Fish and Shellfish Immunology 93 (2019) 17–27

Contents lists available at ScienceDirect

Fish and Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Full length article

Unraveling the evolutionary origin of ELR motif using fish CXC chemokine CXCL8

Krishnakant Gangele[a], Minal Jamsandekar[a], Amit Mishra[b], Krishna Mohan Poluri[a,*]

[a] Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee, 247667, Uttarakhand, India
[b] Cellular and Molecular Neurobiology Unit, Indian Institute of Technology Jodhpur, Jodhpur, 342011, Rajasthan, India

ARTICLE INFO

Keywords:
Chemokines
Interleukin-8
CXCL8
Neutrophil activating chemokines
Molecular evolution
Functional divergence

ABSTRACT

Chemokines are chemotactic proteins involved in host defense through the migration of immune-regulatory cells to the site of infection. Interleukin-8 (CXCL8/IL8) is the most studied "ELR-CXC chemokine/neutrophil activating chemokine (NAC) that regulate neutrophil trafficking during infections and inflammation by binding to its cognate G-protein coupled receptors CXCR1/CXCR2. The "ELR" motif of NAC chemokines is essential for the CXCR1/CXCR2 receptor activation. In order to understand the evolutionary origin of "ELR" motif in the CXC chemokines, a thorough evolutionary study of CXCL8 gene from various fishes and primates was performed. Phylogenetic analysis revealed that the CXCL8 gene can be classified into four distinct lineages (CXCL8-L1a, CXCL8-L1b, CXCL8-L2, and CXCL8-L3), where CXCL8-L1a is the fastest evolving lineage and CXCL8-L3 is the slowest. Selection analysis suggested that The "ELR/DLR" motif containing branches (gadoid and coelacanth) are positively selected. The probable evolutionary trend of "ELR" motif suggested that this motif in ancestor CXCL8 is evolved from the GGR of Lamprey (Agnatha), followed by duplication giving rise to two main motifs in CXCL8 "NXH" in L3 lineage and "ELR/DLR" in L1a/L1b lineages. Although, structural analysis suggested that the overall topology of the CXCL8 proteins is similar, differences do exist at the individual structural elements among the members of different lineages. Functional distance analysis suggested that the CXCL8-L3 lineage is more distant compared to the CXCL8-L1a and L1b lineages from the inferred ancestor. Functional divergence analysis between different lineages suggested that most of the selected residues are important for receptor or glycosaminoglycan binding. Such a functional diversification can be attributed to the novel set of functions adopted by CXCL8 in various species.

1. Introduction

Molecular evolution of proteins is a complex and continuous process that is dependent on various constraints such as molecular structure, functions, and stability. During the process of molecular evolution, a particular gene undergoes selective genetic variations along the axis of time and functional constraints to achieve better fitness [1,2]. The enhanced fitness could be the consequence of either positive selection or diversifying selection of a particular change in the protein and could have differential phenotypic characteristics [3,4]. Phenotypic variations in the proteins can be classified as fluctuations in function, structure, stability, and binding etc [5,6]. The functional and structural diversity of proteins can originate from gene duplication events, as the resultant gene could get neofunctionalized, pseudofunctionalized or might have faced deletion in one of the descendant at the time of speciation [7]. Orthologous genes are those which originated from a single ancestor of two descendent species, and indeed such orthologous evolutionary relationships for multigene families is very crucial to elucidate the origin of the gene/gene families [8,9]. A general trend observed among the protein families is that the proteins involved in the host defense mechanism experience a higher rate of evolution compared to the rest [10].

Chemokines are small molecular weight ubiquitous chemotactic proteins involved in host defense through the migration of immunoregulatory cells to the site of infection during the infection/inflammation [11,12]. They bridge the innate and adaptive immune systems to facilitate elevated immune response [13,14]. The root of chemokine evolution lies in the early ancestor of vertebrates that evolved around 650 mya. They are one of the important genes that evolved with the advent of the adaptive immune system in vertebrates [15]. Chemokines are broadly categorized into four groups based on the position of the N-terminal cysteine residues such as CC, CXC, CX3C, and

* Corresponding author. Department of Biotechnology, Indian Institute of Technology Roorkee (IIT-Roorkee), Roorkee, 247667, Uttarakhand, India.
E-mail addresses: krishfbt@iitr.ac.in, mohanpmk@gmail.com (K.M. Poluri).

https://doi.org/10.1016/j.fsi.2019.07.034
Received 24 May 2019; Received in revised form 12 July 2019; Accepted 12 July 2019
Available online 13 July 2019

C [16]. Among these, the CXC chemokine group has been divided into two subgroups, ELR(+)CXC and ELR(−)CXC chemokines depending upon the presence of ELR motif, preceding the first cysteine residue at the N-terminal [17,18]. In mammals, these two subgroups play different functional roles [19]. ELR(+)CXC chemokines can activate and recruit neutrophils, whereas ELR(−)CXC members can only induce the migration of lymphocytes but cannot activate neutrophil cells [19,20]. Humans possess seven ELR(+)CXC chemokines, among which CXCL8/IL8/Interleukin-8 is the most prominent member [21]. CXCL8 binds to cell surface glycosaminoglycans (GAGs) and regulates their functions by activating two G-protein coupled receptors (GPCRs) CXCR1 and CXCR2 [22,23]. It has a chemotactic property towards several types of known immune cells [24]. It plays an important role in wound healing, angiogenesis, neovascularization, metastasis and inflammation [25–27]. Differential expression of CXCL8 is involved in several pathological conditions such as chronic obstructive pulmonary disease (COPD), human obstructive pancreatitis, psoriasis, rheumatoid arthritis, cystic fibrosis, and many cancers [28,29].

Indeed, chemokines play a major role in the immune defense mechanism of primitive vertebrate species such as fish as they do not have a very well developed adaptive immune system and are greatly dependent on their innate immune system [14,15]. Although the ELR motif is a conserved feature of avian and mammalian ELR(+)CXC chemokines (CXCL8/IL8 protein), this motif is absent in the teleost fish except for haddock (*Melanogrammus aeglefinus*), and Atlantic cod *(Gadus morhua)* [30]. The first non-mammalian CXCL8-like sequence was discovered in the lamprey in 1999, since then several studies have reported a large number of CXC chemokines in the teleost fish, including some novel CXC chemokines whose structural and functional homologs are absent in humans [31–33]. Phylogenetically, CXCL8 like chemokines in fish were categorized into three groups, CXCL8-L1, CXCL8-L2, and novel type lineage as CXCL8-L3 [34,35]. Earlier studies reported that some teleost CXCL8 have substitution form of ELR motif such as DLR/NLH and revealed that they are able to activate the neutrophils [30,36,37]. Plenty of studies have been performed to understand the structure and function of CXC chemokine from human and mouse although fragmentary information is available from the primitive species. In an effort to unravel the evolutionary origins of chemokine multigene family ELR(+)CXC, the current study analyzed the evolution-structure-function information of the CXCL8 protein in the primitive vertebrate (fish) species. Our results evidenced that the branch containing the ELR motif was a result of positive branch selection, and several of the residues exhibiting functional divergence are underlying in the receptor or glycosaminoglycan binding domains of CXCL8.

2. Methodology

2.1. Multiple sequence alignment (MSA)

A total of 76 genes/amino acid sequences of IL8/CXCL8 like chemokine for various fish species were obtained from the sequence databases of ***GenBank*** (http://www.ncbi.nlm.nih.gov/genbank/), ***UniProt*** (http://www.uniprot.org/), and ***Ensembl*** (http://www.ensembl.org/), for which the unique IDs for teleost fish are listed in Table S1 and non-teleost fish are listed in Table S2 of supplementary material. The unique IDs for primate CXCL8 sequences are listed in Table S3. All the sequences were aligned using the multiple sequence alignment tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) by employing default settings for gap extension and gap opening penalty and output as in Pearson/FASTA format. The correctness of the alignments was also confirmed manually considering the conserved cysteine residues at positions C12, C14, C39, and C55 of human IL8 protein sequence.

2.2. Phylogenetic analysis

The output of multiple sequence alignment of CXCL8 genes from Clustal Omega was used as an input data for molecular evolutionary genetic analysis (MEGA6) [38]. The phylogenetic tree of fish CXCL8 was constructed using neighbor-joining method based on p-distance and substitution type as an amino acid in MEGA6. The reliability of the tree was assessed by the bootstrap method using 2000 bootstrap replications with 95% site coverage that resulted in bootstrap proportion for each internal branch of the tree. Further, FigTree3.0 program was used to create the colored graphical representation of the phylogenetic tree [39].

2.3. Conservation analysis

The conservation profile of CXCL8 protein was calculated to obtain the conserved nature of each residue using the lineage wise aligned protein sequences as input by the WebLogo server (https://weblogo.berkeley.edu/logo.cgi).

2.4. Selection analysis

2.4.1. Site selection analysis

The codon-based maximum likelihood (codeml) program in the PAML package was used to study the positive selection in the CXCL8 gene from all existing species of fish. Recombination possibilities were assessed using the GARD program of the Datamonkey server [40]. For selection analyses, codon alignment and neighbor-joining (NJ) tree were used as the input data. Codeml selection analysis was performed using M0 (one ratio), M1 (nearly neutral), M2 (positive selection), M7 (beta) and M8 (beta and ω) for the site selection whereas two ratio models such as alternate and null models for branch and branch site [41]. M0 model had a constant ω-ratio; whereas in M1 and M2 model, ω_0 is varied between 0 and 1 and ω_1 is fixed at 1. M7 and M8 models have distributed ω values. Likelihood ratio test (LRT) of M1 with M2 and M7 with M8 was performed to find out the most significant model. p-values for the test were determined by taking the log likelihood difference between the two models, multiplying by 2 and checking this value in the chi-square (χ^2) distribution table with difference between number of parameters of two models has been considered as degree of freedom. Datamonkey web server was also used to find out the positively selected amino acid sites and the branch in the phylogeny. Datamonkey programs used to identify the sites include: single likelihood ancestor counting (SLAC), a non-neutral evolutionary method for sequence alignment; fixed-effect likelihood (FEL), a statistically rich method where independent dN and dS can be set for each site and fast unbiased bayesian approximation (FUBAR) method, permits large data set for analysis. The p-value was set as 0.05 for SLAC, FEL and the default parameters for the FUBAR [42].

2.4.2. Branch and branch site selection

The branch and branch-site selection analysis were performed using the null and alternative model for each branch considering the omega values for foreground and background branch. All the parameters for analysis were used as described elsewhere [43], and the detailed parameters for selected branches are listed in Table 1. The LRT values between alternate and null models were calculated for each branch. Adaptive branch-site random effect likelihood (aBSREL) method of the datamonkey server was also used to perform branch-site selection analyses [44].

2.5. Homology modeling and structure alignment

The structures of CXCL8 proteins were modeled for most common fish species (Atlantic cod-L1a, Rainbow trout-L1b, Half-smooth tongue sole-L3, West Indian Ocean Coelacanth-2, and Common carp-L2) from

Table 1
Adaptive selection analysis of identified phylogenetic branches.

Tree Branch	*Model*	*Parameters*					*lnL*	*LRT*	*P value*	*Selected Sites*
Branch selection analysis of positively selected lineage										
Full Tree	**One ratio**	$\omega = 0.234$					−9861.09	One ratio vs. Free ratio	< < **0.001**	***Not Allowed***
	Free Ratio						−9744.62	**232.95**		
Coelacanth branch	**Alternate**	$\omega_f = 0.235$					−9860.77	Null vs. Alternate	< **0.05**	***Not Allowed***
		$\omega_b = 0.138$						**5.94**		
	Null	$\omega_f = 0.231$					−9863.76			
		$\omega_b = 1.0$								
Gadoid branch	**Alternate**	$\omega_f = 0.415$					−9857.61	Null vs. Alternate	< < < **0.001**	***Not Allowed***
		$\omega_b = 0.102$						**25.29**		
	Null	$\omega_f = 0.415$					−9870.25			
		$\omega_b = 1.0$								
Branch-Site selection analysis of positively selected sites on predefined branches										
Coelacanth branch		**Class**	**0**	**1**	**2a**	**2b**				
	Alternate						−9692.2	Null vs. Alternate	< **0.05**	1T, 2E, 55Q, 56V
		P	0.614	0.247	0.099	0.039		**4.878**		
		ω_b	0.186	1.0	0.182	1.0				
		ω_f	0.182	1.0	27.22	27.22				
	Null						−9694.64			
		P	0.477	0.193	0.235	0.095				
		ω_b	0.182	1.0	0.182	1.0				
		ω_f	0.182	1.0	1.0	1.0				
L2 branch	**Alternate**						−9686.06	Null vs. Alternate	< **0.001**	1T, 5P, 8Q, 9Q, 18H,19S, 20E, 21P, 22A, 24P, 30S, 43E, 62K, 66I, 67S, 68L, 69 N, 73K
		P	0.372	0.156	0.332	0.139		**14.265**		
		ω_b	0.185	1.0	0.186	1.0				
		ω_f	0.186	1.0	999.0	999.0				
	Null						−9693.19			
		P	0.434	0.178	0.274	0.113				
		ω_b	0.183	1.0	0.183	1.0				
		ω_f	0.183	1.0	1.0	1.0				

P: Proportions, $\boldsymbol{\omega_b}$: ω for background branch, $\boldsymbol{\omega_f}$: ω for foreground branch

each lineage using the SWISS-MODEL. The template structure for each model structure was selected from the available experimental structure (NMR/X-ray) of the CXC family chemokine on the basis of their maximum sequence identity. The structural models were validated by analyzing the torsional angles using Ramachandran maps. For each protein structure, torsional angles were extracted from the pdb file in DSSP format and analyzed using XSSP server (http://www.cmbi.ru.nl/xssp/). To analyze the RMSD values, the backbone C_α atoms of a monomeric structural fold of CXCL8 from individual lineage was aligned against the NMR structure human-CXCL8 using the PyMol software. Comparative 2D contact maps showing C_α contacts between the monomeric human CXCL8 and modeled 3D structures of CXCL8 from different lineages were generated using contact map view (CM view) software with a distance threshold of 6 Å [45].

2.6. Functional divergence calculation

The functional divergence of CXCL8 proteins was analyzed using the Divrge3.0 program. The amino acid MSA and the NJ phylogenetic tree were used as input data for the functional divergence analysis. Four clusters; (a) Ancestral-CXCL8 (b) CXCL8-L1a (c) CXCL8-L1b and (d) CXCL8-L3 were made for further analysis. The functional distances of each lineage from the theoretical inferred ancestor were calculated using the "functional distance analysis" tool of Diverge3.0 [46]. Type-I and Type-II divergence analyses were performed using Gu99 and Type-II divergence tools of Diverge3.0 respectively. The cutoff values for posterior probability were defined by stepwise exclusion of highest scoring residues until theta (θ) becomes zero. The selected residues from Type-I divergence were marked on the 3D CXCL8 structures using the PyMol software.

3. Results

3.1. Sequence and phylogenetic analysis

To gain evolutionary insights into the ELR motif of CXC chemokines, 76 full-length gene sequences of CXCL8 sequences from primitive species (fish) were collected from various databases and analyzed using MSA analysis (Fig. 1 and Figs. S1–S3 of supplementary material). MSA is the most useful tool to get full insight into the sequence similarity and site-specific pattern in a large set of sequences. The aligned CXCL8 protein sequences from all the 76 species showed several peculiar features that are in accordance with the previous studies done on the CXC chemokines of various other species. Upon analyzing the MSA, it is has been found that all the CXCL8 sequences have four conserved cysteine residues that are involved in forming two disulfide bonds required for maintaining the structural fold and functional integrity of CXC chemokine [47,48]. In addition to these four cysteine residues (two disulfide bonds), all 76 aligned sequences of fish CXCL8 comprise of a characteristic residue 'E' at starting of the β2 strand and "WV" motif at C-terminal is conserved throughout the CXCL8, i.e., from ancient (Agnatha) jawless fish lamprey to the fully evolved human.

To understand the detailed classification of CXCL8, a phylogenetic tree was constructed using the neighbor-joining method by including all the possible CXCL8 sequences from different fish species (Fig. 2). The overall pattern of the phylogenetic classification results is consistent with the previously reported CXC classification [34]. All fish CXCL8 sequences are broadly categorized into two groups; (a) teleost fish (62 sequences) and, (b) ancestral fish (14 sequences). The teleost CXCL8 protein sequences are further classified into three lineages based on their phylogenetic evolutionary position. These lineages include the CXCL8-L1 (42 sequences), CXCL8-L2 (2 sequences) and CXCL8-L3 (18 sequences). CXCL8-L1 group is the largest group in the teleost fish for CXCL8 proteins in the constructed phylogenetic tree. CXCL8-L2 proteins are closely placed to the ancestral fish CXCL8 (river lamprey) and

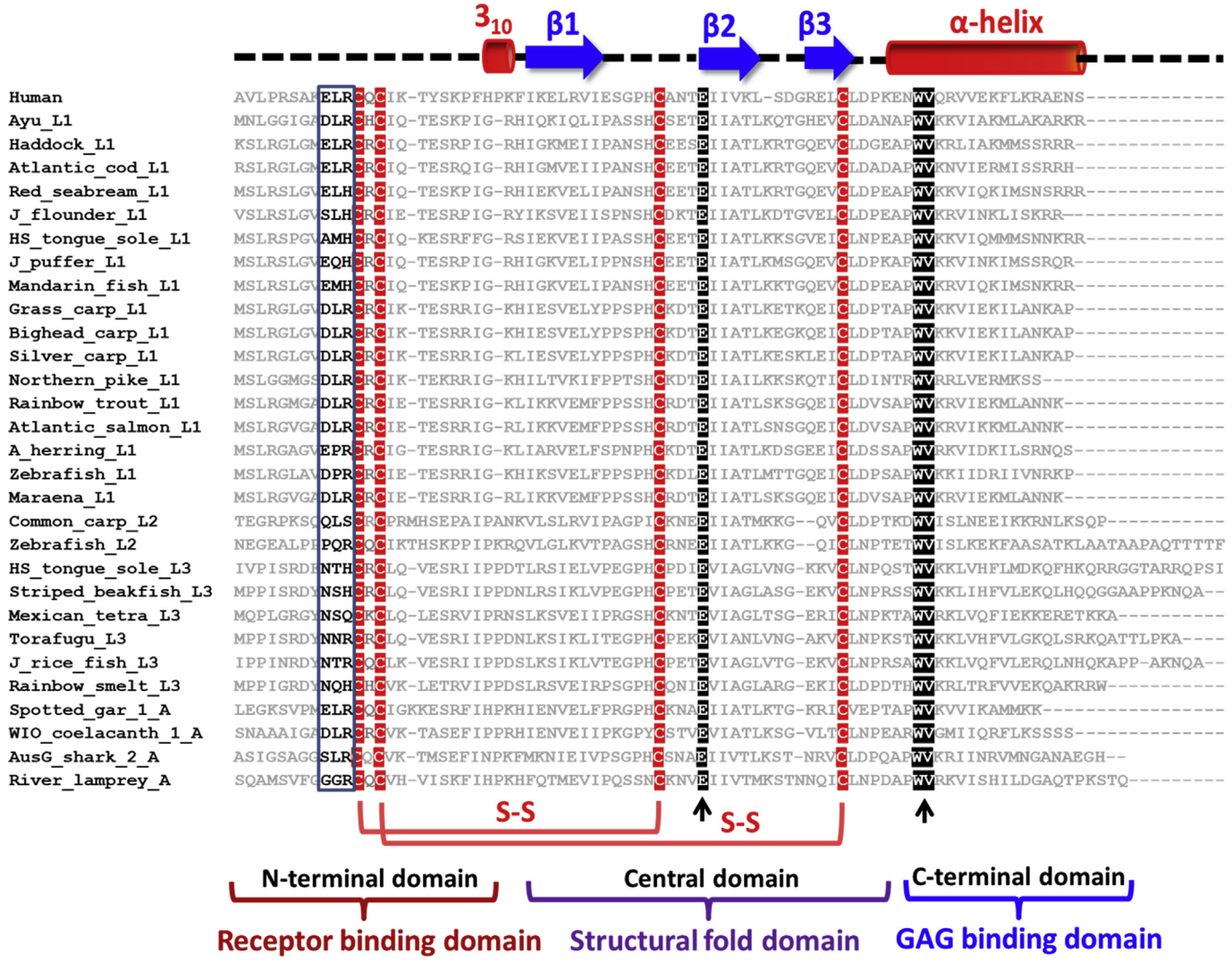

Fig. 1. Amino acid alignment of the CXCL8 protein from selected fish species from all the lineages comprising all major types of "ELR" like motifs; this specific motif at the N-terminal is enclosed in a rectangle. The alignment also includes the CXCL8 sequence of human CXCL8 comprising a fully functional "ELR" motif and the CXCL8 sequence of agnatha (Lamprey) with "GGR" motif. The secondary structures were marked on the top of the sequences according to the NMR based structure of human CXCL8. The full CXCL8 protein sequence had been characterized in three different domains (a) Receptor binding domain (b) Structural fold domain and (c) GAG (glycosaminoglycans) binding domain. C12 and C14 form disulfide bonds with C39 and C55 respectively (highlighted in red color). "E" residue in the structural domain and "WV" motif in the GAG binding domain (highlighted in black color) is the characteristically conserved residues in all CXCL8 like sequences. Alphabets in the selected fish sequences represent: J- Japanese, A- Atlantic, HS- Half-smooth, WIO- West Indian Ocean, and AugG- Australian ghost. The three lineages are marked with L1, L2 and L3. Ancestral fishes are denoted with "A". (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

are exclusively found in the cyprinids, zebrafish and common carp. CXCL8-L3 types of chemokines are novel CXCL8 like sequences in the teleost fish and do not have any homologous sequence in the mammals. Only a few CXCL8-L3 sequences have been reported from fish in previous studies. It is also noticed from the node lengths that, CXCL8-L3 lineage has slow evolving features than CXCL8-L1 lineage and placed near the ancestral fish in the phylogenetic tree. So, one can infer that L1 lineage sequences are the recent/advanced compared to those of the L3 lineage CXCL8 sequences. Our analysis suggested that CXCL8-L1 can be further subdivided into two sub-lineages. These sub-lineages are named as CXCL8-L1a (28 sequences) and CXCL8-L1b (14 sequences). The subdivision is also evident from the divergent node that has strong bootstrap values. It was observed that most species of fish containing CXCL8-L1a dwell in marine habitat while the fish with CXCL8-L1b dwell in freshwater habitat suggesting that the environmental challenges could be one of the reason in diversified evolution of CXCL8 gene considering the *FishBase* database information [49]. Thus our phylogenetic analysis unraveled a distinct lineage and strongly supports the existence of four different lineages of CXCL8 in the teleost fish instead of previously reported three lineages.

3.2. Understanding the evolutionary origin of ELR motif in CXCL8 proteins

The ELR motif of CXC chemokines is a specific motif involved in the CXCR1/CXCR2 receptor binding and activation of neutrophils. To understand the evolutionary origin of the ELR motif in the CXCL8 chemokines, the detailed sequence analysis has been performed on the basis of lineages obtained from the phylogenetic tree. The detailed pattern analysis of this motif was summarized in Table 2. Sequence analysis has shown that this motif is highly variable among the ancestor fish lineage as the patterns of the motifs are "SLR/DLR/ELR/NLR/AFR/GGR/DKH/RAR/ESK/IMR/AIR". Further sequence analysis revealed that the CXCL8-L3 lineage has "NTH/NSH/NSQ/NQH/NNR/NTR" type of motifs forming the most possible consensus sequence of "NXH/R".

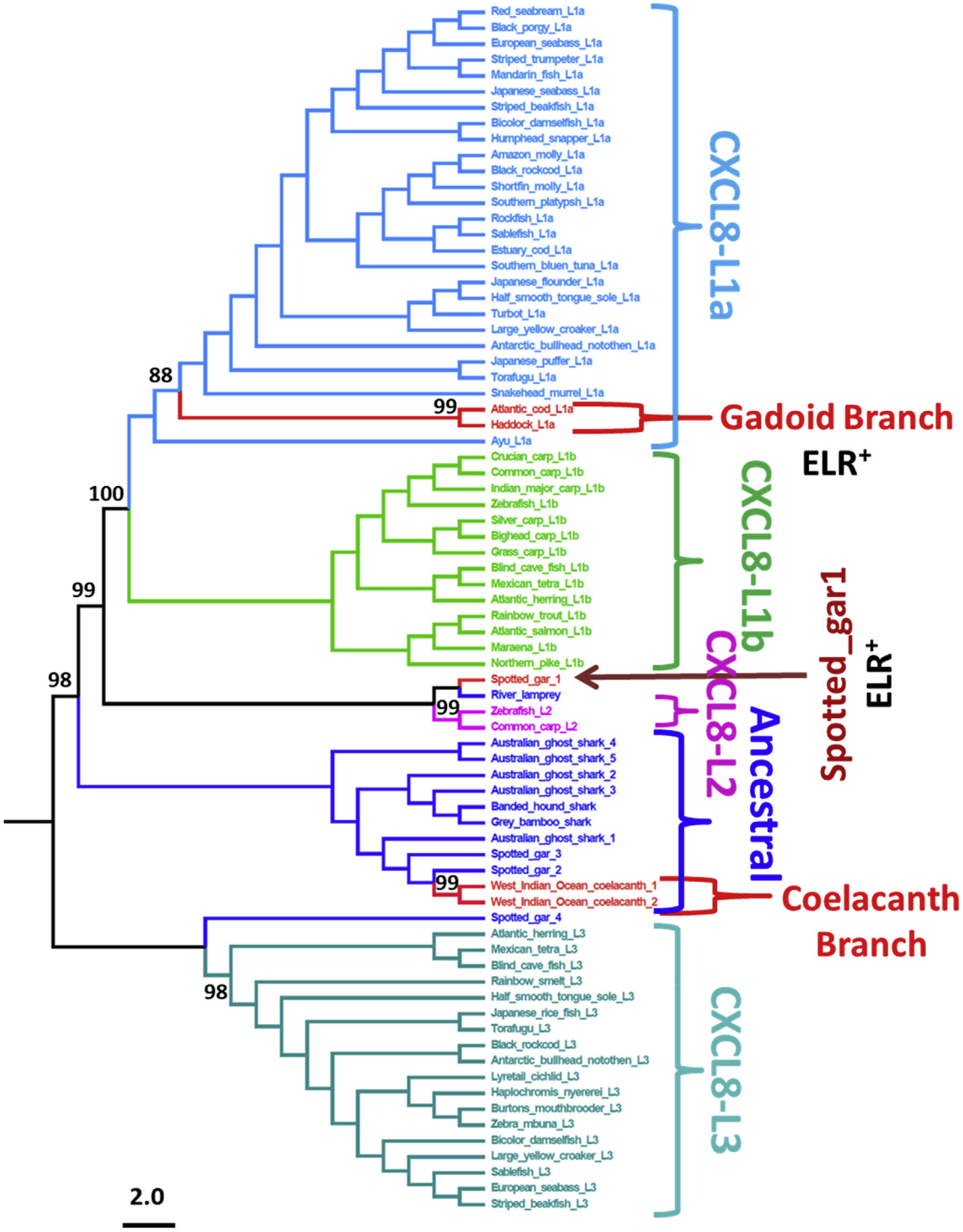

Fig. 2. The neighbor-joining phylogenetic tree of 76 CXCL8 protein sequences from the various fish species. Each lineage of CXCL8 has marked with different colors (CXCL8-L1a: Marine, CXCL8-L1b: Green, Ancestral-CXCL8: Blue, CXCL8-L3: Dark-Cyan. The positively selected lineages (Gadoid and Coelacanth branch) have been marked with the red color. The Gadoid branch and Spotted gar consisting "ELR" motif are highlighted in black color. The numbers on the corresponding nodes represent the bootstrap values, which is the indicator of lineages and sub-lineage bifurcation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Most of the CXCL8-L1a sub-lineage sequences contain "ELH/EMH/EQH" type of motifs, thus forming the consensus sequence "EXH". The other motifs found in CXCL8-L1a are SLH in Japanese flounder, AMH in half-smooth tongue sole. Further, CXCL8-L1a also comprises of species Ayu with DLR motif and gadoids (Atlantic cod and Haddock) with ELR motif. Whereas, CXCL8-L1b lineage has a "DLR/DPR/EPR" type of motifs, forming "D/EXR" as consensus motif, which is closer to the canonical "ELR" motif of mammalian/primate's CXCL8. It is evident from this analysis that the third position of the motif is strictly conserved in all lineages with positively charged residue (R/K/H), except in one sequence of Mexican tetra from L3 lineage that has glutamine. It can be said that possible "XXR" motif from ancestor fish had been evolved and diverged in two directions, (a) "NXH" of CXCL8-L3 lineage and (b) "EXH" of the CXCL8-L1a sub-lineage. Further, this "EXH" motif

Table 2
Conservative nature of ELR motif in different lineages of fish.

SN	CXCL8-Lineage (Total sequence)	Types of ELR Motif (Number of sequence having that particular motif)	Possible Consensus Motif
1	CXCL8-Primates **(16)**	ELR **(16)**	ELR
2	CXCL8-L1a **(28)**	ELH **(18)**/EMH **(3)**/EQH **(2)**/ELR **(2)**/DLR**(1)**/AMH **(1)**/SLH **(1)**/	EXH/R
3	CXCL8-L1b **(14)**	DLR **(7)**/DPR **(4)**/EPR **(3)**	D/EXR
4	CXCL8-L3 **(18)**	NTH **(9)**/NSH **(4)**/NSQ **(2)**/NQH **(1)**/NNR **(1)**/NTR **(1)**	NXH/R
5	CXCL8-Ancestor **(14)**	SLR **(3)**/DLR **(2)**/ELR **(1)**/NLR **(1)** /AFR **(1)**/GGR **(1)**/DKH **(1)**/RAR **(1)**/ESK **(1)**/IMR **(1)**/AIR **(1)**	XXR

may have been evolved into “E/DXR” of CXCL8-L1b sub-lineage. “D/EXR” consists of most evolved residues that are similar to the mammalian "ELR" motif. On the other hand, “NXH/R” motif of CXCL8-L3 lineage is far different from the mammalian ELR motif. The analysis suggested that the complete “ELR” motif is not present in the whole fish CXCL8 sequences except gadoid fish (Atlantic cod and Haddock from CXCL8-L1a lineage) and spotted_gar_1 from ancestor lineage.

3.3. Selection analysis of CXCL8 gene in teleost fish

CXCL8 gene is one of the fast evolving genes in the teleost fish and has undergone duplication and random mutations throughout the evolution process. Further, it is clear from the sequence and phylogenetic analysis that, “ELR” motif has continuously evolved and diversified in CXCL8 protein. In order to understand the natural selection and the role of “ELR” motif in positive selection of particular lineage during the evolution of CXCL8 gene, selection analysis was performed using the maximum likelihood method in codeml program (PAML package) and different methods such as SLAC, FEL, and FUBAR of Datamonkey server. The recombination possibilities were checked by GARD analysis, before proceeding for the selection analysis. We observed that in the CXCL8 sequences no such recombination was observed, which is in line to that of for GRO chemokines [42].

The site selection model of PAML on CXCL8 sequences did not provide any significant positive selection sites except the M8 (β & ω) model. Codeml results depicted that, fish CXCL8 gene has globally evolved under purifying/negative selection pressure (Table S4), i.e., most of the amino acid substitutions of CXCL8 are dominated with the synonymous substitutions ($dN/dS \ll 1$). The purifying nature of fish CXCL8 gene has been also supported by the results from SLAC, FEL and FUBAR programs of Datamonkey server (not shown here). These results suggest that although fish have a diverged range of CXCL8 gene in its genome, the duplicated CXCL8/CXCL8-L proteins have not evolved under the positive site selection. Such a continuous purifying selection is highly justified considering the fact that the CXCL8 structure and function is tightly regulated by evolutionary pressure [50,51]. In order to unravel the major residues that are conserved during the course of evolution process, residue wise conservation profile for all fish sequences in comparison to primates was calculated (Fig. 3). The results illustrated that the ancestral fish have the most variable residues throughout the sequence from N-terminal to C-terminal end, except cysteine and other characteristic conserved residues (E, W, and V) as discussed in previous section 3.1. CXCL8 sequences are mostly conserved in the individual lineage of teleost fish. These results also substantiate the fact that, CXCL8 sequence in fish has gone through the conservative substitution of amino acids during evolution.

Although CXCL8 gene in fish is mostly evolved under the purifying selection pressure, the environment or the habitant can influence the natural selection criteria of a given lineage thus resulting in adaptive selection pressure. Further, to assess the role of adaptive/positive selection pressure acting on a specific lineage of CXCL8, branch selection and branch-site selection models were applied. Branch selection models are useful in identifying adaptive selection if one copy of the duplicated gene has encountered accelerated adaptive evolution and functional divergence [52]. Lineage-specific selective constraint analysis was performed by considering each clade of species taxa as foreground branch under two ratio models. Both branch models and branch-site models were implemented for lineages leading to L1a, L1b, L2, and L3 of teleosts and ancestral CXCL8 clades of Agnatha, Coelacanth, Holocephali, Elasmobranchii, and Holostei. Our results suggested that Coelacanth (ancestral), and Gadoid clade (L1a) are positively selected through branch model with significant LRT values of 5.9 and 25.3 respectively. Further, Coelacanth and CXCL8-L2 clades are selected in the branch site model of both PAML and aBSREL (Table 1). The positively selected branches have been marked with red color in Fig. 2.

Gadoid fish of L1a namely Atlantic cod and Haddock show a very prominent signal for diversifying selection with omega value of the foreground branch as 0.415 and LRT value as 25.29 (Table 1). Gadoids are the only teleost fish having a full “ELR” motif and are selected under the branch selection model. It has been also reported that ELR motif is essential for angiogenesis in the haddock but not in other teleost fish [53]. This strongly suggests that the CXCL8 gene provides a competitive advantage to the gadoid species. In such a case, one can presume that CXCL8 in gadoid has undergone neo-functionalization after its duplication with strong and potent chemotactic activity or with some additional functionality. Coelacanth branch is also selected through the PAML branch selection model with the LRT value of 5.9. Both copies of the CXCL8 gene from Coelacanth fish have DLR motif which is close enough to the ELR of mammals. Coelacanths are the connecting link between fish and tetrapod. Therefore, the adaptive immune system of coelacanth might have developed in such a way that they have the features of both fish and tetrapod. These results are strongly supporting that ELR/DLR motif provide an advantage for the selection of particular lineage in CXCL8.

3.4. Structural and functional divergence analysis of CXCL8

The structure of any protein is strongly dependent on its amino acid sequence. As observed above, fish have a diverged range of CXCL8 sequences in its genome and are evolved under purifying selection. In order to understand the structural (both secondary and tertiary) characteristics of CXCL8 protein from the different types of lineages, structures were modeled. NMR/X-ray structure of CXCL8 protein consists of mainly four parts (a) long unstructured N-terminal loop, (b) 3_{10} helix, (c) three antiparallel β-sheets and (c) C-terminal α-helix (Fig. 4A). The N-terminal loop along with residues from 30s to 40s loop is also involved in receptor interaction. The 3_{10}-helix and the positively charged residues from the C-terminal helix are majorly responsible for glycosaminoglycans binding. The packing interactions within the β-sheets and against the helical hydrophobic residues form the highly stable core, thus stabilizing the overall structure [48]. In order to assess their structural similarities of the fish CXCL8 lineages, the structures were generated using homology modeling and were aligned against the NMR structure of human-CXCL8. The structural folds of these modeled structures were confirmed by analyzing the torsion angles, extracted from PDB files and the detailed comparative analysis of structure was summarized in Table S5. It has been observed that the overall 3D fold of CXCL8 from different lineages is very similar to the human-CXCL8 (Fig. 4B–F). Further, structural differences due to variation in the amino acid composition between the human CXCL8 and fish CXCL8 from

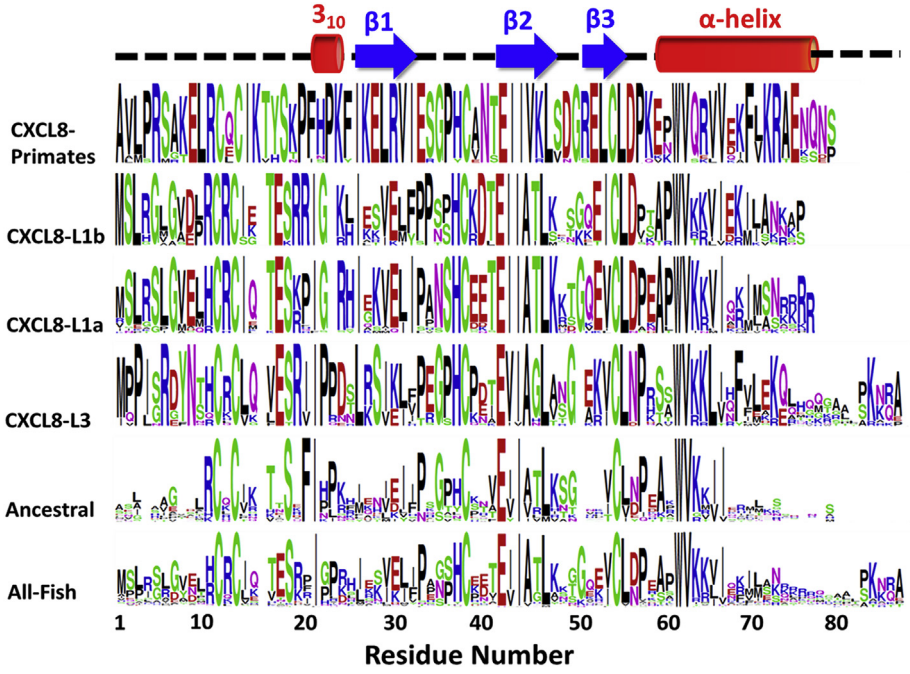

Fig. 3. Comparative graphical representation depicting the residue-wise conservation profiles of different lineages of fish CXCL8 along with the primate CXCL8 sequences. The secondary structure of the CXCL8 has been marked in cartoon format (according to human CXCL8).

different lineages were analyzed using C_α contact map. The comparative contact map analysis suggested that, the human CXCL8 structure shares minimum contact similarity (62.3%) with CXCL8 from Half-smooth tongue sole-L3, whereas maximum contact similarity (70.8%) with the CXCL8 from Common carp-L2 (Table S6). The majority of essential contacts of modeled CXCL8 from different lineages are overlapping with the human counterpart thus corroborating the same structural topology. Differential contacts observed here might be probably dictating the local structural differences, which could lead to the functional divergence as discussed below (Fig. 4G–K). However, a close analysis suggested that all the lineages of fish CXCL8 have a shorter C-terminal α-helix compared to its human counterpart. Further, the comparative analysis suggested that these primitive vertebrate CXCL8 structures do not possess the 3_{10}-helix before the β1 strand except the CXCL8-L2 lineage. The core β-sheet structure and orientation of fish CXCL8 are consistent in all the lineages to those of human, although some extent of variation is observed both the length and the position of the β3 strand.

It is well-known that the functions of the protein are majorly dependent on its structure and sequence. The 3D structure of CXCL8 is almost conserved in all vertebrates. Moreover under evolutionary trait, several of the species retained CXCL8 and its cognate receptors (CXCR1 and CXCR2). This suggests that the important functions of CXCL8 protein can't be easily amended under evolutionary pressure on lineage divergence. Conversely, it is also evident that the multiple isoforms of CXCL8 have attained various other novel functions at species divergence [54,55]. In order to understand the functional divergence of CXCL8 protein in fish, cluster-based functional divergence analysis was performed using program Diverge 3.0. Such analyses identify the possible amino acid sites that lead to the functional divergence of a particular gene family [56]. Functional divergence analysis was performed for identifying the Type-I and Type-II amino acid sites by classifying the sequences into four clusters as, CXCL8-L1a, CXCL8-L1b, CXCL8-L3, and CXCL8-Ancestor. CXCL8-L2 was not considered for the analysis as it comprises of only two sequences and the minimum acceptable limit is 4 sequences. Type-I divergence represents the sites that are conserved in one cluster but are highly variable in another cluster.

Functional distance analysis was performed to understand the altered functional constraint after gene duplication. Fig. 5A suggested that the CXCL8-L3 lineage is more distant from inferred ancestor, whereas the ancestral-CXCL8 lineage is less distant in comparison to the CXCL8-L1a and L1b lineages. Further, phylogenetic and functional distance analysis also suggests that the CXCL8-L1 and CXCL8-L3 were evolved in a distinct direction after the gene duplication and adopted the new set of functions. The outcome of divergence analysis suggested that only Type-I sites are present in fish CXCL8 sequences. The cluster-based divergence analysis was performed to identify the residues that are most likely contributing to the Type-I functional divergence (Table S7). Among the given lineages, the Type-I divergence results identified a total of 13 residues as significant contributors for the CXCL8 functional divergence. The corresponding consensus divergent residues were marked as a colored sphere on the 3D structure of human-CXCL8 (Fig. 5B). The red colored residues (Gln13, Ile15, Tyr18, Lys20, Pro21, and Asp50) are directly involved in the CXCR1/2 receptor interaction and belong to the N-terminal loop and 40s loop. Whereas, green colored residues (Ile33, Asn41, and Thr42) belong to the 30s loop and are interacting indirectly with the receptor [57]. The blue colored residues (Lys59, Gln64, Arg65, and Lys69) exclusively present on the C-terminal helix directly interact with sulfate groups of glycosaminoglycans during the process of neutrophil recruitment [58]. Further, considering the θ and LRT values, the following trend of functional divergence between the lineages has been obtained; L1a-L3 > L1b-Ancestral > L1b-L3 > L3-Ancestral > L1a-Ancestral > L1a-L1b (Table 3). The results suggest that these functionally important diverged residues along with "ELR" motif divergence could have lead CXCL8 molecule in fish to evolve with some additional functions with evolutionary pressure.

4. Discussion

During evolution, gene and genome duplication is the key and primary force for a given organism in order to attain novel functional genes for adaptation in the habitat [59]. Establishing the evolutionary origin and phylogenetic relationship of chemokines between the lower and higher vertebrates has been challenging for decades, although several CC and CXC family chemokines and their cognate receptors have been identified in great numbers in vertebrate genomes [60–62].

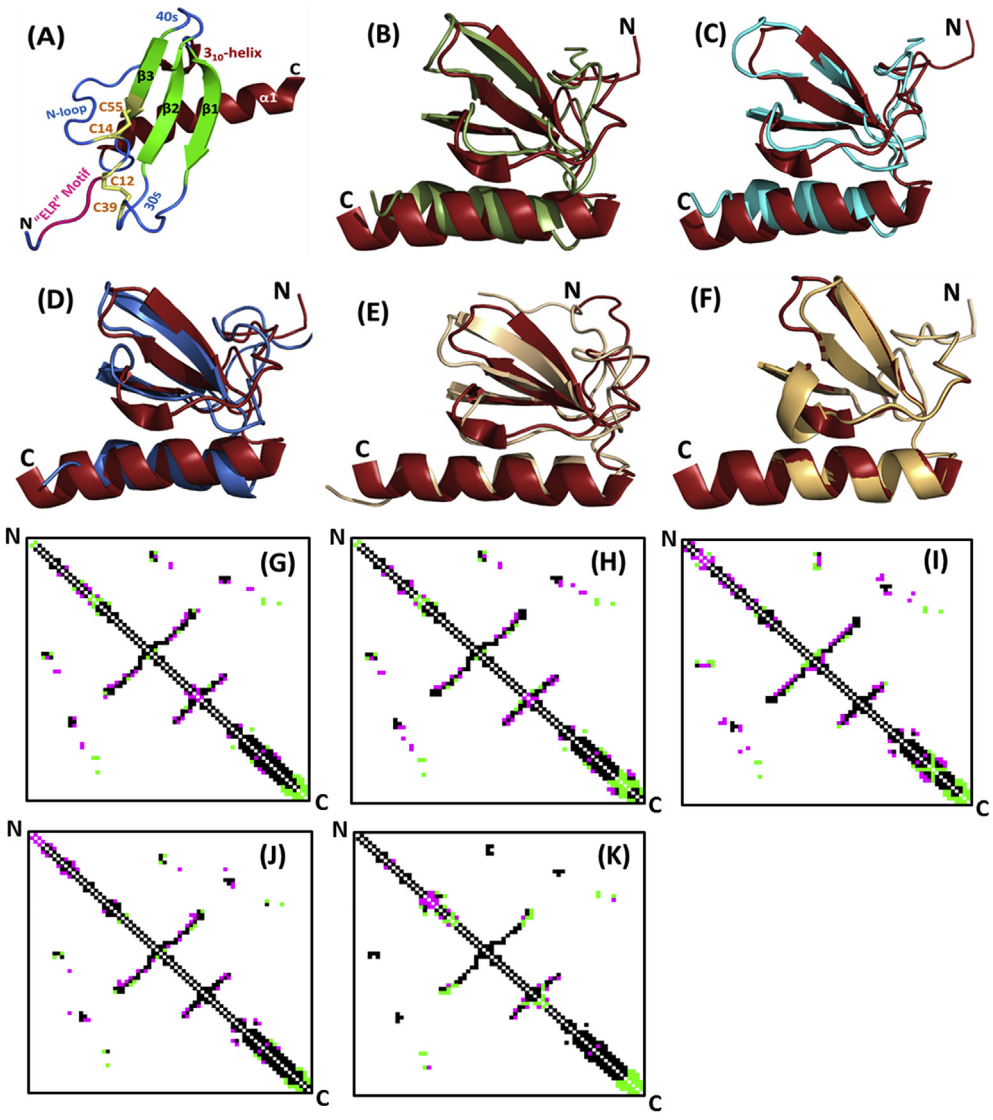

Fig. 4. Comparative structural analysis of fish CXCL8 with human CXCL8; (A) 3D structure of human CXCL8 monomeric unit highlighting the secondary structure elements; (B–F) Backbone alignment of modeled CXCL8 structures from individual lineage to the NMR structure of human-CXCL8 monomeric unit (red colored structure in each image). (B) Atlantic cod-L1a, RMSD: 1.97 Å; (C) Rainbow trout-L1b, RMSD: 2.0 Å; (D) Half-smooth tongue sole-L3, RMSD: 2.33 Å; (E) West Indian Ocean coelacanth-2, RMSD: 1.85 Å; (F) Common carp-L2, RMSD: 0.44 Å; (G–K) Comparative structural contact map between monomer structure of human CXCL8 (Green) and modeled monomeric structure from different lineages of fish CXCL8 (Pink), (G) Atlantic_cod-L1a, (H) hIL8-Rainbow trout_L1b, (I) Half-smooth tongue sole-L3, (J) West Indian Ocean coelacanth-2, and (K) Common carp-L2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

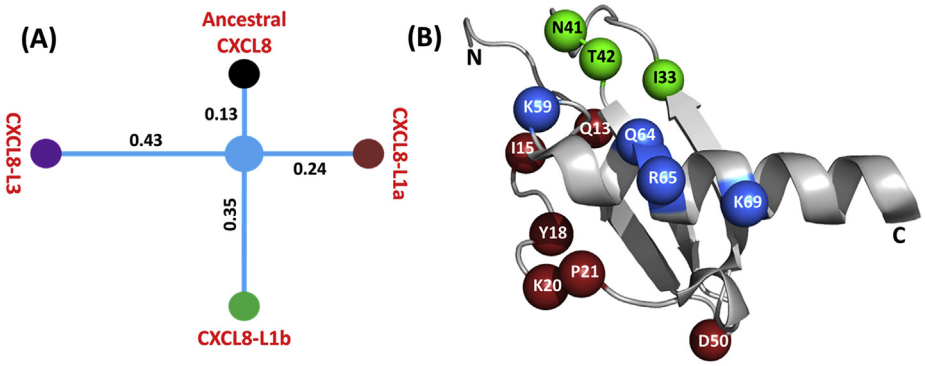

Fig. 5. Functional divergence analysis of CXCL8: (A) Functional distance pattern of different clusters of fish CXCL8 calculated using the Diverge 3.0 program; (B) The significant consensus residues from type-I functional divergence analysis have been marked as sphere on the 3D NMR structure of human CXCL8 monomeric unit (PDBID: 2IL8). Residues, directly and indirectly interacting with receptors were marked with the red and green color respectively, glycosaminoglycans interacting residues were marked with blue color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Type-I functional divergence results of different fish lineages.

Lineage Comparison	L1a/L1b	L1a/L3	L1a/Ancestral	L1b/L3	L1b/Ancestral	L3/Ancestral
MFE Theta	0.30 ± 0.14	0.64 ± 0.16	0.23 ± 0.10	0.49 ± 0.18	0.56 ± 0.18	0.28 ± 0.14
LRT Theta	5.63	33.09	7.74	11.57	21.15	8.32
Z Score	−2.42	−5.13	−2.33	−3.23	−3.69	−2.39
P Value	< 0.01	< 0.0001	< 0.019	< 0.001	< 0.0002	< 0.016

A Large number of mammalian orthologues of chemokines have been identified in the various species of fish. For example, 111 chemokine genes have been identified in the Zebrafish genome due to the multiple events of substitution, insertion, deletion and genome and gene duplication [63]. More than 170 CXC chemokine sequences have been reported from the fish, these also include various novel lineage groups that are homologous to the mammals [34]. In mammals, more than 17 CXC chemokines have been identified and classified in two classes on the basis of ELR motif, CXCL1-8 and CXCL15 (except CXCL4) contains the ELR motif that is essential for neutrophil recruitment, while the rest (CXCL9-17) do not possess this motif [64]. Current understanding on CXCL8 sequences in fish suggests that, the presence of a number of CXCL8 and CXCL8-like (CXCL8L) genes in the fish genome is very unique and is independent of the total rounds of whole genome duplication (WGD).

Lineage-wise sequence analysis suggested that the ancestor fish have gone through two rounds of whole genome duplication (WGD) but species of ancestor subclasses Holocephali (Australian ghostshark) and Holostei (Spotted gar) contain 5 and 4 copies of CXCL8/CXCL8L genes respectively. Although teleost have undergone 3 rounds of WGD, the expected 3 copies of CXCL8 were not identified. It can be presumed that the deletion of few copies of ancestral CXCL8 gene might have happened after third round of WGD in the teleost fishes [65]. Indeed, differential expression profiles and functionalities have been observed among several copies and lineages of CXCL8/CXCL8L proteins of the same species. For example, Japanese flounder CXCL8_L1a targets the CD8 T-cells while CXCL8_L1b is chemotactic to B cells and CD8 T cells [66]. CXCL8-L3 from the large yellow croaker attracts lymphocytes and eosinophils in peripheral blood leukocytes, while CXCL8_L1 express in all types of tissues play crucial role in the inflammatory responses [33,67]. The phylogenetic classification of teleost CXCL8 in the present study evidenced the diversification into four distinct lineages (CXCL8-L1a, CXCL8-L1b, CXCL8-L2 and CXCL8-L3), against three broad lineages (CXCL8-L1, CXCL8-L2 and CXCL8-L3) as identified in previous reports, which can be potentially used to elucidate the evolutionary origins of lineage-specific functionalities [68]. Interestingly, our selection analysis also suggested that the CXCL8 gene has evolved under the strong purifying selection thus supporting the fact that the structure and the function of this unique chemokine are tightly regulated by evolutionary pressure [69]. It is well established that functionally important and highly expressing genes mostly evolve under the purifying selection so that the important function of protein should not be affected [70]. Several other chemokines such as GRO chemokines and chemokine receptor CCR5 faced purifying selection pressure is the major force of evolution [42,71].

It has been also reported that, although the ELR motif is essential for the CXCL8 receptor activation, its presence is not sufficient for receptor activation, thus suggesting that along with ELR, some other structural elements are also essential for neutrophil recruitment. It has also been reported the third residue of this motif is overall conserved with the positively charged residue (K/R/H). These findings are supported by mutational studies, showing that activity had mostly affected by the third residue in the following order Arg ≫ Glu > Leu in human CXCL8 [72]. Recent study on amphibian (*Xenopus laevis*) CXCL8 orthologs (CXCL8a and CXCL8b) also suggested that the ortholog CXCL8a containing an ELR motif is more potent chemoattractant compared to CXCL8b that lacks ELR motif [73]. Moreover, the ELR-containing CXCL8a isoform exhibited similar functional features to those of associated higher vertebrate CXCL8 homologs. This observation suggested that in amphibians, one copy (CXCL8a) getting evolved towards higher vertebrates and another copy (CXCL8b) retained primitive CXCL8 features as deciphered in fish sequences. Comparison of these *X. laevis* CXCL8 features with those of fish and primate CXCL8 establishes that neofunctionaliztion has been occurred during the evolution of CXCL8 gene. Our analysis comprehensively elucidated that, the complete ELR motif is majorly absent from CXCL8 like protein in the fish except for Gadoid fish lineage and spotted gar-1 from subclass Holostei. These results are in line with various mutagenesis studies which revealed that a complete ELR motif is not essential for the neutrophil recruitment in the fish, and suggested for an intermediate motif during evolution [30]. The current study hierarchically delineates that the ELR-like motif of ancestor fish had evolved from the GGR of Lamprey, then it underwent duplication and evolved in two directions (a) NXH/R of L3 lineage and (b) ELR/DLR motif of L1 lineage. This observation is well corroborated with the previously reported evolutionary hypothesis trend of the ELR-like motif [36]. Indeed, our branch and branch site selection analysis indicated that Coelacanth (ancestral) with “DLR”, and Gadoid clade (L1a) with complete “ELR” motif have positively selected and might have provided with the advantage. Coelacanth and CXCL8-L2 clades are selected in the branch site model and “DLR” motif in the Coelacanth supposed to be one of the reasons for selection. One more advantage could be that the coelacanth is connecting link between the fish and tetrapod. A strong selection of CXCL8-L2 lineage under the branch-site model can be reasoned to that, it has some unique functional sites that provide a selective advantage to cyprinids over the other teleost. Such a positive branch selection was also reported for the lysozyme-c gene of colobine monkey from the set of 19 primate lysozyme c sequence [74].

During evolution, it has been generally observed that the structure of the protein is more conserved than its sequence. For example, different lysozymes starting from bacteriophages to mammals adopt similar 3D fold albeit great sequence variations [75]. Although the evolution of chemokine genes (such as GRO genes, CCL2, and CCR5) is tightly regulated by their functional importance, structural conservation with sequence variation is commonly observed [42,71,76]. Our analysis presented a great similarity between the sequence and the structures of various CXCL8 chemokine proteins from different lineages with high resemblance of contact maps, thus suggesting that most of the genes are derived from a common ancestor through gene duplication and deletion. The most common way of acquiring a new function via duplication is when one of the duplicated copies is retained to maintain the original functions and the other copy evolves to acquire a novel function. The example of such a process can be found in the β tubulin superfamily from the fungi where functionally important residues are evolved under type-II divergence and gained differential functions in contrast to others which retained the original function [77]. Another best-studied case of functional divergence is of different sub-families of bestrophin *trans*-membrane proteins, where it is evidenced that the majority of divergent residues are functionally important and belong to the loop or C-terminal region that are accessible to the soluble ligands [78]. In line with these reports, lineage-based functional divergence analysis of CXCL8 chemokine suggested that each lineage of CXCL8 possess functional differences supported by the substitution of functionally important amino acid. These in-silico functional divergence results can be considered for further experimental investigations to elucidate the lineage-specific functional differences of CXCL8 chemokines. It is worth noting that the specific functional divergence trends were also observed between the GRO chemokines of primates and rodents [79].

The residue level structural binding studies performed between the chemokines and their cognate receptor(s)/glycosaminoglycans in human/mice have provided important insights into the domains/motif/residues that are responsible for their interactions [80–82]. The functional divergence residues obtained in the current study are very much in line with these reported experimental results. It has been characterized that the N-terminal unstructured loop of CXCL8 majorly participates in receptor interaction. Positively charged residues mainly from the C–terminal helix and from other loops are responsible for glycosaminoglycan interaction. Various mutagenesis-based structural and biochemical studies have proposed that ELR motif, other residues from the dynamic N-loop, first disulfide bond, and 30s loop are mainly involved in receptor binding [57]. For receptor binding, the major interacting residues of the dynamic N-loop of human CXCL8 are Ile15,

Thr17, Tyr18, Ser19, Lys20, Lys25, His23, and Phe26; whereas, positively charged residues (His23, Lys25, Arg65, Lys69, Lys72, and Arg73) make interactions with cell surface GAGs [58]. Our functional divergence studies on the fish CXCL8 lineages have suggested that several of the corresponding residues for receptor binding and GAG-binding are selected under type-I functional divergence, thus suggesting for a greater functional diversity of these primitive vertebrate species.

5. Concluding remarks

In summary, the current study has redefined the lineage based classification of CXC chemokines based on the CXCL8 gene sequences of fish genome. A clear distinction has been seen in the L1 lineage and has been named as L1a and L1b lineages. The updated lineage based phylogenetic classification can help to understand the classification and evolution of different chemokine families in the lower vertebrates. The selection analyses suggested that the CXCL8 has been evolved under purifying selection. Branch site analysis suggested that the species containing ELR motif has been positively selected, and the molecular evolutionary trend of the "ELR" motif established the evolutionary relationship of CXCL8 between the primate and lower vertebrates, thus providing insights into the origin and hierarchy of this specific motif. Further, the study on CXCL8 evidenced for a conserved structural nature and a great extent of functional divergence through subtle sequence-specific amino acid changes in the lower vertebrates. Such an evolutionary mechanism has generated a great repertoire of CXCL8/CXCL8 like genes that perform differential physiological functions in addition to their conserved cell migration roles, which can potentially compensate for the poor development of their acquired immune system.

Acknowledgements

KMP acknowledge the receipt of Grant CRG/2018/001329 from SERB-DST, and DBT-IYBA fellowship – BT/07/IYBA/2013–19, KG acknowledges the receipt of MHRD fellowship from IIT-Roorkee.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsi.2019.07.034.

References

[1] G. Caetano-Anolles, M. Wang, D. Caetano-Anolles, J.E. Mittenthal, The origin, evolution and structure of the protein world, Biochem. J. 417 (2009) 621–637.
[2] N.S. de Groot, M.T. Burgas, C.N. Ravarani, A. Trusina, S. Ventura, M.M. Babu, The fitness cost and benefit of phase-separated protein deposits, Mol. Syst. Biol. 15 (2019) e8075.
[3] H.A. Orr, Fitness and its role in evolutionary genetics, Nat. Rev. Genet. 10 (2009) 531–539.
[4] J.I. Boucher, D.N. Bolon, D.S. Tawfik, Quantifying and understanding the fitness effects of protein mutations: laboratory versus nature, Protein Science, vol. 25, a publication of the Protein Society, 2016, pp. 1219–1226.
[5] M. Nei, The new mutation theory of phenotypic evolution, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 12235–12242.
[6] K. Gulati, K. Gangele, D. Kumar, K.M. Poluri, An inter-switch between hydrophobic and charged amino acids generated druggable small molecule binding pocket in chemokine paralog CXCL3, Arch. Biochem. Biophys. 662 (2019) 121–128.
[7] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science 290 (2000) 1151–1155.
[8] G.A. Dover, Molecular drive in multigene families: how biological novelties arise, spread and are assimilated, Trends Genet. 2 (1986) 159–165.
[9] M. Nei, A.P. Rooney, Concerted and birth-and-death evolution of multigene families, Annu. Rev. Genet. 39 (2005) 121–152.
[10] D. Zelus, M. Robinson-Rechavi, M. Delacre, C. Auriault, V. Laudet, Fast evolution of interleukin-2 in mammals and positive selection in ruminants, J. Mol. Evol. 51 (2000) 234–244.
[11] M. Baggiolini, P. Loetscher, Chemokines in inflammation and immunity, Immunol. today 21 (2000) 418–420.
[12] K.M. Poluri, Chemokines: the holy messengers of leukocyte trafficking, Austin J. Biotechnol. Bioeng. 1 (2014) 1–3.
[13] A.D. Luster, The role of chemokines in linking innate and adaptive immunity, Curr. Opin. Immunol. 14 (2002) 129–135.
[14] C. Esche, C. Stellato, L.A. Beck, Chemokines: key players in innate and adaptive immunity, J. Investig. Dermatol. 125 (2005) 615–628.
[15] M.E. DeVries, A.A. Kelvin, L. Xu, L. Ran, J. Robinson, D.J. Kelvin, Defining the origins and evolution of the chemokine/chemokine receptor system, J. Immunol. 176 (2006) 401–415.
[16] A. Zlotnik, O. Yoshie, The chemokine superfamily revisited, Immunity 36 (2012) 705–716.
[17] C. Bizzarri, A.R. Beccari, R. Bertini, M.R. Cavicchia, S. Giorgini, M. Allegretti, ELR+ CXC chemokines and their receptors (CXC chemokine receptor 1 and CXC chemokine receptor 2) as new therapeutic targets, Pharmacol. Ther. 112 (2006) 139–149.
[18] R.M. Strieter, P.J. Polverini, S.L. Kunkel, D.A. Arenberg, M.D. Burdick, J. Kasper, et al., The functional role of the ELR motif in CXC chemokine-mediated angiogenesis, J. Biol. Chem. 270 (1995) 27348–27357.
[19] M.O. Huising, E. Stolte, G. Flik, H.F.J. Savelkoul, B.M.L. Verburg-van Kemenade, CXC chemokines and leukocyte chemotaxis in common carp (Cyprinus carpio L.), Dev. Comp. Immunol. 27 (2003) 875–888.
[20] K.A. McInnis, A. Britain, R.N. Lausch, J.E. Oakes, Human corneal epithelial cells synthesize ELR− α-chemokines in response to proinflammatory mediators, Ocul. Immunol. Inflamm. 15 (2007) 295–302.
[21] K. Rajarathnam, M. Schnoor, R.M. Richardson, S. Rajagopal, How do chemokines navigate neutrophils to the target site: dissecting the structural mechanisms and signaling pathways, Cell. Signal. 54 (2018) 69–80.
[22] R. Horuk, Chemokine receptors, Cytokine Growth Factor Rev. 12 (2001) 313–335.
[23] K. Gulati, K.M. Poluri, Mechanistic and therapeutic overview of glycosaminoglycans: the unsung heroes of biomolecular signaling, Glycoconj. J. 33 (2016) 1–17.
[24] N. Mukaida, Interleukin-8: an expanding universe beyond neutrophil chemotaxis and activation, Int. J. Hematol. 72 (2000) 391–398.
[25] J. Heidemann, H. Ogawa, M.B. Dwinell, P. Rafiee, C. Maaser, H.R. Gockel, et al., Angiogenic effects of interleukin 8 (CXCL8) in human intestinal microvascular endothelial cells are mediated by CXCR2, J. Biol. Chem. 278 (2003) 8508–8515.
[26] K. Inoue, J.W. Slaton, B.Y. Eve, S.J. Kim, P. Perrotte, M.D. Balbay, et al., Interleukin 8 expression regulates tumorigenicity and metastases in androgen-independent prostate cancer, Clin. Cancer Res. 6 (2000) 2104–2119.
[27] R. Strieter, S. Kunkel, V. Elner, C. Martonyi, A. Koch, P. Polverini, et al., Interleukin-8. A corneal factor that induces neovascularization, Am. J. Pathol. 141 (1992) 1279.
[28] H. Ha, B. Debnath, N. Neamati, Role of the CXCL8-CXCR1/2 axis in cancer and inflammatory diseases, Theranostics 7 (2017) 1543.
[29] I. Gilowska, CXCL8 (interleukin 8)–the key inflammatory mediator in chronic obstructive pulmonary disease? Postępy Higieny Medycyny Doświadczalnej 68 (2014) 842–850.
[30] Z. Cai, C. Gao, Y. Zhang, K. Xing, Functional characterization of the ELR motif in piscine ELR+CXC-like chemokine, Mar. Biotechnol. 11 (2009) 505–512.
[31] A.M. Najakshin, L.V. Mechetina, B.Y. Alabyev, A.V. Taranin, Identification of an IL-8 homolog in lamprey (Lampetra fluviatilis): early evolutionary divergence of chemokines, Eur. J. Immunol. 29 (1999) 375–382.
[32] L. Chen, C. He, P. Baoprasertkul, P. Xu, P. Li, J. Serapion, et al., Analysis of a catfish gene resembling interleukin-8: cDNA cloning, gene structure, and expression after infection with Edwardsiella ictaluri, Dev. Comp. Immunol. 29 (2005) 135–142.
[33] C. Li, C.L. Yao, Molecular and expression characterizations of interleukin-8 gene in large yellow croaker (Larimichthys crocea), Fish Shellfish Immunol. 34 (2013) 799–809.
[34] J. Chen, Q. Xu, T. Wang, B. Collet, Y. Corripio-Miyar, S. Bird, et al., Phylogenetic analysis of vertebrate CXC chemokines reveals novel lineage specific groups in teleost fish, Dev. Comp. Immunol. 41 (2013) 137–152.
[35] L.M. Van Der Aa, M. Chadzinska, E. Tijhaar, P. Boudinot, B.L. Verburg-van Kemenade, CXCL8 chemokines in teleost fish: two lineages with distinct expression profiles during early phases of inflammation, PLoS One 5 (2010) e12384.
[36] C. Zhonghua, G. Chunpin, Z. Yong, X. Kezhi, Z. Yaou, Cloning and bioactivity analysis of a CXC ligand in black seabream Acanthopagrus schlegeli: the evolutionary clues of ELR+ CXC chemokines, BMC Immunol. 9 (2008) 66.
[37] M. Seppola, A.N. Larsen, K. Steiro, B. Robertsen, I. Jensen, Characterisation and expression analysis of the interleukin genes, IL-1beta, IL-8 and IL-10, in Atlantic cod (Gadus morhua L.), Mol. Immunol. 45 (2008) 887–897.
[38] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[39] A. Rambaut, FigTree v1. 4 (2012).
[40] S.L.K. Pond, S.D. Frost, Datamonkey: rapid detection of selective pressure on individual sites of codon alignments, Bioinformatics 21 (2005) 2531–2533.
[41] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591.
[42] K. Gulati, M. Jamsandekar, K.M. Poluri, Mechanistic Insights into Molecular Evolution of Species-specific Differential Glycosaminoglycan Binding Surfaces in Growth-Related Oncogene Chemokines vol. 4, Royal Society open science, 2017, p. 171059.
[43] Z. Yang, R. Nielsen, Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages, Mol. Biol. Evol. 19 (2002) 908–917.
[44] M.D. Smith, J.O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, S.L. Kosakovsky Pond, Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection, Mol. Biol. Evol. 32 (2015) 1342–1353.
[45] C. Vehlow, H. Stehr, M. Winkelmann, J.M. Duarte, L. Petzold, J. Dinse, et al., CMView: interactive contact map visualization and analysis, Bioinformatics 27 (2011) 1573–1574.
[46] X. Gu, Y. Zou, Z. Su, W. Huang, Z. Zhou, Z. Arendsee, et al., An update of DIVERGE software for functional divergence analysis of protein family, Mol. Biol. Evol. 30 (2013) 1713–1719.

[47] G.M. Clore, E. Appella, M. Yamada, K. Matsushima, A.M. Gronenborn, Three-dimensional structure of interleukin 8 in solution, Biochemistry 29 (1990) 1689–1696.
[48] E.T. Baldwin, I.T. Weber, R. St Charles, J.-C. Xuan, E. Appella, M. Yamada, et al., Crystal structure of interleukin 8: symbiosis of NMR and crystallography, Proc. Natl. Acad. Sci. 88 (1991) 502–506.
[49] R. Froese, D. Pauly, FishBase. Fisheries Centre, University of British Columbia, 2010.
[50] B.-Y. Liao, N.M. Scott, J. Zhang, Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins, Mol. Biol. Evol. 23 (2006) 2072–2080.
[51] D.P. Wall, A.E. Hirsh, H.B. Fraser, J. Kumm, G. Giaever, M.B. Eisen, et al., Functional genomic analysis of the rates of protein evolution, Proc. Natl. Acad. Sci. 102 (2005) 5483–5488.
[52] C. Pegueroles, S. Laurie, M.M. Alba, Accelerated evolution after gene duplication: a time-dependent process affecting just one copy, Mol. Biol. Evol. 30 (2013) 1830–1842.
[53] Y. Corripio-Miyar, S. Bird, K. Tsamopoulos, C.J. Secombes, Cloning and expression analysis of two pro-inflammatory cytokines, IL-1 beta and IL-8, in haddock (Melanogrammus aeglefinus), Mol. Immunol. 44 (2007) 1361–1373.
[54] S. Brugman, M. Witte, R.C. Scholman, M.R. Klein, M. Boes, E.E. Nieuwenhuis, T lymphocyte-dependent and -independent regulation of Cxcl8 expression in zebrafish intestines, J. Immunol. 192 (2014) 484–491.
[55] S. de Oliveira, A. Lopez-Munoz, F.J. Martinez-Navarro, J. Galindo-Villegas, V. Mulero, A. Calado, Cxcl8-l1 and Cxcl8-l2 are required in the zebrafish defense against Salmonella Typhimurium, Dev. Comp. Immunol. 49 (2015) 44–48.
[56] E.A. Gaucher, X. Gu, M.M. Miyamoto, S.A. Benner, Predicting functional divergence in protein evolution by site-specific rate shifts, Trends Biochem. Sci. 27 (2002) 315–321.
[57] P.R.B. Joseph, K.V. Sawant, A. Isley, M. Pedroza, R.P. Garofalo, R.M. Richardson, et al., Dynamic conformational switching in the chemokine ligand is essential for G-protein-coupled receptor activation, Biochem. J. 456 (2013) 241–251.
[58] G.S. Kuschert, A.J. Hoogewerf, A.E. Proudfoot, C-w Chung, R.M. Cooke, R.E. Hubbard, et al., Identification of a glycosaminoglycan binding surface on human interleukin-8, Biochemistry 37 (1998) 11193–11201.
[59] S. Magadum, U. Banerjee, P. Murugan, D. Gangapur, R. Ravikesavan, Gene duplication as a major force in evolution, J. Genet. 92 (2013) 155–161.
[60] E. Peatman, Z. Liu, Evolution of CC chemokines in teleost fish: a case study in gene duplication and implications for immune diversity, Immunogenetics 59 (2007) 613–623.
[61] A. Alejo, C. Tafalla, Chemokines in teleost fish species, Dev. Comp. Immunol. 35 (2011) 1215–1222.
[62] M.O. Huising, R.J.M. Stet, C.P. Kruiswijk, H.F.J. Savelkoul, B.M. Lidy Verburg-van Kemenade, Molecular evolution of CXC chemokines: extant CXC chemokines originate from the CNS, Trends Immunol. 24 (2003) 306–312.
[63] H. Nomiyama, K. Hieshima, N. Osada, Y. Kato-Unoki, K. Otsuka-Ono, S. Takegawa, et al., Extensive expansion and diversification of the chemokine gene family in zebrafish: identification of a novel chemokine subfamily CX, BMC Genomics 9 (2008) 222.
[64] K.J. Laing, C.J. Secombes, Chemokines, Dev. Comp. Immunol. 28 (2004) 443–460.
[65] J. Inoue, Y. Sato, R. Sinclair, K. Tsukamoto, M. Nishida, Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling, Proc. Natl. Acad. Sci. 112 (2015) 14918–14923.
[66] B. Zhao, T. Katagiri, H. Kondo, I. Hirono, Comparative analysis of two types of CXCL8 from Japanese flounder (Paralichthys olivaceus), Dev. Comp. Immunol. 52 (2015) 37–47.
[67] S. Zhou, Y. Mu, J. Ao, X. Chen, Molecular characterization and functional activity of CXCL8_L3 in large yellow croaker Larimichthys crocea, Fish Shellfish Immunol. 75 (2018) 124–131.
[68] X. Wang, G. Ma, R. Zhang, L. Liu, J. Zhu, H. Zhu, Molecular characterization and biological functioning of interleukin-8 in Siberian sturgeon (Acipenser baeri), Fish Shellfish Immunol. 90 (2019) 91–101.
[69] J. Siltberg-Liberles, J.A. Grahnen, D.A. Liberles, The evolution of protein structures and structural ensembles under functional constraint, Genes 2 (2011) 748–762.
[70] S. Subramanian, S. Kumar, Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome, Genetics 168 (2004) 373–381.
[71] Y. Zhang, O.A. Ryder, Y. Zhang, Sequence evolution of the CCR5 chemokine receptor gene in primates, Mol. Biol. Evol. 16 (1999) 1145–1154.
[72] I. Clark-Lewis, B. Dewald, T. Geiser, B. Moser, M. Baggiolini, Platelet factor 4 binds to interleukin 8 receptors and activates neutrophils when its N terminus is modified with Glu-Leu-Arg, Proc. Natl. Acad. Sci. 90 (1993) 3574–3577.
[73] D.V. Koubourli, A. Yaparla, M. Popovic, L. Grayfer, Amphibian (Xenopus laevis) interleukin-8 (CXCL8): a perspective on the evolutionary divergence of granulocyte chemotaxis, Front. Immunol. 9 (2018) 2058.
[74] C.-B. Stewart, J.W. Schilling, A.C. Wilson, Adaptive evolution in the stomach lysozymes of foregut fermenters, Nature 330 (1987) 401.
[75] M. Grütter, L. Weaver, B. Matthews, Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? Nature 303 (1983) 828.
[76] K.J. Metzger, M.A. Thomas, Evidence of positive selection at codon sites localized in extracellular domains of mammalian CC motif chemokine receptor proteins, BMC Evol. Biol. 10 (2010) 139.
[77] Z. Zhao, H. Liu, Y. Luo, S. Zhou, L. An, C. Wang, et al., Molecular evolution and functional divergence of tubulin superfamily in the fungal tree of life, Sci. Rep. 4 (2014) 6746.
[78] V.M. Milenkovic, T. Langmann, R. Schreiber, K. Kunzelmann, B.H. Weber, Molecular evolution and functional divergence of the bestrophin protein family, BMC Evol. Biol. 8 (2008) 72.
[79] K. Gulati, K. Gangele, N. Agarwal, M. Jamsandekar, D. Kumar, K.M. Poluri, Molecular cloning and biophysical characterization of CXCL3 chemokine, Int. J. Biol. Macromol. 107 (2018) 575–584.
[80] S.J. Allen, S.E. Crown, T.M. Handel, Chemokine: receptor structure, interactions, and antagonism, Annu. Rev. Immunol. 25 (2007) 787–820.
[81] Z. Johnson, A. Proudfoot, T. Handel, Interaction of chemokines and glycosaminoglycans: a new twist in the regulation of chemokine function with opportunities for therapeutic intervention, Cytokine Growth Factor Rev. 16 (2005) 625–636.
[82] K.M. Poluri, P.R.B. Joseph, K.V. Sawant, K. Rajarathnam, Molecular basis of glycosaminoglycan heparin binding to the chemokine CXCL1 dimer, J. Biol. Chem. 288 (2013) 25143–25153.