GENOME-WIDE IDENTIFICATION AND CHARACTERIZATION OF DL_X GENES IN STRONGYLOCENTROTUS PURPURATUS

MAZHAR, M. W.[1*] – MAHMOOD, J.[1] – SAIF, S.[1] – WAQAS, N.[1] – SIKANDAR, M.[1]

[1] *Department of Biotechnology and Bioinformatics, Government College University, Faisalabad, Pakistan.*

**Corresponding author*
e-mail: waqarmazhar63[at]gmail.com

(Received 10th April 2021; accepted 17th June 2021)

Abstract. DLX is a group of genes that is part of the homeobox-containing superfamily, which is involved in a variety of morphogenetic processes. Dlx genes occur as several paralogues in the invertebrate genome, resulting from tandem replication accompanied by whole-genome duplication. Despite the growing number of genome sequence tools in various vertebrates, embryological study, Dlx expression and function remains a challenge in terms of species diversity. Dlx gene family is in between the multigene families encoding for homeodomain-containing the transcription factors that are expressed earlier in the cells for the cranial neural crest and later in craniofacial mesenchyme and skeletal tissues. Gene structure analysis shows the number of exons and introns by a 3-dimensional structure of the Dlx gene. The Dlx gene family exons and introns are found to be very similar during this analysis. The distribution of genes across chromosomes was very varied. Collectively the newly discovered genes provide data for manipulating the Dlx genome to develop organs in vertebrates.
Keywords: *distal-less, homeobox genes, vertebrates, hematopoiesis*

Introduction

The Dlx gene family is the earliest known gene family, it comprises a highly conserved family of homeobox genes homologous to distal-less (D-II) gens of drosophila. It is important for developing insect limbs, its expression is maintained throughout the limb development. This family is said to have a role in developing features of jaws and limbs. Dlx is well preserved among species (Merlo et al., 2004). It is a member of the homeobox-containing superfamily which is involved in morphogenetic processes, multiple paralogues of the Dlx gene occur in the invertebrate genome as a result of tandem replication followed by whole-genome duplication. Both Dlx genes are expressed in spatially and temporally restricted patterns in craniofacial primordia basal telencephalon as well as in distal regions of growing appendages. The Dlx gene is used to test the hypothesis of separate development programme for nerve and visceral skull sheath cartilage and skin bones by comparing the expression pattern of six Dlx gene during skull formation in zebra fish ranging from 1 DPF TO 15 DPF. Dlx gene is included among the multigene households encoding for homeo domain containing transcription elements, While Dlx genes are linked to ectodermal and neural-crest related tissues, they also show expression in the mesoderm (Kraus and Lufkin, 2006).

A homeobox present in each of the six DLX genes is similar to the one found in the insect Distal-less (Dll) gene. The six DLX genes are grouped into three tail-to-tail biogenic pairs and located on chromosomes with HOX clusters (DLX5/DLX6) syntenic to the HOXA cluster), (DLX1/DLX2), HOXD cluster syntenic (Stock, 2005), and (DLX3/DLX4), HOXD cluster syntenic (Azeem et al., 2020). During embryonic

development, DLX genes regulate appendage and craniofacial morphogenesis, as well as the differentiation of reproductive organs in adults, they regulate bone homeostasis and tissue integrity. The DLX5 protein is made up of 289 amino acids and has a molecular weight of 31.5 kDa. The homeobox protein, which has low similarity to the distal N-terminus, and the homeodomain, which has a high similarity to the distal N-terminus, is the two motifs in the protein. Dlx5 and Dlx6 are expressed by differentiated osteoblasts. In the mouse germline, Dlx1, Dlx2, Dlx3, and Dlx5 are involved in craniofacial structure, sensory organ morphogenesis, osteogenesis, and placenta formation. In a study of gene inactivation, however, there was no effect on limb development.

The initial tandem duplication resulted in a pair of Dlx genes being bound in pairs after a pair of chain links and arthropods diverged, but before the envelope and vertebrates diverged, according to a genetic study of the Dlx gene sequence in the form of its chromosomal arrangement. The number of chromosomal events then doubled, resulting in four distinct classes of bony fish and tetrapods with Hox gene signatures. In mammals, a pair of Dlx genes belonging to the Hox family may have been missing. We can't tell whether the replication is independent or not. And keep bone vertebrates in their ancestral state to understand why zebrafish have more Dlx genes than mammals. The discovery of a connection between these additional zebrafish Dlx genes and the Hox group should aid in the solution of this issue (*Figure 1*).

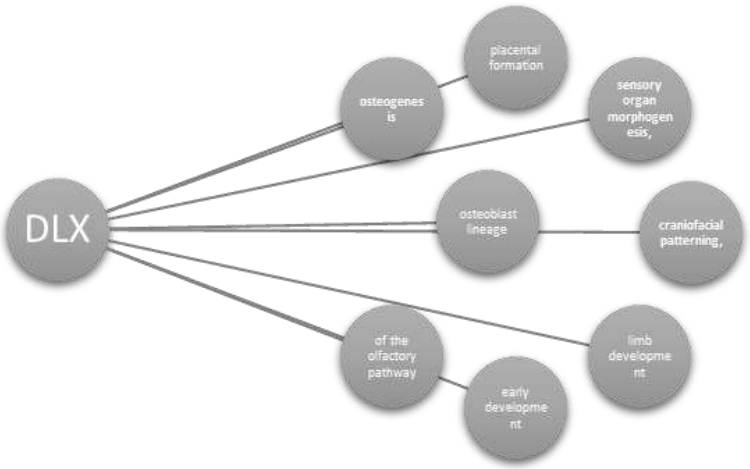

Figure 1. *Major activities referred to the Dl_X superfamily.*

Materials and Methods

Retrieval of DNA and protein sequence of the gene DLX in Homo sapiens from all NCBI databases, FASTA format of the Dlx protein, and BLAST-P was run for Homo sapiens genome database to find similar sequences from protein databases. To perform this protein sequence of *distal less gene* Strongylocentrotus purpuratus (purple sea urchin) was given as a query sequence. Target sequences are then used in further databases like PFAM databases to find out the conserved domains and motifs.

Identification of fomains and motifs

The sequence of Dlx protein was given as a query sequence and downloaded all target sequences to analyze the conserved motifs and domains of the protein sequences

among are strongylocentrotus purpuratus, Anneissia japonica, parasteatoda tepidriorum, and Asterias Rubens at NCBI online database Pfam (EMBL Official Portal, 2021) and SMART (SMART Official Portal, 2021) (*Table 1*). This step was done to remove sequences in which there were no conserved domains and motifs that are essential for the Dlx protein to perform its functions normally. Only one conserved domain was identified *Homeodomain* (PF00046) is a member of the superfamily CL0123. All of the variants were carefully examined with an E-value less than 1e-10 and those with lengthy ORF were chosen for further data analysis using various databases (EMBL Official Portal, 2021; NCBI Official Portal, 2021; SMART Official Portal, 2021). Position of the chromosome, length of the protein, and cDNA and genomic sequences were observed by using the NCBI database.

Table 1. *Proposed nomenclature and important features of Dl_X gene.*

Proposed names	Gene locus	Protein accession no	RNA accession no	Exons	Chromosome no.	Orf length	Amino acid length	Start of genomic location	Conserved domains in protein sequence
Strongylocentrotus purpuratus	Unknown	NP-001123282 1	XM-001129810.1	3	Unplaced scaffold	1326	441	55287731	Pfam0046
Anneissia japonica	LOC117104046	XP-033100592.1	XM-033244470 1.1	3	Unplaced scaffold	852	283	1272681	Pfam00046 Pfam09770
Asterias rubens	LOC117288453	XP-033625212.1	XM-003769321.1	3	Unplaced scaffold	1110	369	3640497	Pfam00046
Parasteatoda Tepidrioum	LOC104750100	NP-001310748 1.1	NM-001323381 9.1	4	Unplaced scaffold	1020	339	4639660	Pfam0046

Gene structure and phylogenetic analysis of DL_X gene family

Dlx gene structure was observed by using an online tool named Gene structure display sever to determine the position of introns and exons and a three-dimensional structure of the protein was obtained (Ahmad et al., 2020; Azeem et al., 2020). By using the bioinformatics tool MEGA7 with default parameters all the sequences from different organism like strongylocentrotus purpuratus, paracentrotus lividus, peronella japonica, patiria Miniata, ptychodera Flava, Balanoglossus misakiensis, schizocardium californicum, Balanoglossus simodensis and Asterias rubens were aligned to perform multiple alignment and pairwise alignment in order to report the conserved domains in Dlx gene members. MEGA7 tool was used to coordinate sequences and a perfect unrooted evolutionary tree was built on the basis of alignments of sequences (Stock, 2005). Tree topology and branch lengths are from a maximum likelihood analysis, with midpoint rooting. Nodes with single numbers indicate the results of a bootstrap maximum likelihood analysis with 100 replicates. Even after that, the Dlx gene family was classified based on their evolutionary similarity. A separate instrument was used to measure the molecular weight and isoelectric point toolbox (Expasy Official Portal, 2021). Evolutionary analysis was used to name the subfamilies of the Dlx gene based on their homologs. The evolutionary relationship among the families was performed using the online Time Tree Server (Azeem et al., 2020).

Results and Discussion

The Dlx family has six members, which are organised into three convergently transcribed (arrows) loci. Each pair is close a Hox cluster: DLX1 and DLX2 on human chromosome 2, DLX5 and DLX6 on human chromosome 7, and DLX3 and DLX4 on human chromosome 17 with the HoxB cluster.

Pfam

It is used for functional annotation of genomic data. It enables us to submit protein sequences to search for matches to families in the database. It gives greater weight to matches at conserved sites, allowing better homology detection as shown in *Figure 2* to *Figure 5*. The Dlx gene has only one important domain and 16 insignificant domains, according to Pfam (a wide set of protein families each defined by multiple sequence alignments and hidden Markov models) (*Figure 3*).

Figure 2. *The platform of Pfam.*

Figure 3. *Conserve domain of super family CL0123.*

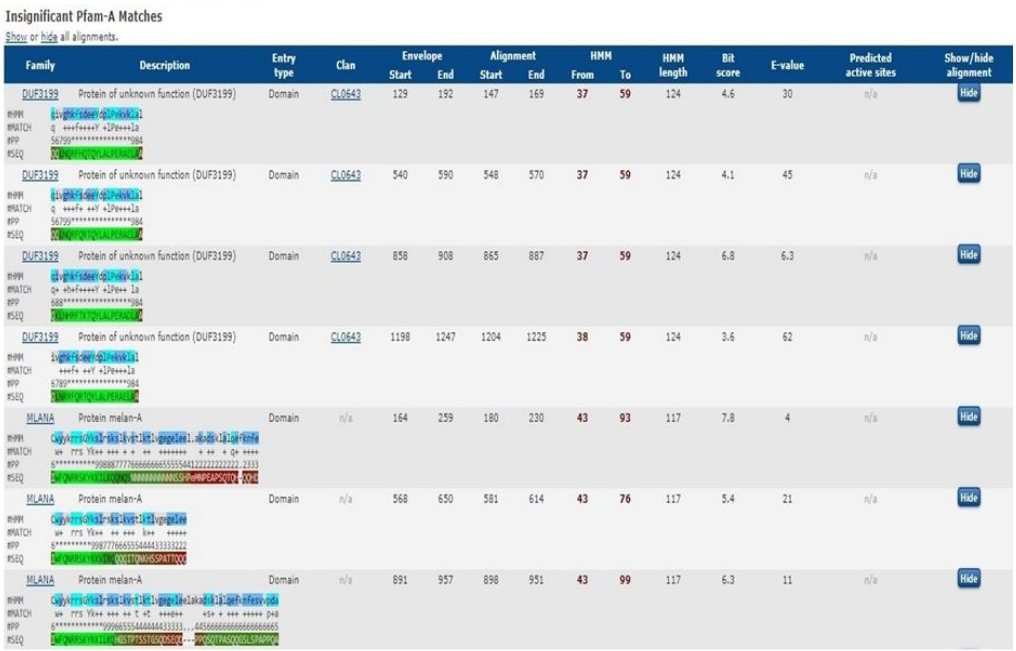

Figure 4. *Insignificant matches of Dl_X gene family.*

Figure 5. *Insignificant matches of Dl_X gene family.*

Genome structure display server

The visualization of Dlx gene features such as the composition and location of exons, introns, and conserved domains were done using the Gene Structure Display Server (GSDS) as shown in *Figure 6*.

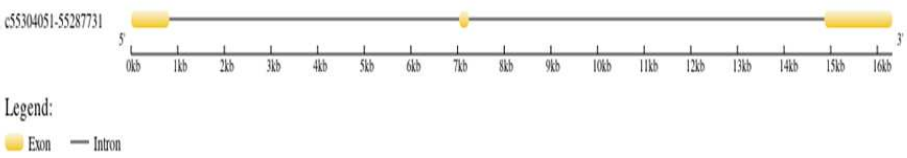

Figure 6. *Exon introns analysis of Dl_X gene.*

Mega 7

A phylogenetic tree, also called phylogeny, is a diagram that shows how various plants, organisms, or genes derived from a common ancestor. Phylogenies are useful for organising biological diversity records, structuring classifications, and gaining insight into evolutionary events (Baum, 2008). Phylogenetic analysis at the molecular level should be figured out. Gene distribution across chromosomes, on the other hand, was extremely diverse. The newly discovered genes can hold the key to manipulating the Dlx genome to grow limbs. The Dlx gene's evolutionary relationships with closely related species are depicted in the phylogenetic tree as shown in *Figure 7*.

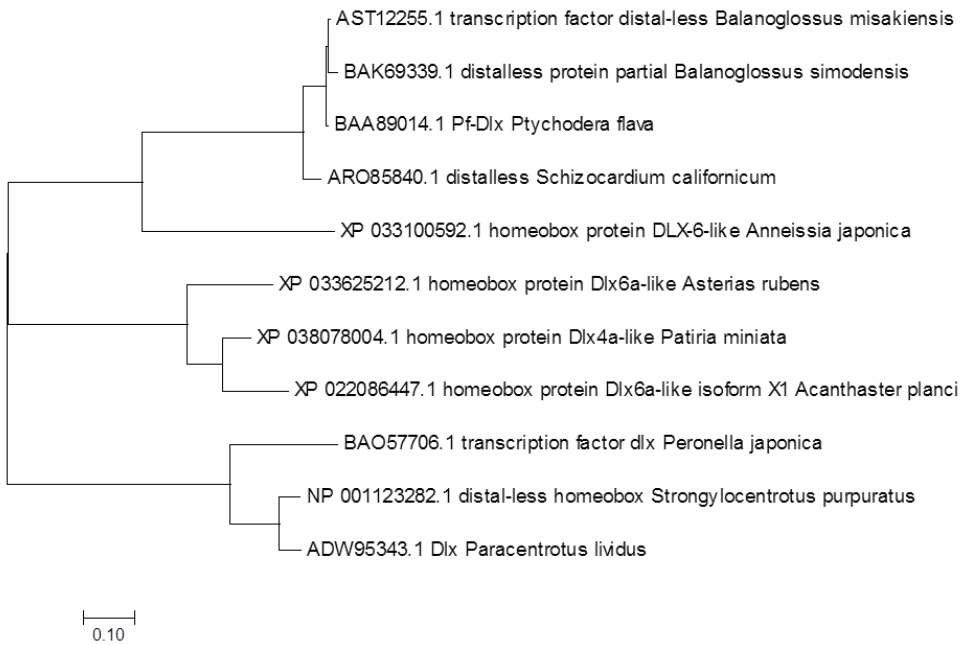

Figure 7. *Phylogenetic tree of Dl_X gene family.*

Serial cloner

This tool is used for sequence analysis of proteins and DNA. It has multiple applications including the translation of ORF .serial cloning of protein sequence and genome sequence of Dlx gene is performed and the graphic map of the plasmid is obtained. This tool could also perform virtual PCR. Results of the graphic map of protein and genome sequences are illustrated below as shown in *Figure 8* and *Figure 9*.

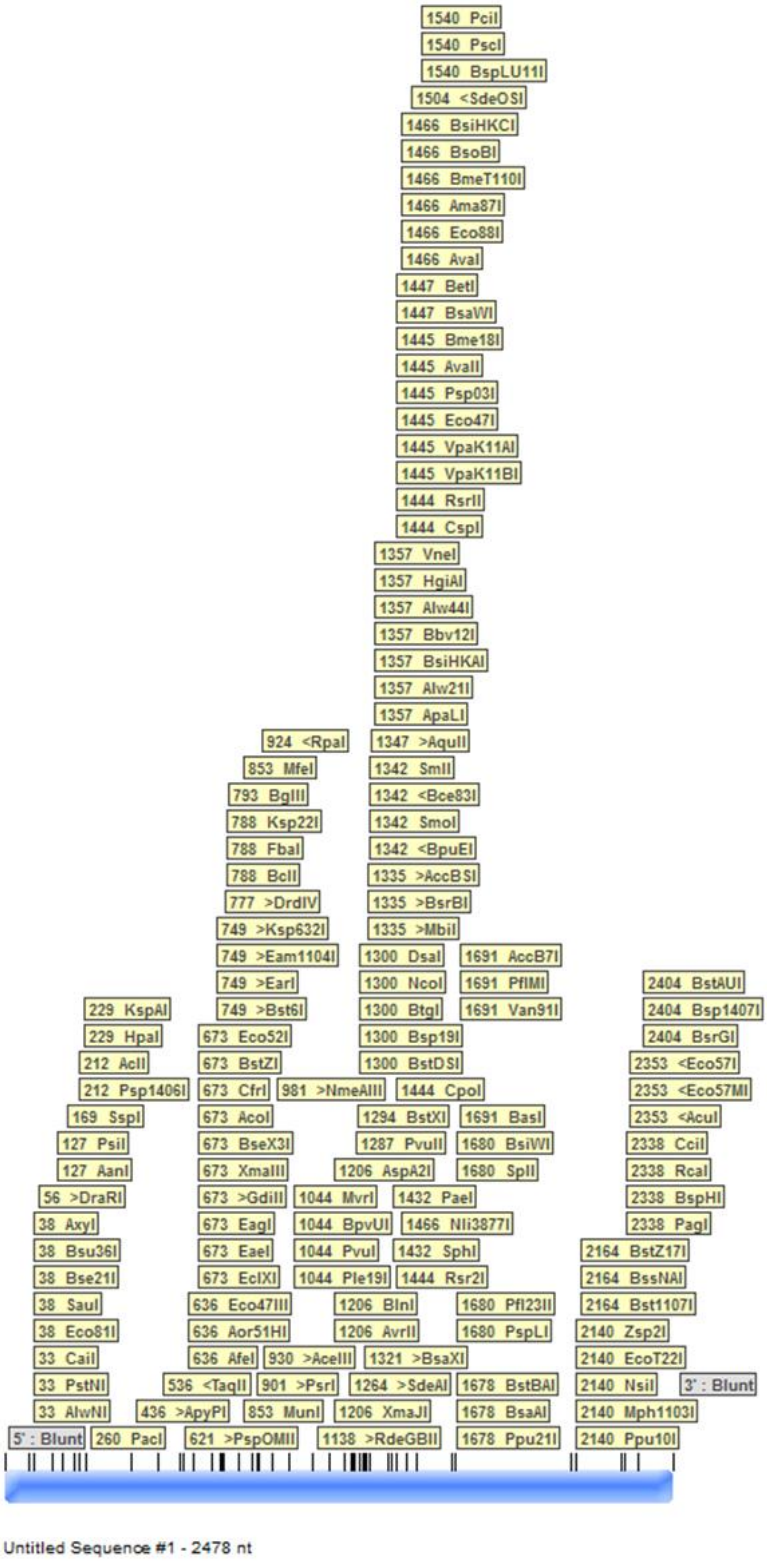

Figure 8. *Genomic map of MRNA sequence.*

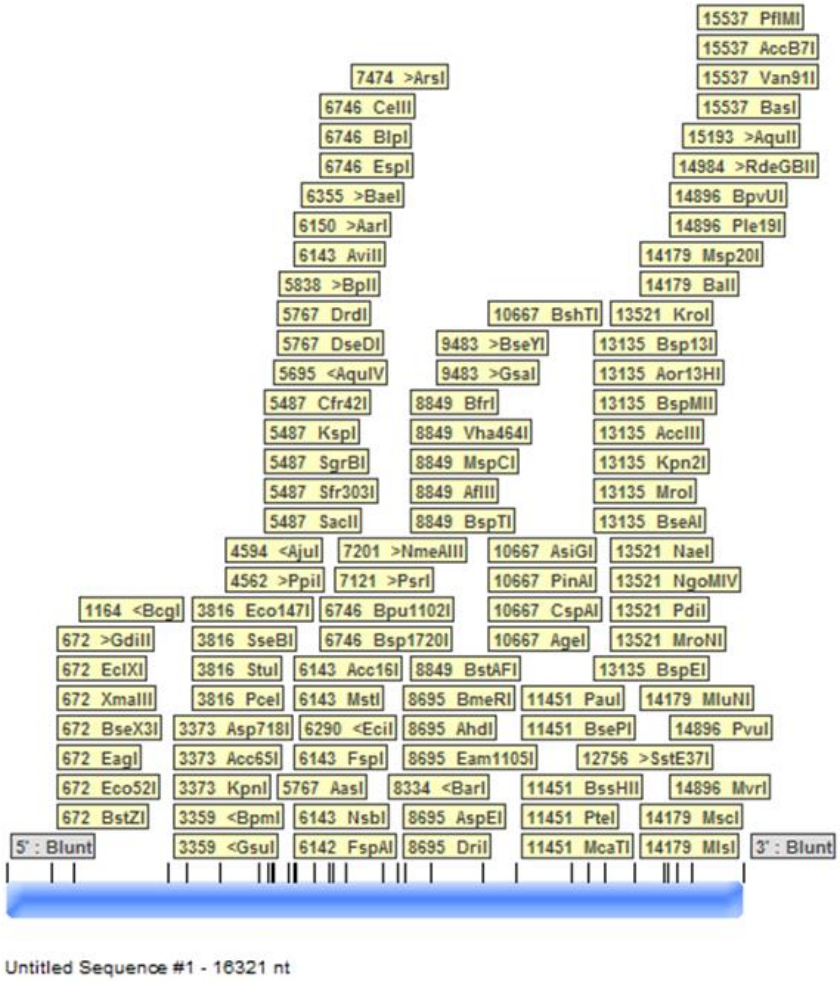

Figure 9. *Genomic map of Dl_X gene.*

Conclusion

The earliest known gene family expressed in developing limbs is distal less, and its expression is retained during the development of the limb. The Dlx genes, which are distal-less homologs, are expressed in developing appendages in at least six phyla, including chordates, indicating that Dlx function is needed for normal appendage development in the animal kingdom. Recent research has implicated the Dlx genes of vermin in several other developmental processes, ranging from neurogenesis to hematopoiesis. In this analysis, the exon-intron structure of the Dlx gene family was found to be very similar. The distribution of genes across chromosomes, on the other hand, was quite varied. The newly discovered genes could provide a wealth of information on how to manipulate the Dlx genome to grow limbs. The evolutionary relationships of the Dlx gene with closely related species are depicted in the phylogenetic study.

Acknowledgement

This research study is self-funded.

Conflict of interest

There are no conflict of interest involve any parties in this research study.

REFERENCES

[1] Ahmad, B., Azeem, F., Ali, M.A., Nawaz, M.A., Nadeem, H., Abbas, A., Batool, R., Atif, R.M., Ijaz, U., Nieves-Cordones, M., Chung, G. (2020): Genome-wide identification and expression analysis of two component system genes in Cicer arietinum. – Genomics 112(2): 1371-1383.

[2] Azeem, F., Tahir, H., Ijaz, U., Shaheen, T. (2020): A genome-wide comparative analysis of bZIP transcription factors in G. arboreum and G. raimondii (Diploid ancestors of present-day cotton). – Physiology and Molecular Biology of Plants 12p.

[3] Baum, D. (2008): Reading a phylogenetic tree: the meaning of monophyletic groups. – Nature Education 1(1): 190-197.

[4] EMBL Official Portal (2021): The Pfam online databased. – European Molecular Biology Laboratory Official Portal. Available on:
http://pfam.xfam.org/

[5] Expasy Official Portal (2021): Measuring the molecular weight and isoelectric point toolbox. – Expassy Official Portal. Available on:
https://www.expasy.org/

[6] Kraus, P., Lufkin, T. (2006): Dlx homeobox gene control of mammalian limb and craniofacial development. American Journal of Medical Genetics Part A 140(13): 1366-1374.

[7] Merlo, G.R., Zerega, B., Paleari, L., Trombino, S., Mantero, S., Levi, G. (2004): Multiple functions of Dlx genes. – International Journal of Developmental Biology 44(6): 619-626.

[8] NCBI Official Portal (2021): Conserved domains. – NCBI Official Portal. Available on:
https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

[9] SMART Official Portal (2021): Online databased Pfam. – SMART Official Portal. Available on:
http://smart.embl-heidelberg.de/

[10] Stock, D.W. (2005): The Dlx gene complement of the leopard shark, Triakis semifasciata, resembles that of mammals: implications for genomic and morphological evolution of jawed vertebrates. – Genetics 169(2): 807-817.