Available online at www.sciencedirect.com
ScienceDirect
Genomics 88 (2006) 381–385
GENOMICS
www.elsevier.com/locate/ygeno

Short Communication

CC chemokines in zebrafish: Evidence for extensive intrachromosomal gene duplications

Eric Peatman, Zhanjiang Liu *

Fish Molecular Genetics and Biotechnology Laboratory, 203 Swingle Hall, Department of Fisheries and Allied Aquacultures, and Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, AL 36849, USA

Received 12 January 2006; accepted 23 March 2006
Available online 12 May 2006

Abstract

Chemokines are a family of structurally related chemotactic cytokines that regulate the migration of leukocytes. CC chemokines represent the largest subfamily of chemokines, with 28 genes in mammals. In recent studies in channel catfish, *Ictalurus punctatus*, we identified 26 distinct CC chemokine transcripts and obtained the genomic sequences and structures of 23 CC chemokine genes. However, without the availability of similar sets of CC chemokines in closely related species or a sequenced genome in catfish, it was difficult to make inferences as to the origins and modes of duplication of these molecules or to analyze conserved synteny between teleost and mammalian CC chemokines. Here, we have identified as many as 46 loci in the zebrafish genome that encode putative CC chemokines. The zebrafish CC chemokines are highly clustered on several chromosomes and show evidence of extensive, species-specific intrachromosomal duplications.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Chemokine; Cytokine; Immunity; Fish; Genome duplication; Zebrafish

Chemokines are crucial to host defense, conducting immunosurveillance under homeostasis as well as stimulating the recruitment, activation, and adhesion of cells to sites of infection or injury [1–3]. They are structurally related small peptides, with the majority containing four conserved cysteine residues. Based on the arrangement of these conserved cysteine residues [4], chemokines were divided into four subfamilies: CXC (α), CC (β), C, and CX3C. CC chemokines constitute the largest subfamily of chemokines with 28 CC chemokines identified from mammalian species [5]. The largest number of CC chemokines found in a single species is 24 from humans, missing orthologues to the murine CCL6, CCL9/CCL10, and CCL12.

Fish represent a transition point on the phylogenetic spectrum between species possessing only innate immunity (i.e., invertebrates) and species depending heavily on adaptive immunity (i.e., mammals). Identification of CC chemokine orthologues from fish, therefore, is an important step in analyzing the ancient underpinnings of immunity and the inflammatory response. Progress toward this goal has been slow, hampered by low sequence conservation and rapid intraspecies duplication and divergence [6]. We previously identified 26 distinct CC chemokine transcripts from catfish and, in subsequent studies, obtained the genomic sequences and structures of 23 CC chemokine genes [6–9]. Fingerprinting analysis of the CC chemokine-containing BAC clones suggested that the catfish CC chemokine genes were extensively clustered within the genome and highly duplicated, with the majority present in three or more genomic copies [8]. The presence of such a large and diverse family of CC chemokines in catfish challenged our expectations as to the sophistication of the innate immune system in lower vertebrates. However, without the availability of similar sets of CC chemokines in closely related species, it was difficult to make inferences as to the origin and modes of duplication of these molecules or to predict whether similarly large chemokine families were present across the teleost radiation. We turned, therefore, to the model species zebrafish, *Danio rerio*, to gain a better understanding of the diversity and origins of CC chemokines. Here, we conducted extensive analysis using EST resources as well as the draft genome sequence of zebrafish and report the

* Corresponding author. Fax: +1 334 844 9208.
E-mail address: zliu@acesag.auburn.edu (Z. Liu).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.03.014

Table 1
Genomic locations of the putative CC chemokines of zebrafish, given by contig and approximate location within the contig

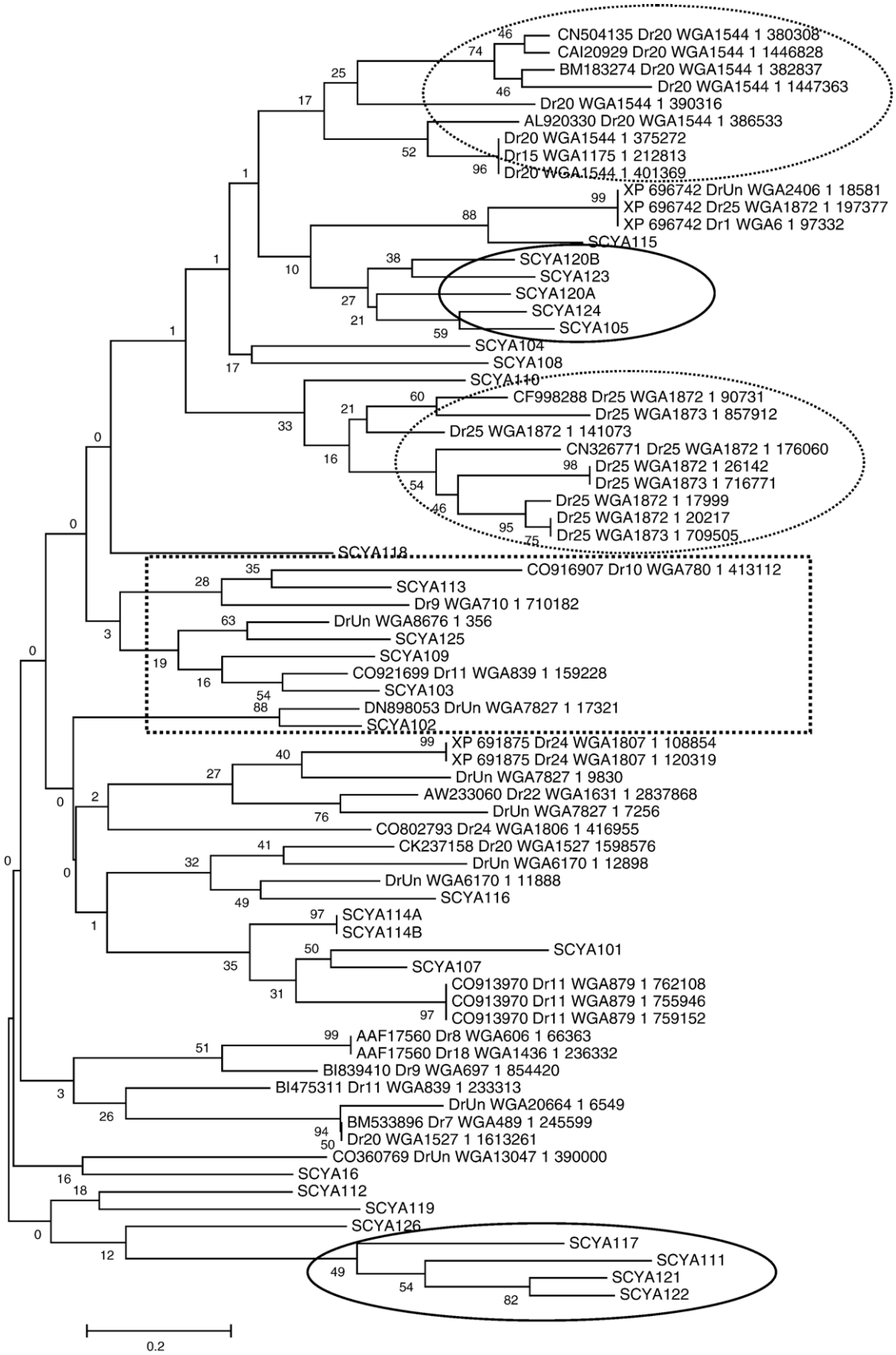
Genomic contig	Location within contig	Matching query or no transcript (NT)	BLAST ID	*E* value
Dr1_WGA6_1	97332–95874	XP_696742	CCL3 (*Sus scrofa*)	4×10^{-11}
Dr7_WGA489_1	245599–246124	BM533896	CCL12 (*Rattus norvegicus*)	5×10^{-6}
Dr8_WGA606_1	66363–66292	AAF17560	CCL27 (*Bos taurus*)	4×10^{-5}
Dr9_WGA697_1	854420–854063	BI839410	CCL28 (*R. norvegicus*)	2×10^{-6}
Dr9_WGA710_1	710182–710307	NT	CC chemokine SCYA109 (*Ictalurus furcatus*)	3×10^{-5}
Dr10_WGA780_1	413112–413041	CO916907	CCL19 (*Mus musculus*)	8×10^{-9}
Dr11_WGA839_1	159228–159595	CO921699	CCL25 (*S. scrofa*)	3×10^{-9}
Dr11_WGA839_1	233313–234458	BI475311	CCL21 (*Homo sapiens*)	0.10
Dr11_WGA879_1	755946–751957 759152–758731 762108–761868	CO913970	CC chemokine SCYA101 (*I. furcatus*)	1×10^{-19}
Dr15_WGA1175_1	212813–213121	NT	CCL2 (*M. musculus*)	1×10^{-8}
Dr18_WGA1436_1	236332–229356	AAF17560	CCL27 (*B. taurus*)	4×10^{-5}
Dr20_WGA1527_1	1598576–1597254	CK237158	CCL20 (*Canis familiaris*)	3×10^{-5}
Dr20_WGA1527_1	1613261–1612801	NT	CCL12 (*R. norvegicus*)	5×10^{-6}
Dr20_WGA1544_1	375272–375580	NT	CCL2 (*M. musculus*)	1×10^{-8}
Dr20_WGA1544_1	380308–380913	CN504135	CCL8 (*S. scrofa*)	6×10^{-10}
Dr20_WGA1544_1	382837–383483	BM183274	CCL2 (*Sigmodon hispidus*)	2×10^{-13}
Dr20_WGA1544_1	386533–386601	AL920330	CCL2 (*Si. hispidus*)	7×10^{-9}
Dr20_WGA1544_1	390316–390677	NT	CCL8 (*B. taurus*)	3×10^{-8}
Dr20_WGA1544_1	401369–401061	NT	CCL2 (*M. musculus*)	1×10^{-8}
Dr20_WGA1544_1	1446828–1446504	CAI20929	CCL2 (*Si. hispidus*)	1×10^{-9}
Dr20_WGA1544_1	1447363–1447069	NT	Novel protein (*Danio rerio*)	8×10^{-9}
Dr22_WGA1631_1	2837868–2838293	AW233060	CCL20 (*M. musculus*)	0.016
Dr24_WGA1806_1	416955–414904	CO802793	CCL20 (*H. sapiens*)	8×10^{-8}
Dr24_WGA1807_1	108854–109580 120319–120459	XP_691875	CCL20 (*M. musculus*)	0.003
Dr25_WGA1872_1	17999–18296	NT	CCL24 (*H. sapiens*)	0.001
Dr25_WGA1872_1	20217–20499	NT	CCL3 (*H. sapiens*)	0.005
Dr25_WGA1872_1	26142–26514	NT	CCL13 (*H. sapiens*)	0.036
Dr25_WGA1872_1	90731–90087	CF998288	CCL5 (*Felis catus*)	2×10^{-6}
Dr25_WGA1872_1	141073–140787	NT	CCL13 (*C. familiaris*)	0.001
Dr25_WGA1872_1	176060–174162	CN326771	CCL2 (*M. musculus*)	8×10^{-5}
Dr25_WGA1872_1	197377–195807	XP_696742	CCL3 (*S. scrofa*)	4×10^{-11}
Dr25_WGA1873_1	709505–709802	NT	CCL24 (*H. sapiens*)	0.001
Dr25_WGA1873_1	716771–717167	NT	CCL2 (*R. norvegicus*)	0.023
Dr25_WGA1873_1	857912–857589	NT	CCL22 (*B. taurus*)	2×10^{-6}
DrUn_WGA2406_1	18581–18474	XP_696742	CCL3 (*S. scrofa*)	4×10^{-11}
DrUn_WGA6170_1	11888–12241	NT	CCL20 (*M. musculus*)	6×10^{-5}
DrUn_WGA6170_1	12898–13263	NT	CCL20 (*M. musculus*)	4×10^{-5}
DrUn_WGA7827_1	7256–7006	NT	CC-1 (*D. rerio*)	6×10^{-12}
DrUn_WGA7827_1	9830–9405	NT	CC-1 (*D. rerio*)	2×10^{-22}
DrUn_WGA7827_1	17321–16477	DN898053	CCL21(*B. taurus*)	0.001
DrUn_WGA8676_1	356–520	NT	CCL25 (*S. scrofa*)	2×10^{-4}
DrUn_WGA13047_1	39000–38881	CO360769	CCL19 (*M. musculus*)	8×10^{-9}
DrUn_WGA20664_1	6549–7045	NT	SCYA105 (*P. chilotes*)	0.055

If the sequences at the genomic location matched a *Danio* transcript perfectly (or with a few SNPs), that transcript's GenBank accession number is given. A cutoff of 0.10 was imposed on BLASTp searches using the translated coding sequences to allow identification of the short, divergent sequences. If no mammalian hits qualified, the top fish sequence was taken.

identification of 46 putative CC chemokine genes. Zebrafish, like catfish, shows evidence of extensive intrachromosomal duplications of the CC chemokine genes.

Initial mining of the dbEST database by tBLASTn searches using existing fish CC chemokines as queries resulted in the identification of 102 zebrafish ESTs as putative CC chemokine

Fig. 1. Phylogenetic tree drawn from a ClustalW-generated multiple sequence alignment of amino acid sequences using the neighbor-joining method within the MEGA (3.0) package. Data were analyzed using Poisson correction and gaps were removed by complete deletion. The topological stability of the neighbor-joining trees was evaluated by 1000 bootstrapping replications, and the bootstrapping values are indicated by numbers at the nodes. Zebrafish sequence names used their GenBank accession number (where available), the chromosome and contig to which they were localized, and the approximate location within the contig of the start of the coding sequences. Catfish CC chemokines were named SCYA 101–126 as previously described [6,7]. Circles indicate examples of large clades of likely duplicated catfish (solid circles) and zebrafish (dotted circles) CC chemokines that do not include sequences from the other species. The dotted box indicates several smaller clades that include both catfish and zebrafish chemokines as discussed in the text.


CN504135 Dr20 WGA1544 1 380308
CAI20929 Dr20 WGA1544 1 1446828
BM183274 Dr20 WGA1544 1 382837
Dr20 WGA1544 1 1447363
Dr20 WGA1544 1 390316
AL920330 Dr20 WGA1544 1 386533
Dr20 WGA1544 1 375272
Dr15 WGA1175 1 212813
Dr20 WGA1544 1 401369
XP 696742 DrUn WGA2406 1 18581
XP 696742 Dr25 WGA1872 1 197377
XP 696742 Dr1 WGA6 1 97332
SCYA115
SCYA120B
SCYA123
SCYA120A
SCYA124
SCYA105
SCYA104
SCYA108
SCYA110
CF998288 Dr25 WGA1872 1 90731
Dr25 WGA1873 1 857912
Dr25 WGA1872 1 141073
CN326771 Dr25 WGA1872 1 176060
Dr25 WGA1872 1 26142
Dr25 WGA1873 1 716771
Dr25 WGA1872 1 17999
Dr25 WGA1872 1 20217
Dr25 WGA1873 1 709505
SCYA118
CO916907 Dr10 WGA780 1 413112
SCYA113
Dr9 WGA710 1 710182
DrUn WGA8676 1 356
SCYA125
SCYA109
CO921699 Dr11 WGA839 1 159228
SCYA103
DN898053 DrUn WGA7827 1 17321
SCYA102
XP 691875 Dr24 WGA1807 1 108854
XP 691875 Dr24 WGA1807 1 120319
DrUn WGA7827 1 9830
AW233060 Dr22 WGA1631 1 2837868
DrUn WGA7827 1 7256
CO802793 Dr24 WGA1806 1 416955
CK237158 Dr20 WGA1527 1598576
DrUn WGA6170 1 12898
DrUn WGA6170 1 11888
SCYA116
SCYA114A
SCYA114B
SCYA101
SCYA107
CO913970 Dr11 WGA879 1 762108
CO913970 Dr11 WGA879 1 755946
CO913970 Dr11 WGA879 1 759152
AAF17560 Dr8 WGA606 1 66363
AAF17560 Dr18 WGA1436 1 236332
BI839410 Dr9 WGA697 1 854420
BI475311 Dr11 WGA839 1 233313
DrUn WGA20664 1 6549
BM533896 Dr7 WGA489 1 245599
Dr20 WGA1527 1 1613261
CO360769 DrUn WGA13047 1 390000
SCYA16
SCYA112
SCYA119
SCYA126
SCYA117
SCYA111
SCYA121
SCYA122
0.2

transcripts. Cluster analysis assigned these EST clones into 16 distinct clusters or singletons. The 16 sequences were confirmed to be members of the CC chemokine family by reciprocal BLASTX searches (Table 1) and by the presence of four conserved cysteine residues as previously described for catfish [6]. In this process of BLAST searches, four additional CC chemokine protein sequences that had already been annotated were identified, bringing the pool of zebrafish CC chemokines to 20.

BLAST searches were conducted to localize the putative CC chemokines on the zebrafish draft genome sequence. A total of 46 genomic locations were among the top hits of the BLAST searches, larger than twice the number of queries (Table 1). The CC chemokine queries hit on genomic locations within 12 chromosomes plus unassigned chromosomal segments. The in silico Southern analysis suggested that the zebrafish CC chemokines were highly clustered in a handful of chromosomes, with chromosomes 20 and 25 containing the largest number of CC chemokines, with 10 loci each.

Of the 46 genomic locations identified to code for putative CC chemokine genes, 26 had perfect matches with the original 20 queries. Six CC chemokine loci were duplicated copies. CO913970 had three perfect matches within a single contig, Dr11_WGA879_1, suggesting triplication of the CC chemokine gene at the locus. Similarly, XP_691875 had two perfect matches within a single contig, Dr24_WGA1807_1, also suggesting tandem gene duplication within this genomic environ. For the queries AAF17560 and XP_696742, their perfect matches were found on contigs belonging to different chromosomes, representing potentially duplicated chromosome regions or gene duplication on different chromosomes. In addition to the perfect matches, 20 genomic locations were identified to contain CC chemokine-like genes that appeared to encode additional novel CC chemokines for which expressed transcripts are not available in the dbEST database (Table 1).

A phylogenetic analysis was conducted using the putative zebrafish CC chemokine amino acid sequences in the context of their genomic locations (Fig. 1). A strong correlation between chromosomal locations and sequence similarity was apparent. The presence of clades of CC chemokines from the same or nearby genomic contigs was highly suggestive of intrachromosomal gene duplications. The most dramatic of these patterns was that of extensive intrachromosomal duplications on zebrafish chromosomes 20 and 25. These clusters of highly similar CC chemokine sequences account for close to half of all the genomic loci encoding zebrafish CC chemokines. Phylogenetic analysis of the combined zebrafish and catfish CC chemokines also revealed a clear pattern of rapid species-specific duplications after divergence (Fig. 1). The large CC chemokine clusters in zebrafish and catfish are exclusive of sequences of the other species, forming their own clades, characteristic of species-specific multiplication of CC chemokines in a local genomic environ. In contrast, the nonclustered, nonduplicated CC chemokines from each species often fall into the same clades, albeit with low bootstrapping support.

A prominent theory of genome evolution holds that two genome duplications occurred early in vertebrate history, followed by another whole genome duplication event in ray-finned fishes after the split from mammals and before the teleost radiation [10,11]. The most familiar example given to support this theory has been the comparison of *hox* gene complex numbers across vertebrate species. Genome-wide analyses of zebrafish and pufferfish have lent partial support to the whole genome duplication theory [12,13]. Smaller scale studies, however, continue to demonstrate the prevalence of independent, species-specific tandem duplications [14–19]. Gloriam and colleagues [18] described a particularly extensive expansion of the trace amine receptor family in zebrafish resulting from multiple intrachromosomal duplication events. Zebrafish was found to possess 57 such receptors compared to fewer than 10 from the teleost *Takifugu rubripes* and from humans.

Clearly, such marked expansion of gene families has a powerful impact on genome composition and complicates attempts to establish syntenic relationships between species within such families. The majority of the zebrafish CC chemokines identified in this study also appeared to be derived from local duplication events, and syntenic relationships with the human, mouse, and chicken chromosomal segments containing CC chemokines could not be established. BLASTX searches on 100-kb genomic regions surrounding the zebrafish CC chemokine loci failed to identify similar sets of genes existing on more than one chromosome and did not hit on genes orthologous to those found surrounding the CC chemokines in other species.

Duplication within the zebrafish CC chemokines has not been confined exclusively within chromosomes, however. Phylogenetic analysis (Fig. 1) and pair-wise comparisons of 20-kb regions surrounding the CC chemokine loci using NCBI's BLAST2seq program revealed that a small number of loci have highly similar or identical coding sequences on one or more chromosomes. In some cases, the region of similarity extended beyond the gene-encoding area into either repetitive elements or sequences with no similarity to known proteins. The prevalence of repetitive and/or transposable elements surrounding almost all the zebrafish CC chemokine loci was notable and may provide clues to the mechanisms by which this family has rapidly expanded [20,21]. On the whole, however, large-scale duplications of genomic loci, often associated with chromosomal duplication, were not observed. Similarly, sequence analysis of genomic contigs containing multiple duplicated copies of zebrafish CC chemokines found that repeat regions were largely confined to the genes themselves and did not extend far beyond the coding regions. These results suggested that the intrachromosomal duplications may have occurred in fairly short segments. Future efforts are needed to determine the mechanisms of CC chemokine diversification in zebrafish and whether neo- or subfunctionalization has allowed the maintenance of such a large number of duplicated genes.

Acknowledgments

This project was supported by National Research Initiative Grant 2006-35616-16685 from the USDA Cooperative State Research, Education, and Extension Service under the Animal

Genome Tools and Resources Program and in part by a Specific Cooperative Agreement with the USDA ARS Aquatic Animal Health Laboratory under Contract 58-6420-5-030.

References

[1] T.S. Olson, K. Ley, Chemokines and chemokine receptors in leukocyte trafficking, Am. J. Physiol. Regul. Integr. Comp. Physiol. 283 (2002) R7–R28.

[2] B. Moser, P. Loetscher, Lymphocyte traffic control by chemokines, Nat. Immunol. 2 (2001) 123–128.

[3] K.J. Laing, C.J. Secombes, Chemokines, Dev. Comp. Immunol 28 (2004) 443–460.

[4] P.M. Murphy, M. Baggiolini, I.F. Charo, C.A. Hebert, R. Horuk, K. Matsushima, L.H. Miller, J.J. Oppenheim, C.A. Power, International Union of Pharmacology, XXII. Nomenclature for chemokine receptors, Pharmacol. Rev. 52 (2000) 145–176.

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

[6] E. Peatman, P. Baoprasertkul, C. He, P. Li, H. Kucuktas, Z. Liu, *In silico* identification and expression analysis of 12 novel CC chemokines in catfish, Immunogenetics 57 (2005) 409–419.

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

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

[9] B. Bao, E. Peatman, P. Xu, P. Baoprasertkul, G. Wang, Z. Liu, Characterization of 23 CC chemokine genes and analysis of their expression in channel catfish (*Ictalurus punctatus*), Dev. Comp. Immunol. (in press) (Electronic publication ahead of print, 2005 Nov 21).

[10] P.W. Holland, J. Garcia-Fernandez, N.A. Williams, A. Sidow, Gene duplications and the origins of vertebrate development, Development (Suppl.) (1994) 125–133.

[11] A. Amores, A. Force, Y.L. Yan, L. Joly, C. Amemiya, A. Fritz, R.K. Ho, J. Langeland, V. Prince, Y.L. Wang, M. Westerfield, M. Ekker, J.H. Postlethwait, Zebrafish hox clusters and vertebrate genome evolution, Science 282 (1998) 1711–1714.

[12] J.S. Taylor, I. Braasch, T. Frickey, A. Meyer, Y. Van de Peer, Genome duplication, a trait shared by 22000 species of ray-finned fish, Genome Res. 13 (2003) 382–390.

[13] I.G. Woods, C. Wilson, B. Friedlander, P. Chang, D.K. Reyes, R. Nix, P.D. Kelly, F. Chu, J.H. Postlethwait, W.S. Talbot, The zebrafish gene map defines ancestral vertebrate chromosomes, Genome Res. 15 (2005) 1307–1314.

[14] J.C. Dugas, J. Ngai, Analysis and characterization of an odorant receptor gene cluster in the zebrafish genome, Genomics 71 (2001) 53–65.

[15] M. Robinson-Rechavi, O. Marchand, H. Escriva, P.L. Bardet, D. Zelus, S. Hughes, V. Laudet, Euteleost fish genomes are characterized by expansion of gene families, Genome Res. 11 (2001) 781–788.

[16] W.S. Modi, *CCL3L1* and *CCL4L1* chemokine genes are located in a segmental duplication at chromosome 17q12, Genomics 83 (2004) 735–738.

[17] E. Kountikov, M. Wilson, S. Quiniou, N. Miller, W. Clem, E. Bengten, Genomic organization of the channel catfish CD45 functional gene and CD45 pseudogenes, Immunogenetics 57 (2005) 374–383.

[18] D.E. Gloriam, T.K. Bjarnadottir, Y.L. Yan, J.H. Postlethwait, H.B. Schioth, R. Fredriksson, The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish, Mol. Phylogenet. Evol. 35 (2005) 470–482.

[19] S.D. Eastman, T.H.P. Chen, M.M. Falk, T.C. Mendelson, M.K. Iovine, Phylogenetic analysis of three complete gap junction gene families reveals lineage-specific duplications and highly supported gene classes, Genomics 87 (2006) 265–274.

[20] M.P. Coleman, L. Conforti, E.A. Buckmaster, A. Tarlton, R.M. Ewing, M.C. Brown, M.F. Lyon, V.H. Perry, An 85-kb tandem triplication in the slow Wallerian degeneration (Wlds) mouse, Proc. Natl. Acad. Sci. USA 95 (1998) 9985–9990.

[21] M. Mallo, A novel possible mechanism for the genesis of genomic duplications and its experimental test, J. Mol. Evol. 61 (2005) 390–397.