Evolutionary Analysis and Motif Discovery in Rhodopsin from Vertebrates

Kumar Akash[1] and Dwivedi Vivek Dhar[2]*

[1]Department of Bioinformatics, Uttaranchal College of Science and Technology, Dehradun, INDIA
[2]Forest Pathology Division, Forest Research Institute, Dehradun, INDIA

Abstract

In the present investigation, total twenty different protein sequences of rhodopsin from different organisms of vertebrates were obtained from GenPept database and only 347 characters of each sequence were considered for motif discovery, motif family analysis and phylogenetic analysis. Three different motifs were discovered by MEME program. The Pfam analysis of these motifs result revealed that two motifs belonged to 7 transmembrane receptor family. Two major clusters of all retrieved sequences were obtained after phylogenetic analysis.

Keywords: Evolutinary analysis, Motif discovery, rhodopsin, vertebrates, GenPept.

Introduction

Rhodopsins are biological pigments in photoreceptor cells of the retina that is responsible for the first events in the perception of light. Rhodopsin belongs to G-protein-coupled receptor family and is extremely sensitive to light, enabling vision in low-light conditions[1]. Exposed to light, the pigment immediately photobleaches, and it takes about 45 minutes to regenerate fully in humans[2]. Rhodopsin consists of the protein moiety opsin and a reversibly covalently bound cofactor, retinal. Opsin, a bundle of seven transmembrane helices connected to each other by protein loops, binds retinal, which is located in a central pocket on the seventh helix at a lysine residue. Retina lies horizontally with relation to the membrane. Each outer segment disc contains thousands of visual pigment molecules. Retinol is produced in the retina from Vitamin A, from dietary beta-carotene. Isomerization of 11-*cis*-retinal into all-*trans*-retinal by light induces a conformational change (bleaching) in opsin, continuing with metarhodopsin II, which activates the associated G protein transducin and triggers a Cyclic Guanosine Monophosphate, second messenger and cascade[3,4]. Considering the above facts the study of amino acid sequences of rhodopsin from different organisms of vertebrates is quit challenging. In this communication we performed the *In silico* analysis including motif identification, pfam analysis and phylogenetic analysis of various sequences of rhodopsin from vertebrates.

Material and Methods

Twenty different protein sequences of rhodopsin from different organisms of vertebrates were searched in GenPept database and randomly selected. All the selected sequences were opened in MEGA4 program and trimmed from end positions to make them for equal length[5]. Total 347 characters were taken for phylogenetic analysis and motif discovery. Motifs were identified in sequences using MEME program[6]. All motifs were subjected to their family analysis using Pfam Database[7]. The Multiple Sequence Alignment was performed using CLASTAL-W program before phylogenetic tree construction[8]. The phylogenetic analysis was performed by UPGMA method using MEGA4 program.

Results and Discussion

All retrieved sequences along with their accession number, species name and origin is listed in table-1. All the sequences were trimmed from end side to make them for equal lengths. Motif discovery result revealed that three motifs were discovered (figure-1). Figure-2, figure-4 and figure-6 are showing the sites of bock one, two and three respectively. Figure-3, figure-5 and figure-7 are showing the locations of motif one, two and three in each rhodopsin sequences. Figure-8 is showing the combined block diagram all Motifs locations of each block. The Pfam analysis showed that motif second and third was belonging to 7 transmembrane receptor family. The Pfam entry of first motif was not found. The multiple sequence alignment result showed some conserved regions in all aligned sequences.

Two major sequences clusters were obtained by phylogenetic analysis. Cluster I consisted of 15 species and further divided in two subcluters. Subcluster I consisted of 11 species namely *Rattus norvegicus*, *Mus musculus, Canis lupus familiaris, Felis catus, Cricetulus griseus, Homo sapiens, Sus scrofa, Cavia porcellus, Bos taurus, Gallus gallus* and *Taeniopygia guttata*. Subcluster II consisted of 4 species namely *Xenopus laevis, Xenopus tropicalis, Rana pipiens* and *Rana catesbeiana*. Cluster II contained 5 species namely *Sargocentron tiere, Plecoglossus altivelis, Poecilia reticulata, Danio rerio* and *Cyprinus carpio*.

Table-1
Organism name and accession number of all retrieved sequences from GenPept of rhodopsin from vertebrates

S.No.	Organism name	Accession number
1.	*Plecoglossus altivelis*	BAB88652.1
2.	*Homo sapiens*	NP_000530.1
3.	*Bos taurus*	NP_001014890.1
4.	*Rattus norvegicus*	NP_254276.1
5.	*Xenopus laevis*	NP_001080517.1
6.	*Sus scrofa*	NP_999386.1
7.	*Canis lupus familiaris*	NP_001008277.1
8.	*Xenopus tropicalis*	NP_001090803.1
9.	*Danio rerio*	NP_571159.1
10.	*Gallus gallus*	NP_001025777.1
11.	*Taeniopygia guttata*	NP_001070163.1
12.	*Rana pipiens*	AAB24265.1
13.	*Mus musculus*	NP_663358.1
14.	*Cavia porcellus*	NP_001166556.1
15.	*Cyprinus carpio*	AAB33306.1
16.	*Rana catesbeiana*	AAB35478.1
17.	*Poecilia reticulata*	ABI33107.1
18.	*Felis catus*	NP_001009242.1
19.	*Cricetulus griseus*	NP_001231336.1
20.	*Sargocentron tiere*	AAB39534.1

DISCOVERED MOTIFS

Motif Overview

Motif 1
- 2.9e-623
- 18 sites

Motif 2
- 3.8e-750
- 18 sites

Motif 3
- 4.7e-734
- 18 sites
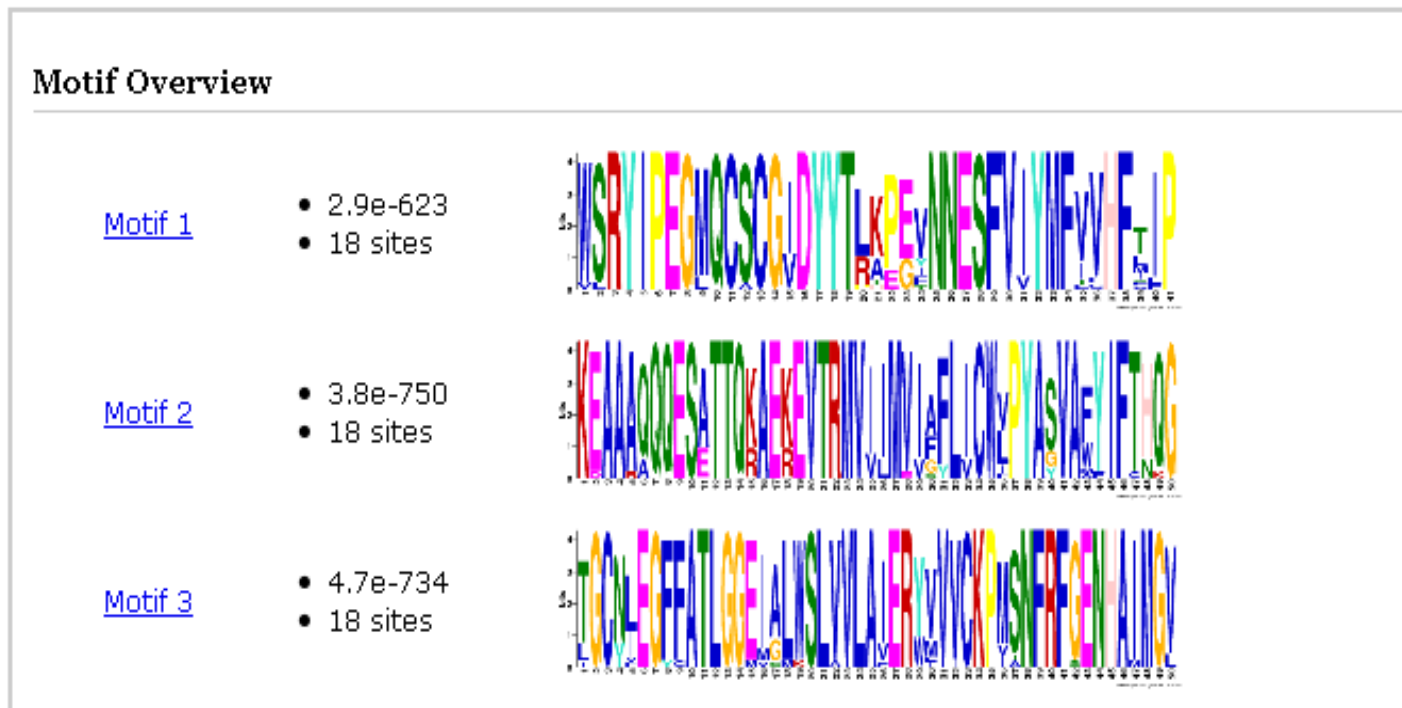
Figure-1
Conserved motifs of rhodopsin

Name	Start	*p*-value	Sites
Felis	175	1.26e-53	LACAAPPLVG WSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIP MIVIFFCYGQ
Cavia	175	1.26e-53	LACAAPPLVG WSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIP MIIIFFCYGQ
Mus	175	1.26e-53	LACAAPPLVG WSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIP MIVIFFCYGQ
Rattus	175	1.26e-53	LACAAPPLVG WSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIP MIVIFFCYGQ
Taeniopygia	175	3.09e-53	LACAAPPLFG WSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIP LSIIFFCYGN
Cricetulus	175	5.49e-53	LACAAPPLVG WSRYIPEGMQCSCGVDYYTLKPEVNNESFVIYMFVVHFTIP LIVIFFCYGQ
Xenopus	175	5.49e-53	LSCAAPPLFG WSRYIPEGMQCSCGVDYYTLKPEVNNESFVIYMFVVHFTIP LIVIFFCYGR
Rana	175	2.27e-52	LACAVPPLFG WSRYIPEGMQCSCGVDYYTLKPEVNNESFVIYMFVVHFLIP LIIISFCYGR
Gallus	175	3.77e-52	MACAAPPLFG WSRYIPEGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIP LAVIFFCYGN
Homo	175	2.19e-51	LACAAPPLAG WSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIP MIIIFFCYGQ
Sus	175	1.85e-50	LACAAPPLVG WSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFSIP LVIIFFCYGQ
Danio	175	1.20e-48	CSCAVPPLVG WSRYIPEGMQCSCGVDYYTRTPGVNNESFVIYMFIVHFFIP LIVIFFCYGR
Bos	175	1.34e-48	LACAAPPLVG WSRYIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIP LIVIFFCYGQ
Cyprinus	175	2.57e-48	CTCAVPPLVG WSRYIPEGMQCSCGVDYYTRAPGYNNESFVIYMFLVHFIIP LIVIFFCYGR
Plecoglossus	175	1.26e-47	AACAVPPLFG WSRYIPEGMQCSCGIDYYTRAPGFNNESFVVYMFIVHFTLP LTVVTFCYGR
Canis	185	5.35e-47	WSSLLSHSPL VLRYIPEGMQCSCGIDYYTLKPEINNESFVIYMFVVHFAIP MIVIFFCYGQ
Poecilia	175	5.31e-46	NSCAAPPLLG WSRYIPEGMQCSCGVDYYTRAEGFNNESFVVYMFICHFLIP LVVVFFCYGR
Sargocentron	172	4.97e-44	LACTVPPLVG WSRYIPEGMQCACGIDYYTRAEGYNNESFVIYMFTFHFLFP MFIIFFCYGR

Figure-2
Site of Block one
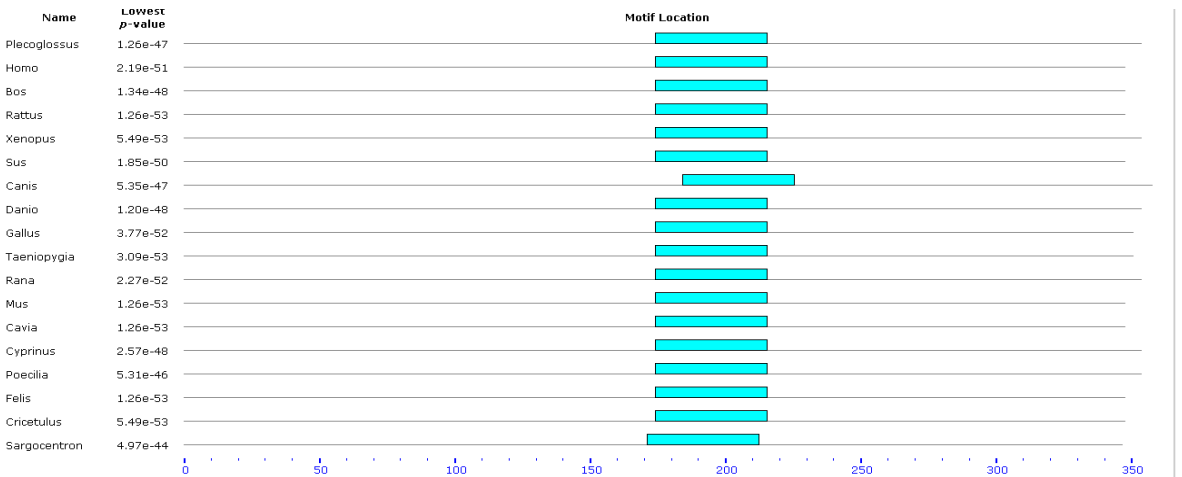
Figure-3
Block One Show the Motif Location in each rhodopsin **sequences**

Name	Start	*p*-value	Sites
Felis	231	2.26e-63	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQG SNFGPIFMTL
Canis	241	2.26e-63	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQG SDFGPIFMTL
Homo	231	2.26e-63	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQG SNFGPIFMTI
Mus	231	6.87e-63	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWLPYASVAFYIFTHQG SNFGPIFMTL
Bos	231	2.61e-62	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQG SDFGPIFMTI
Sus	231	4.01e-62	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWLPYASVAFYIFTHQG SDFGPIFMTI
Cavia	231	8.13e-62	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAAYIFTHQG SNFGPIFMTV
Rattus	231	1.07e-61	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWLPYASVAMYIFTHQG SNFGPIFMTL
Rana	231	1.21e-61	FCYGRLVCTV KEAAAQQQESATTQKAEKEVTRMVIIMVIFFLICWVPYAYVAFYIFTHQG SEFGPIFMTV
Taeniopygia	231	1.94e-61	FCYGNLVCTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTNQG SDFGPIFMTI
Gallus	231	1.94e-61	FCYGNLVCTV KEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTNQG SDFGPIFMTI
Danio	231	2.55e-60	FCYGRLVCTV KEAARQQQESETTQRAEREVTRMVIIMVIAFLICWLPYAGVAWYIFTHQG SEFGPVFMTL
Xenopus	231	8.31e-60	FCYGRLLCTV KEAAAQQQESATTQKAEKEVTRMVVIMVVFFLICWVPYAYVAFYIFTHQG SDFGPVFMTV
Cyprinus	231	6.41e-59	FCYGRLVCTV KDAAAQQQESETTQRAEREVTRMVVIMVIGFLICWIPYASVAWYIFTHQG SEFGPVFMTV
Cricetulus	231	8.91e-59	FCYGQLVFTV KEAAAQQQESATTQKAEKEVTRMVILMVVFFLICWFPYAGVAFYIFTHQG SNFGPIFMTL
Poecilia	231	4.91e-58	FCYGRLLCAV KEAAAAQQESETTQRAEREVTRMVVIMVIGFLVCWIPYASVAWYIFTHQG SEFGPLFMTV
Sargocentron	228	1.55e-54	FCYGRLLCAV KEAAAAQQESETTQRAEREVTRMVILMVIGYLVCWLPYASVAWFIFTHKG SEFGPLFMAV
Plecoglossus	231	4.86e-53	FCYGRLLCTV KEAAAQQQESETTQRAEREVTRMVVLMEISYLVCWLPYASVAWYIFCNQG SEFGPVFMTA

Figure-4
Site of Block Two

International Research Journal of Biological Sciences _____ ISSN 2278-3202

Vol. **2(7),** 6-11, July **(2013)** *Int. Res. J. Biological Sci.*
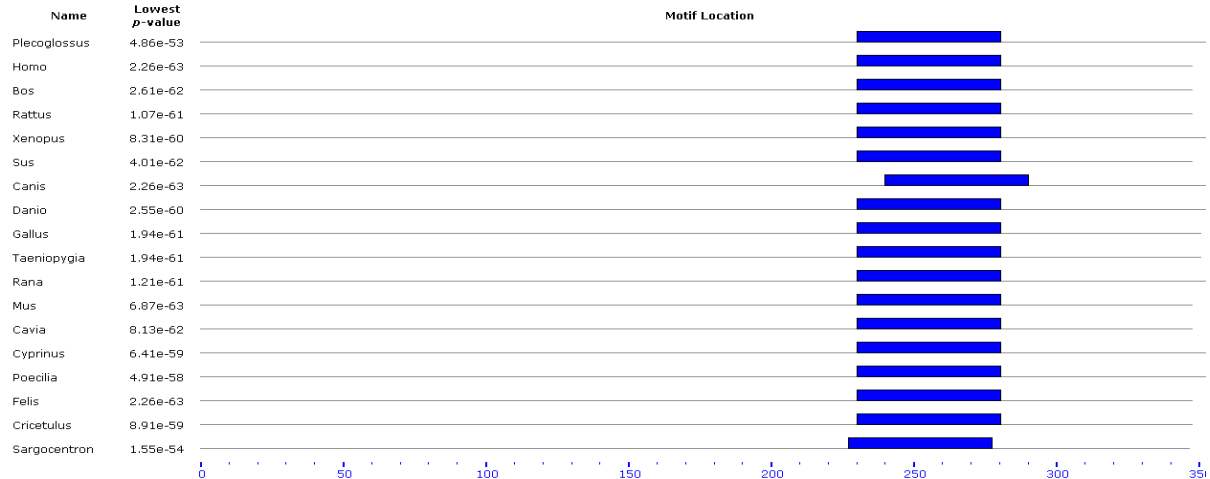
Figure-5
Block Two Show the Motif Location in each rhodopsin **sequences**
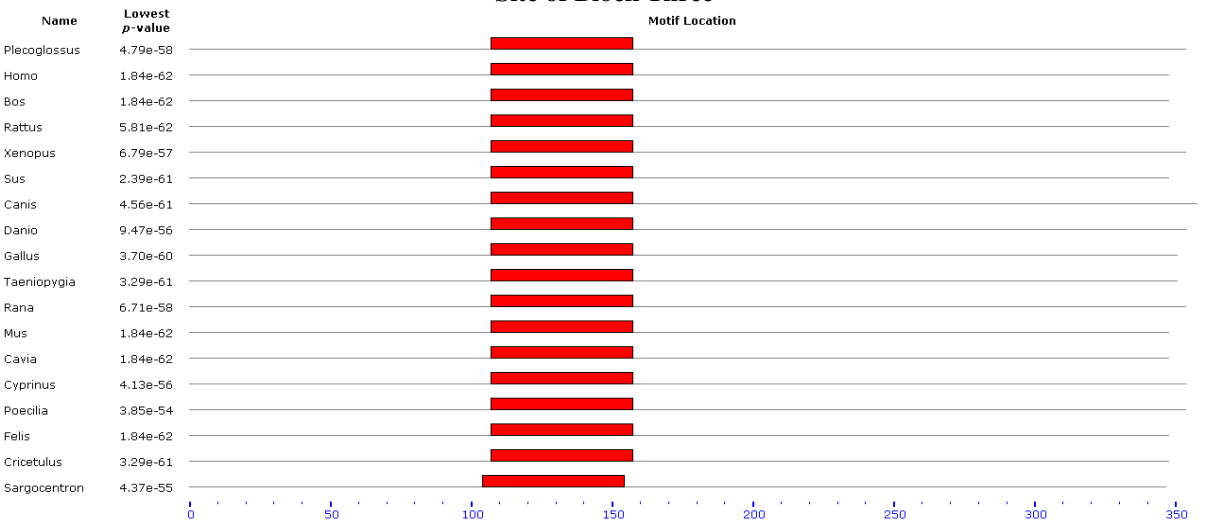

Figure-6
Site of Block Three

Figure-7
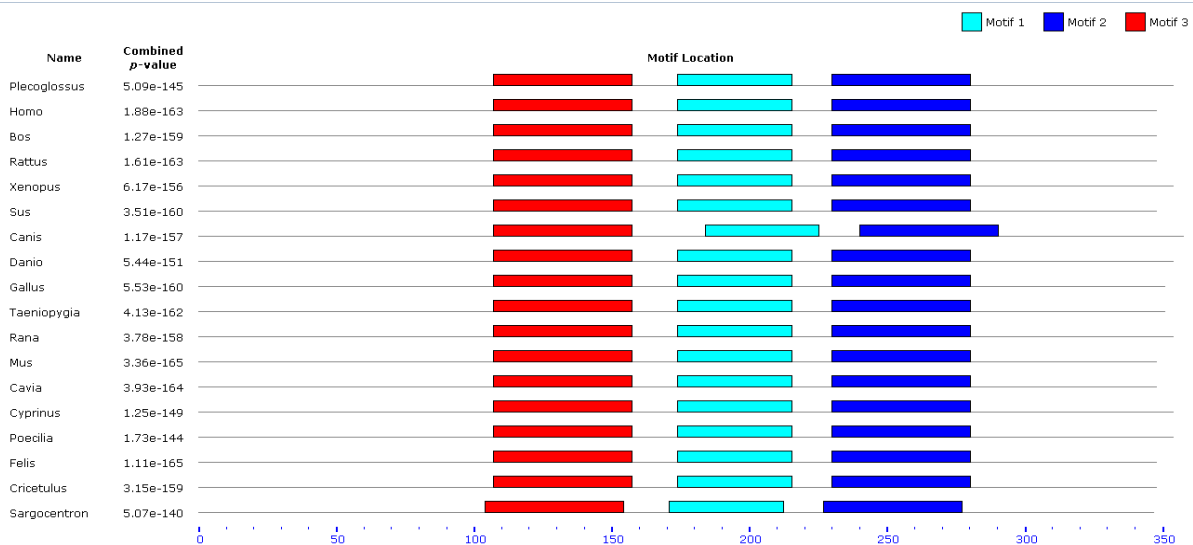**Block Three Show the Motif Location in each** rhodopsin **sequences**

Figure- 8
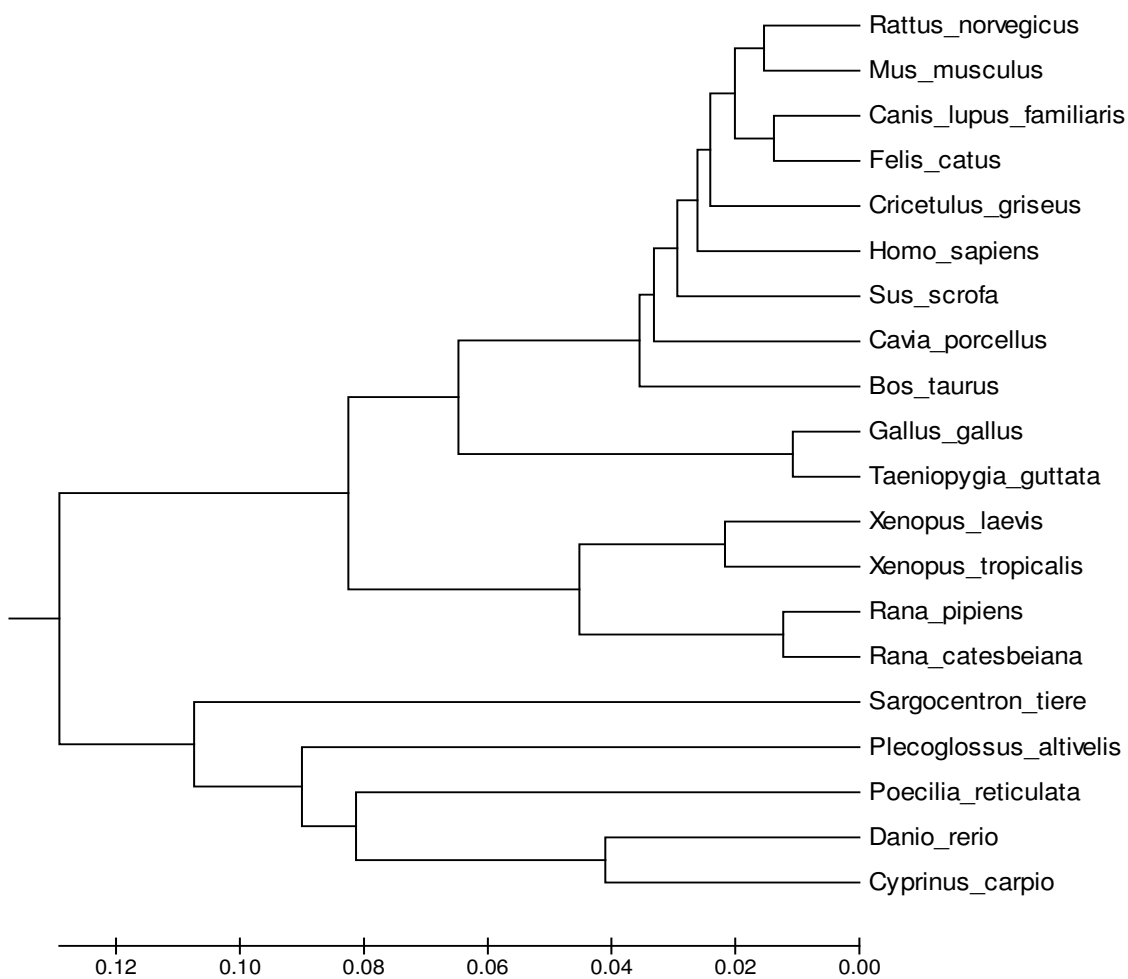**Combined block diagram show the Motif location of each block**

Figure-9
Phylogenetic analysis of all retrieved sequences of rhodopsin from vertebrates

Conclusion

Motifs identification in a group of related sequences of rhodopsin showed the evolutionary relationships of functional features among different organisms of vertebrates. Three motifs belonging to 7 transmembrane receptor family was identified. This suggests that these motifs have an important function in the evolution of rhodopsin in vertebrates. Two major sequence clusters were obtained by phylogenetic analysis. This suggests that the sequences of cluster I is more closely related in comparison to sequences of cluster II. This classification can significantly contribute in the understanding of the evolutionary relationships between the species at sequence level[9-11].

References

1. Humphries P., Kenna P. and Farrar G.J., On the molecular genetics of retinitis pigmentosa, *Science,* **256 (5058)**, 804–8**(1992)**

2. Edwards S.C., Involvement of cGMP and calcium in the photoresponse in vertebrate photoreceptor cells, *The Journal of the Florida Medical Association,* **82(7),** 485–8 **(1995)**

3. Maghtheh M., Gregory C., Inglehearn C., *et al.*, Rhodopsin mutations in autosomal dominant retinitis pigmentosa, *Hum. Mutat.,* **2(4)**, 249–55 **(1993)**

4. Garriga P., Manyosa J., The eye photoreceptor protein rhodopsin. Structural implications for retinal disease, *FEBS Lett.*, **528(1–3)**, 17–22 (2002)

5. Kumar S., Dudley J., Nei M., and Tamura K., MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences**,** *Briefings in Bioinformatics*, **9**, 299-306 **(2008)**

6. Bailey T.L. and Elkan C., Unsupervised learning of multiple motifs in biopolymers using expectation maximization, Mach Learn, **21(51)**, 80-33 **(1995)**

7. Punta M., Coggill P.C., Eberhardt R.Y., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer E.L.L., Eddy S.R., Bateman A., and Finn R.D. The Pfam Protein Families Database, Nucleic *Acids Research Database* **(2012)**

8. Thompson J.D., Higgins D.G., Gibson T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice, *Nucleic Acids Res*., **22(22)**, 4673-80 **(1994)**

9. Dwivedi V.D., Arora S., Kumar A. and Mishra S.K., Computational analysis of xanthine dehydrogenase enzyme from different source organisms, *Network Modeling Analysis in Health Informatics and Bioinformatics,* DOI : 10.1007/s13721-013-0029-7**(2012)**

10. Dhar D.V., Tanuj S., Kumar M.S. and Kumar P.A., Insights to Sequence Information of Lactoylglutathione Lyase Enzyme from Different Source Organisms, *I. Res. J. Biological Sci.,* **1(6),** 38-42 **(2012)**

11. Dhar D.V., Tanuj S., Amit P. and Kumar M.S., Insights to Sequence Information of Alpha Amylase Enzyme from Different Source Organisms, *International Journal of Advanced Biotechnology and Bioinformatics*, **1(1),** 87-91 **(2012)**