Immunogenetics (2007) 59:613–623
DOI 10.1007/s00251-007-0228-4

REVIEW

Evolution of CC chemokines in teleost fish: a case study in gene duplication and implications for immune diversity

Eric Peatman · Zhanjiang Liu

Received: 9 April 2007 / Accepted: 11 April 2007 / Published online: 31 May 2007

Abstract Chemokines are a superfamily of cytokines responsible for regulating cell migration under both inflammatory and physiological conditions. CC chemokines are the largest subfamily of chemokines, with 28 members in humans. A subject of intense study in mammalian species, the known functional roles of CC chemokines ligands in both developmental and disease conditions continue to expand. They are also an important family for the study of gene copy number variation and tandem duplication in mammalian species. However, little is known regarding the evolutionary origin and status of these ligands in primitive vertebrates such as teleost fish. In this paper, we review the evolution of the teleost fish CC chemokine gene family, noting evidence of widespread tandem gene duplications and examining the implications of this phenomenon on immune diversity. Through extensive phylogenetic analysis of the CC chemokine sets of four teleost species, zebrafish, catfish, rainbow trout, and Atlantic salmon, we identified seven large groups of CC chemokines. It appeared that several major groups of CC chemokines are highly related including the CCL19/21/25 group, the CCL20 group, CCL27/28 group, and the fish-specific group. In the three remaining groups that contained the largest number of members, the CCL17/22 group, the MIP group, and the MCP group, similarities among species members were obscured by rapid, tandem duplications that may contribute to immune diversity.

Keywords Chemokine · Cytokine · Immunity · Evolution · Gene duplication · Catfish · Salmon · Zebrafish · Trout

E. Peatman · Z. Liu (✉)
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program
of Cell and Molecular Biosciences, Aquatic Genomics Unit,
Auburn University,
203 Swingle Hall,
Auburn, AL 36849, USA
e-mail: zliu@acesag.auburn.edu

Introduction

Cytokines coordinate the tremendous complexity of chemical signals needed to initiate and maintain vertebrate immune responses. Cytokines and their receptors control the timing, location, and intensity of the pathogen response, initiate expression changes in a broad array of genes, and recruit leukocyte cell types to sites of infection or injury. Cytokine function, therefore, is intimately linked with the progression and outcome of a wide variety of infectious and inflammatory diseases. Chemotactic cytokines, or chemokines, are a family of small proteins that regulate cell migration under both inflammatory and normal physiological conditions. They share characteristic structures, with the majority containing four conserved cysteine residues. Based on the arrangement of the first two of these conserved cysteine residues (Bacon et al. 2003), chemokines were divided into four subfamilies: CC, CXC, C, and CX3C. CC chemokines, distinguished by adjacent cysteine residues in a conserved position, make up the largest subfamily of ligands with 28 members in humans and primarily attract mononuclear cell types. The ligand–receptor relations of the CC chemokine family in mammals are promiscuous; many receptors can bind multiple ligands, and some ligands can bind to multiple receptors (Laing and Secombes 2004a; Zlotnik et al. 2006).

Studies of the function and origins of the regulatory mechanisms of immunity are necessary for a complete

understanding of complex immune responses. In mammals, CC chemokines comprise a significant portion of the cytokine repertoire and coordinate the migration of diverse leukocyte populations. Conversely, invertebrate species, although possessing many innate immune components, do not appear to encode CC chemokine ligands or receptors (DeVries et al. 2006). Until recently, little has been known regarding the evolutionary origin of these ligands in ancestral vertebrates such as teleost fish. Questions remain, therefore, whether the diversity of chemokine ligands present in higher vertebrates is conserved in species possessing more rudimentary immune systems. Research into early CC chemokines has taken on a greater importance with the realization of the fluid nature of chemokine ligand genes. While it has been noted in the comparison of the human and mouse genomes that immune-related genes are evolving more quickly than other types of genes, chemokines are under particularly strong selection and are one of the eight most rapidly changing proteins and domains (Waterston et al. 2002). This strong selection pressure has manifested itself in independent tandem gene duplications that have arisen in both human and mouse since their divergence. This rapid generation of species-specific immune diversity may reflect a response to the different 'infectious experiences' encountered by the two species (Zlotnik et al. 2006).

The study of fish CC chemokines provides an important point of comparison with these findings in mammals, providing an evolutionary perspective on chemokine gene duplications and their contributions to the complexity of immune responses in vertebrates. In this paper, we review the evolution of the teleost fish CC chemokine gene family, noting evidence of widespread tandem gene duplications and examining the implications of this phenomenon on immune diversity.

CC chemokine classifications—spatial and functional

Just as chemokines can be divided into four subfamilies based on their structural motifs and their targeted cell types, subgroups of mammalian CC chemokines can be classified by certain spatial and functional characteristics. Understanding these classifications is important for the analysis of the teleost CC chemokines in the context of the mammalian family. In mammals, the broadest functional classification system divides CC chemokines into inflammatory and homeostatic groups (Moser et al. 2004; Zlotnik et al. 2006) based on their expression patterns. Inflammatory chemokines are expressed in immune-related cell types only after appropriate stimulation, whereas homeostatic chemokines are expressed in discrete locations under normal physiological conditions. This division is widely acknowledged to be overly simplistic and, as a result, a third "dual function" category is quickly gaining membership. Another level of classification (and complexity) is added by studying the genomic organization of the CC chemokines. The majority of human CC chemokines are located in two large clusters on distinct regions of human chromosome 17. Similar clusters occur on mouse chromosome 11. The remaining CC chemokine genes are located separately in small 2–3 gene clusters or are not clustered (Yoshie et al. 2001). The two large clusters of CC chemokines, the monocyte chemotactic protein group (MCP) and the macrophage inflammatory protein group (MIP), contain the majority of the "inflammatory" CC chemokines, while most "homeostatic" or dual function CC chemokines are in small clusters or non-clustered. Molecular evidence points to tandem gene duplications as the source of many of the MCP and MIP CC chemokines in human and mouse (Zlotnik et al. 2006). These ligands multiplied from a smaller ancestral set sometime after the divergence of the human and mouse species. As a result, these cluster chemokines share significantly higher sequence identities with their ancestral paralogues than they do with the equivalent group of chemokines from another species. On the other hand, small cluster or non-clustered CC chemokines in human and mouse, on the whole, have not multiplied to the same extent, and orthologies between these members in each species can be concretely established.

Teleost CC chemokine datasets

Broad surveys of molecular evolution in vertebrate species have traditionally favored comparisons with the fish model species zebrafish (*Danio rerio*) and the green spotted pufferfish (*Tetraodon nigroviridis*) because of the availability of their genome sequences (Woods et al. 2005). However, much of the pioneering work for defining the teleost immune system has not come from the model species but rather from important aquaculture species. A growing worldwide aquaculture industry has been confronted by a host of bacterial and viral pathogens to fish that were virtually unknown in natural conditions. To address these disease issues, systematic characterization of the immune responses of several aquaculture species has been made a priority. As a result of these efforts, large expressed sequence tag (EST) databases have become available from aquaculture species such as Atlantic salmon (*Salmo salar*; Davey et al. 2001; Martin et al. 2002; Rise et al. 2004; von Schalburg et al. 2005; Hagen-Larsen et al. 2005), rainbow trout (*Oncorhynchus mykiss*; Rexroad et al. 2003; Govoroun et al. 2006), and channel catfish and blue catfish (*Ictalurus punctatus* and *Ictalurus furcatus*; Li et al. 2007), as well as from the model species zebrafish (Zeng and

Gong 2002; Lo et al. 2003; Song et al. 2004). These EST resources contain much of the immune gene machinery of these species needed to survive in pathogen-rich environments and are, therefore, a plentiful source for comparative immunology.

Regarding chemokines, ESTs have allowed the rapid detection of the small (<100 amino acids), divergent ligand genes in fish that had proved difficult to isolate by hybridization or PCR, or even to detect by bioinformatic analysis in sequenced genomes. Following the discoveries of a small number of CC chemokines in rainbow trout, carp, and Japanese flounder (Dixon et al. 1998; Liu et al. 2002; Fujiki et al. 1999; Kono et al. 2003; Khattiya et al. 2004; Mackenzie et al. 2004), larger EST surveys identified increasing numbers of CC chemokines in the transcriptomes of teleost species. Seven new chemokines from two cichlid fish, *Paralabidochromis chilotes* and *Melanochromis auratus*, and catshark *Scyliorhinus canicula* were reported (Kuroda et al. 2003), followed by sequencing and analysis of 14 CC chemokines from catfish (He et al. 2004) and 15 additional trout CC chemokines (Laing and Secombes 2004b). Subsequently, 12 more distinct CC chemokines were identified from catfish (Peatman et al. 2005) for a total of at least 26 CC chemokines in a teleost species. All 26 CC chemokines were localized to clones of a catfish genomic bacterial artificial chromosome (BAC) library. Of these, complete genomic sequences and structures were obtained for 23 CC chemokine genes (Bao et al. 2006). During gene sequencing, two additional CC chemokine genes, SCYA114B and SCYA120B, were identified in tandem with the targeted genes and possessed nearly identical amino acid sequences to the targets. To obtain additional teleost CC chemokine sets for comparative and phylogenetic analysis, the zebrafish ESTs and genome sequence and a set of 431,774 ESTs from Atlantic salmon were searched for CC chemokines using previously identified CC chemokines as queries in tBLASTn searches. In zebrafish, as many as 46 loci in the zebrafish genome were found to encode putative CC chemokines, most previously unannotated (Peatman and Liu 2006). A total of 466 salmon ESTs were identified by basic local alignment search tool (BLAST) searches as candidates potentially coding for CC chemokines. After clustering and translation, 30 putative salmon CC chemokines were selected for further analysis in this study.

Phylogenetic analysis of teleost CC chemokines

Phylogenetic analysis of chemokines has been considered problematic due to the fact that the short amino acid sequences often provide an insufficient number of comparison sites to establish accurate branching patterns. Rapid divergence and independent duplication events within each species further complicate the identification of orthologs. Additionally, surveys of chemokine evolution must balance the desire for complete analysis with the practical difficulties of interpreting and presenting phylogenetic trees constructed with over 100 sequences (Shields 2003; DeVries et al. 2006; Peatman et al. 2006; Zlotnik et al. 2006). Previously, CC chemokine phylogenies have been constructed utilizing sequence information from several teleost species (Laing and Secombes 2004b; He et al. 2004; Liu and Peatman 2006; Peatman and Liu 2006; DeVries et al. 2006). For the purpose of this review, we conducted phylogenetic analysis with the four large teleost CC chemokine sets described above from zebrafish, rainbow trout, catfish, and Atlantic salmon (Table 1). While small numbers of additional CC chemokine sequences are available from other teleost species, we limited our analysis to a set of sequences likely to be representative of the larger patterns of conservation and divergence in the gene family. Constraints on tree size and readability made it necessary to exclude CC chemokines from model species such as *T. nigroviridis* and *Gallus gallus* from the phylogenetic tree presented in this review, although these too were analyzed. Amino acid sequences of 28 catfish CC chemokines (SCYA-), 30 Atlantic salmon CC chemokines (Salmo), 17 rainbow trout CC chemokines (CK-), and 46 zebrafish CC chemokines (Dr-) were aligned by ClustalW with 28 mammalian CC chemokines. Three short zebrafish sequences that lacked the final conserved cysteine were eliminated from the analysis, leaving only sequences with all four conserved cysteines. Phylogenetic trees were constructed from the ClustalW generated multiple sequence alignment. The presented tree (Fig. 1) was constructed using the neighbor-joining method (Saitou and Nei 1987), and gaps were removed by pairwise deletion.

Seven groups of CC chemokines were evident from the phylogenetic analysis (Fig. 1) and were named the CCL20 group, CCL27/28 group, MCP group, MIP group, CCL17/22 group, CCL19/21/25 group, and the Fish (specific) CC group, based on their mammalian membership. Bootstrapping support of these larger clades is low, as expected from previous studies comparing vertebrate chemokines (He et al. 2004; Wang et al. 2005; DeVries et al. 2006). However, bootstrapping values of fish and human subclades were often significant (>50%). Additional support for the overall topology of the tree was gained by constructing separate trees using different algorithms (minimum evolution and maximum parsimony) and also by comparing the present tree with previously published smaller phylogenies containing teleost CC chemokines.

As expected, the orthologous relationships within the large cluster MIP and MCP groups of CC chemokines were largely obscured by the rapid and radiated multiplication of these genes. Within the MCP group, the human MCP members,

Table 1 Teleost CC chemokines used for phylogenetic analysis

Teleost CC chemokine	Species	Reference
CK1	*Oncorhynchus mykiss*	Dixon et al. (1998)
CK2	*Oncorhynchus mykiss*	Liu et al. (2002)
CK3	*Oncorhynchus mykiss*	Laing and Secombes (2004a, b)
CK4A	*Oncorhynchus mykiss*	
CK4B	*Oncorhynchus mykiss*	
CK5A	*Oncorhynchus mykiss*	
CK5B	*Oncorhynchus mykiss*	
CK6	*Oncorhynchus mykiss*	
CK7A	*Oncorhynchus mykiss*	
CK7B	*Oncorhynchus mykiss*	
CK8A	*Oncorhynchus mykiss*	
CK8B	*Oncorhynchus mykiss*	
CK9	*Oncorhynchus mykiss*	
CK10	*Oncorhynchus mykiss*	
CK11	*Oncorhynchus mykiss*	
CK12A	*Oncorhynchus mykiss*	
CK12B	*Oncorhynchus mykiss*	
Dr1_WGA6_97332	*Danio rerio*	Peatman and Liu (2006)
Dr7_WGA489_245599	*Danio rerio*	
Dr8_WGA606_66363	*Danio rerio*	
Dr9_WGA697_854420	*Danio rerio*	
Dr10_WGA780_413112	*Danio rerio*	
Dr11_WGA839_159228	*Danio rerio*	
Dr11_WGA839_233313	*Danio rerio*	
Dr11_WGA879_755946	*Danio rerio*	
Dr11_WGA879_759152	*Danio rerio*	
Dr11_WGA879_762108	*Danio rerio*	
Dr15_WGA1175_212813	*Danio rerio*	
Dr18_WGA1436_236332	*Danio rerio*	
Dr20_WGA1527_1598576	*Danio rerio*	
Dr20_WGA1527_1613261	*Danio rerio*	
Dr20_WGA1544_1446828	*Danio rerio*	
Dr20_WGA1544_1447363	*Danio rerio*	
Dr20_WGA1544_375272	*Danio rerio*	
Dr20_WGA1544_380308	*Danio rerio*	
Dr20_WGA1544_382837	*Danio rerio*	
Dr20_WGA1544_386533	*Danio rerio*	
Dr20_WGA1544_401369	*Danio rerio*	
Dr22_WGA1631_2837868	*Danio rerio*	
Dr24_WGA1806_416955	*Danio rerio*	
Dr24_WGA1807_108854	*Danio rerio*	
Dr24_WGA1807_120319	*Danio rerio*	
Dr25_WGA1872_141073	*Danio rerio*	
Dr25_WGA1872_176060	*Danio rerio*	
Dr25_WGA1872_17999	*Danio rerio*	
Dr25_WGA1872_197377	*Danio rerio*	
Dr25_WGA1872_20217	*Danio rerio*	
Dr25_WGA1872_26142	*Danio rerio*	
Dr25_WGA1872_90731	*Danio rerio*	
Dr25_WGA1873_709505	*Danio rerio*	
Dr25_WGA1873_716771	*Danio rerio*	
Dr25_WGA1873_857912	*Danio rerio*	
DrUn_WGA13047_39000	*Danio rerio*	
DrUn_WGA20664_6549	*Danio rerio*	
DrUn_WGA2406_18581	*Danio rerio*	
DrUn_WGA6170_11888	*Danio rerio*	

Table 1 (continued)

Teleost CC chemokine	Species	Reference
DrUn_WGA6170_12898	*Danio rerio*	
DrUn_WGA7827_17321	*Danio rerio*	
DrUn_WGA7827_7256	*Danio rerio*	
DrUn_WGA7827_9830	*Danio rerio*	
SalmoBG935738	*Salmo salar*	NA
SalmoCO471983	*Salmo salar*	
SalmoDW566039	*Salmo salar*	
SalmoDY692162	*Salmo salar*	
SalmoDY704818	*Salmo salar*	
SalmoDY725280	*Salmo salar*	
SalmoDY728991	*Salmo salar*	
SalmoDY730515	*Salmo salar*	
SalmoEG760122	*Salmo salar*	
SalmoEG766286	*Salmo salar*	
SalmoEG788483	*Salmo salar*	
SalmoEG794131	*Salmo salar*	
SalmoEG810240	*Salmo salar*	
SalmoEG816357	*Salmo salar*	
SalmoEG818960	*Salmo salar*	
SalmoEG823993	*Salmo salar*	
SalmoEG835932	*Salmo salar*	
SalmoEG837555	*Salmo salar*	
SalmoEG840880	*Salmo salar*	
SalmoEG851286	*Salmo salar*	
SalmoEG856447	*Salmo salar*	
SalmoEG861420	*Salmo salar*	
SalmoEG865207	*Salmo salar*	
SalmoEG867584	*Salmo salar*	
SalmoEG873956	*Salmo salar*	
SalmoEG874392	*Salmo salar*	
SalmoEG876131	*Salmo salar*	
SalmoEG879192	*Salmo salar*	
SalmoEG930049	*Salmo salar*	
SalmoEG940598	*Salmo salar*	
SCYA101	*Ictalurus punctatus*	He et al. (2004); Peatman et al. (2005); Bao et al. (2006)
SCYA102	*Ictalurus punctatus*	
SCYA103	*Ictalurus punctatus*	
SCYA104	*Ictalurus punctatus*	
SCYA105	*Ictalurus punctatus*	
SCYA106	*Ictalurus punctatus*	
SCYA107	*Ictalurus punctatus*	
SCYA108	*Ictalurus punctatus*	
SCYA109	*Ictalurus punctatus*	
SCYA110	*Ictalurus punctatus*	
SCYA111	*Ictalurus punctatus*	
SCYA112	*Ictalurus punctatus*	
SCYA113	*Ictalurus punctatus*	
SCYA114A	*Ictalurus punctatus*	
SCYA114B	*Ictalurus punctatus*	
SCYA115	*Ictalurus punctatus*	
SCYA116	*Ictalurus punctatus*	
SCYA117	*Ictalurus punctatus*	
SCYA118	*Ictalurus punctatus*	
SCYA119	*Ictalurus punctatus*	
SCYA120A	*Ictalurus punctatus*	
SCYA120B	*Ictalurus punctatus*	

Table 1 (continued)

Teleost CC chemokine	Species	Reference
SCYA121	*Ictalurus punctatus*	
SCYA122	*Ictalurus punctatus*	
SCYA123	*Ictalurus punctatus*	
SCYA124	*Ictalurus punctatus*	
SCYA125	*Ictalurus punctatus*	
SCYA126	*Ictalurus punctatus*	

CCL1, CCL2, CCL7, CCL8, CCL11, and CCL13, were tightly clustered. CCL24, not a member of the MCP genomic cluster in mammals, also fell within this group, consistent with previous studies (Laing and Secombes 2004b; Kuroda et al. 2003). The MCP group contains a large number of *D. rerio* sequences that, due to the availability of the zebrafish genome sequences, provide additional information about chromosomal position. The two large *Danio* MCP subclades are located in tandem duplications on *Dr*20 and *Dr*25 (Fig. 1). Another *Danio* sequence is located on chromosome 15. The sequences comprising the cluster on chromosome 25 have low sequence homologies ($>10^{-3}$, *E* value) with mammalian CC chemokines, the exception being Dr25_WGA1872_176060, which is most like CCL2 from mouse (8×10^{-5}). It is possible that this locus represents the founder gene that translocated to this site from another location (possibly chromosome 15 or 20) and then underwent tandem gene duplications and divergence. The other *Danio* cluster contains CC chemokine sequences with the majority bearing stronger similarity to mammalian CCL2 and all highly similar to one another. With the exception of a sequence from chromosome 15, all these zebrafish CC chemokines are in the same genomic contig within chromosome 20. Interestingly, while this zebrafish cluster has no clear teleost orthologs among the sequences used, the *Dr*25 MCP cluster appeared to be related to a divergent group of catfish CC chemokines. These catfish CC chemokines, by hybridization to BAC clones and fingerprinting analysis, were found to be in close proximity to each other in the catfish genome (Peatman et al. 2006). In the middle of the MCP group are salmon and trout sequences that share strongest similarities with mammalian CCL13 sequences. These sequences are noteworthy in that they have no clear equivalents in zebrafish and catfish, and may, therefore, have originated after the divergence of the salmonids from lower teleost species.

The MIP group of CC chemokines in Fig. 1 is also indicative of a complex pattern of duplication and divergence. Of the large mammalian MIP cluster, human CCL5 is phylogenetically closest to fish CC chemokines as previously noted (Laing and Secombes 2004b). Hughes et al. (2007) suggested that CCL5 was the prototypic MIP family chemokine based on his analysis of chicken CC chemokines. Catfish, zebrafish, trout, and salmon all have one or more members in the clade containing human CCL5, and the fish members share high bootstrapping support (89%). Additionally, a large catfish CC chemokine group falls within the MIP group of Fig. 1. The seven catfish CC chemokines share highest BLAST similarities to mammalian CCL3, CCL4, and CCL14 and, with the exception of SCYA105, are known to be clustered within the catfish genome (Peatman et al. 2006).

The CCL17/22 group is the most problematic of the seven CC chemokine groups. In mammals, CCL17 and CCL22 are clustered together on a different chromosome from the large MCP and MIP clusters. They are described as dual-function and share the same receptor, CCR4 (Zlotnik et al. 2006). Larger, species-specific subclades of teleost CC chemokines, seen in the MCP and MIP groups, are not present here. Additionally, the zebrafish loci are from several different chromosomes, and the catfish CC chemokines within this group are not known to be clustered in the catfish genome. These facts suggest that these CC chemokines may be distributed more broadly through teleost genomes, and while they have diverged from ancestral sequences, duplication rates have been slow. Teleost sequences within this group share low similarities with both MCP and MIP groups from mammals. Further functional analysis is needed to determine the evolutionary relationships between these CC chemokines and others.

The CCL19/21/25 group is the first of three groups showing higher evolutionary conservation between teleost fish and mammalian CC chemokines. The CCL19/21/25, CCL20, and CCL27/28 groups all contain homeostatic or dual-function CC chemokines from mammals that are either non-clustered or in mini-clusters. The number of CC chemokines from teleost fish in each of these groups is also notably lower than in the large cluster, "inflammatory" groups. Members of these groups in zebrafish are on different chromosomes from members in the highly duplicated MIP and MCP groups, as is the case in mammalian

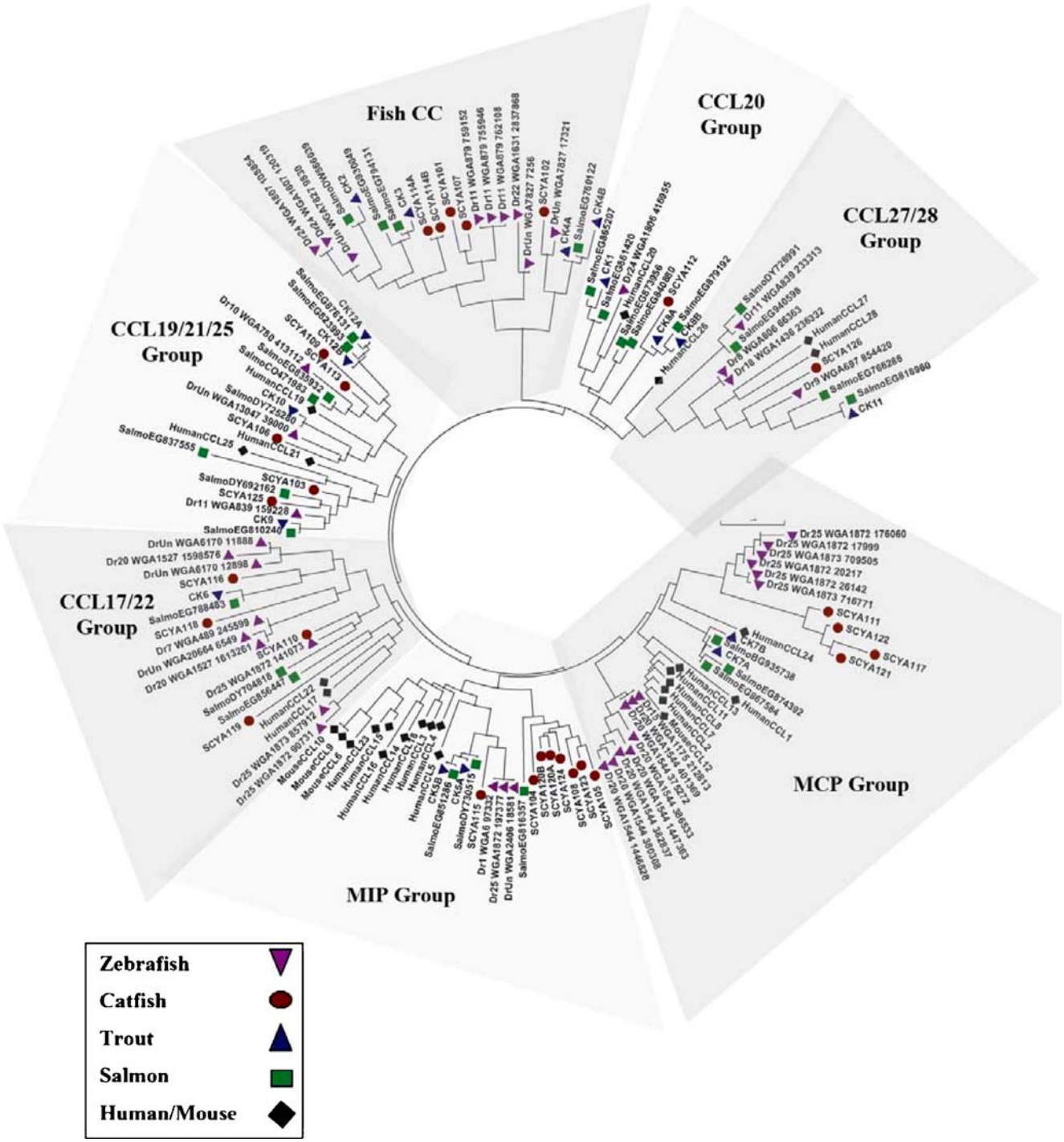

Fig. 1 Phylogenetic tree drawn from a ClustalW-generated multiple sequence alignment of CC chemokine amino acid sequences using the neighbor-joining method within the MEGA 3.0 package (Kumar et al. 2004). Data were analyzed using Poisson correction, and gaps were removed by pairwise deletion. Zebrafish, catfish, trout, salmon, and human/mouse sequences are denoted on the tree using the color and shape as shown in the *legend*. Zebrafish CC chemokine sequences were named based on genomic loci according to assembly Zv5, using chromosome number, contig number, and approximate start of coding sequences, as presented previously (Peatman and Liu 2006). Established nomenclature for catfish (SCYA–) and trout (CK–) CC chemokines was used in this study. Atlantic salmon putative CC chemokines were named using *Salmo* followed by the EST accession number

species. These facts suggest that these chemokine ligands are critical for normal physiological functions within each organism and are, therefore, under considerably higher evolutionary pressure to maintain sequence and structure. They have been maintained at discrete loci away from the recombination hotspots of the inflammatory CC chemokines. In the CCL19/21/25 group, the studied teleost fish have 1–3 genes most closely related to CCL25, with lower similarity to CCL21. The branching pattern indicates that ancient vertebrates had multiple CCL25/21-like genes before the diversification of teleost fish species. The CCL19 subclade in this group is particularly well established with fairly high bootstrapping support (47%). Here too, it appears that at least three ancient CCL19 forms existed before teleost diversification and have been maintained in fish (Fig. 1) with limited levels of gene duplication and loss (perhaps related to salmonid tetraploidization) following the divergence of the teleost species. Interestingly, while only one copy of CCL19 exists in humans, mouse has three pseudogene copies of CCL19 copies in addition to its functional copy (Zlotnik et al. 2006). Fingerprinting indicated that multiple SCYA106–SCYA109 copies (both most similar to CCL19) are together in multiple locations in the catfish genome (Peatman et al. 2006).

The CCL20 group is the smallest of those presented in this paper. Its teleost members consistently clade with human CCL20 regardless of tree type or settings. In both human and mouse, CCL20 is non-clustered, indicating evolutionary pressure for conservation. Only a single gene corresponding to CCL20 was found from zebrafish and catfish, whereas expansion of CCL20-like genes appears to have occurred in both rainbow trout (three genes) and salmon (five genes).

Teleost membership in the CCL27/28 group is also stable regardless of the algorithm used for tree construction. CCL27 and CCL28 in humans share the same receptor, CCR10. At least two gene copies of CCL27/28-like molecules appear to have existed in primitive fish. Four copies were detected from zebrafish, each on a different chromosome. Four copies were also identified from salmon. Although only a single CCL27/28 gene has been reported from catfish and trout, this may simply be a reflection of the smaller EST collections of these two species at present.

The last major CC chemokine group identified by phylogenetic analysis is named in this paper the "Fish CC" chemokine group. Recognized early in the analysis of fish CC chemokines (Laing and Secombes 2004b; He et al. 2004), this group contains fish CC chemokines that share negligible similarities with any mammalian CC chemokines. Bootstrapping support of the clade is extremely high (92%) when compared with similarly sized clades of CC chemokines from several teleost species. At least two ancestral subclades are apparent in this group, one larger with evidence of multiple species-specific duplication events after divergence and a smaller one (Fig. 1). The "Fish CC" group may represent a subset of ancestral CC chemokines that still carry on important functional roles in teleost fish and have, therefore, been maintained in fish genomes. Present expression evidence indicates that these genes may carry on homeostatic roles, as they are constitutively expressed in both trout and catfish (Laing and Secombes 2004b; Peatman et al. 2006). In higher vertebrates, these homeostatic roles may have become irrelevant or have been assumed by other CC chemokines, reducing evolutionary pressure for the maintenance of these genes.

Present shortcomings and potential future solutions

Given the difficulties associated with phylogenetic analysis of the CC chemokine family, other avenues of comparison between the CC chemokines of different species are needed. One potential method is the analysis of CC chemokine receptors that are believed to be more conserved than their ligands. However, a search of the zebrafish genome (DeVries et al. 2006) found that zebrafish lacks clear orthologs of CC receptors 1, 2, 3, 5, 8, and 10. This list encompasses all the mammalian MCP and MIP ligand receptors, the two large groups of CC chemokines for which orthologs are most difficult to determine. Expression analysis offers the potential of dividing large teleost CC chemokine families into homeostatic and inducible categories. This has been conducted with mixed results to date in trout and catfish (Laing and Secombes 2004b; Peatman et al. 2006). However, the earlier divisions between these two groups in mammals are rapidly breaking down as chemokine expression is tested in a wider variety of cells, tissues, and disease conditions. Currently, researchers of teleost fish are unable to test expression patterns for such large gene families in more than a handful of tissues and cell types due to funding limitations. For example, a teleost CC chemokine that appears homeostatic based on the lack of inducibility in the head kidney and spleen, may subsequently be found to have an inflammatory function in the skin or thymus.

Genomic localization currently holds the greatest potential for comparing teleost CC chemokine datasets with each other and with those from mammalian species. This too is currently limited to the handful of fish with assembled genome sequences or to labs with physical mapping resources (Bao et al. 2006; Peatman et al. 2006). Attempts in this study to trace CC chemokine gene radiation from ancestral proto-chromosomes through homologous regions of zebrafish and *Tetraodon* chromosomes to their mammalian chromosomal counterparts were unsuccessful (Woods

et al. 2005). There was no evidence of large patterns of conserved synteny indicating block duplications between CC chemokine loci of zebrafish, *Tetraodon*, and human. Nevertheless, as physical maps are constructed and/or genomes of more teleost species are sequenced, genome loci-anchored comparisons between more closely related species should provide valuable information on duplication and divergence rates and may allow the identification of an ancestral chemokine gene in fish. In closely related species, a high level of syntenic conservation is expected.

In the present analysis, we used a mixed dataset of unanchored ESTs from trout and salmon, BAC-contig anchored catfish genes, and genome-anchored zebrafish chemokine loci. The zebrafish data provides a crucial "complete" picture from a single species, as well as allowing tandem duplications and gene clusters to be identified, although its genome sequence assembly is not finalized, and genome annotation is far from being complete. It is likely that a number of additional CC chemokines will be discovered in the future from trout, salmon, and catfish through further EST sequencing and/or genome sequencing. On the other hand, a number of CC chemokine loci identified from zebrafish had no apparent matching transcript among the zebrafish ESTs, raising questions as to whether all the duplicated loci are expressed. Identifying potential CC chemokine paralogs from EST sets can also be problematic, as highly similar EST sequences encoding for identical proteins can represent duplicated genes within the genome (Peatman et al. 2006). Therefore, a conservative approach to identifying unique CC chemokine transcripts from ESTs is likely to underestimate the number of CC chemokine genes within a given species' genome, while tracing remnants of chemokine-related sequences in genome sequences could achieve the opposite, overestimated number of chemokine genes. It is, therefore, important to generate sufficient EST resources to support genome annotations.

Naming of non-mammalian CC chemokines continues to be an unresolved issue. Catfish CC chemokines (Peatman et al. 2005) were annotated based on a teleost-naming system proposed by Kuroda et al. (2003). Under this system, CC chemokines are named by the species name and SCYA101–SCYA1XX depending on the number reported. This system postpones assigning mammalian chemokine names to fish CC chemokines until further concrete information is accumulated, or orthologies are established. Trout CC chemokines continue to use an older system of CK1-CKXX. Most zebrafish CC chemokines were previously unannotated in genome assembly Zv5, and we, therefore, identified them based on their genomic locations (Peatman and Liu 2006). For continuity, we used the same names in this study. The current zebrafish assembly, Zv6, has automatically assigned many of these sequences with names based on sequence similarities that fail to reflect their duplicated nature. DeVries et al. (2006) in a survey of chemokine sequences named zebrafish chemokines based on their closest human equivalent, which may or may not survive the orthology test with time. Similar issues have arisen in the naming of chicken CC chemokines (Hughes et al. 2007). Determining identities of novel teleost CC chemokines by BLAST is wrought with uncertainty given the low sequence conservation levels and widespread gene duplication. Only when all members of CC chemokines are completely identified, may a mutual best hit approach be applicable. We propose again in this paper that fish CC chemokines be named using a nomenclature distinct from that of mammalian CC chemokines until further genomic and functional information is available. Furthermore, an international conference or workshop should be held in the near future to develop a shared nomenclature for fish cytokines.

Chemokine gene variation and immune diversity

While the tremendous diversity of CC chemokine ligands in teleost fish has been established, little is currently known about the contributions of these molecules to immune and physiological functions. From a smaller, shared ancestral set, teleost species have evolved larger, unique repertoires of CC chemokines, likely through radiated tandem duplication. While this phenomenon makes it difficult to trace evolutionary history, it also raises exciting questions as to how different CC chemokine sets may coordinate specific immune responses for each species depending on aquatic environment. Do the tandem duplications of CC chemokine loci allow customization of inflammatory responses based on new pathogen threats to a given species (Zlotnik et al. 2006)? In humans, chemokine gene sequence and copy number variations are linked to different immune susceptibilities between individuals and between populations. For example, the number of duplicated copy numbers of human CCL3 and CCL4 an individual possesses is inversely correlated with HIV/AIDS susceptibility (Nibbs et al. 1999; Gonzalez et al. 2005; Modi et al. 2006). It is of great interest whether similar chemokine gene variations among and within teleost species similarly contribute to phenotypic differences in disease response. As single nucleotide polymorphism (SNP) assays are implemented in several aquaculture species in the near future to search for disease-linked quantitative trait loci (QTL), the contributions of chemokine loci to teleost immune responses may become clearer. Large-scale functional and gene-profiling experiments in teleost fish have already begun to provide additional information regarding teleost CC chemokine expression patterns (Gonzalez et al. 2007; Sanchez

et al. 2007; Peatman et al. 2007), and this trend should accelerate as genomic tools become more widely available to researchers. This information will help to evaluate the similarities and differences between the chemokine repertoires of different teleost species and aid in the study of how they have adapted to different pathogen environments.

Conclusions

The recent availability of large CC chemokine datasets from a range of teleost fish species has allowed new insights into the evolution and diversity of this important gene family. Levels of diversity of CC chemokine ligands appear to vary between teleost fish species, but in many cases, may be greater than those seen in higher vertebrates. Correlations between CC chemokine genomic location and evolutionary patterns, first observed in mammals, appear to also exist in teleost fish, with concrete evidence existing from zebrafish and catfish for species-specific tandem gene duplications occurring at clustered loci. Current expression evidence is insufficient to conclusively connect CC chemokine genomic environs (clustered vs non-clustered) with inflammatory or homeostatic function. However, the majority of clustered, duplicated teleost CC chemokines share closer phylogenetic relationships with inflammatory CC chemokines of mammals, indicating that these genes may represent novel immune adaptations. Establishing one-to-one orthologous relationships between teleost and mammalian inflammatory CC chemokines will remain largely unsuccessful due to their rapid, ongoing evolution. On the other hand, much higher levels of sequence conservation among "homeostatic" CC chemokines have been maintained from fish to higher vertebrates, likely reflecting a shared need for these genes for mediating normal physiological functions. Further sequencing, functional and expression-based experiments in teleost fish should clarify the conserved and novel roles of this diverse gene family.

Acknowledgment Our research is supported by USDA CSREES grants under the National Research Initiative Animal Genome Program and NOAA Sea Grant Gulf Oyster Industry Program. We appreciate the efforts of the Aquaculture Genome research community for their submission of ESTs into GenBank that made this work possible.

References

Bacon K, Baggiolini M, Broxmeyer H, Horuk R, Lindley I, Mantovani A, Matsushima K, Murphy P, Nomiyama H, Oppenheim J, Rot A, Schall T, Tsang M, Thorpe R, Van Damme J, Wadhwa M, Yoshie O, Zlotnik A, Zoon K (2003) Chemokine/chemokine receptor nomenclature. Cytokine 21:48–49

Bao B, Peatman E, Peng X, Baoprasertkul P, Wang G, Liu Z (2006) Characterization of 23 CC chemokine genes and analysis of their expression in channel catfish (*Ictalurus punctatus*). Dev Comp Immunol 30:783–796

Davey GC, Caplice NC, Martin SA, Powell R (2001) A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags. Gene 263:121–130

DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ (2006) Defining the origins and evolution of the chemokine/chemokine receptor system. J Immunol 176:401–415

Dixon B, Shum B, Adams EJ, Magor KE, Hedrick RP, Muir DG, Parham P (1998) CK-1, a putative chemokine of rainbow trout (*Oncorhynchus mykiss*). Immunol Rev 166:341–348

Fujiki K, Shin DH, Nakao M, Yano T (1999) Molecular cloning of carp (*Cyprinus carpio*) CC chemokine, CXC chemokine receptors, allograft inflammatory factor-1, and natural killer cell enhancing factor by use of suppression subtractive hybridization. Immunogenetics 49:909–914

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434–1440

Gonzalez SF, Chatziandreou N, Nielsen ME, Li W, Rogers J, Taylor R, Santos Y, Cossins A (2007) Cutaneous immune responses in the common carp detected using transcript analysis. Mol Immunol 44:1664–1679

Govoroun M, Le Gac F, Guiguen Y (2006) Generation of a large scale repertoire of Expressed Sequence Tags (ESTs) from normalised rainbow trout cDNA libraries. BMC Genomics 7:196

Hagen-Larsen H, Laerdahl JK, Panitz F, Adzhubei A, Hoyheim B (2005) An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (*Salmo salar*). BMC Genomics 6:171

He C, Peatman E, Baoprasertkul P, Kucuktas H, Liu Z (2004) Multiple CC chemokines in channel catfish and blue catfish as revealed by analysis of expressed sequence tags. Immunogenetics 56:379–387

Hughes S, Poh TY, Bumstead N, Kaiser P (2007) Re-evaluation of the chicken MIP family of chemokines and their receptors suggests that CCL5 is the prototypic MIP family chemokine, and that different species have developed different repertoires of both the CC chemokines and their receptors. Dev Comp Immunol 31:72–86

Khattiya R, Ohira T, Hirono I, Aoki T (2004) Identification of a novel Japanese flounder (*Paralichthys olivaceus*) CC chemokine gene and an analysis of its function. Immunogenetics 55:763–769

Kono T, Kusada R, Kawahara E, Sakai M (2003) The analysis of immune responses of a novel CC-chemokine gene from Japanese flounder *Paralichthys olivaceus*. Vaccine 21:446–457

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150–163

Kuroda N, Uinuk-ool TS, Sato A, Samonte IE, Figueroa F, Mayer WE, Klein J (2003) Identification of chemokines and a chemokine receptor in cichlid fish, shark, and lamprey. Immunogenetics 54:884–895

Laing KJ, Secombes CJ (2004a) Chemokines. Dev Comp Immunol 28:443–460

Laing KJ, Secombes CJ (2004b) Trout CC chemokines: comparison of their sequences and expression patterns. Mol Immunol 41:793–808

Li P, Peatman E, Wang S, Feng J, He C, Baoprasertkul P, Xu P, Kucuktas H, Nandi S, Somridhivej B, Serapion J, Simmons M, Turan C, Liu L, Muir W, Dunham R, Brady Y, Grizzle J, Liu Z (2007) Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. BMC Genomics (in press)

Liu Z, Peatman E (2006) Chemokines in fish: a rapidly expanding repertoire. In: Veskler BA (ed) Focus on immunology research. Nova, NY, pp 121–144

Liu L, Fujiki K, Dixon B, Sundick RS (2002) Cloning of a novel rainbow trout (*Oncorhynchus mykiss*) CC chemokine with a fractalkine-like stalk and a TNF decoy receptor using cDNA fragments containing Au-rich elements. Cytokine 17:71–81

Lo J, Lee S, Xu M, Liu F, Ruan H, Eun A, He Y, Ma W, Wang W, Wen Z et al (2003) 15000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. Genome Res 13:455–466

Mackenzie S, Liarte C, Iliev D, Planas JV, Tort L, Goetz FW (2004) Characterization of a highly inducible novel CC chemokine from differentiated rainbow trout (*Oncorhynchus mykiss*) macrophages. Immunogenetics 56:611–615

Martin SA, Caplice NC, Davey GC, Powell R (2002) EST-based identification of genes expressed in the liver of adult Atlantic salmon (*Salmo salar*). Biochem Biophys Res Commun 293:578–585

Modi WS, Lautenberger J, An P, Scott K, Goedert JJ, Kirk GD, Buchbinder S, Phair J, Donfield S, O'Brien SJ, Winkler C (2006) Genetic variation in the CCL18-CCL3-CCL4 chemokine gene cluster influences HIV Type 1 transmission and AIDS disease progression. Am J Hum Genet 79:120–128

Moser B, Wolf M, Walz A, Loetscher P (2004) Chemokines: multiple levels of leukocyte migration control. Trends Immunol 25:75–85

Nibbs RJ, Yang J, Landau NR, Mao JH, Graham GJ (1999) LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity. J Biol Chem 274:17478–17483

Peatman E, Liu Z (2006) CC chemokines in zebrafish: evidence for extensive intrachromosomal gene duplications. Genomics 88:381–385

Peatman E, Bao B, Baoprasertkul P, Liu Z (2005) In silico identification and expression analysis of 12 novel CC chemokines in catfish. Immunogenetics 57:409–419

Peatman E, Bao B, Peng X, Baoprasertkul P, Brady Y, Liu Z (2006) Catfish CC chemokines: genomic clustering, duplications, and expression after bacterial infection with *Edwardsiella ictaluri.* Mol Genet Genomics 275:297–309

Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang S, Somridhivej B, Dunham R, Liu Z (2007) Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram negative bacterium. Dev Comp Immunol. DOI 10.1016/j.dci.2007.03.003

Rexroad CE, 3rd, Lee Y, Keele JW, Karamycheva S, Brown G, Koop B, Gahr SA, Palti Y, Quackenbush J (2003) Sequence analysis of a rainbow trout cDNA library and creation of a gene index. Cytogenet Genome Res 102:347–354

Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR et al (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. Genome Res 14:478–490

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sanchez E, Coll J, Tafalla C (2007) Expression of inducible CC chemokines in rainbow trout (*Oncorhynchus mykiss*) in response to a viral haemorrhagic septicemia virus (VHSV) DNA vaccine and interleukin 8. Dev Comp Immunol 31(9):916-926

Shields DC (2003) Molecular evolution of CXC chemokines and receptors. Trends Immunol 24:355

Song HD, Sun XJ, Deng M, Zhang GW, Zhou Y, Wu XY, Sheng Y, Chen Y, Ruan Z, Jiang CL et al (2004) Hematopoietic gene expression profile in zebrafish kidney marrow. Proc Natl Acad Sci USA 101:16240–16245

von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF (2005) Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. BMC Genomics 6:126

Wang J, Adelson DL, Yilmaz A, Sze SH, Jin Y, Zhu JJ (2005) Genomic organization, annotation, and ligand-receptor inferences of chicken chemokines and chemokine receptor genes based on comparative genomics. BMC Genomics 6:45

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS (2005) The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res 15:1307–1314

Yoshie O, Imai T, Nomiyama H (2001) Chemokines in immunity. Adv Immunol 78:57–110

Zeng S, Gong Z (2002) Expressed sequence tag analysis of expression profiles of zebrafish testis and ovary. Gene 294:45–53

Zlotnik A, Yoshie O, Nomiyama H (2006) The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome Biol 7:243