Original Research Article

A Phylogenetic study of Plant, Animal and Microbial Hemoglobin using MEGA

N.Dhurigai*, Rachel Regi Daniel and S.Pitchumani Angayarkanni

Department of Botany &Microbiology, Lady Doak College, Madurai, India
*Corresponding author

A B S T R A C T

K e y w o r d s Hemoglobin, Leg- hemoglobin, Phylo- genetics, Symbiotic and Non symbiotic	Hemoglobin molecule is found ubiquitously in eukaryotes and many prokaryotes (Dordas *et al*., 2003). The gene encoding hemoglobin is thus very old, going back to the ancestor common to essentially all life on this earth. The aim of this study is to draw a phylogenetic tree with amino acid sequences of the protein hemoglobin of selected plants, animals and microbes using MEGA6 (Molecular Evolutionary Genetics Analysis) software and to estimate the relationships among the taxa and their hypothetical common ancestor. The analysis involved 47 amino acid sequences of hemoglobin from two bacteria, one alga, one fungus, four animals and 31 plants. The evolutionary history was inferred by using the Maximum Likelihood method and reliability of the phylogenetic tree was estimated using bootstrap method with bootstrap replicates. Animal hemoglobin subunits (α and β) together formed a separate clade. Hemoglobin of Vascular plants (symbiotic and non symbiotic) formed a separate clade. Ancestor for all vascular plant hemoglobin both symbiotic and non symbiotic was found to be the hemoglobin of Non vascular plant *Physcomitrella patens*. As per the tree hemoglobin sequence of *Photobacterium* and green unicellular alga *Chlamydomonas* seem to be the common ancestors for hemoglobin of all the other organisms analysed. Surprisingly both of these organisms are aquatic in nature.

Introduction

Hemoglobin is an ancient class of molecule uniting almost all forms of life on this planet. Whenever the term hemoglobin is mentioned oxygen transport within erythrocytes in the circulatory system of blood is remembered. The name hemoglobin comes from the globular structure associated with heme prosthetic group which binds oxygen. It may be therefore surprising that non-blood containing plants also encode hemoglobin. (Gupta *et al*., 2011) What surprises most people is the ubiquitousness and varied functions of hemoglobins not just in the animal world. Plants use hemoglobins, as do fungi, protists and bacteria. The discovery of hemoglobins in virtually all kingdoms of organisms has shown that the ancestral gene for hemoglobin is ancient, and that hemoglobins can serve additional functions besides transport of oxygen

between tissues, ranging from intracellular oxygen transport to catalysis of redox reactions (Hardison, 1998) However, hemoglobins not only bind oxygen but also other ligands such as nitric oxide (NO), carbon monoxide (CO), hydrogen sulfide (H_2S) and even organic molecules such as certain membrane lipids (Gupta *et al*., 2011)

The concentration of hemoglobin in erythrocytes is in the order of ten millimolars whilst that of myoglobin in the skeletal muscle is in the sub millimolar range. In plants, hemoglobin was first studied within symbiotic nitrogen-fixing root nodules, where ''leghemoglobin'' (legHb) plays a key role transporting free oxygen away from the oxygen-sensitive-nitrogenase enzyme. The concentration of legHb can reach 0.7 mM in nodules rendering them with their characteristic red color. In case of plants that produce hemoglobins with a non-symbiotic role the concentrations are usually in the range of 5–20 µM upon induction, too low for plants to be red. (Gupta *et al*., 2011) In plants, hemoglobins were first identified in species, which could fix nitrogen via symbiosis with bacteria. Recent findings suggested that another class of hemoglobins termed as non-symbiotic hemoglobins were present throughout the plant kingdom and were expressed differentially during plant development. Limited data available suggested that non-symbiotic hemoglobins were involved in hypoxic stress and oversupply of nutrients.

 Recently non-symbiotic hemoglobin genes were identified from both nitrogen and non-nitrogen fixing dicot and from monocot species(Reddy, 2007) Thus, two different types of hemoglobin have been discovered in plants, a non symbiotic type that is widely distributed and perhaps ubiquitous among species and a symbiotic type that is induced upon nodulation. (Hardison, 1998).

Plant hemoglobins constitute a diverse group of hemeproteins and evolutionarily belong to three different classes. Class 1 hemoglobins possess an extremely high affinity to oxygen and their main function consists in scavenging of nitric oxide (NO) at very low oxygen levels. Class 2 hemoglobins have a lower oxygen affinity and they facilitate oxygen supply to developing tissues. Symbiotic hemoglobins in nodules have mostly evolved from class 2 hemoglobins. Class 3 hemoglobins are truncated and represent a clade with a very low similarity to class 1 and 2 hemoglobins. They may regulate oxygen delivery at high O2 concentrations. (Gupta *et al*., 2011)

Comparative analysis of amino acid sequences is an increasingly important component of biological research (Lipman *et al.,* 1989). Accurate multiple sequence alignments of proteins are very important to several areas of computational biology and provide an understanding of phylogenetic history of domain families, their identification and classification (Chakrabarthi *et al*., 2006).

Phylogenetic analysis is sometimes regarded as being an intimidating, complex process that requires expertise and years of experience. In fact, it is a fairly straightforward process that can be learned quickly and applied effectively. Several steps are required to produce a phylogenetic tree from molecular data for novices.

The Molecular Evolutionary Genetics Analysis (MEGA) software is used to implement all those steps, thereby eliminating the need to learn several programs, and to deal with multiple file formats from one step to another (Tamura *et al.,* 2011). The first step, identification of a set of homologous sequences and downloading those sequences, is

implemented by MEGA's own browser built on top of the Google Chrome toolkit. For the second step, alignment of those sequences, MEGA offers two different algorithms: ClustalW and MUSCLE. For the third step, construction of a phylogenetic tree from the aligned sequences, MEGA offers many different methods.

Here the maximum likelihood method, beginning with MEGA's Models feature, which permits selecting the most suitable substitution model was illustrated. Finally, MEGA provides a powerful and flexible interface for the final step, actually drawing the tree for publication. Here a step-by-step protocol is presented in sufficient detail to allow a novice to start with a sequence of interest and to build a publication-quality tree illustrating the evolution of an appropriate set of homologs of that sequence (Hall, 2013).

Thus a phylogenetic study of Plant, animal and microbial hemoglobin using the software MEGA would be useful to analyze the evolutionary relationship of the organisms. This Paper aims to provide the phylogenetic relationship between selected plants, animals and microbes based on the amino acid sequence of the protein hemoglobin using MEGA.

Materials and Methods

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (Tamura, Stecher, Peterson, Filipski, and Kumar 2013).

Step 1: Acquiring the sequences

For the present study 47 amino acid sequences were chosen from the link http://www.ncbi.nlm.nih.gov/protein of NCBI protein data base.

Step 2: Aligning the sequences

The sequences were aligned using CLUSTAL W method.

Step 3: Estimating the Tree

The phylogenetic tree was constructed using Maximum Likelihood method. And the reliability of the tree was estimated by bootstrapping with bootstrap replicates.

Results and Discussion

The sequences listed in Table I were selected from Protein Database of NCBI and Multiple sequence alignment was performed with CLUSTAL W method of MEGA 6 version.

Molecular phylogenetic analyses of alignments of hemoglobins (Symbiotic and non-symbiotic) of selected plant with representative microbial and animal hemoglobins was carried out using Maximum likelihood method. Bootstrapping was done to test the reliability of the tree with 300 boot strap replicates.

The phylogenetic tree in the Figure I revealed the following results.

Hemoglobin of plants (symbiotic and non symbiotic) formed a separate clade which in turn gave rise to two major Nested clades.

All the plants with class I hemoglobins formed one nested clade of plants.

Symbiotic hemoglobins namely the Leghemoglobin (Class II plant hemoglobins) of all 13 plants along with Class II hemoglobin of 2 non symbiotic plants namely *Brassica napus* & *Arabidopsis lyrata subsp. lyrata* formed another nested clade.

As expected Class I and Class II hemoglobin of plants formed two nested clades though within a common clade of Plant Hemoglobins. This could be because of the differences in their function as stated by Gupta *et al*., 2011. Class 1 hemoglobins possess an extremely high affinity to oxygen and their main function consists in scavenging of nitric oxide (NO) at very low oxygen levels. Class 2 hemoglobins have a lower oxygen affinity and they facilitate oxygen supply to developing tissues. Symbiotic hemoglobins in nodules have mostly evolved from class 2 hemoglobins. As such they can be considered as a subclass of Class 2 hemoglobins of plants.

Animal hemoglobin subunits together formed a separate clade with alpha and beta globins forming the nested clades. This report is in accordance with statement given by Goodman *et al*. 1987. He reported that the amino acid sequences of the α- and β-globins were approximately 50% identical, regardless of which vertebrate species was the source, arguing that these two genes would have descended from a common ancestor approximately 450 million years ago, in the ancestral jawed vertebrate.

As per the tree hemoglobin sequences of Microbes form the Out group. *Photobacterium* and green unicellular alga *Chlamydomonas* were found to be the common ancestors for hemoglobin of all the other organisms analysed. Surprisingly both of these organisms are aquatic in nature. The hemoglobin of the terrestrial organisms namely the fungi *Colletotricum falcatum* and the bacterium *Rhizobium leguminosarum* seem to have evolved from the hemoglobin of aquatic organisms.
Bootstrap percentages indicated the reliability of the cluster descending from that node; the higher the number, more

reliable was the estimate of taxa that descend from that node. Nodes with percentage greater than 70 were considered more reliable and the rest were not considered. Thus it revealed which part of tree we should be trusted. (Hall, 2013). Based on this Bootstrap analysis was done with 300 replicates. And the nodes with more than 50% were displayed in the phylogenetic tree shown in Fig I.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model [1]. The tree with the highest log likelihood (-6081.0606) was shown. The percentage of trees in which the associated taxa clustered together was shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.8179)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.0000% sites). The analysis involved 47 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 132 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [2]. The bootstrap consensus tree inferred from 300 replicates is taken to represent the evolutionary history of the taxa analyzed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (300 replicates) was shown next to the branches.

Table.I List of Organisms used for the Phylogenetic analysis

	Plant: Symbiotic hemoglobin sequences
1.	>gi|77994689|gb|ABB13622.1| leghemoglobin [*Astragalus sinicus*]
2.	>gi|495289|gb|AAA18503.1| leghemoglobin [*Canavalia lineata*]
3.	>gi|169995|gb|AAA33980.1| leghemoglobin [*Glycine max*]
4.	>gi|11230993|dbj|BAB18108.1| leghemoglobin [*Lotus japonicus*]
5.	>gi|2921626|gb|AAC04853.1| leghemoglobin [*Lupinus luteus*]
6.	>gi|166388|gb|AAA32657.1| leghemoglobin [*Medicago sativa*]
7.	>gi|19663|emb|CAA40899.1| leghemoglobin [*Medicago truncatula*]
8.	>gi|169351|gb|AAA33767.1| leghemoglobin [*Phaseolus vulgaris*]
9.	>gi|2723412|dbj|BAA24088.1| leghemoglobin [*Pisum sativum*]
10.	>gi|20953|emb|CAA46704.1| leghemoglobin [*Psophocarpus tetragonolobus*]
11.	>gi|21375|emb|CAA32044.1| leghemoglobin [*Sesbania rostrata*]
12.	>gi|122087146|sp|P02232.2|LGB1_VICFA [*Vicia faba*]
13.	>gi|1177057|gb|AAA86756.1| leghemoglobin I [*Vigna unguiculata*]
	Plant: Non Symbiotic hemoglobin sequences
14.	>gi|475566186|gb|EMT14978.1| Non-symbiotic hemoglobin [*Aegilops tauschii*]
15.	>gi|84993584|dbj|BAE75956.1| nonsymbiotic hemoglobin [*Alnus firma*]
16.	>gi|297330635|gb|EFH61054.1| non-symbiotic hemoglobin 2 [*Arabidopsis lyrata subsp. lyrata]*
17.	>gi|2581783|gb|AAB82769.1| class 1 non-symbiotic hemoglobin [*Arabidopsis thaliana*]
18.	>gi|559807528|gb|AHB20278.1| non-symbiotic hemoglobin class 1 [*Beta vulgaris subsp. vulgaris*]
19.	>gi|15809392|gb|AAK07741.1| class 2 non-symbiotic hemoglobin [*Brassica napus*]
20.	>gi|18015|emb|CAA37898.1| hemoglobin [*Casuarina glauca*]
21.	>gi|1276977|gb|AAA97887.1| nonsymbiotic hemoglobin [*Glycine max*]
22.	>gi|62548111|gb|AAX86687.1| non-symbiotic hemoglobin protein [*Gossypium hirsutum*]
23.	>gi|2071976|gb|AAB70097.1| hemoglobin [*Hordeum vulgare*]
24.	>gi|77799654|dbj|BAE46739.1| nonsymbiotic hemoglobin [*Lotus japonicus*]
25.	>gi|11095158|gb|AAG29748.1|AF172172_1 non-symbiotic hemoglobin [*Medicago sativa*]
26.	>gi|657388440|gb|KEH30375.1| Non-symbiotic hemoglobin [*Medicago truncatula*]
27.	>gi|125661837|gb|ABN49927.1| hemoglobin [*Myrica gale*]
28.	>gi|2058498|gb|AAC49882.1| hemoglobin 1 [*Oryza sativa* Indica Group]
29.	>gi|14701802|gb|AAK72229.1| hemoglobin 1 [*Oryza sativa* Japonica Group]
30.	>gi|7658243|gb|AAF66104.1|AF218049_1 non-symbiotic hemoglobin [*Physcomitrella patens*]
31.	>gi|30909306|gb|AAP37043.1| nonsymbiotic hemoglobin [*Raphanus sativus*]
32.	>gi|12963875|gb|AAK07676.1| non-symbiotic hemoglobin class 1 [*Solanum lycopersicum]*
33.	>gi|27085253|gb|AAN85431.1| hemoglobin [*Solanum tuberosum*]
34.	>gi|377643998|gb|AFB70892.1| non-symbiotic hemoglobin [*Vigna radiata*]
35.	>gi|7271869|gb|AAF44664.1|AF236080_1 hemoglobin [*Zea mays*]
	Animal Hemoglobin sequences
36.	>gi|49169791|ref|NP_990820.1| hemoglobin subunit beta [*Gallus gallus*]
37.	>gi|52138655|ref|NP_001004376.1| hemoglobin subunit alpha-A [*Gallus gallus*]
38.	>gi|27819608|ref|NP_776342.1| hemoglobin subunit beta [*Bos taurus*]
39.	>gi|116812902|ref|NP_001070890.2| hemoglobin subunit alpha [*Bos taurus*]
40.	>gi|28558|emb|CAA23752.1| hemoglobin alpha chain [*Homo sapiens*]
41.	>gi|183830|gb|AAA52634.1| beta hemoglobin [*Homo sapiens*]
42.	>sp|P0CH25|HBA1_CAPHI Hemoglobin subunit alpha-1 OS=Capra hircus GN=HBA1 PE=1

	SV=1	
43.	>sp|P02078|HBBC_CAPHI Hemoglobin subunit beta-C OS=Capra hircus GN=HBBC PE=2 SV=;2	
Hemoglobin sequence of Bacteria, fungi and Algae		
44.	>gi|489640169|ref|WP_003544609.1| hemoglobin [*Rhizobium leguminosarum*]	
45.	>gi|441429904|gb|ELR67355.1| hemoglobin [*Photobacterium sp.* AK15]	
46.	>gi|380471629|emb|CCF47182.1| hemoglobin [*Colletotrichum higginsianum*]	
47.	>gi|10835657|pdb|1DLY|A Chain A, X-Ray Crystal Structure Of Hemoglobin From The Green Unicellular Alga *Chlamydomonas eugametos*	

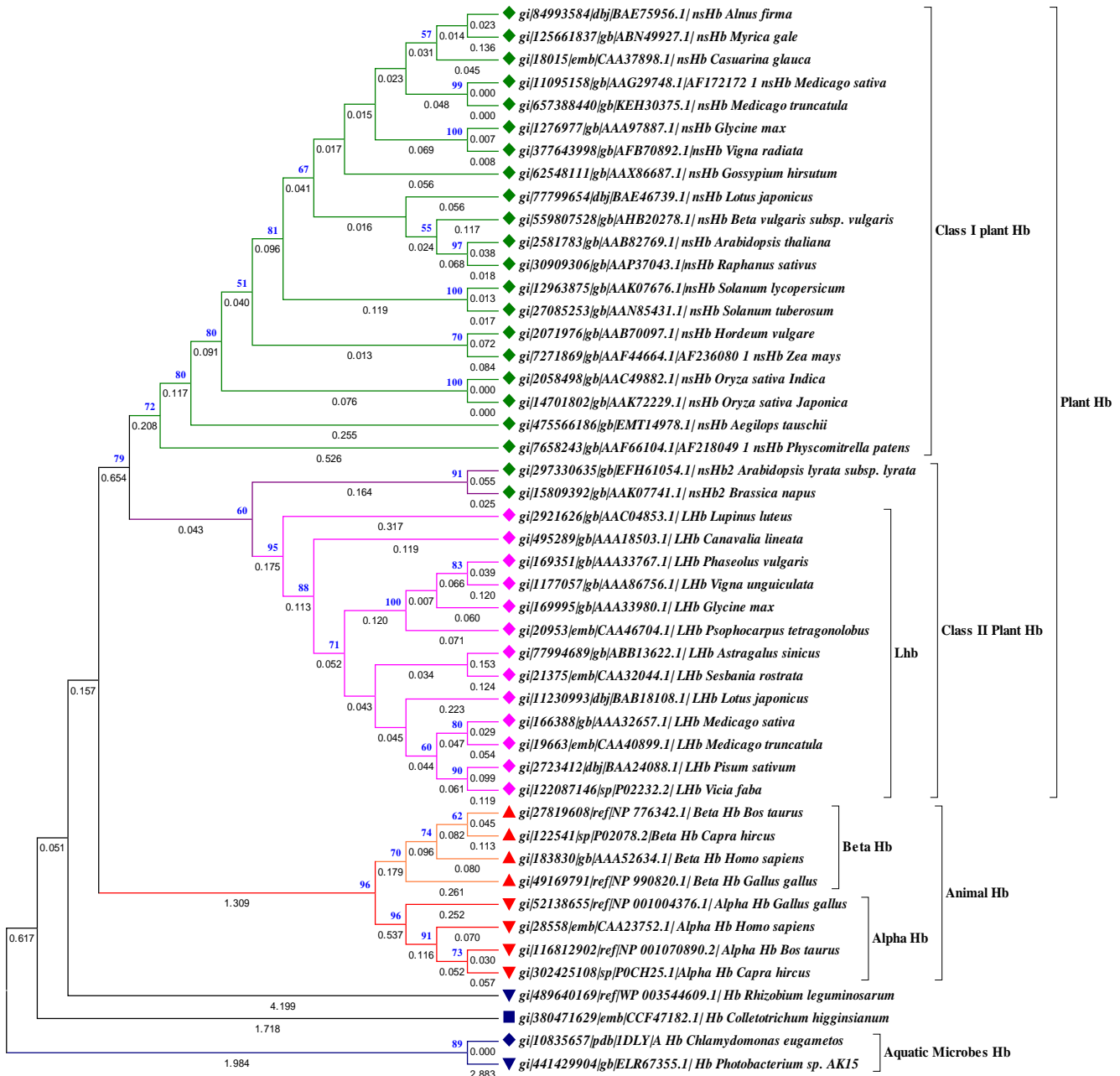## Figure.I Molecular Phylogenetic analysis by Maximum Likelihood method

Table.II Common features of the organisms in the phylogenetic tree

S.No	Name of Organism	Kingdom	Taxonomic Order	Nature of Hemoglobin	Comments
1.	*Alnus firma*	Plantae	Fagales	Non-symbiotic Hb	Nitrogen Fixing Non Legumes
2.	*Casuarina glauca*				
3.	*Myrica gale*				
4.	*Medicago sativa*		Fabales		Nitrogen Fixing legumes
5.	*Medicago truncatula*				
6.	*Glycine max*				
7.	*Vigna radiate*				
8.	*Gossypium hirsutum*		Malvales		Cotton
9.	*Lotus japonicas*		Fabales		Nitrogen Fixing legumes
10.	*Beta vulgaris subsp. vulgaris*		Caryophyllales		Beet
11.	*Arabidopsis thaliana*		Brassicales		Members of Brassicales
12.	*Raphanus sativus*				
13.	*Solanum lycopersicum*		Solanales		Solanales members
14.	*Solanum tuberosum*				
15.	*Hordeum vulgare*		Poales		Monocots
16.	*Zea mays*				
17.	*Oryza sativa Indica Group*				
18.	*Oryza sativa Japonica Group*				
19.	*Aegilops tauschii*				
20.	*Physcomitrella patens*		Funariales		Non Vascular Plant
21.	*Arabidopsis lyrata subsp. lyrata*		Brassicales	Non-symbiotic Class 2 Hb	Members of Brassicales
22.	*Brassica napus*				
23.	*Lupinus luteus*		Fabales	Symbiotic Lhb	Leguminous plants
24.	*Canavalia lineata*				
25.	*Astragalus sinicus*				
26.	*Sesbania rostrata*				
27.	*Lotus japonicus*				
28.	*Glycine max*				
29.	*Psophocarpus tetragonolobus*				
30.	*Phaseolus vulgaris*				
31.	*Vigna unguiculata*				
32.	*Medicago sativa*				
33.	*Medicago truncatula*				
34.	*Pisum sativum*				
35.	*Vicia faba*				
36.	*Bos Taurus*	Animalia	Artiodactyla	Hemoglobin - β subunit	Vertebrates
37.	*Capra hircus*				
38.	*Homo sapiens*		Primates		
39.	*Gallus gallus*		Galliformes		
40.	*Gallus gallus*			Hemoglobin - α subunit	
41.	*Homo sapiens*		Primates		
42.	*Bos Taurus*		Artiodactyla		
43.	*Capra hircus*				
44.	*Rhizobium leguminosarum*	Bacterium	Rhizobiales	Hemoglobin	Terrestrial
45.	*Colletotrichum higginsianum*	Fungi	Glomerellales		
46.	*Chlamydomonas eugametos*	Viridiplantae	Volvocales		Aquatic
47.	*Photobacterium sp. (AK15)*	Bacterium	Vibrionales		

Significance of this phylogenetic analysis is that it has proved Hemoglobins of all organisms have a common ancestry. The results support taxonomic classification in its own way. Clear partition of Kingdoms is seen and they form the main clades. And the nested clades are formed according to the functional difference and evolutionary order.

References

Chakrabarti, S., Lanczycki, C. J., Panchenko, A. R., Przytycka, T. M., Thiessen, P. A. and Bryant, S. H. 2006. State of the art: refinement of multiple sequence alignments. BMC Bioinform. 7:499

Dordas, C., Rivoal, J. and Hill, R.D. 2003. Plant haemoglobins, nitric oxide and hypoxic stress. Ann. Bot.91, 173–178.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Goodman, M., Czelusniak, J., Koop, B., Tagle, D. and Slightom, J. (1987). Globins: A case study in molecular phylogeny. Cold Spring Harbor Symp. quant. Biol. 52, 875–890.

Gupta, K. J., Hebelstrup, K. H., Mur, L. A. and Igamberdiev, A.U. 2011b. Plant hemoglobins: important players at the crossroads between oxygen and nitric oxide. FEBS Letters 585: 3843-3849.

Hall, B. G. 2013. Building Phylogenetic Trees from Molecular Data with MEGA Mol Biol Evol 30 (5): 1229-1235.

Hardison, R. C. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression J. Exp. Biol., 201, pp. 1099–1117

Lipman, D. J., Altschul, S. F. and Kececioglu, J. D. 1989. A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. USA 86, 4412-4415

Reddy, D. M. R. 2007. Evolutionary trace analysis of plant haemoglobins: implications for site-directed mutagenesis. Bioinformation 1(9): 370-375.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28: 2731–2739.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30: 2725-2729.

Whelan, S. and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.Molecular Biology and Evolution 18:691-699.