Evolutionary Analysis and Motif Discovery in Pinopsin from Vertebrates

Akash Kumar[1], Mohd. Zakir Khawaja[2*],Vivek Dhar Dwivedi [3*]

[1]*Department of Bioinformatics,* [2]*Department of Biotechnology ,*
[1,2]*Uttaranchal College of Science & Technology, Dehradun*
[3]*Forest Pathology Division, Forest Research Institute, Dehradun*
welkinakash@gmail.com

Abstract

In the present investigation, total nine protein sequences of Pinopsin 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 where minimum motif width was 6 and maximum motif width was 50. All three discovered motifs were aligned using MAST tool which revealed the similarity between all of three submitted motif's sequence. The motif matches shown have a position p-value less than 0.0001. Each of the following 9 sequences has an E-value less than 10. Two major sequence clusters were constructed by phylogenetic analysis.

1. Introduction

Pinopsin are one of the photoreceptor protein which have absorption maximum at 470 nm . It belongs to Opsin protein family and found in pineal region and in brain .Opsin protein are sensitive to light and changes their structure when light fall on it . It have two broad division with further several subdivision. Type I opsins are found in prokaryotes, whereas animals use Type II opsins .Vertibrate opsin comes under Type II opsin and further subdivided based on their evolutionary history and several other related factor . Opsin start a cascade of activity when exposed to light which leads or boost the overall result of opsin (Maghtheh et al., 1993; Garriga et al., 2002). There are few opsin which are absent from vertibrates sucha as go/gs opsin. In vertebrate opsin Rod opsin are used for night vision and found in rod photoreceptor cell while cone opsin responsible for colour vision. In Vertibrate opsins , rhodopsins belongs to G-protein-coupled receptor family and is extremely sensitive to light, enabling vision in low-light conditions (Humphries et al., 1992). When light fall to it the pigment immediately photobleaches and takes about 45 minutes to regenerate fully in humans (Edwards et al., 1995). Considering the above points the study of amino acid sequences of pinopsin from different organisms of vertebrates is quit challenging. In this communication we performed the *In silico* analysis including motif identification and phylogenetic analysis of various sequences of pinopsins from vertebrates and show more clear view for evolution of pinopsin , and we hope that it will be helpful for new discoveries.

2. Materials and Methods

Nine amino acid sequences of pinopsin 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 (Kumar et al., 2008; Thompson et al., 1994). Motifs were identified in sequences using MEME program (Bailey and Elkan, 1995). All three discovered motifs were aligned using MAST program (Bailey and Gribskov, 1998) to judge the similarity between three motifs discovered. The Multiple Sequence Alignment was performed using MUSCLE program and CLUSTAL-W (Thompson et al., 1994) program before we construct phylogenetic tree . The phylogenetic analysis was performed by UPGMA method using MEGA4 program (Kumar et al., 2008).

3. Result and Discussion

The accession number of retrieved sequences along with the species name and origin is listed in Table-1. 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 pinopsin sequences. Figure-8 is showing the combined block diagram all Motifs locations of each block. Figure-9 showing the similarity among discovered motif sequences. Each of the following 9 sequences has an E-value less than 10 (Figure-10). The motif matches shown have a position p-value less than 0.0001(Figure-8). 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 8 species and further divided in two subcluster. Subcluster I consisted of 4 species namely Trichoplax adhaerens , Xenopus (Silu-

rana) tropicalis , Bufo japonicas , Uta stansburiana and Subcluster II consisted of 4 species namely Phelsuma sundbergi longinsulae , Iguana iguana , Iguana iguana , Columba livia , Podarcis siculus. Cluster II consist of Gallus gallus .

Table 1. *Organism name and accession number of all retrieved sequences from GenPept of pinopsin from vertebrates*

Serial no.	Organism name	Accession number
1	Trichoplax adhaerens	EDV22958.1
2	Xenopus (Silurana) tropicalis	NP_998830.1
3	Bufo japonicus	AAF12820.1
4	Uta stansburiana	AAZ79905.1
5	Phelsuma sundbergi longinsulae	BAA90297.1
6	Iguana iguana	BAM28747.1
7	Columba livia	EMC80590.1
8	Podarcis siculus	AAY34940.2
9	Gallus gallus	NP_990740.1

Fig.1. *Conserved motifs of pinopsin*

Fig.2. *Site of Block one*

Fig.3. *Block One Show the Motif Location in each pinopsin sequences*

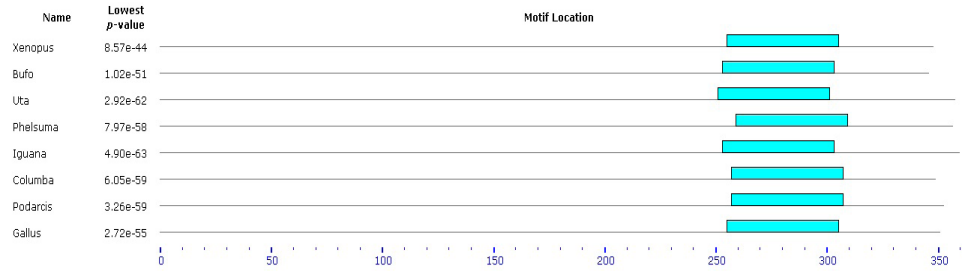


Fig.4. *Site of Block Two*

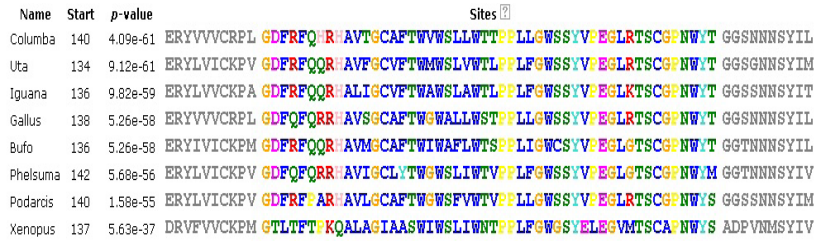


Fig.5. *Block Two Show the Motif Location in each pinopsin sequences*

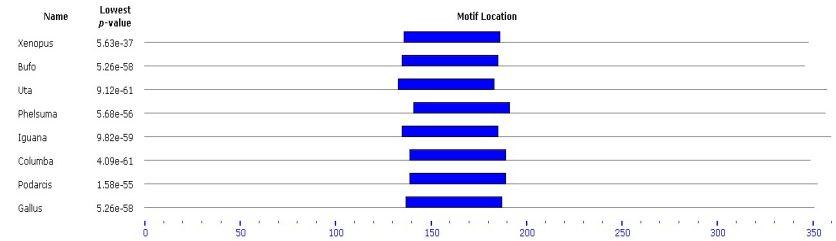


Fig.6. *Site of Block Three*

Fig.7. *Block Three Show the Motif Location in each pinopsin sequences*

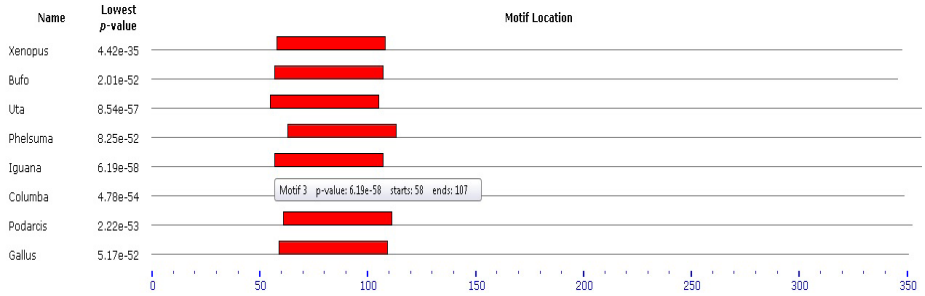


Fig.8. *Combined block diagram show the Motif location of each block*

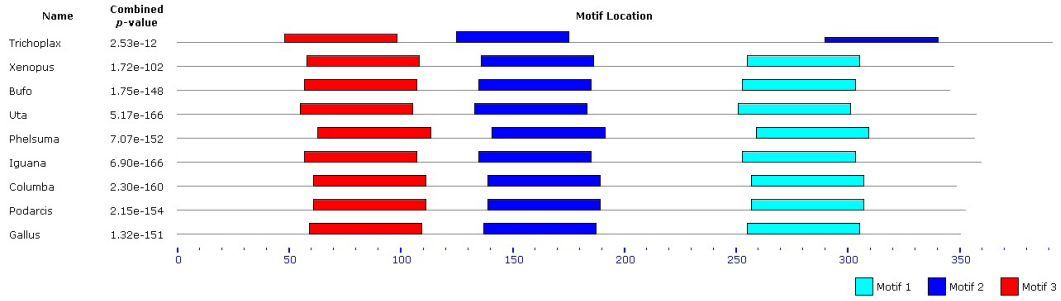
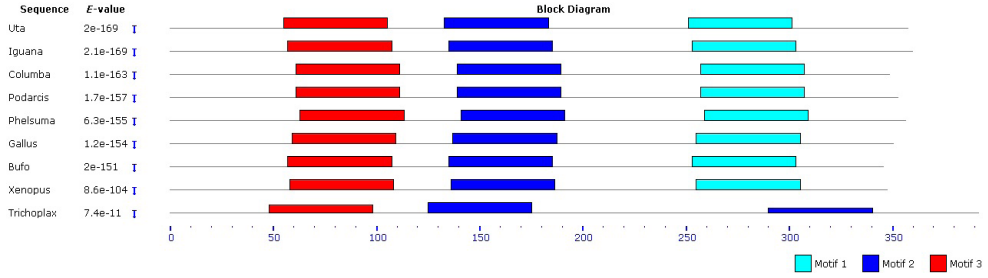


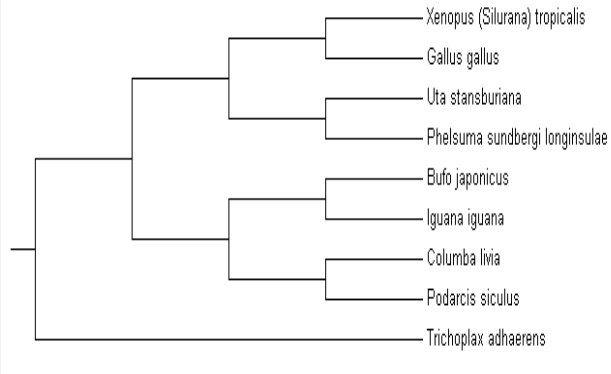
Fig.9. *Best possible match diagram showing similarity between submitted Motif*

Motif	Width	Best possible match	Similarity 1	2	3
1	50	CWLPYATFAMVVATNKDLVIQPALASLPSYFSKTATVYNPIIYVFMNKQF	-	0.15	0.18
2	50	GDFRFQQRHAVIGCAFTWGWSLIWTTPPLFGWSSYVPEGLRTSCGPNWYT	0.15	-	0.14
3	50	KKLRSPLNYILVNLAIADLLVTFFGSTISFHNNIYGFFVFGKRACEFEGF	0.18	0.14	-

Fig.10. *Above block diagram shows the E-value among all Nine sequences submitted*



Phylogenetic analysis of all retrieved sequences of Pinopsin from vertebrates



4. Conclusion

Motifs identification and similarity in a group of related sequences of pinopsin showed the evolutionary relationships of function features among different organisms of vertebrates. This suggests that these motifs have an important function in the evolution of pinopsin 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 significantly contributes in the understanding of the evolutionary relationships between the species at molecular level (Dwivedi et al., 2012 2013) and presents an very exciting picture of new discoveries which may show the evolution in general.

5. References

1. Humphries P., Kenna P., 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., Elkan C., Unsupervised learning of multiple motifs in biopolymers using expectation maximization, Mach Learn 21 (51), 80-33 (1995)
7. 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)
8. Timothy L. Bailey and Michael Gribskov, "Combining evidence using p-values: application to sequence homology searches", Bioinformatics, 14(1):48-54, 1998. .
9. Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", **Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology**, pp. 28-36, AAAI Press, Menlo Park, California, 1994.
10. Dwivedi V.D., Sharma T., Pandey A. and Mishra S.K., Insights to Sequence Information of Alpha amylase Enzyme from Different Source Organisms, I.J.A.B.B. 1(1), 87-91 (2012)
11. 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(2013).