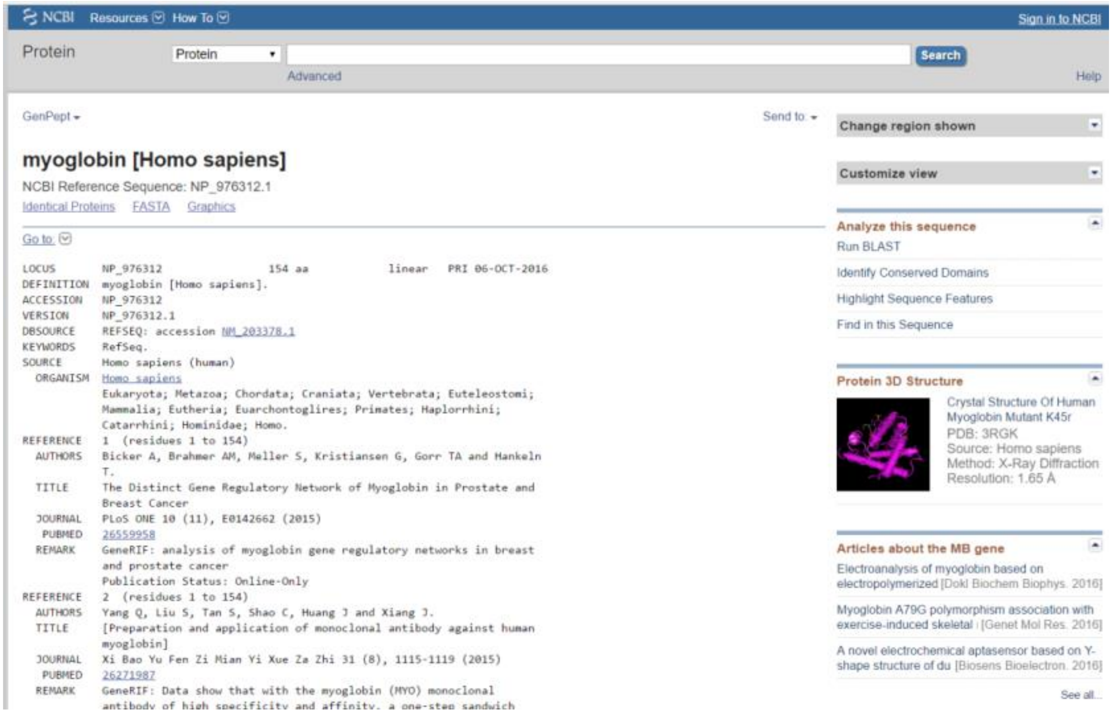
FIGURE 3: output of PSI BLAST showing alignment of query sequence

Here, red indicate almost similar sequences, pink indicates less similar sequence.From here the similar proteins with identity value of similar proteins ranging from70% to 97% were selected depending upon their e-value .

TABLE 1: Showing the results of PSI BLAST of protein Hemoglobin in *clostridium perfringes*

PROTEIN	**e-VALUE**	**IDENTITY**	**ACCESSION NO.**
Bacitracin resistance protein BacA *[Romboutsia timonensis]*	3e-69	72%	WP_071121120
Bacitracin resistance protein BacA *[Intestinibacter bartlettii]*	5e-69	73%	WP_071121120.1
Hemoglobin *[Clostridium chauvoei]*	4e-67	70%	WP_021876824.1
Bacitracin resistance protein BacA *[Lachnospiraceae bacterium mt14]*	1e-64	67%	WP_053982519.1
Bacitracin resistance protein BacA *[Candidatus Arthromitus sp. SFB-turkey]*	4e-65	68%	WP_066760512.1
Bacitracin resistance protein BacA [*Fusobacterium mortiferum]*	2e-58	59%	WP005886941
Bacitracin resistance protein BacA [*Terrisporobacter othiniensis]*	2e-66	70%	WP_039679269.1
Hemoglobin *[Vitreoscilla stercoraria]*	2e-65	67%	WP_019959060.1
Soluble cytochrome O [uncultured Clostridium sp.]	5e-68	70%	SCH00953
Bacterial hemoglobin [*Clostridium perfringens SM101]*	3e-97	97%	ABG86855.1
Nitric oxide dioxygenase *[Psychrobacillus psychrodurans]*	2e-51	61%	SFM87783.1

The table shows that the proteins having e- value between 1e-64 to 5e- 65 were selected having identity between 61 % to 73 %.



myoglobin [Homo sapiens]

NCBI Reference Sequence: NP_976312.1

```
LOCUS       NP_976312                154 aa            linear   PRI 06-OCT-2016
DEFINITION  myoglobin [Homo sapiens].
ACCESSION   NP_976312
VERSION     NP_976312.1
DBSOURCE    REFSEQ: accession NM_203378.1
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 154)
  AUTHORS   Bicker A, Brahmer AM, Meller S, Kristiansen G, Gorr TA and Hankeln
            T.
  TITLE     The Distinct Gene Regulatory Network of Myoglobin in Prostate and
            Breast Cancer
  JOURNAL   PLoS ONE 10 (11), E0142662 (2015)
   PUBMED   26559958
  REMARK    GeneRIF: analysis of myoglobin gene regulatory networks in breast
            and prostate cancer
            Publication Status: Online-Only
REFERENCE   2  (residues 1 to 154)
  AUTHORS   Yang Q, Liu S, Tan S, Shao C, Huang J and Xiang J.
  TITLE     [Preparation and application of monoclonal antibody against human
            myoglobin]
  JOURNAL   Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 31 (8), 1115-1119 (2015)
   PUBMED   26271987
  REMARK    GeneRIF: Data show that with the myoglobin (MYO) monoclonal
            antibody of high specificity and affinity, a one-step sandwich
```

Crystal Structure Of Human Myoglobin Mutant K45r
PDB: 3RGK
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 1.65 Å

Articles about the MB gene

Electroanalysis of myoglobin based on electropolymerized [Dokl Biochem Biophys. 2016]

Myoglobin A79G polymorphism association with exercise-induced skeletal [Genet Mol Res. 2016]

A novel electrochemical aptasensor based on Y-shape structure of du [Biosens Bioelectron. 2016]

FIGURE 4: Showing NCBI homepage for query Myoglobin in *Homo sapiens*

PSI blast Iteration 1

Job title: ref|NP_976312.1| (154 letters)

RID BDEMDERD013 (Expires on 03-02 19:11 pm)
Query ID NP_976312.1
Description myoglobin [Homo sapiens]
Molecule type amino acid
Query Length 154

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.6.1+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures] [Multiple alignment] [MSA viewer]

New Analyze your query with SmartBLAST

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

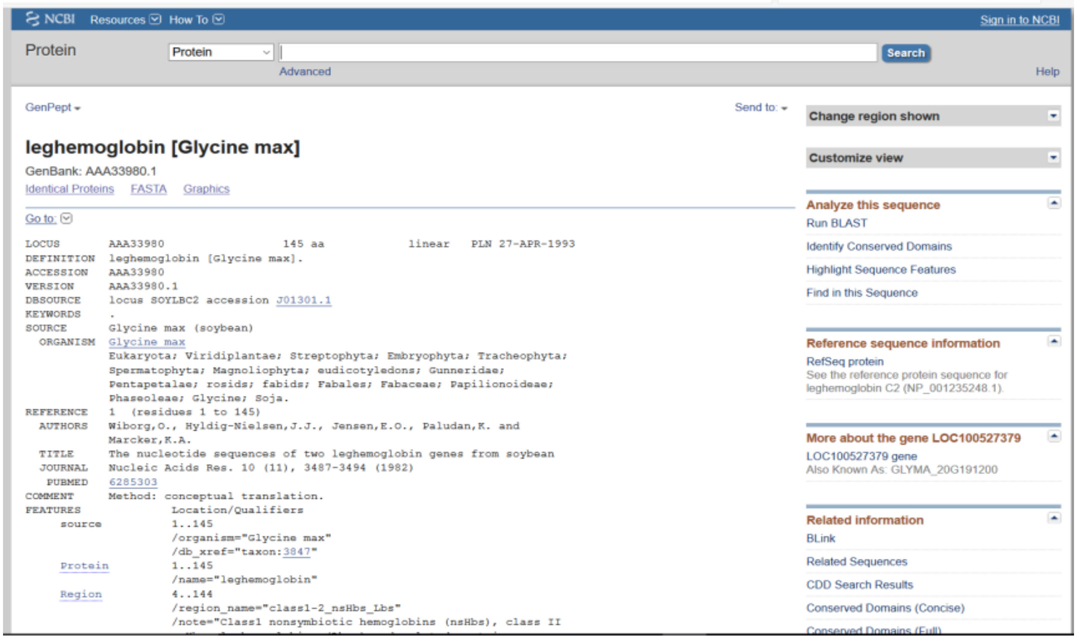
Mouse over to see the title, click to show alignments

FIGURE 5 : Output of PSI BLAST showing alignment of query sequence where, red indicate totally similar sequences.

TABLE 2: Showing the results of PSI BLAST of protein Myoglobin in *Homo sapiens*

PROTEIN	**E-VALUE**	**IDENTITY**	**ACCESSION NO.**
Myoglobin(*Sus scrofa)*pig	1e-100	94%	NP_999401.1
Myoglobin(*Ursus maritimus*) polar bear	4e-99	92%	NP_001288305.1
Myoglobin(*Castor canadensis*) American beaver	8e-97	90%	XP_020029521.1
Myoglobin(*Ochotona ladacensis)* Ladak Pika	5e-96	90%	AFN89860.1
Myoglobin(*Eospalax baileyi)* Zokar	1e-95	88%	AFN89857.1
Myoglobin(*Ochotona erythrotis)* *Chinese red pika*	2e-95	89%	AFX00022.1
Myoglobin(*Scapanus orarius)* Coast mole	3e-95	90%	AGM75749.1
Myoglobin(*Monodon monoceros)* Narwhal	4e-95	88%	AGM7575.1
Myoglobin(*Callorhinus ursinus)* Northern Fur Seal	7e-95	88%	AGM7576.1
Myoglobin(*Hexa protodon liberinsis)* pygmy hippopotamus	4e-94	90%	AGM75737.1
Myoglobin(*Dugogdugon)* Sea cow	7e-95	87%	AGM75766.1

The table shows that the proteins having e- value between 1e-95 to 7e- 95 were selected having identity between 94 % to 87% for further analysis.



leghemoglobin [Glycine max]

GenBank: AAA33980.1

```
LOCUS       AAA33980                 145 aa            linear   PLN 27-APR-1993
DEFINITION  leghemoglobin [Glycine max].
ACCESSION   AAA33980
VERSION     AAA33980.1
DBSOURCE    locus SOYLBC2 accession J01301.1
KEYWORDS    .
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae;
            Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae;
            Phaseoleae; Glycine; Soja.
REFERENCE   1  (residues 1 to 145)
  AUTHORS   Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E.O., Paludan,K. and
            Marcker,K.A.
  TITLE     The nucleotide sequences of two leghemoglobin genes from soybean
  JOURNAL   Nucleic Acids Res. 10 (11), 3487-3494 (1982)
   PUBMED   6285303
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..145
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
     Protein         1..145
                     /name="leghemoglobin"
     Region          4..144
                     /region_name="class1-2_nsHbs_Lbs"
                     /note="Class1 nonsymbiotic hemoglobins (nsHbs), class II
```

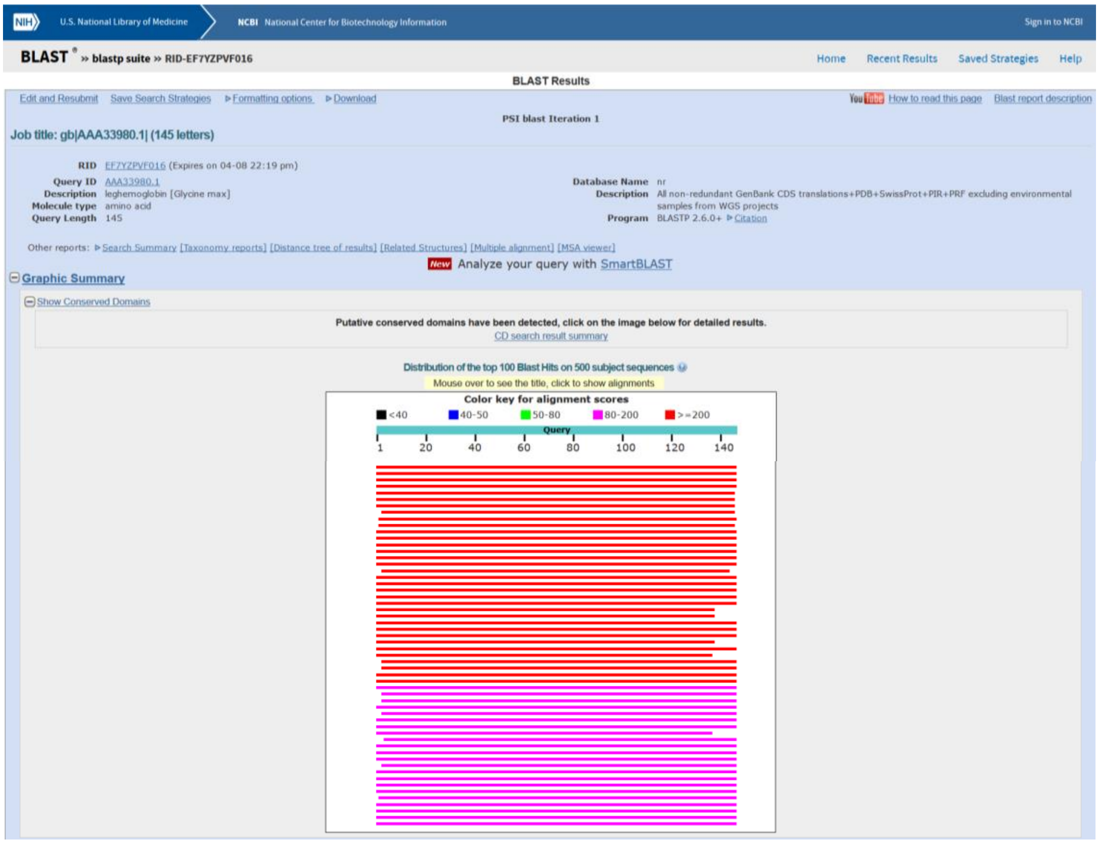
Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence

Reference sequence information
RefSeq protein
See the reference protein sequence for leghemoglobin C2 (NP_001235248.1).

More about the gene LOC100527379
LOC100527379 gene
Also Known As: GLYMA_20G191200

Related information
BLink
Related Sequences
CDD Search Results
Conserved Domains (Concise)
Conserved Domains (Full)

FIGURE 6: Showing NCBI page for query Leghemoglobin in *Glycine max*



PSI blast Iteration 1

Job title: gb|AAA33980.1| (145 letters)

RID EF7YZPVF016 (Expires on 04-08 22:19 pm)
Query ID AAA33980.1
Description leghemoglobin [Glycine max]
Molecule type amino acid
Query Length 145

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.6.0+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures] [Multiple alignment] [MSA viewer]

New Analyze your query with SmartBLAST

Putative conserved domains have been detected, click on the image below for detailed results.
CD search result summary

Distribution of the top 100 Blast Hits on 500 subject sequences

Mouse over to see the title, click to show alignments

FIGURE 7: Output of PSI BLAST showing alignment of query sequence where, red indicate totally similar sequences, pink indicates almost similar sequence.

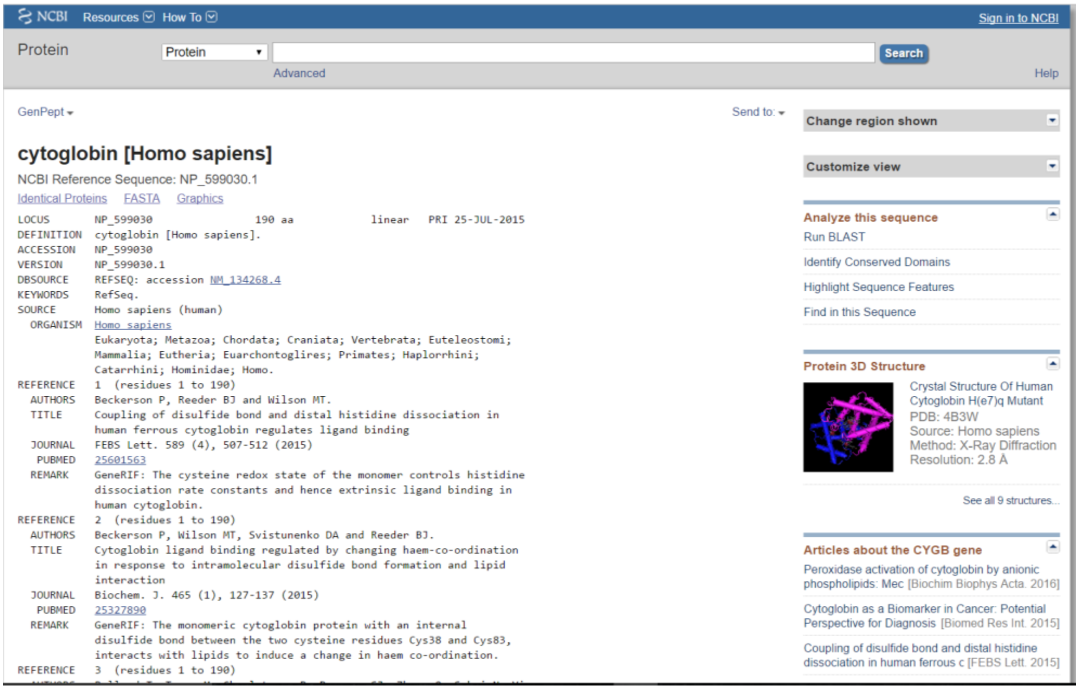
TABLE 3: Showing results of PSI BLAST of protein Leghemoglobin in *Glycine max*

PROTEIN	e-VALUE	IDENTITY	ACCESSION NO.
Leghemoglobin C2(*Glycine max)*Soyabean	2e-77	79%	NP001235248.1
Leghemoglobin(*Sesbania rostara)* semi-aquatic leguminous tree	1e-62	70%	ABC40722.1
Leghemoglobin Lb120-1 (*Medicgo truncatula)*Barrelclover	9e-61	71%	XP003588689.1
Leghemoglobin(*Lotus japenicus)*Birdsfoot trefoil	4e-57	65%	BAE46737.1
Leghemoglobin(*Astragalus sinicus)*Milkvetch	4e-57	64%	ABB13622.1
Leghemoglobin K (*Vicia faba)* Broadbean	4e-56	67%	CAA90869.1
Non symbiotic Hemoglobin 2 (*Gossypium arboretum)* Asian cotton	2e-43	41%	KHG23471.1
Hemoglobin 2 (*zea mays)* Corn	2e-31	48%	NP_001105819.1
Non symbiotic Hemoglobin(*Quercus petraea)* Oak tree	2e-28	46%	ABO93466.1
Hemoglobin hb1 (Triticum aestivum) Wheat	3e-24	41%	AAN85432.1

The table shows that the proteins having e- value between 1e-62 to 9e- 61 were selected having identity between 41% to 79%.

Results showing the query sequence family Leghemoglobin were retrieved in FASTA format from NCBI after PSI BLAST.11 proteins were retrieved from each family

3. LEGHEMOGLOBIN (*Glycine max)*



cytoglobin [Homo sapiens]

NCBI Reference Sequence: NP_599030.1

Identical Proteins FASTA Graphics

```
LOCUS       NP_599030                190 aa            linear   PRI 25-JUL-2015
DEFINITION  cytoglobin [Homo sapiens].
ACCESSION   NP_599030
VERSION     NP_599030.1
DBSOURCE    REFSEQ: accession NM_134268.4
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 190)
  AUTHORS   Beckerson P, Reeder BJ and Wilson MT.
  TITLE     Coupling of disulfide bond and distal histidine dissociation in
            human ferrous cytoglobin regulates ligand binding
  JOURNAL   FEBS Lett. 589 (4), 507-512 (2015)
   PUBMED   25601563
  REMARK    GeneRIF: The cysteine redox state of the monomer controls histidine
            dissociation rate constants and hence extrinsic ligand binding in
            human cytoglobin.
REFERENCE   2  (residues 1 to 190)
  AUTHORS   Beckerson P, Wilson MT, Svistunenko DA and Reeder BJ.
  TITLE     Cytoglobin ligand binding regulated by changing haem-co-ordination
            in response to intramolecular disulfide bond formation and lipid
            interaction
  JOURNAL   Biochem. J. 465 (1), 127-137 (2015)
   PUBMED   25327890
  REMARK    GeneRIF: The monomeric cytoglobin protein with an internal
            disulfide bond between the two cysteine residues Cys38 and Cys83,
            interacts with lipids to induce a change in haem co-ordination.
REFERENCE   3  (residues 1 to 190)
```

Analyze this sequence
- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

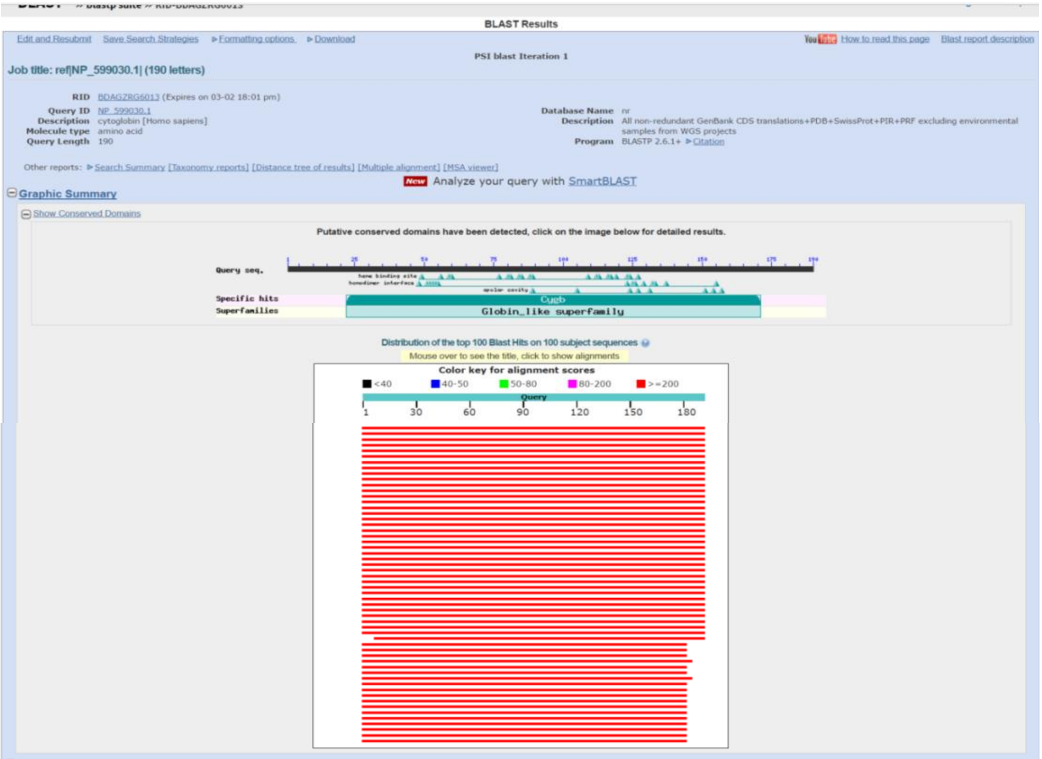
Protein 3D Structure

Crystal Structure Of Human Cytoglobin H(e7)q Mutant
PDB: 4B3W
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.8 Å

See all 9 structures...

Articles about the CYGB gene
- Peroxidase activation of cytoglobin by anionic phospholipids: Mec [Biochim Biophys Acta. 2016]
- Cytoglobin as a Biomarker in Cancer: Potential Perspective for Diagnosis [Biomed Res Int. 2015]
- Coupling of disulfide bond and distal histidine dissociation in human ferrous c [FEBS Lett. 2015]

FIGURE 25: Showing the NCBI result of query protein Cytoglobin in *Homo sapiens*



PSI blast Iteration 1

Job title: ref|NP_599030.1| (190 letters)

RID BDAGZRG6013 (Expires on 03-02 18:01 pm)
Query ID NP_599030.1
Description cytoglobin [Homo sapiens]
Molecule type amino acid
Query Length 190

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.6.1+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment] [MSA viewer]

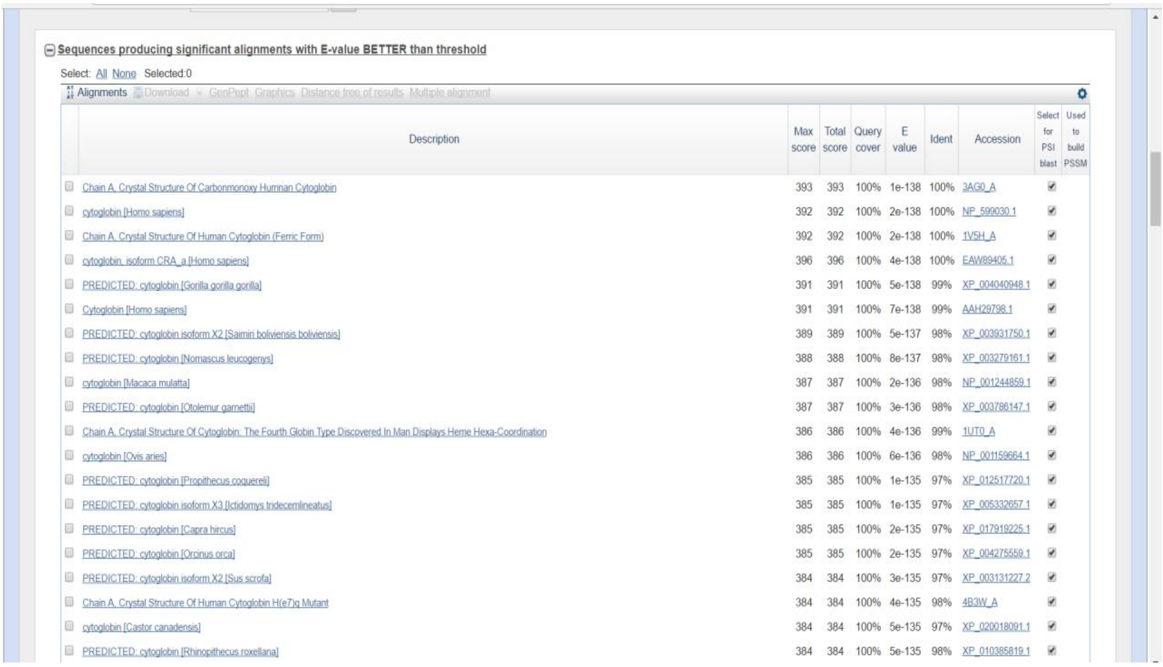
New Analyze your query with SmartBLAST

Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

FIGURE 26: Output of PSI BLAST showing alignment of query sequence where, red indicate totally similar sequences


⊟ Sequences producing significant alignments with E-value BETTER than threshold

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
☐	Chain A, Crystal Structure Of Carbonmonoxy Humnan Cytoglobin	393	393	100%	1e-138	100%	3AG0_A	☑	
☐	cytoglobin [Homo sapiens]	392	392	100%	2e-138	100%	NP_599030.1	☑	
☐	Chain A, Crystal Structure Of Human Cytoglobin (Ferric Form)	392	392	100%	2e-138	100%	1V5H_A	☑	
☐	cytoglobin, isoform CRA_a [Homo sapiens]	396	396	100%	4e-138	100%	EAW89405.1	☑	
☐	PREDICTED: cytoglobin [Gorilla gorilla gorilla]	391	391	100%	5e-138	99%	XP_004040948.1	☑	
☐	Cytoglobin [Homo sapiens]	391	391	100%	7e-138	99%	AAH29798.1	☑	
☐	PREDICTED: cytoglobin isoform X2 [Saimiri boliviensis boliviensis]	389	389	100%	5e-137	98%	XP_003931750.1	☑	
☐	PREDICTED: cytoglobin [Nomascus leucogenys]	388	388	100%	8e-137	98%	XP_003279161.1	☑	
☐	cytoglobin [Macaca mulatta]	387	387	100%	2e-136	98%	NP_001244859.1	☑	
☐	PREDICTED: cytoglobin [Otolemur garnettii]	387	387	100%	3e-136	98%	XP_003786147.1	☑	
☐	Chain A, Crystal Structure Of Cytoglobin: The Fourth Globin Type Discovered In Man Displays Heme Hexa-Coordination	386	386	100%	4e-136	99%	1UT0_A	☑	
☐	cytoglobin [Ovis aries]	386	386	100%	6e-136	98%	NP_001159664.1	☑	
☐	PREDICTED: cytoglobin [Propithecus coquereli]	385	385	100%	1e-135	97%	XP_012517720.1	☑	
☐	PREDICTED: cytoglobin isoform X3 [Ictidomys tridecemlineatus]	385	385	100%	1e-135	97%	XP_005332657.1	☑	
☐	PREDICTED: cytoglobin [Capra hircus]	385	385	100%	2e-135	97%	XP_017919225.1	☑	
☐	PREDICTED: cytoglobin [Orcinus orca]	385	385	100%	2e-135	97%	XP_004275559.1	☑	
☐	PREDICTED: cytoglobin isoform X2 [Sus scrofa]	384	384	100%	3e-135	97%	XP_003131227.2	☑	
☐	Chain A, Crystal Structure Of Human Cytoglobin H(e7)q Mutant	384	384	100%	4e-135	98%	4B3W_A	☑	
☐	cytoglobin [Castor canadensis]	384	384	100%	5e-135	97%	XP_020018091.1	☑	
☐	PREDICTED: cytoglobin [Rhinopithecus roxellana]	384	384	100%	5e-135	98%	XP_010385819.1	☑	

FIGURE 27: Output of PSI BLAST shows similar protein in different species of query protein Cytoglobin in *Homo sapiens*

TABLE 4: Showing the results of PSI BLAST of protein Cytoglobin in *Homo sapiens*

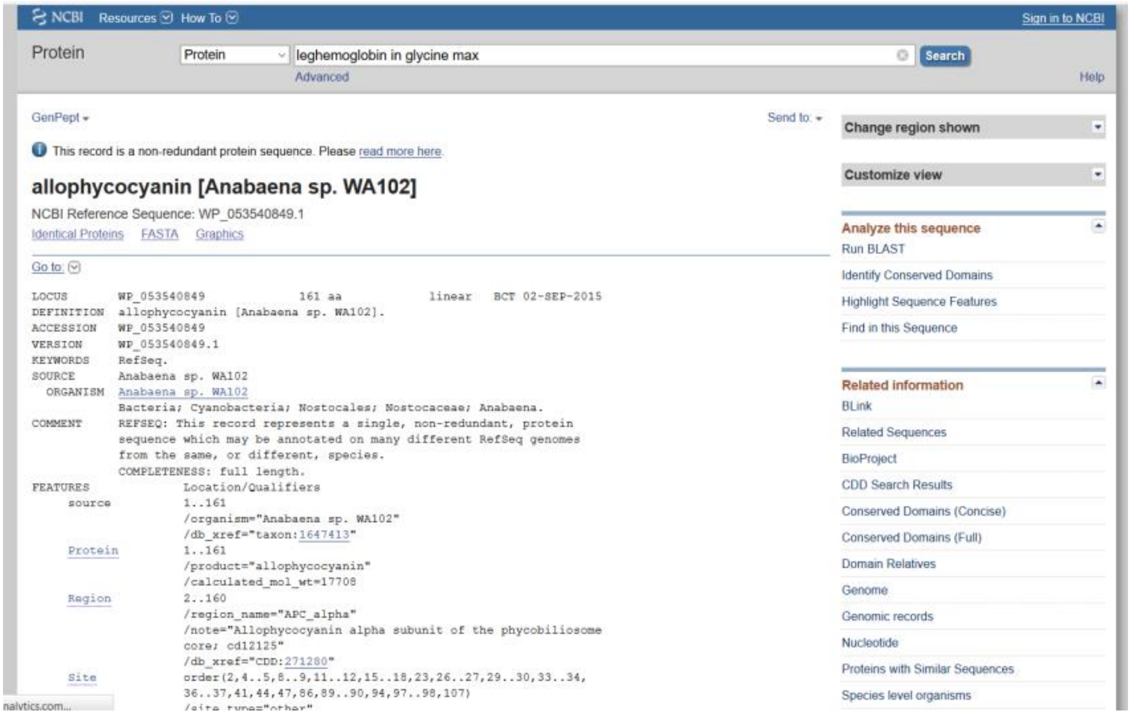
Protein	**e-value**	**Identity**	**Accesion no.**
Cytoglobin(*Ovies aries)* *Sheep*	6e-136	98%	NP_001159664.1
Cytoglobin(*Bos Taurus)* *Cow*	5e-135	97%	NP_001193649.1
Cytoglobin(*Bos grunnies)* *Yak*	1e-133	96%	AHB86985.1
Cytoglobin(*Phodopus sungorus)* *Hamster*	3e-122	95%	3AX11509.1
Cytoglobin(*Mus musculus)* *House mouse*	5e-122	95%	NP_089482.1
Cytoglobin(*Nannospalax judali)* Mole rat	4e-120	96%	CAL91964.1
Cytoglobin(*Rattus norvegics)* *Brown rat*	4e-120	94%	NP_570100.1
Cytoglobin(*Nannosplax galli)* *Blind mole rat*	5e-119	95%	CAL91963.1

Cytoglobin(*Canislupus familiaris)*Dog	3e-117	95%	NP_001071055.1
Cytoglobin(*Pteropus alecto)* *Black flying fox*	6e-112	97%	ELK12333.1
Cytoglobin(*Macaca mulatta)* *Hen*	2e-136	98%	NP001244859.1

The table shows that the proteins having e- value between 1e-133 to 6e- 136 were selected having identity value between similar proteins is 94% to 98%.

Results showing the query sequence of cytoglobin were retrieved in FASTA format from NCBI after PSI BLAST.11 proteins were retrieved from each family

4. CYTOGLOBIN (*Homo sapiens)*



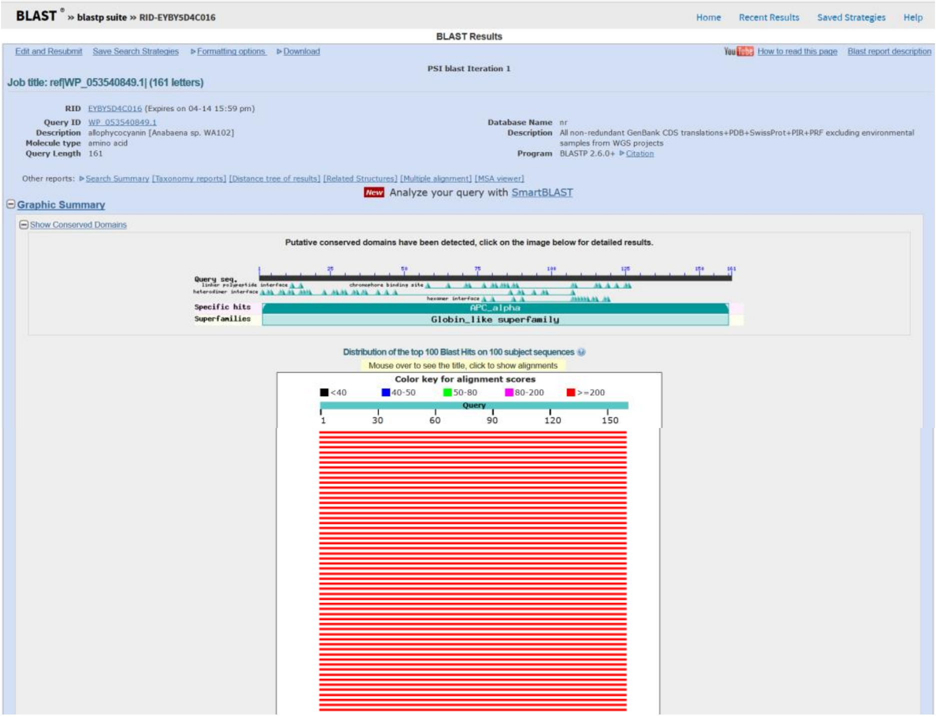
This record is a non-redundant protein sequence. Please read more here.

allophycocyanin [Anabaena sp. WA102]

NCBI Reference Sequence: WP_053540849.1

```
LOCUS       WP_053540849             161 aa            linear   BCT 02-SEP-2015
DEFINITION  allophycocyanin [Anabaena sp. WA102].
ACCESSION   WP_053540849
VERSION     WP_053540849.1
KEYWORDS    RefSeq.
SOURCE      Anabaena sp. WA102
  ORGANISM  Anabaena sp. WA102
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
COMMENT     REFSEQ: This record represents a single, non-redundant, protein
            sequence which may be annotated on many different RefSeq genomes
            from the same, or different, species.
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..161
                     /organism="Anabaena sp. WA102"
                     /db_xref="taxon:1647413"
     Protein         1..161
                     /product="allophycocyanin"
                     /calculated_mol_wt=17708
     Region          2..160
                     /region_name="APC_alpha"
                     /note="Allophycocyanin alpha subunit of the phycobiliosome
                     core; cd12125"
                     /db_xref="CDD:271280"
     Site            order(2,4..5,8..9,11..12,15..18,23,26..27,29..30,33..34,
                     36..37,41,44,47,86,89..90,94,97..98,107)
                     /site_type="other"
```

FIGURE 28 : Showing the NCBI result of query protein Allophycocyanin*Anabena sp.*



BLAST® » blastp suite » RID-EYBY5D4C016

BLAST Results

PSI blast Iteration 1

Job title: ref|WP_053540849.1| (161 letters)

RID: EYBY5D4C016 (Expires on 04-14 15:59 pm)

Query ID: WP_053540849.1
Description: allophycocyanin [Anabaena sp. WA102]
Molecule type: amino acid
Query Length: 161

Database Name: nr
Description: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program: BLASTP 2.6.0+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures] [Multiple alignment] [MSA viewer]

New Analyze your query with SmartBLAST

Graphic Summary

Show Conserved Domains

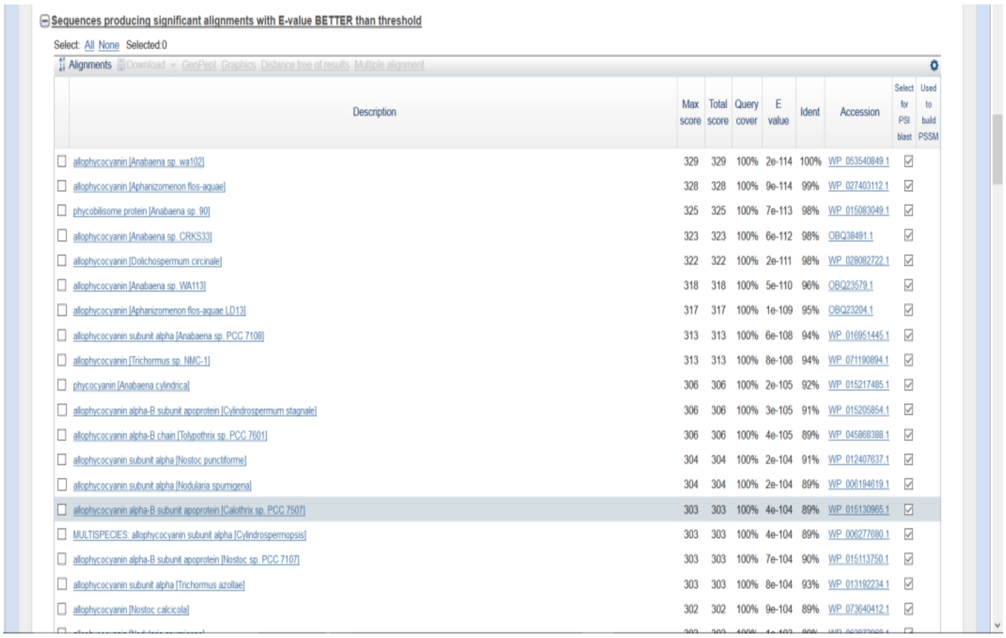
Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

FIGURE 29: Output of PSI BLAST showing alignment of query sequence where, red indicate totally similar sequences



Sequences producing significant alignments with E-value BETTER than threshold

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
☐	allophycocyanin [Anabaena sp. wa102]	329	329	100%	2e-114	100%	WP_053540849.1	☑	
☐	allophycocyanin [Aphanizomenon flos-aquae]	328	328	100%	9e-114	99%	WP_027403112.1	☑	
☐	phycobilisome protein [Anabaena sp. 90]	325	325	100%	7e-113	98%	WP_015083049.1	☑	
☐	allophycocyanin [Anabaena sp. CRKS33]	323	323	100%	6e-112	98%	OBQ38491.1	☑	
☐	allophycocyanin [Dolichospermum circinale]	322	322	100%	2e-111	98%	WP_028082722.1	☑	
☐	allophycocyanin [Anabaena sp. WA113]	318	318	100%	5e-110	96%	OBQ23579.1	☑	
☐	allophycocyanin [Aphanizomenon flos-aquae LD13]	317	317	100%	1e-109	95%	OBQ23204.1	☑	
☐	allophycocyanin subunit alpha [Anabaena sp. PCC 7108]	313	313	100%	6e-108	94%	WP_016951445.1	☑	
☐	allophycocyanin [Trichormus sp. NMC-1]	313	313	100%	8e-108	94%	WP_071190894.1	☑	
☐	phycocyanin [Anabaena cylindrica]	306	306	100%	2e-105	92%	WP_015217485.1	☑	
☐	allophycocyanin alpha-B subunit apoprotein [Cylindrospermum stagnale]	306	306	100%	3e-105	91%	WP_015205854.1	☑	
☐	allophycocyanin alpha-B chain [Tolypothrix sp. PCC 7601]	306	306	100%	4e-105	89%	WP_045868388.1	☑	
☐	allophycocyanin subunit alpha [Nostoc punctiforme]	304	304	100%	2e-104	91%	WP_012407637.1	☑	
☐	allophycocyanin subunit alpha [Nodularia spumigena]	304	304	100%	2e-104	89%	WP_006194619.1	☑	
☐	allophycocyanin alpha-B subunit apoprotein [Calothrix sp. PCC 7507]	303	303	100%	4e-104	89%	WP_015130965.1	☑	
☐	MULTISPECIES: allophycocyanin subunit alpha [Cylindrospermopsis]	303	303	100%	4e-104	89%	WP_006277680.1	☑	
☐	allophycocyanin alpha-B subunit apoprotein [Nostoc sp. PCC 7107]	303	303	100%	7e-104	90%	WP_015113750.1	☑	
☐	allophycocyanin subunit alpha [Trichormus azollae]	303	303	100%	8e-104	93%	WP_013192234.1	☑	
☐	allophycocyanin [Nostoc calcicola]	302	302	100%	9e-104	89%	WP_073640412.1	☑	

FIGURE 30 : Output of PSI BLAST shows similar proteins in different species of query protein Allophycocyanin *Anabena sp.*

TABLE 5: Showing PSI BLAST results of protein Allophycocynin

PROTEIN	e-VALUE	IDENTITY	ACCESION NO.
Allophycocyanin Subunit alpha (Geitlerima sp. pcc 7105)	5e-73	62%	WP017663120.1
Allophycocyanin (*Leptolynbya sp. Jsc-1)*	5e-72	58%	WP036003913
Allophycocyanin B (*Thermosynechoccus elongates)*	3e-70	58%	WP0115739.1
Allophycocyanin subunit alpha (*Microcystis aeruginosa)*	5e-70	63%	WP002738559.1
Allophycocyanin subunit alpha (*Spirulin subsalsa)*	5e-76	61%	WP017303467.1
Allophycocyanin subunit B(*Synechocystis sp. Pcc 6803)*	3e-69	54%	YP009313632.1
Allophycocyanin gamma subunit (Helminthodadia australis)	5e-62	54%	YP009313632.1
Allophycocyanin subunit alpha (*Galdieria sulphuraria)*	5e-50	47%	YP009051103.1
Allophycocyanin B(*Gloebacter violaceus)*	2e-48	49%	WP011141181.1

The table shows that the proteins having e- value between 2e-48 to 5e- 76 were selected having identity value between similar proteins 67% to 97%.

Results showing the query sequence family Leghemoglobin were retrieved in FASTA format from NCBI after PSI BLAST.11 proteins were retrieved from each family

5.ALLOPHYCOCYANIN (*Anabena sp.)*

TABLE 6: Showing prediction of domains of 11 Truncated hemoglobin proteins in different species by using tool Pfam

PROTEIN	DOMAIN	e-VALUE	START	END
Bacitracin resistance protein BacA *[Romboutsia timonensis]*	Globin	7e-15	6	103
Bacitracin resistance protein BacA *[Intestinibacter bartlettii*	Globin	1.1e-13	6	103
Hemoglobin *[Clostridium chauvoei]*	Globin	2.3e-14	6	103
Bacitracin resistance protein BacA *[Lachnospiraceae bacterium mt14]*	Globin	1.2e-15	6	103
Bacitracin resistance protein BacA *[Candidatus Arthromitus sp. SFB-turkey]*	Globin	2.1e-12	7	103
Bacitracin resistance protein BacA [*Fusobacterium mortiferum]*	Globin	2.1e-12	6	103
Bacitracin resistance protein BacA [*Terrisporobacter othiniensis]*	Globin	8e-14	6	103
Hemoglobin *[Vitreoscilla stercoraria]*	Globin	5e-14	6	103
Soluble cytochrome O [uncultured Clostridium sp.]	Globin	1.4e-13	6	103
Bacterial hemoglobin [*Clostridium perfringens SM101]*	Globin	3.4e-14	6	103
Nitric oxide dioxygenase *[Psychrobacillus psychrodurans]*	Globin	1.6e-11	6	103

The results shows that the conserved domain in truncated hemoglobin is globin and having e value between 1.1e-13 to 8e-14. and start and end point of region lies between 6-103

TABLE 7: Showing prediction of domains of 11 proteins of Myoglobin in different species by using tool Pfam

PROTEIN	DOMAIN	E-VALUE	START	END
Myoglobin(*Sus scrofa)* Pig	Globin	2.8e-21	7	113
Myoglobin(*Ursus maritimus*) Polar bear	Globin	5.7e-22	7	113
Myoglobin(*Castor canadensis*) American beaver	Globin	3e-21	7	113
Myoglobin(*Ochotona ladacensis)* Ladak pika	Globin	2.1e-23	7	113
Myoglobin(*Eospalax baileyi)* Zokar	Globin	1.5e-23	7	113
Myoglobin(*Ochotona erythrotis)* Chinese red pika	Globin	8.9e-24	7	113
Myoglobin(*Scapanus orarius)* Coast mole	Globin	4.8e-21	7	113
Myoglobin(*Monodon monoceros)* Narwal	Globin	7e-21	7	113
Myoglobin(*Callorhinus ursinus)* Northen fur seal	Globin	7.9e-22	7	113
Myoglobin(*Hexa protodon liberinsis)* Pygmy hippotamus	Globin	1.5e-19	7	113
Myoglobin(*Dugogdugon)* Sea cow	Globin	3.6e-20	7	113

The results shows that the conserved domain in myoglobin is globin and having e value between 1.5e-23 to 8e-24. and start and end point of region lies between 7-113

TABLE 8: Showing prediction of domains of 11 proteins of Leghemoglobin in different species by using tool Pfam

PROTEIN	DOMAIN	e-VALUE	START	END
Leghemoglobin C2(Glycine *max)*Soybean	Globin	3.2e-13	8	111
Leghemoglobin(*Sesbania rostara)* Semi aquatic leguminous tree	Globin	3.7e-17	8	114
Leghemoglobin Lb120-1(*Medicgo truncatula)* Barrel clover	Globin	1.6e-16	7	111
Leghemoglobin(*Lotus japenicus)* Birds foot trefoil	Globin	1.2e-15	7	111
Leghemoglobin(*Astragalus sinicus)*Milkvetch	Globin	2e-18	7	113
Leghemoglobin K (*Vicia faba)* Broadbean	Globin	9.9e-16	7	111
Non symbiotic Hemoglobin 2 (*Gossypium arboretum)* Asian cotton	Globin	1.3e-19	7	116
Hemoglobin 2 (*zea mays)* Corn	Globin	1e-19	7	117
Non symbiotic Hemoglobin(*Quercus petraea)* Oak tree	Globin	9.8e-18	14	123
Hemoglobin hb1 (Triticum aestivum)Wheat	Globin	1e-16	14	123
Leghemoglobin (*Cajarnus cajan)*Wheat	Globin	1.4e-12	8	111

The results shows that the conserved domain in leghemoglobin is globin and having e value between 1e-19 to 9.8e-18 and start and end point of region lies between 8-123

TABLE 9: Showing prediction of domains of 11 proteins of Cytoglobin in different species by using tool Pfam

PROTEIN	DOMAIN	e-VALUE	START	END
Cytoglobin(*Ovies aries)* Sheep	Globin	4.7e-25	23	132
Cytoglobin(*Bos Taurus)* Cow	Globin	6.3e-25	23	132
Cytoglobin(*Bos grunnies)*Hamster	Globin	7.6e-24	23	132
Cytoglobin(*Phodopus sungorus)*Mouse	Globin	7.9e-25	23	132
Cytoglobin(*Mus musculus)*Mole rat	Globin	1.6e-24	23	132
Cytoglobin(*Nannospalax judali)* Blind mole rat	Globin	7.1e-25	23	132
Cytoglobin(*Rattus norvegics)*Brown rat	Globin	4.8e-25	23	132
Cytoglobin(*Nannosplax galli)*Blind mole rat	Globin	8.6e-24	23	132
Cytoglobin(*Canis lupus Familiaris)*Dog	Globin	2.5e-25	23	132
Cytoglobin(*Pteropus alecto)*Black flying fox	Globin	6.6e-25	23	132
Cytoglobin(*Macaca mulatta)*Hen	Globin	5.2e-25	23	132

The results shows that the conserved domain in cytoglobin is globin and having e value between 1.6e-24 to 8.6e-24. and start and end point of region lies between 23-132

TABLE 10: Showing prediction of domains of 11 proteins of Allophycocyanin in different species by using tool Pfam

PROTEIN	DOMAIN	E-VALUE	START	END
Allophycocyanin Subunit alpha (Geitlerima sp. pcc 7105)	Phycobilisome	1.3e-55	6	160
Allophycocyanin (*Leptolynbya sp. Jsc-1)*	Phycobilisome	6e-61	6	160
Allophycocyanin B (*Thermosynechoccus elongates)*	Phycobilisome	8.8e-63	6	160
Allophycocyanin subunit alpha (*Microcystis aeruginosa)*	Phycobilisome	3.4e-61	16	170
Allophycocyanin subunit alpha (*Spirulin subsalsa)*	Phycobilisome	6.7e-60	6	160
Allophycocyanin subunit B(*Synechocystis sp. Pcc 6803)*	Phycobilisome	2.5e-61	6	160
Allophycocyanin gamma subunit (Helminthodadia australis)	Phycobilisome	1.2e-59	6	160
Allophycocyanin subunit alpha (*Galdieria sulphuraria)*	Phycobilisome	5.2e-54	6	160
Allophycocyanin B(*Gloebacter violaceus)*	Phycobilisome	3e-58	6	160
Allophycocyanin subunit alpha (*Mastigocladopsis repens*)	Phycobilisome	1.7e-59	6	160
Allophycocyanin alpha B chain (*Cyanobium sp.)*	Phycobilisome	7.9e-52	6	160

The results shows that the conserved domain in allophycocyanin is Phycobilisome and having e-value between 1.2e-13 to 8.8e-63 and start and end point of region lies between 6-160.

PHYSIOCHEMICAL PROPERTIES USING PROTPARAM TOOL

TABLE 11: Showing physiochemical properties of Truncated hemoglobin by Protparam l

PROTEIN	**NO. OF A.A**	**MOL. WEIGHT**	**pI**	**AI**	**II**	**GRAVY**
Bacitracin resistance protein BacA *[Romboutsia timonensis]*	155	17481.32	5.38	101.87	38.75	-0.161
Bacitracin resistance protein BacA *[Intestinibacter bartlettii]*	155	17409.17	5.23	105.68	28.74	-0.094
Hemoglobin *[Clostridium chauvoei]*	145	16243.89	6.31	103.66	25.01	0.158
Bacitracin resistance protein BacA *[Lachnospiraceae bacterium mt14]*	145	15883.38	5.31	107.66	34.13	-0.054
Bacitracin resistance protein BacA *[Candidatus Arthromitus sp. SFB-turkey]*	144	15963.63	5.41	108.33	24.85	-0.059
Bacitracin resistance protein BacA [*Fusobacterium mortiferum]*	146	16060.63	5.43	105.62	31.68	-0.044
Bacitracin resistance protein BacA [*Terrisporobacter othiniensis]*	155	17535.44	5.59	107.55	33.21	-0.109
Hemoglobin *[Vitreoscillastercoraria]*	146	15774.34	5.27	113.01	39.34	0.234
Soluble cytochrome O [uncultured Clostridium sp.]	155	17525.37	5.40	103.74	34.59	-0.115

Bacterial hemoglobin [*Clostridium perfringens SM101]*	144	16250.82	5.63	100.21	39.59	-0.228
Nitric oxide dioxygenase *[Psychrobacillus psychrodurans]*	395	43425.46	5.18	41.37	39.17	-0.082

TABLE 12: Showing physiochemical properties of Myoglobin by Protparam tool

PROTEIN	**NO. OF A.A**	**MOL. WEIGHT**	**pI**	**AI**	**II**	**GRAVY**
Myoglobin(*Sus scrofa)* Pig	154	17084.61	6.76	84.94	20.50	-0.444
Myoglobin(*Ursus maritimus*) Polar bear	154	17176.78	7.96	88.12	13.15	-0.487
Myoglobin(*Castor canadensis*) American beaver	154	17168.64	8.66	90.00	19.34	-0.400
Myoglobin(*Ochotona ladacensis)*Ladak pika	154	17104.64	7.11	84.94	16.41	-0.486
Myoglobin(*Eospalax baileyi)*Ladak Pika	154	17104.63	7.07	87.47	12.68	-0.489
Myoglobin(*Ochotona erythrotis)*Zokar	154	17081.61	7.07	84.29	16.96	-0.486
Myoglobin(*Scapanus orarius)*Chinese red pika	154	16993.54	7.94	86.17	18.09	-0.409
Myoglobin(*Monodon monoceros)*Narwal	154	17201.85	8.72	89.42	15.30	-0.436
Myoglobin(*Callorhinus ursinus)*Northen fur seal	154	17338.04	8.98	88.12	19.48	-0.514
Myoglobin(*Hexa protodon liberinsis)* Pygmy hippopotamus	154	17231.80	6.80	87.47	18.44	-0.457

Myoglobin(*Dugogdugon)* Sea cow	154	17191.92	7.17	90.00	24.84	-0.336

TABLE 13: Showing physiochemical properties of Leghemoglobin by Protparam tool

PROTEIN	**NO. OF A.A**	**MOL. WEIGHT**	**pI**	**AI**	**II**	**GRAVY**
Leghemoglobin C2(*Glycine max)*Soybean	145	15494.64	5.38	21.87	9.66	0.043
Leghemoglobin(*Sesbania rostara)* Semi aquatic leguminous tree	149	16031.29	5.66	18.41	92.35	0.006
Leghemoglobin Lb120-1(*Medicgo truncatula)* Barrelclover	146	15752.16	6.29	34.97	92.88	0.037
Leghemoglobin(*Lotus japenicus)* Birds foot trefoil	146	15390.62	6.40	27.26	91.10	0.115
Leghemoglobin(*Astragalus sinicus)* MMilkvetch	148	15853.11	6.06	32.27	93.65	0.052
Leghemoglobin K (*Vicia faba)*Broadbean	146	15851.34	5.32	33.24	98.22	0.085
Non symbiotic Hemoglobin 2 (*Gossypium arboretum)* Asian cotton	159	18084.71	5.57	33.59	82.90	-0.426
Hemoglobin 2 (*zea mays)* Corn	191	20646.32	5.02	38.17	74.40	-0.239
Non symbiotic Hemoglobin(*Quercus petraea)*Oak tree	161	17913.81	8.57	30.39	85.34	-0.130
Hemoglobin hb1 (Triticum aestivum)Wheat	162	18144.13	8.67	29.68	85.62	-0.138
Leghemoglobin (*Cajanus cajan)*Pigeon pea	146	15652.81	6.29	21.33	90.34	-0.055

TABLE 14: Showing physiochemical properties of cytoglobin by Protparam tool

PROTEIN	**NO. OF A.A**	**MOL. WEIGHT**	**pI**	**AI**	**II**	**GRAVY**
Cytoglobin(*Ovies aries)* Sheep	190	21416.61	6.32	84.16	48.43	-0.273
Cytoglobin(*BosTaurus)* Cow	190	21459.63	6.32	83.63	48.43	-0.301
Cytoglobin(*Bos grunnies)Yak*	190	21399.60	6.32	83.63	48.83	-0.281
Cytoglobin(*Phodopus sungorus)*Hamster	190	21474.61	6.60	81.56	42.38	-0.301
Cytoglobin(*Mus musculus)*House mouse	190	21465.66	6.32	83.63	46.03	-0.282
Cytoglobin(*Nannospalax judali)*Mole rat	190	21533.67	6.10	83.11	45.83	-0.330
Cytoglobin(*Rattus norvegics)*Brown rat	190	21496.71	6.31	85.16	43.36	-0.314
Cytoglobin(*Nannosplax galli)*Blind mole rat	190	21579.79	6.98	83.11	47.55	-0.342
Cytoglobin(*Canis lupus familiaris)*Dog	183	20704.78	7.02	85.25	51.24	-0.261
Cytoglobin(*Pteropus alecto)* Black flying cat	180	20588.72	6.98	85.60	43.36	-0.314
Cytoglobin(*Macaca mulatta)*Hen	190	21415.58	6.32	83.63	46.68	-0.288

TABLE 15: Showing physiochemical properties of Allophycocyanin by Protparam tool

PROTEIN	**NO. OF A.A**	**MOL. WEIGHT**	**pI**	**AI**	**II**	**GRAVY**
Allophycocyanin Subunit alpha (Geitlerima sp. pcc 7105)	164	18149.84	5.51	87.96	44.64	-0.269

Allophycocyanin (*Leptolynbya sp. Jsc-1)*	161	18142.79	5.17	93.35	46.03	-0.211
Allophycocyanin B (*Thermosynechoccus elongates)*	171	17911.61	5.29	87.89	42.08	-0.262
Allophycocyanin subunit alpha (*Microcystis aeruginosa)*	161	19228.0	4.97	91.49	37.68	-0.294
Allophycocyanin subunit alpha (*Spirulin subsalsa)*	161	18011.67	6.33	99.38	36.18	-0.126
Allophycocyanin subunit B(*Synechocystis sp. Pcc 6803)*	161	17904.65	5.30	91.49	29.00	-0.253
Allophycocyanin gamma subunit (Helminthodadia australis)	161	17923.45	5.44	110.93	43.34	-0.047
Allophycocyanin subunit alpha (*Galdierisulphuraria)*	161	18109.12	7.70	96.34	40.84	-0.050
Allophycocyanin B(*Gloebacter violaceus)*	161	17380.75	4.90	99.38	37.94	-0.105
Allophycocyanin subunit alpha (*Mastigocladopsis repens*)	161	17548.01	4.88	95.69	31.36	-0.068
Allophycocyanin alpha B chain (*Cyanobium sp.)*	161	18632.84	5.30	89.21	31.69	-0.180

Results shows physiochemical parameters of 55 proteins of globin-like superfamily. server provides directly calculated values of pi/MW(isoelectric point, molecular weight),percentage of amino acid, instability index (II) aliphatic index (AI) and GRAVY(Grand average of hydrophobicity).Instability index give information about stability of one protein in test tube. A protein whose instability index is smaller than 40 is predicted as stable and above 40 is unstable.so, according to above result 13 proteins all proteins of cytoglobin subfamily and 2 proteins of allophycyanin are unstable.GRAVY value of protein is sum of hydropathy of all amino acids divided by no. of residues in sequence. So in above results Myoglobin(*Callorhinus ursinus)*is most hydropathic.AI constitute the aliphatic index refers to a relative volume occupied by aliphatic side chains(alanine, valine, leucine and isoleucine).Increased AI results into hydrophobic interactions and thus may be regarded as a positive factor for increase of thermostability of globular proteins.Hemoglobin *[Vitreoscilla stercoraria]*has highest value.

TABLE 16: Prediction of location of proteins of truncated hemoglobin by using PSORT II tool.protein sites are predicted as cytoplasmic and mitochondrial.

PROTEIN	ACCESSION NO.	LOCATION
Bacitracin resistance protein BacA *[Romboutsia timonensis]*	WP_071121120	Cytoplasmic
Bacitracin resistance protein BacA *[Intestinibacter bartlettii]*	WP_071121120.1	Cytoplasmic
Hemoglobin *[Clostridium chauvoei]*	WP_021876824.1	Cytoplasmic
Bacitracin resistance protein BacA *[Lachnospiraceae bacterium mt14]*	WP_053982519.1	Cytoplasmic
Bacitracin resistance protein BacA *[Candidatus Arthromitus sp. SFB-turkey]*	WP_066760512.1	Cytoplasmic
Bacitracin resistance protein BacA [*Fusobacterium mortiferum]*	WP005886941	Cytoplasmic
Bacitracin resistance protein BacA [*Terrisporobacter othiniensis]*	WP_039679269.1	Mitochondrial
Hemoglobin *[Vitreoscilla stercoraria]*	WP_019959060.1	Cytoplasmic
Soluble cytochrome O [uncultured Clostridium sp.]	SCH00953	Cytoplasmic
Bacterial hemog lobin [*Clostridium*	ABG86855.1	Cytoplasmic

perfringens SM101]		
Nitric oxide dioxygenase *[Psychrobacillus psychrodurans]*	SFM87783.1	Cytoplasmic

TABLE 17: Prediction of location of proteins of Myoglobin by using PSORT II tool

PROTEIN	**ACCESSION NO.**	**LOCATION**
Myoglobin(*Sus scrofa)* Pig	NP_999401.1	Cytoplasm
Myoglobin(*Ursus maritimus*) Polar bear	NP_001288305.1	Cytoplasm
Myoglobin(*Castor canadensis*) American beaver	XP_020029521.1	Cytoplasm
Myoglobin(*Ochotona ladacensis)* Ladak pika	AFN89860.1	Cytoplasm
Myoglobin(*Eospalax baileyi)* Zokar	AFN89857.1	Cytoplasm
Myoglobin(*Ochotona erythrotis)* Chinese red pika	AFX00022.1	Cytoplasm
Myoglobin(*Scapanus orarius)*Coast mole	AGM75749.1	Cytoplasm
Myoglobin(*Monodon monoceros)*Narwal	AGM7575.1	Cytoplasm
Myoglobin(*Callorhinus ursinus)*Northen fur seal	AGM7576.1	Cytoplasm
Myoglobin(*Hexa protodon liberinsis)* Pygmy hippotamus	AGM75737.1	Cytoplasm
Myoglobin(*Dugogdugon)* Sea cow	AGM75766.1	Cytoplasm

TABLE 18: Prediction of location of proteins of Leghemoglobin by using PSORT II tool.protein sites are predicted as cytoplasmic.

PROTEIN	**ACCESSION NO.**	**LOCATION**
Leghemoglobin C2(Glycine *max*Soybean	NP001235248.1	Cytoplasmic
Leghemoglobin(*Sesbania rostara)* Semiaquatic leguminous tree	ABC40722.1	Cytoplasmic
Leghemoglobin Lb120-1(*Medicgo truncatula)*Barrelclover	XP003588689.1	Cytoplasmic
Leghemoglobin(*Lotus japenicus)*Birds foot trefoil	BAE46737.1	Cytoplasmic
Leghemoglobin(*Astragalus sinicus)*Milkvetch	ABB13622.1	Cytoplasmic
Leghemoglobin K (*Vicia faba)*Broadbean	CAA90869.1	Cytoplasmic
Non symbiotic Hemoglobin 2 (*Gossypium arboretum)* Asian cotton	KHG23471.1	Cytoplasmic
Hemoglobin 2 (*zea mays)* Corn	NP_001105819.1	Cytoplasmic
Non symbiotic Hemoglobin(*Quercus petraea)*Oak tree	ABO93466.1	Cytoskeletal
Hemoglobin hb1 (Triticum aestivum)Wheat	AAN85432.1	Cytoplasmic
Leghemoglobin (*Cajarnus cajan)*Pigeon pea	XP_020222796.1	Cytoplasmic

TABLE 19: Prediction of location of proteins of Cytoglobin by using PSORT II tool protein sites are predicted as cytoplasm and mitochondria.

PROTEIN	ACCESSION NO.	LOCATION
Cytoglobin(*Ovies aries)* Sheep	NP_001159664.1	Cytoplasm
Cytoglobin(*Bos Taurus)* Cow	NP_001193649.1	Cytoplasm
Cytoglobin(*Bos grunnies)* Yak	AHB86985.1	Cytoplasm
Cytoglobin(*Phodopus sungorus)*Hamstere	3AX11509.1	Cytoplasm
Cytoglobin(*Mus musculus)* House mouse	NP_089482.1	Cytoplasm
Cytoglobin(*Nannospalax judali)*Mole rat	CAL91964.1	Cytoplasm
Cytoglobin(*Rattus norvegics)*Brown rat	NP_570100.1	Mitochondria
Cytoglobin(*Nannosplax galli)*Blind mole rat	CAL91963.1	Cytoplasm
Cytoglobin(*Canis lupus familiaris)*Dog	NP_001071055.1	Cytoplasm
Cytoglobin(*Pteropus alecto)*Black flying fox	ELK12333.1	Cytoplasm
Cytoglobin(*Macaca mulatta)*Hen	NP001244859.1	Cytoplasm

TABLE 20: Prediction of location of proteins of Allophycocyanin by using PSORT II tool.protein sites are predicted as cytoplasm and mitochondria.

PROTEIN	ACCESSION NO.	LOCATION
Allophycocyanin Subunit alpha (Geitlerima sp. pcc 7105)	WP017663120.1	Cytoplasm

Allophycocyanin (*Leptolynbya sp. Jsc-1)*	WP036003913	Cytoplasm
Allophycocyanin B (*Thermosynechoccus elongates)*	WP0115739.1	Cytoplasm
Allophycocyanin subunit alpha (*Microcystis aeruginosa)*	WP002738559.1	Cytoplasm
Allophycocyanin subunit alpha (*Spirulin subsalsa)*	WP017303467.1	Cytoplasm
Allophycocyanin subunit B(*Synechocystis sp. Pcc 6803)*	YP009313632.1	Cytoplasm
Allophycocyanin gamma subunit (Helminthodadia australis)	YP009313632.1	Cytoplasm
Allophycocyanin subunit alpha (*Galdieria sulphuraria)*	YP009051103.1	Cytoplasm
Allophycocyanin B(*Gloebacter violaceus)*	WP011141181.1	Cytoplasm
Allophycocyanin subunit alpha (*Mastigocladopsis repens*)	WP017317206.1	Cytoplasm
Allophycocyanin alpha B chain (*Cyanobium sp.)*	SB0422901	Cytoplasm

PHYLOGENETIC TREES OF QUERY PROTEINS

1. Truncated hemoglobin

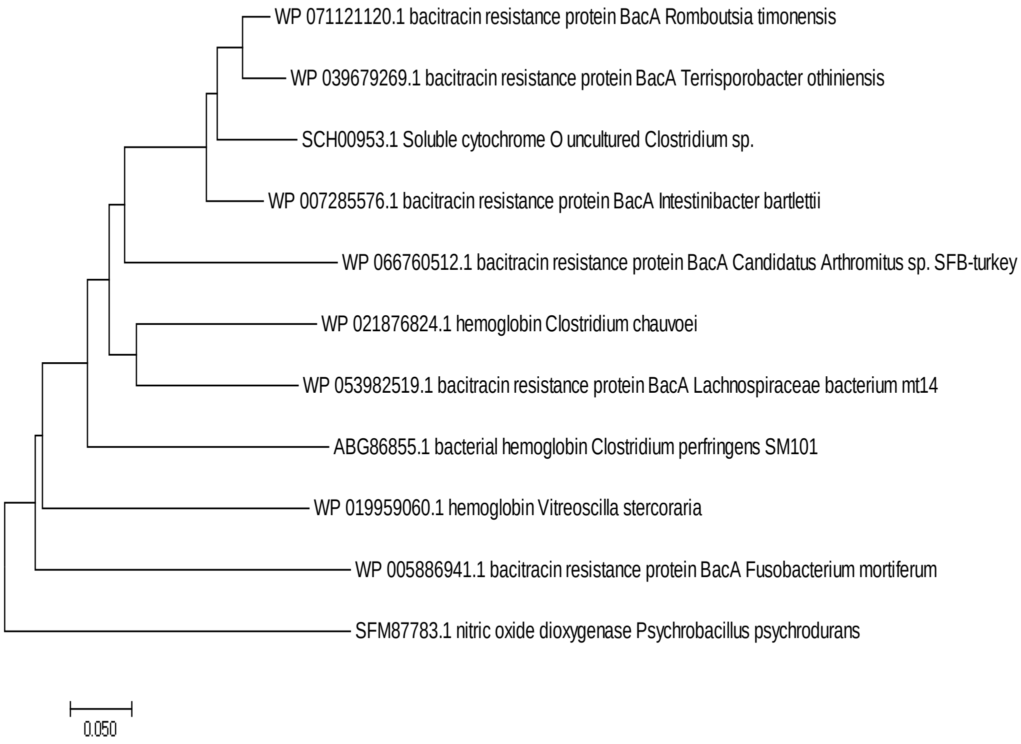


FIGURE 32: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Truncated hemoglobin protein among various species. The analysis involved 11 amino acid sequences. Here, *Romboutsia timonensis and Terrisporobacter othniesis* are closely related but *Vitreosolcilla stercoraria* is distantly related.

2.MYOGLOBIN

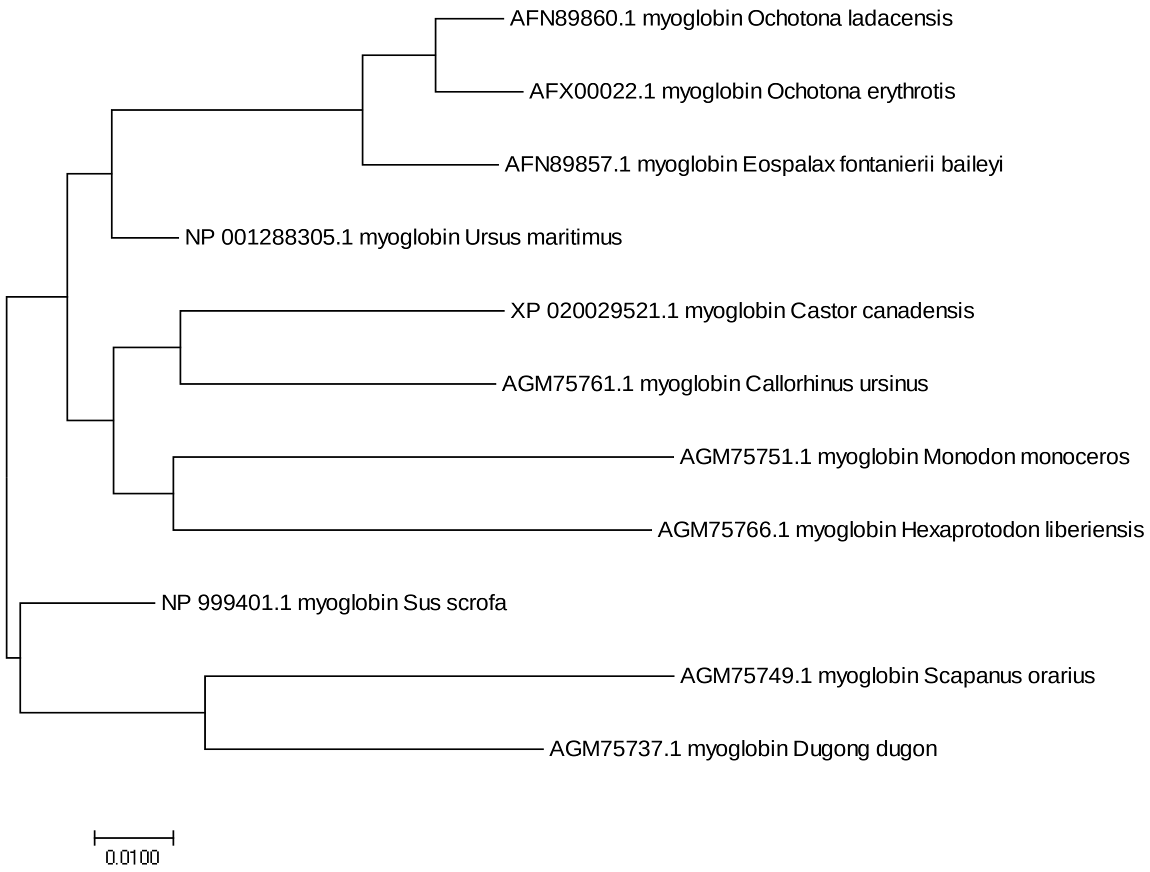


FIGURE 33: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Myoglobin protein among various species. The analysis involved 11 amino acid sequences.Here,*Castor Canadensis* and *callorhinus ursinus* are closely related.

3.LEGHEMOGLOBIN

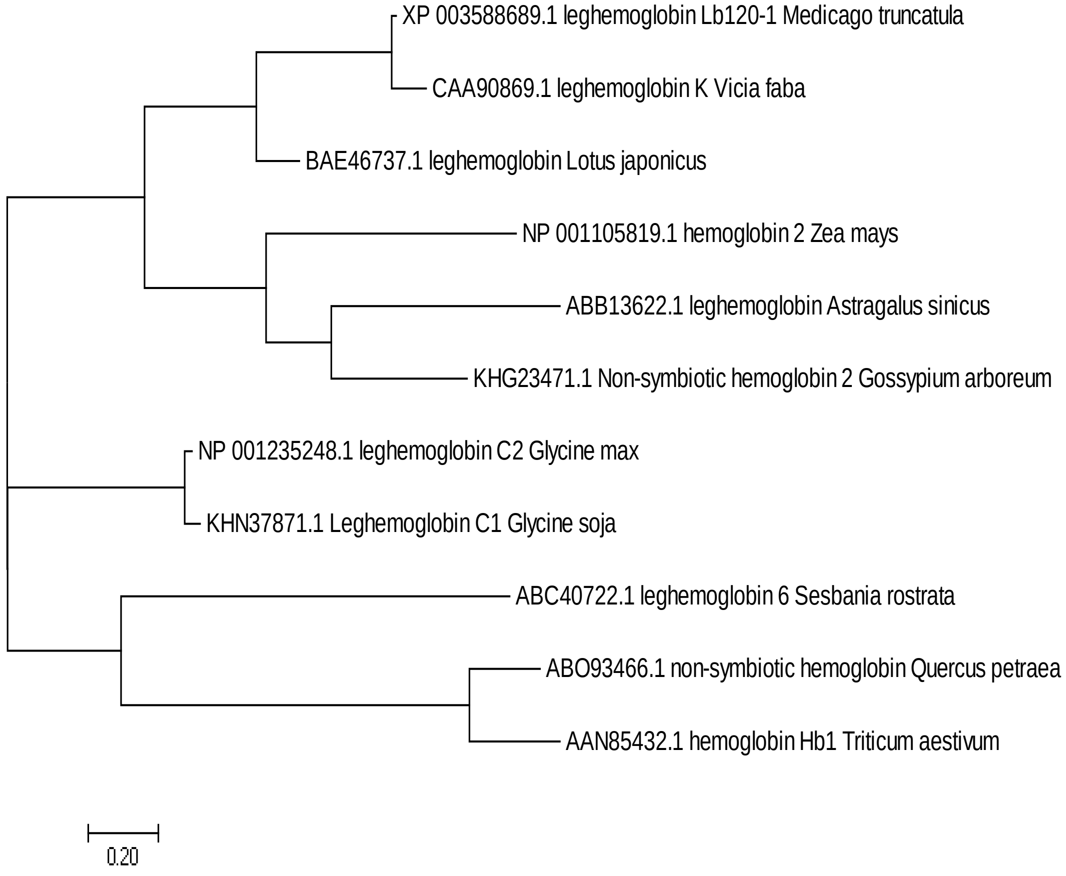


FIGURE 34: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Leghemoglobin protein among various species. The analysis involved 11 amino acid sequences.Here, *Vicia faba* and *Lotus japonicas* are closely related but *Triticum aestivum* is distantly related.

4. CYTOGLOBIN

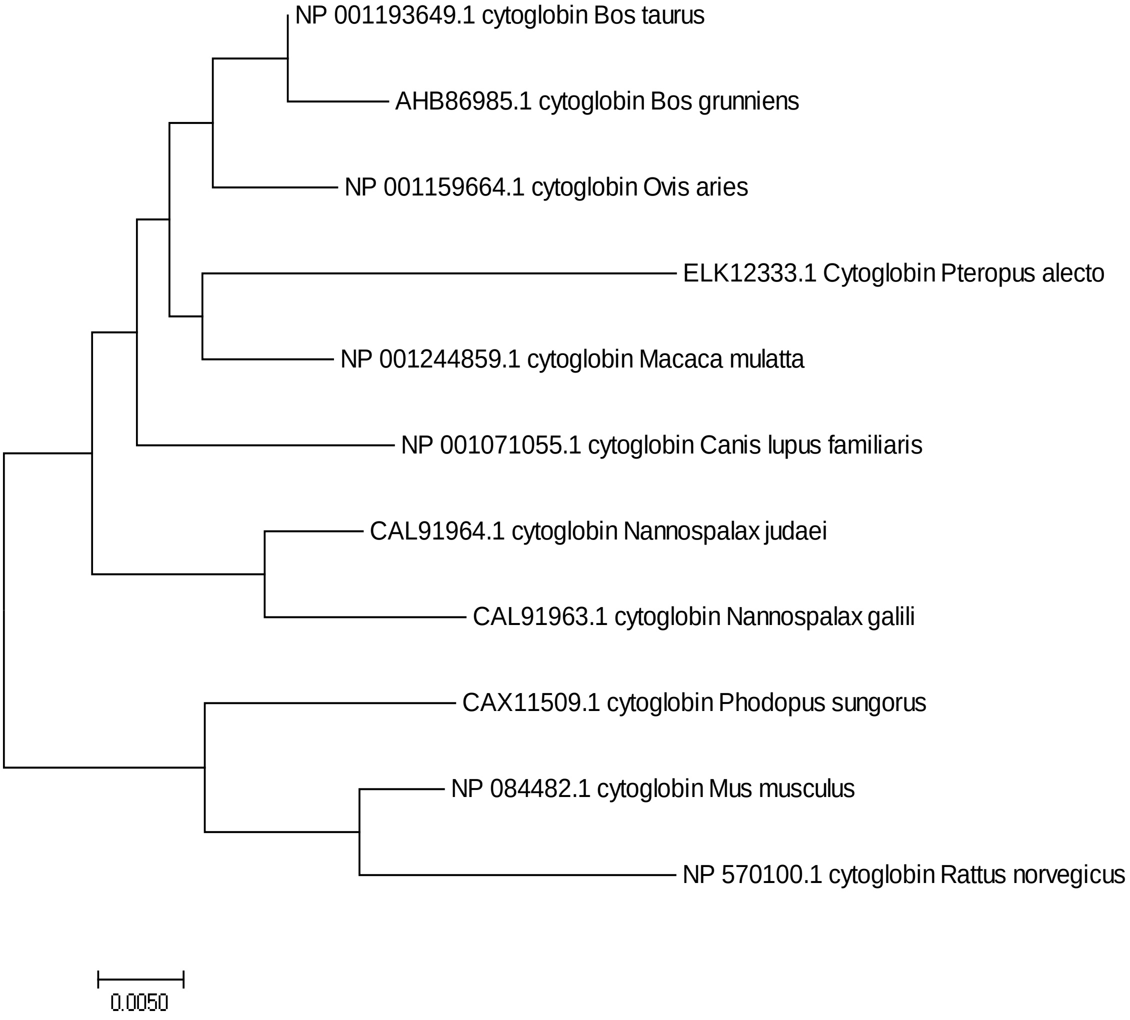


FIGURE 35: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Cytoglobin protein among various species. The analysis involved 11 amino acid sequences. Here *mus Musculus* and *Rattus norvegicus* are closely related.

5.ALLOPHYCOCYANIN

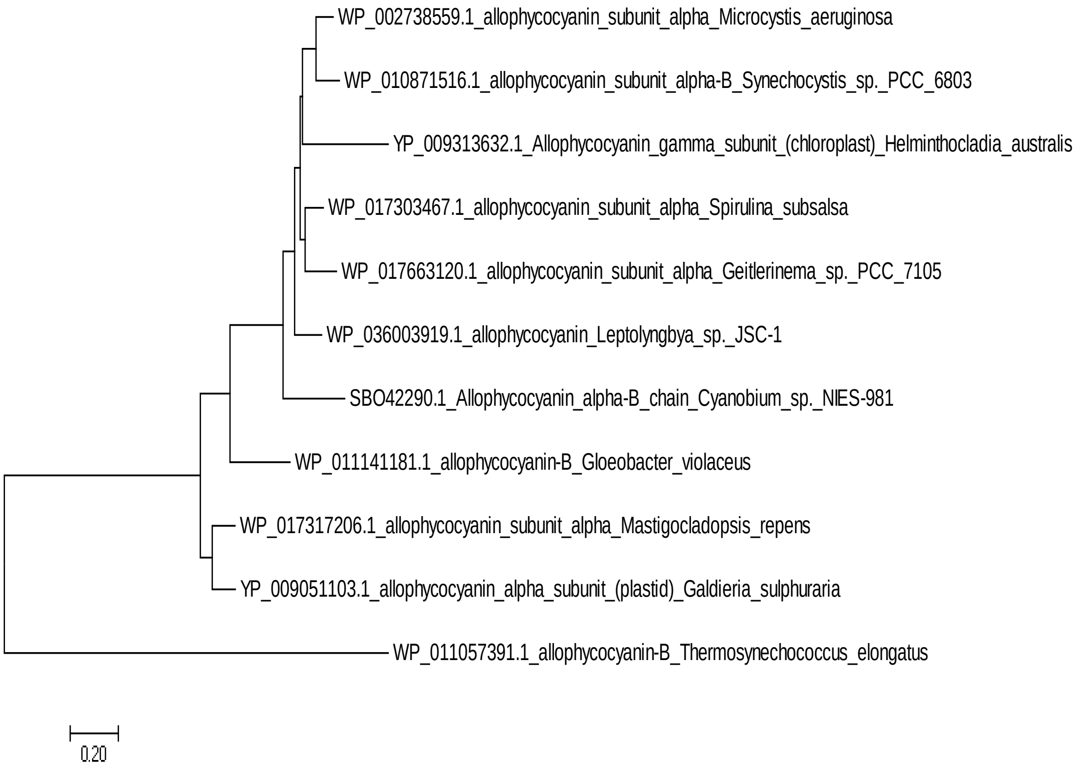


FIGURE 36: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Allophycoyanin protein among various species. The analysis involved 11 amino acid sequences. Here *Spirullina subsalsa* and *Geitlerinma* sp are closely related but *Thermosynechococcus elongates* is distantly related.

Results showing the nucleotidequery sequence families of GLOBIN–LIKE SUPERFAMILY retrieved in FASTA format from NCBI of different species.

1.Truncated Hemoglobin

PHYLOGENETIC TREES OF QUERY NUCLEOTIDE SEQUENCES

1. TRUNCATED HEMOGLOBIN

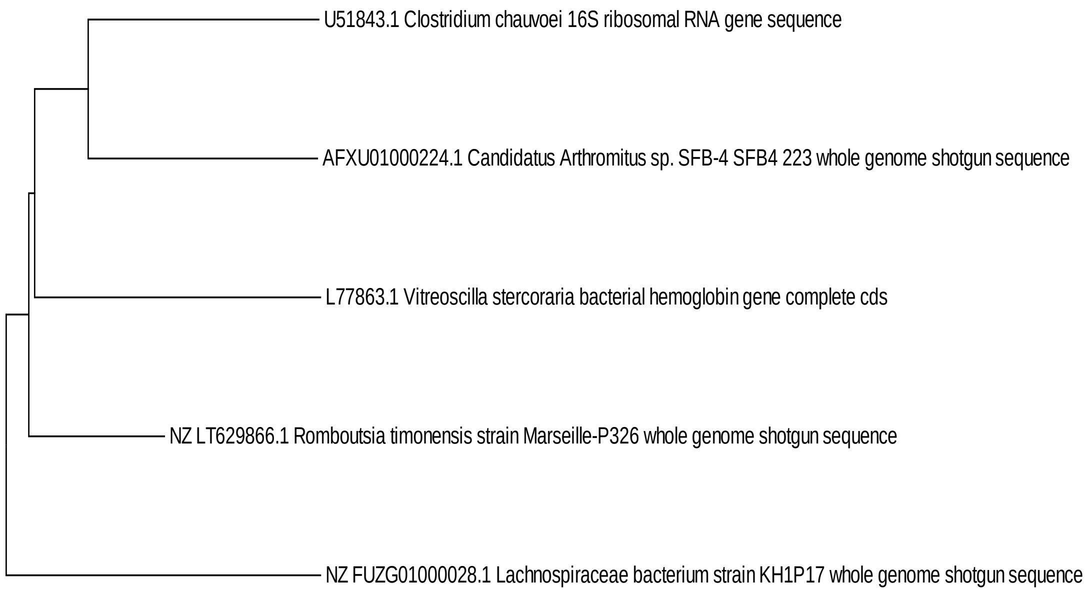


FIGURE 37 : Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Truncated hemoglobin nucleotides among various species. The analysis involved five nucleotide sequences. Here *Clostridium chavoei* and *Candidatus Artromitus* sp are closely related but Lachnospiraceae bacterium strain KH1P17 is distantly related.

2. MYOGLOBIN

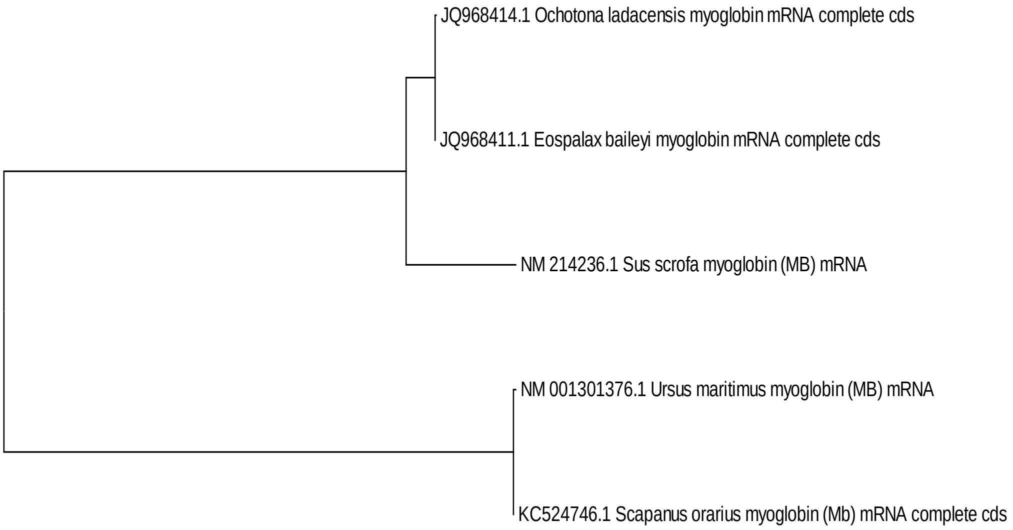


FIGURE 38: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Myoglobin nucleotides among various species. The analysis involved five nucleotide sequences. Here, *Ursus martimus* abd *Scapanus orarius* are closely related but Sus scrofa is distantly related

3. LEGHEMOGLOBIN

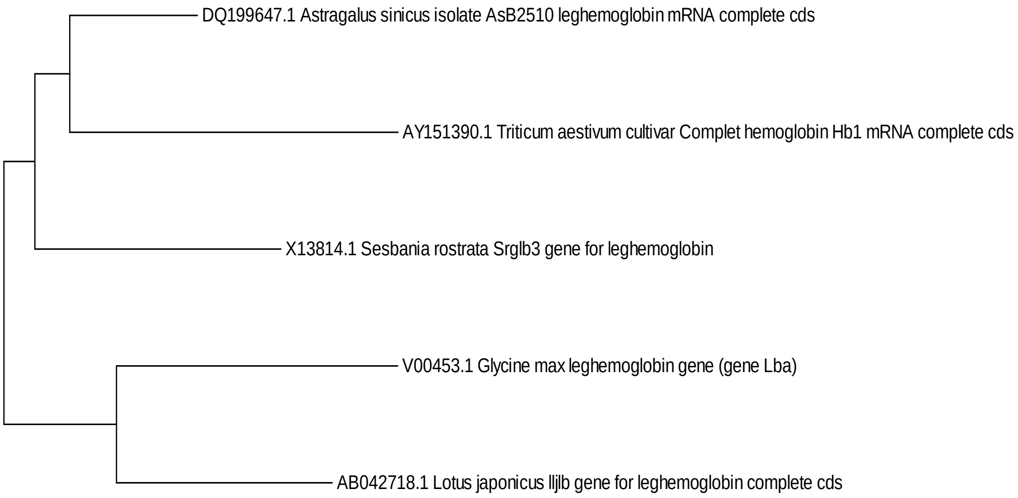


FIGURE39: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of

Leghemoglobin nucleotides among various species. The analysis involved five nucleotide sequences. Here *Glycine max* and *Lotus japonicas* are closely related.

4. CYTOGLOBIN

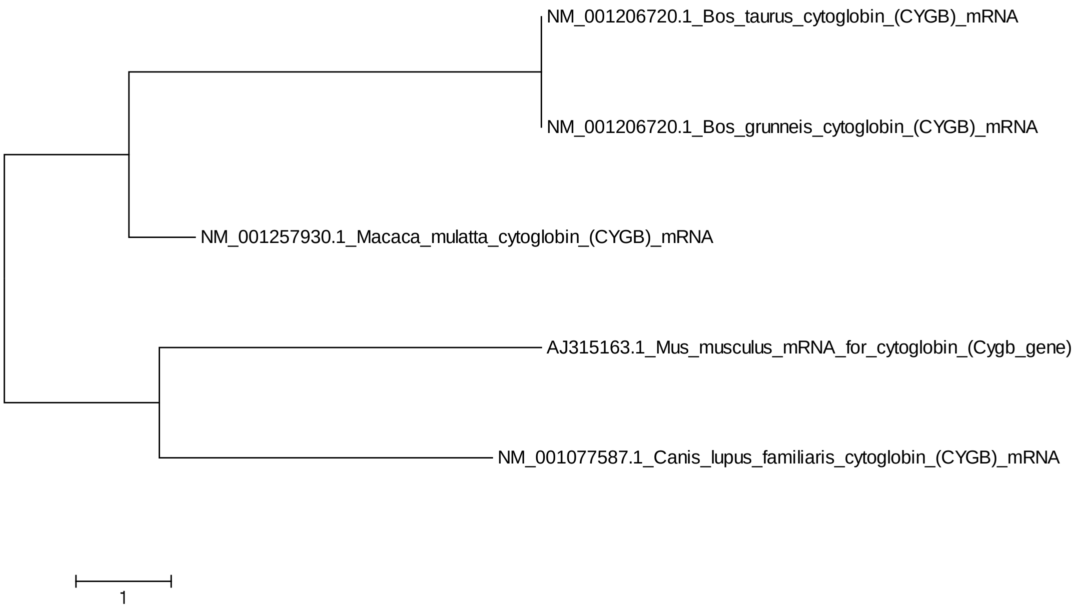


FIGURE 40: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Cytoglobin nucleotides among various species. The analysis involved five nucleotide sequences.Here *Bos Taurus* and *Bos grunneis* are closely related.

5. **ALLOPHYCOCYANIN**

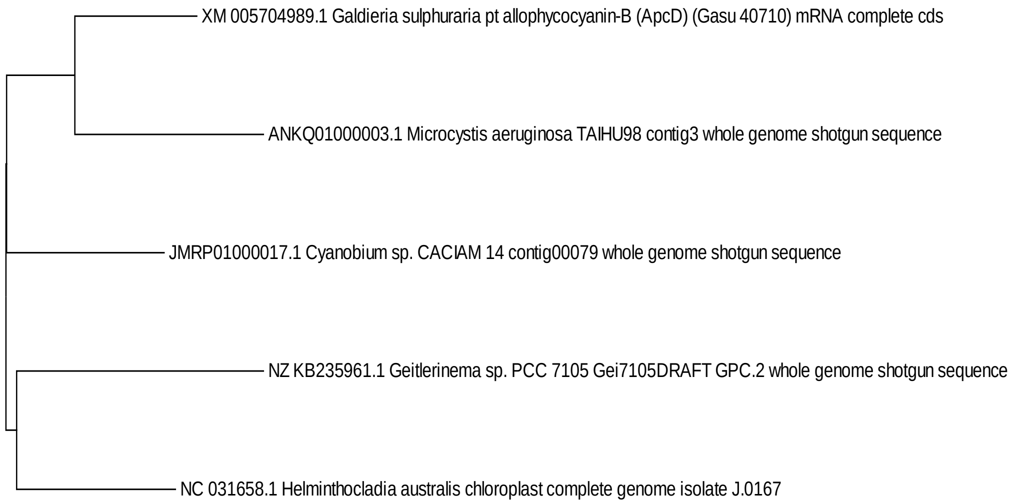


FIGURE 41: Comparitive analysis of globin proteins among various species were performed by Neighbour Joining method in MEGA 7 to know evolutionary history of Allophycoyanin nucleotides among various species. The analysis involved five nucleotide sequences. Here *Galdieria sulphuria* and *Microcystis aeruginosa* are closely related but Geitlerinema sp. PCC 7105 is distantly related.

DISCUSSION:

Protein sequence comparison has become one of the most powerful tools for charactering the protein sequences(structural and functional prediction) because of enarmaus amount of information that is conserved in the evolutionary process.The globins are a complex family of proteins.In our work we have studied evolutionary history of globins.Our phylogenetic analysis suggest that globins appeared early in history of life and are present in all three kingdoms of life. Although, we have analysed conventional divergence pattern among species in phylogenetic analysis at gene and protein level. The functionality of proteins also varies as depicted by variation in pI values, physiochemical properties. So, phylogenetic analysis at protein level will provide us important insights into the history of changing functions of these proteins.

ACKNOWLEDGEMENT:

Thanks to Gurpreet Kaur, Sherry Bhalla and Bhoomika Sharma for their support and guidance.

REFERENCES:

1. Balasubramanian L, Mishra M and Srinivasan N (2015) structure based phylogenetic analysis of lipocalin superfamily. *PLoS One*

2. Giangiacome Laura, Ilari Andew, Alberlo baffi, Morea Veronica and Chiancone (2005) The truncated oxygen-avid hemoglobin from *Bacillus subtilis. The journal of biological chemistry* **280**:9192-9202

3. Lindberg DA (2000) Internet access to national library of medicine (PDF). *Eff Clin Pract* **3(5)**:256-260

4. Murzin AG, Brenner SE, Hubbard T and Chothia C (1995) Scop:a structural and classification of proteins database for the investigation of sequences and structures. *J mol Biol* **247:**536-40

5. Nakai K (2000) *Protein* sorting signals and prediction of subcellular location. *Adv Protein Chem* **54**:277-344

6. Wilkins MR, Gasteige E, Bairoch A, Sanchez JC and Williams K (1999) Protein identification and analysis tool in ExPAsy Server.*Methods mol biol* **112**:531-552

7. Xie LK and Yang SH (2016) Brain globins in physiology and pathology. *Med Gas Res* **6**:154-163