ORIGINAL PAPER

Comprehensive clarification of two paralogous interleukin 4/13 loci in teleost fish

**Maki Ohtani · Nobuhiro Hayashi ·
Keiichiro Hashimoto · Teruyuki Nakanishi ·
Johannes Martinus Dijkstra**

Abstract Interleukins 4 and 13 (IL-4 and IL-13) are related cytokines important for Th2 immune responses and encoded by adjacent genes on human chromosome 5. Efforts were made previously to detect these genes in fish, but research was hampered by a lack of sequence conservation. A *Tetraodon nigroviridis* (green spotted pufferfish) gene was annotated as IL-4 by Li et al. (*Mol Immunol*, 44:2078–2086, 2007), but this annotation was not well substantiated. However, the present study concludes that the reported pufferfish gene belongs to the IL-4/13 lineage indeed, while also describing an additional IL-4/13 copy in a paralogous genomic region. Our analyses of IL-4/13 loci in fish describe (1) genomic region history, (2) characteristic intron–exon organization, (3) deduced IL-4/13 molecules for several teleost fish species, (4) IL-4/13 lineage-specific protein motifs including a cysteine pair (pair 1), and (5) computer software predictions of a type I cytokine fold. Teleost IL-4/13 molecules have an additional cysteine pair (pair 2) or remnants thereof, which is absent in mammalian IL-4 and IL-13. We were unable to determine if the teleost IL-4/13 genes are orthologous to either IL-4 or IL-13, or if these mammalian genes separated later in evolution.

M. Ohtani · T. Nakanishi
Laboratory of Fish Pathology,
Department of Veterinary Medicine, Nihon University,
1866 Kameino,
Fujisawa, Kanagawa 252-8510, Japan

N. Hayashi · K. Hashimoto · J. M. Dijkstra (✉)
Institute for Comprehensive Medical Science,
Fujita Health University,
Toyoake, Aichi 470-1192, Japan
e-mail: dijkstra@fujita-hu.ac.jp

Introduction

Cytokines are intercellular messengers that mediate immune responses. Although often hampered by very poor sequence conservation, a number of fish cytokines have been described (Secombes et al. 2001). Among these are several type I cytokines such as interleukin 12 (IL-12; Yoshiura et al. 2003) and IL-2 (Bird et al. 2005). These recent successes were possible because fish genomic databases became available and identification of genes could be supported by their relative position in the genome (synteny analysis).

Type I cytokines fold into a characteristic "up-up-down-down" bundle of four α-helices, A to D, and share related receptors (Bazan 1990; Rozwarski et al. 1994; Leonard 2003; Huising et al. 2006; Liongue and Ward 2007). They can be further divided into the short-chain and the long-chain subfamilies. The short-chain subfamily members are typically less than 150 amino acids, have α-helices that are less than 20 residues long, and have β-strands positioned between helices A and B and between helices C and D. The short-chain subfamily includes IL-2, IL-3, IL-4, IL-5, IL-7, IL-9, IL-13, IL-15, IL-21, granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage colony-stimulating factor (M-CSF), stem cell factor (SCF), and thymic stromal lymphopoietin (TSLP). Based on similar receptors, further distinction can be made into the IL-2 subfamily (IL-2, IL-4, IL-7, IL-9, IL-13, IL-15, IL-21, TSLP) and the IL-3 subfamily (IL-3, IL-5, GM-CSF; Liongue and Ward 2007). Except for IL-7, M-CSF, and SCF, the human genes for the short-chain type I molecules

are situated on chromosome arms 4q (IL-2, IL-15, IL-21) and 5q (IL-3, IL-4, IL-5, IL-9, IL-13, GM-CSF, TSLP). HuChr4 and huChr5q are paralogues from before teleost fish and tetrapods separated, probably as a result of a whole-genome duplication (WGD) event (Ohno 1970; Pebusque et al. 1998; Morgan et al. 1999, Wraith et al. 2000; Gu et al. 2002; Dehal and Boore 2005; Nakatani et al. 2007). Although the huChr4–huChr5q paralogy is important for the interpretation of the evolution of type I cytokines, to our knowledge, this has not been appreciated in cytokine literature. Fish orthologues of the three short-chain type I cytokine genes on huChr4q have been convincingly identified (Bird et al. 2005; Bei et al. 2006; Wang et al. 2007), whereas that was not done for any of the huChr5q-situated family members. The short-chain type I cytokine genes are very poorly conserved, and previous comparison with huChr5q regions was not extensive enough to pinpoint the expected sites in teleost genomes (Huising et al. 2006).

IL-4 is intensively searched after in current fish immunology science. That is because IL-4 in mammals is typifying and pivotal for Th2 differentiation of stimulated T helper cells (Howard et al. 1982; Murphy and Reiner 2002; Ansel et al. 2006) and because hallmark genes of Th1 responses as IL-12 (Yoshiura et al. 2003) and IFNγ (Zou et al. 2004) have been found in fish already. IL-4 binding to stimulated Th cells silences the IFNγ locus and enhances expression from the IL-4 locus (autocrine loop), whereas IFNγ has the reciprocal effect (Ansel et al. 2006). Functional evidence for Th1 responses in fish is accumulating (Zou et al. 2005; Takizawa et al. 2008), but there are no solid indications for Th2 responses yet.

IL-4 and IL-13 genes are organized in head-to-tail fashion in tetrapod genomes (e.g., Avery et al. 2004) and are closely related structurally and functionally (Minty et al. 1993; Zurawski and de Vries 1994; Chomarat and Banchereau 1998; Moy et al. 2001; Wynn 2003). Both Th2 cytokines uniquely act through binding of a receptor having an IL-4 receptor (IL-4R) α chain, thereby uniquely activating signal transducer and activator of transcription 6 protein. However, besides the IL-4Rα chain, the IL-4R and IL-13R also have distinct components (IL-2Rγc and IL-13Rα1 chain, respectively), and whereas IL-4 can functionally bind either IL-4R or IL-13R, IL-13 function needs IL-13R (Murata et al. 1998; Nelms et al. 1999). T cells express IL-4R but do not express functional amounts of IL-13R, enabling only IL-4 to affect them directly (Zurawski and de Vries 1994; Obiri et al. 1995; de Vries 1998). IL-4 and IL-13 both promote IgE antibody production and inhibit inflammatory cytokine production, but only IL-4 has a pivotal negative effect on IFNγ production, whereas IL-13 is expressed at higher levels and more important for resistance to a number of diseases (Minty et al. 1993; Wynn 2003).

There has been an IL-4 annotation for the fish Tetraodon (*Tetraodon nigroviridis*; green spotted pufferfish; Li et al. 2007). This IL-4 annotation was based on head-to-tail organization with a RAD50 gene and proximity of a gene of the Shroom family (alias "apical") as found in chicken and mammals, an intron–exon organization characteristic for short-chain type I cytokines, and an expression pattern agreeing with immune function. Unfortunately, Li and coworkers failed to align the molecules properly with tetrapod IL-4 or to discuss structural and functional motifs, despite such necessity for identification because of the very low homology (<15% amino acid identity between tetrapods and Tetraodon). Li et al. (2007) also failed to compare their detected sequence with IL-13 or other related genes situated in this genomic region of tetrapod species. Therefore, IL-4 identity of the gene was questioned by many researchers, including us. The present article, however, concludes after extensive analysis of the genome history and structural characteristics of the encoded molecules that Li and coworkers found an orthologue of IL-4 and/or IL-13 indeed. In addition, a new IL-4/13 locus was found in a fish-specific paralogous region. The present study constitutes the first detailed analysis of Th2 cytokine loci in fish.

Materials and methods

Database searching and genetic software analysis

Genes were found in the Ensembl genome databases (http://www.ensembl.org/index.html) for chicken (release 2.1), stickleback (first release), Tetraodon (release 7), medaka (first release), and zebrafish (version Zv7), by means of tblastn similarity searches (default settings) using human molecule fragments that were encoded by single (partial) exons. Then, the matching Ensembl genome fragments plus the surrounding region (~200 kb) were analyzed for the presence of genes. Similarly, Fugu genes were found using the Joint Genome Institute *Fugu rubripes* v4.0 database (http://genome.jgi-psf.org/Takru4/Takru4.info.html). Most genes were only roughly identified by blastx analysis using the National Center for Biotechnology Information database (NCBI; http://www.ncbi.nlm.nih.gov/). However, for the prediction of IL-4/13 molecules and molecules that were compared by phylogenetic analysis (supplementary Fig. 1), a variety of methods was used, including gene prediction software (http://genes.mit.edu/GENSCAN.html) and comparison with expressed sequence tags (ESTs) in the NCBI database as well as with homologous sequences. Sequence alignment for phylogenetic tree analysis was performed by computer software (GENETYX version 7.0.6, Software Development, Tokyo, Japan) and modified

by hand, and phylogenetic tree analysis (GENETYX) was performed by the neighbor-joining method based on these alignments. Regions upstream of teleost IL-4/13 open reading frames (ORFs) were screened for promoter elements by TFSEARCH software (http://www.cbrc.jp/research/db/TFSEARCH.html). Only GATA-3 binding motifs recognized by this software were discussed in the present article; although the consensus sequence for GATA-3 binding is (A/T)GATA(A/G), exceptions to this rule were found (Ko and Engel 1993). The prediction of leader
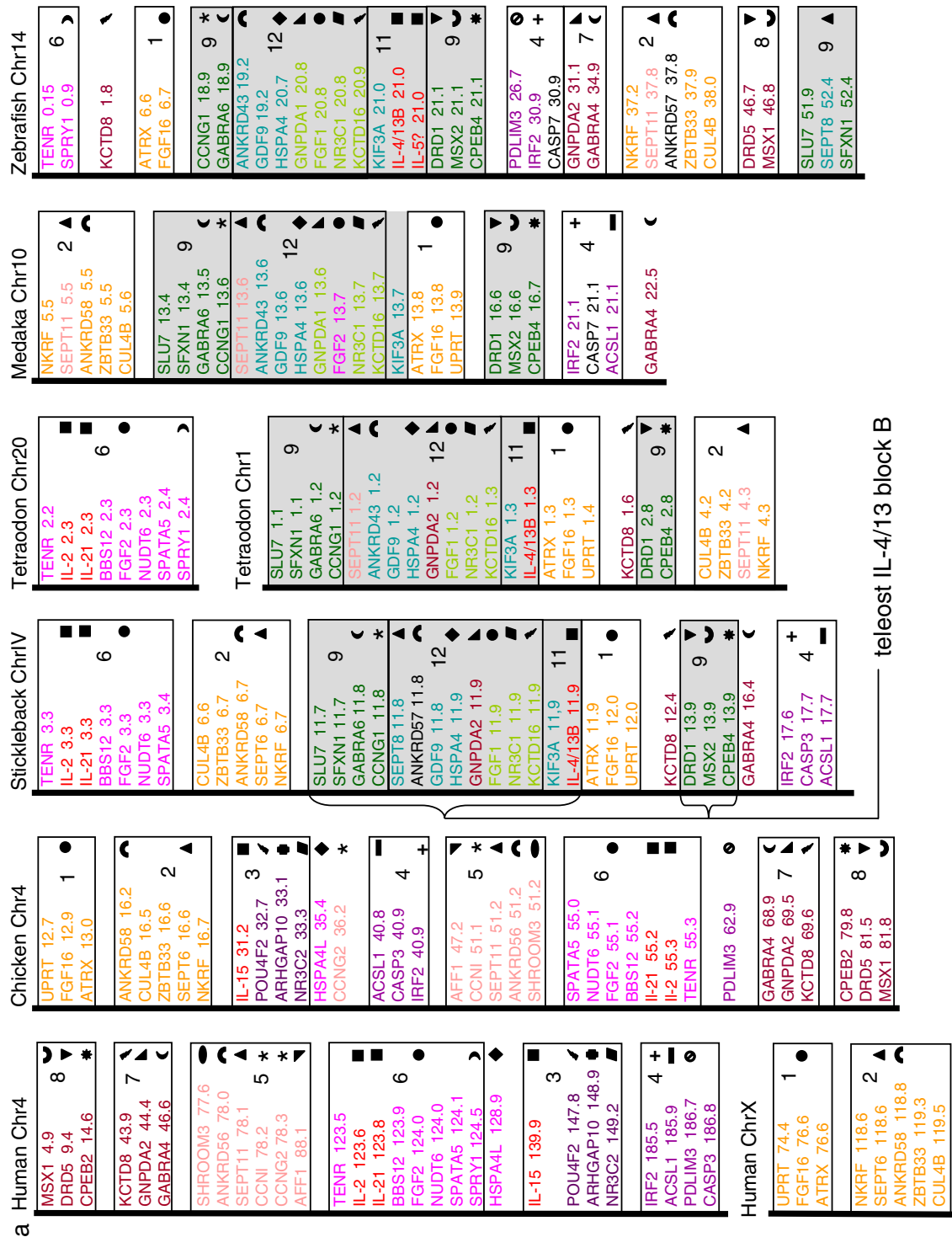
Fig. 1 Chromosome maps of genes informative on *IL-4/13* region history. **a** Chromosomes with C1 signatures. **b** Chromosomes with C2 signatures and without pronounced C1 signatures. Numbers express the chromosome position in megabases and were derived from the databases "Genecards" (human) and "Ensembl" (all others). Similar genes are indicated with *identical symbols*
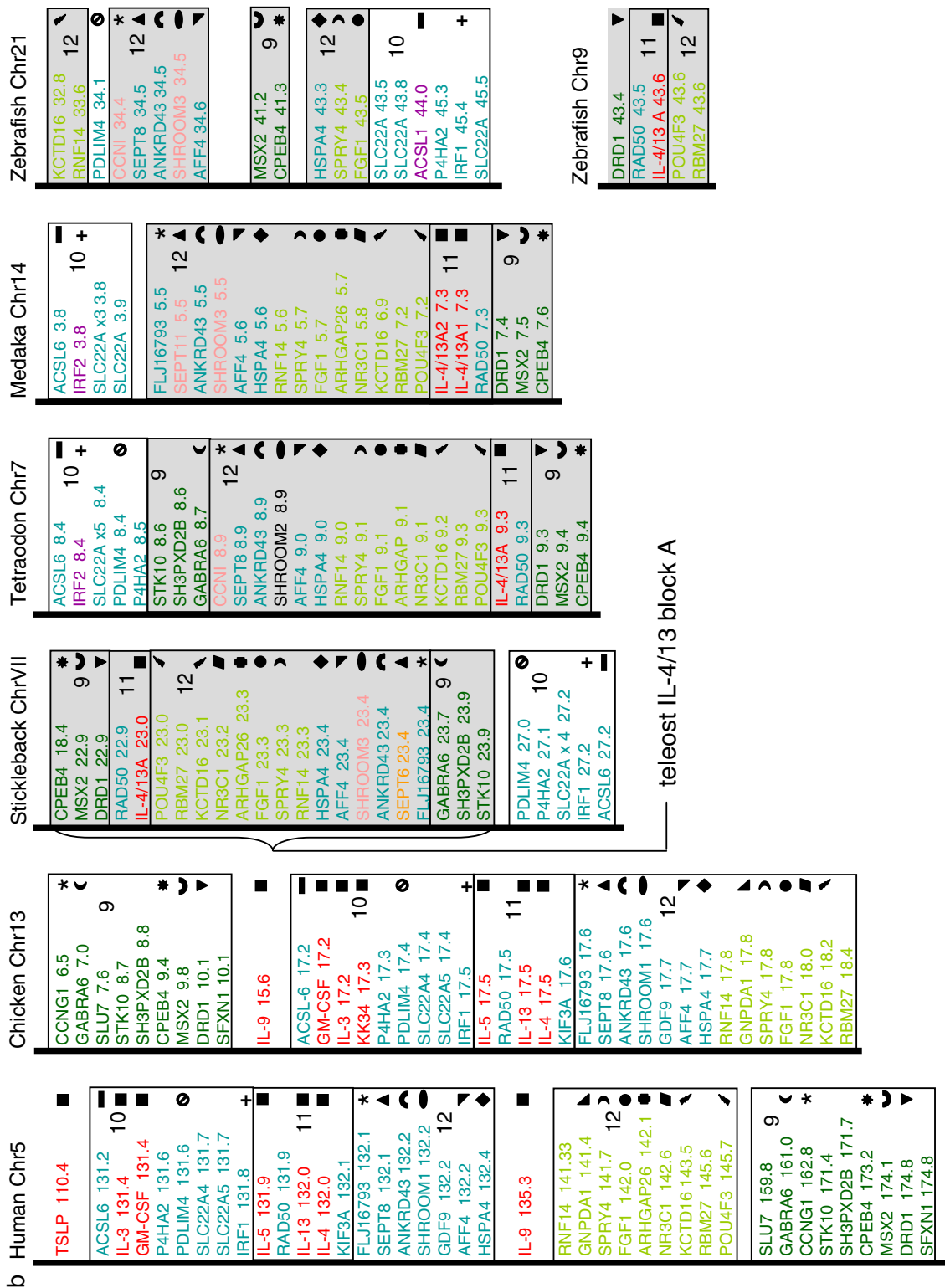
Fig. 1 Continued

peptides of IL-4/13 lineage molecules was performed by SignalP software (http://www.cbs.dtu.dk/services/SignalP/). NCBI Genbank accession numbers used for the analysis of short-helix type I cytokines were the following: human IL-4, M23442; mouse IL-4, NM_021283; chicken IL-4,

NM_001007079; human IL-13, AF043334; mouse IL-13, NM_008355; chicken IL-13, AJ621735; human IL-2, U25676; mouse IL-2, NM_008366; chicken IL-2, NC_006091; fugu IL-2, AJ749303; human IL-21, NM_021803; mouse IL-21, NM_021782; chicken IL-21,

NC_006091; human IL-5, NM_000879; mouse IL-5, NM_010558; chicken IL-5, NM_001007084; human GM-CSF, NM_000758; mouse GM-CSF, NM_009969; chicken GM-CSF, NM_001007078.

Gene nomenclature

The nomenclature of human genes in Fig. 1 and supplementary Table 1 is in agreement with the GeneCards online database (http://www.genecards.org). Chicken gene nomenclature was based on similarity and synteny with the human genes. Teleost gene nomenclature was based on top matches (lowest expectation value) with human molecules in the NCBI database upon blasting of large (~200 kb) genomic stretches including the entire gene (blastx analysis; matrix BLOSUM62, gap existence cost 11, gap extension cost 1).

Experimental fish models

Adult medaka (*Oryzias latipes*) obtained from a pet shop were maintained at 20°C and fed commercial foods. Three fish were injected intraperitoneally with concanavalin A type IV (Sigma-Aldrich), 10 ng/g body weight. Three hours after stimulation, the kidney was collected, and total ribonucleic acid (RNA) was isolated using "RNAiso" (TAKARA). Total RNA isolated by TRIzol (Invitrogen) from the kidney of adult zebrafish (*Danio rerio*), 7 days after intraperitoneal injection with 10^4 PFU of spring viraemia of carp virus strain A-1, was kindly provided by Dr. Sun Jingjing, Sun Yat-Sen University,Guangzhou, China.

cDNA synthesis, PCR, and sequence analysis

Total RNA isolates were reverse-transcribed into complementary deoxyribonucleic acid (cDNA) using *oligo-dT RBam24* primer (Table 1) and SuperScript™ III RNase H⁻ Reverse Transcriptase (Invitrogen). Gene-specific polymerase chain reaction (PCR) amplifications were performed in 25-μl reaction mixtures: 0.5 μM of forward and reverse primers, 0.5 μl cDNA template, 2.5 μl 10× buffer, 0.2 mM deoxyribonucleotide triphosphates, 1.5 mM MgCl₂, and 0.5 U Platinum *Taq* DNA polymerase (Invitrogen). Primers are listed in Table 1. The PCR conditions were: 95°C for 2 min, then 35 cycles (30 s at 95°C, 30 s at 60°C, and 30 s at 72°C), followed by 5 min at 72°C. For medaka IL-4/13A2 amplification, initially, 3′ rapid amplification of cDNA ends (RACE) was performed using primer *me-IL4/13A2-F1* in conjunction with primer *RBam24*. Then, the diluted product was used for amplification with primers *me-IL4/13A2-F2* plus *me-IL4/13A2-R1*, and this procedure was repeated using primers *me-IL4/13A2-F3* plus *me-IL4/13A2-R2*. For zebrafish IL-4/13A amplification, a 5′ part was amplified with primers *zf-IL4/13A-F1* plus *zf-IL4/13A-R1*. An overlapping 3′ part was amplified by 3′RACE, using primer *zf-IL-4/13A-F2* in conjunction with primer *RBam24*. For zebrafish IL-4/13B amplification, this gene was amplified by a single 3′RACE reaction using primers *zf-IL4/13B-F1* and *RBam24*. To determine the sequence of PCR products, all PCR products were ligated into pCR®2. TOPO® vector (Invitrogen) and transformed into competent *Escherichia coli* cells (Invitrogen). Plasmid DNA was purified and sequenced by ABI 3130 Genetic analyzer (Applied Biosystems). GenBank accession numbers of the determined sequences are the following: medaka *IL-4/13A2*, AB375406; zebrafish *IL-4/13A*, AB375404; zebrafish *IL-4/13B*, AB375405.

Tertiary structural modeling

The initial prediction of the likelihood of tertiary structures was performed with 3D-PSSM software, which compares sequences against a large database of known protein structures (http://www.sbg.bio.ic.ac.uk/~3dpssm/; Kelley et al. 2000); for zebrafish IL-4/13A, this retrieved human IL-13 as top match. With the tertiary structure of human IL-13 (PDB 1GA3) as the template and an initial sequence alignment as in supplementary Fig. 4, a tertiary structural

Table 1 Primer sequences

Primer	Sequence
me-IL4/13A2-F1	5′-GTTCCTTCAATCGTCAACAA
me-IL4/13A2-F2	5′-ATGGATCATTCTGCACTTGT
me-IL4/13A2-F3	5′-GACACTACGAGCCAGTTTGT
me-IL4/13A2-R1	5′-AAGGCAGCAACAAACAAGT
me-IL4/13A2-R2	5′-TTACAGCAGTCTTGGAGGAA
zf-IL4/13A-F1	5′-GGTATACCTCTAGGCCTTCAGAC
zf-IL4/13A-F2	5′-CACCAAGACGAGAAAAACTG
zf-IL4/13A-R1	5′-AAAAACACAAAACCACCAGA
zf-IL4/13B-F1	5′-CCTGAAGATCTCAACATCTGG
RBam24	5′-AGCACTCTCCAGCCTCTCACCGAG
oligo-dT RBam24	5′-AGCACTCTCCAGCCTCTCACCGAGTTTTTTTTTTTTTTTTTT

model of the zebrafish IL-4/13A was constructed by a homology-modeling method using Homology module of Insight II 2000 (Accerlys, Cambridge, UK). The geometry of the models was optimized by Discover 3 module of Insight II 2000 (Accerlys) using a consistent valence force field (Furuhashi et al. 2000). Atomic charges for potentially charged groups were estimated assuming a pH of 7.0. After taking the initial atom coordinate, another restrained molecular dynamics and energy minimization protocol was performed using distance restraints of each pairs of the cystein residues to construct the two predicted disulfide bridges. Then, to search the most stable arrangement of the predicted structure, the system was equilibrated through 5 ps of number/volume/temperature molecular dynamics at 298 K. The program Discover 3 module of Insight II 2000 was used again for energy minimization and the molecular dynamic calculations.

Results and discussion

Initial screening for probable teleost IL-4 sites: comparison of IL-2 and IL-4 genomic regions between human, chicken, and fish

To detect teleost IL-4/13 lineage loci, teleost genomic sequence databases were searched for homologues of human genes neighboring IL-2 and IL-4. The IL-2 regions were included because IL-2, IL-21, IL-4, and IL-13 are related cytokines in paralogous gene clusters (IL-2 family cytokines, FGF, SPRY; Fig. 1), and this paralogy should be helpful to understand the site of teleost IL-4/13 loci. Supplementary Table 1 lists top similarities of selected human Chr4-, Chr5-, and ChrX-encoded molecules with the translated genomic sequences of chicken, stickleback, Tetraodon, medaka, and zebrafish (Ensembl database; tblastn analysis). Figure 1 shows the chromosomal organization of these genes, with the designations of teleost genes based on top similarities with human sequences (NCBI database; blastx analysis) and colors allowing rapid comparison with the human genome map. Beware that top similarity does not necessarily mean orthology and that Fig. 1 is not an attempt to provide the best possible nomenclature for the fish genes. Orthologous gene blocks in Fig. 1 are numbered according to the chicken genome because of the rather conservative genome arrangement in this species (Hillier et al. 2004; Nakatani et al. 2007). Chromosome location or scaffold information of some gene blocks that were not mapped (yet) to the teleost chromosomes depicted in Fig. 1 (e.g., block 3) can readily be retrieved from supplementary Table 1.

The general findings depicted in Fig. 1 are as follows: (1) Stickleback chromosome IV (stChrIV), Tetraodon

chromosomes 1 plus 20 (teChr1, teChr20), medaka chromosome 10 (meChr10), and zebrafish chromosome 14 (zeChr14) have segments that are orthologous with huChr4/huChrX/chChr4 (ch=chicken) as well as segments orthologous with huChr5/chChr13; (2) stChrVII, teChr7, meChr14, and zeChr21/zeChr9 only have the huChr5/chChr13 signature. Although lacking the details of Fig. 1 and to our knowledge without the stickleback analysis, these basic patterns of chromosome orthologies have been reported before (Liu et al. 2002; Jaillon et al. 2004; Woods et al. 2005; Kasahara et al. 2007; Nakatani et al. 2007). Separation of chChr4 gene blocks among huChr4 and huChrX (Rabie et al. 2004) and of zeChr14 gene blocks among teChr1 and teChr20 (Woods et al. 2005) have been reported as well, the separations probably being derived traits. It remains to be determined if the zeChr9 cluster depicted in Fig. 1 truly maps to that chromosome, because in an earlier draft of the zebrafish genome sequence (Ensembl Zv6), it mapped to zeChr21, which much better agrees with the findings in other fishes.

Closely related genes are indicated in Fig. 1 with identical symbols and reveal syntenies between gene blocks 2, 5, and 12, between 3 and 12, between 4 and 10, between 6 and 12, between 7 and 12, and between 8 and 9. These syntenies derived from both intrachromosomal and interchromosomal segmental duplications, many of the latter presumably as part of WGD events (see below). Previous reports on paralogy between huChr4 and huChr5q mentioned, among others, the gene couples *MSX1/MSX2*, *DRD5/DRD1*, *GABRA4/GABRA6*, *FGF2/FGF1*, *SPRY1/SPRY4*, and *IRF2/IRF1* (Fig. 1; Pebusque et al. 1998; Morgan et al. 1999; Wraith et al. 2000; Itoh and Ornitz 2004).

Teleost fish have two paralogous IL-4/13 regions; probable implications of whole-genome duplications

A disproportional large number of gene duplications in vertebrate genomes were mapped to the period between the separation from cephalochordata (amphioxus) and the split between cartilaginous and bony fish, the incidence peaking around 600 million years ago (MYA; Gu et al. 2002; McLysaght et al. 2002; Robinson-Rechavi et al. 2004; Vandepoele et al 2004; Blomme et al. 2006). Although there remains discussion on the relative contribution of various modes of duplication in that period, most researchers now agree on one and probably two WGDs (Ohno 1970; Gu et al. 2002; McLysaght et al. 2002; Vandepoele et al. 2004; Dehal and Boore 2005; Blomme et al. 2006; Nakatani et al. 2007). Quadruplicated syntenies on human chromosomes 4, 5, 8, and 10 serve as prime example for the two rounds of the WGD theory (Wolfe 2001; Dehal and Boore 2005). Nakatani et al. (2007) designated the ancestral chromosomes of huChr4 and huChr5q as "C1" and "C2,"

respectively, and below we will follow that nomenclature. Nakatani et al. (2007) hypothesized that the two rounds of WGD in early vertebrates resulted in the four chromosomes C0 (huChr10/13 regions), C1 (huChr4), C2 (huChr5q), and C3 (huChr7/8 regions) from a single C chromosome. Phylogenetic tree analyses (Pebusque et al. 1998; Morgan et al. 1999) suggest that the C1–C2 duplication occurred after their separation from C0/C3.

It is now commonly accepted that in an ancestor of all teleost fish, ~350 MYA, an additional WGD event took place (fish-specific round 3; FS3R; Amores et al. 1998; Taylor et al. 2003; Jaillon et al. 2004; Vandepoele et al. 2004; Woods et al. 2005; Kasahara et al. 2007; Nakatani et al. 2007). Teleosts form the vast majority of extant fish species. For reasons explained below, from hereon, the combination of blocks 9, 11, and 12 on stChrVII/teChr7/ meChr14/zeChr21/9 and stChrIV/teChr1/meChr10/zeChr14 are mentioned as teleost IL-4/13 blocks A and B, respectively (gray regions in Fig. 1). The duplicated *MSX2*, *NR3C1*, *FGF1*, and *HSPA4* genes in teleost IL-4/ 13 blocks A and B (Fig. 1) have been listed as argumentative for FS3R (Liu et al. 2002; Taylor et al. 2003; Kasahara et al. 2007). The reciprocal loss of most duplicated genes in teleost IL-4/13 blocks A and B (*RAD50*, *KIF3A*, *POU4F3*, *RBM27*, *ARHGAP26*, *SPRY4*, *RNF14*, *GNPDA1*, *AFF4*, *SHROOM*, *GDF9*, *FLJ16793*, *CCNG1*, *SFXN1*, *SH3PXD2B*, *STK10*, and *SLU7*; Fig. 1) agrees with the reported massive loss of FS3R-derived gene duplicates (Jaillon et al. 2004; Woods et al. 2005; Brunet et al. 2006; Sémon and Wolfe 2007). Kasahara et al. (2007) hypothesized that large parts of the FS3R-duplicated C1 and C2 chromosomes fused together, followed by a number of smaller intrachromosomal rearrangements (Fig. 2b). They based this hypothesis solely on syntenies and top similarities, which do not exclude the possibility that stretches such as IL-4/13 block B appear as C2 only because they were lost from the corresponding tetrapod C1 regions. However, phylogenetic analyses by other authors and us support their theory, the general finding being that the proteins encoded by the paralogous genes in teleost IL-4/13 blocks A and B share more similarity with each other than with their tetrapod orthologues (MSX2: Ekker et al. 1997; KCTD16: Gamse et al. 2005; Septin 8: Cao et al. 2007; NR3C1: Stolte et al. 2006; examples in supplementary Fig. 1). The FS3R-history/C2-identity of teleost IL-4/ 13 block B is even better argued by unique organization features shared with teleost IL-4/13 block A, such as the relative position and orientation of block 11 and the division of block 9 (Fig. 1). Figure 2a schematically depicts the organization of the IL-4/13 region in a hypothetical C2 chromosome of an ancestor to both teleosts and tetrapods. The Fig. 2a model is based on: (1) simplicity; (2) similar genes, such as for the IL-4, IL-13, IL-5, GM-CSF, and IL-3
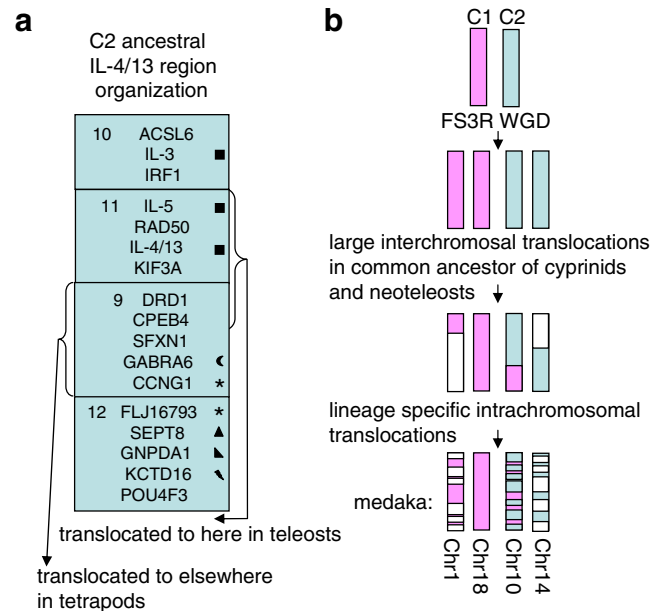
Fig. 2 *IL-4/13* region history. **a** Hypothetical ancestral C2 organization. Gene blocks are indicated with a few representative genes. Gene families supportive of the history model are indicated with *symbols* as in Fig. 1. **b** C1 and C2 history in teleost fish according to Kasahara et al. (2007) and exemplified by medaka. The drawing is only schematic. The *white blocks* represent several other ancestral chromosomes

cytokines, or for the FLJ16793 (alias "similar to cyclin I") and CCNG1 (cyclin G1) cyclins, often originate from nearby tandem duplications (Read et al. 2004); (3) the hypothetical ancestral C2 organization shares syntenies with C1 blocks 5 (*SEPT11–CCNI–CCNG2*) and 7 (*GABRA4–GNPDA2–KCTD8*); (4) the *DRD1* and *SFXN1* genes are very close in tetrapod block 9. The comparison between neoteleosts (stickleback, Tetraodon, medaka) and zebrafish (a cyprinid fish) indicates that the insertion of gene block 1 into IL-4/13 block A occurred only after these teleosts separated.

Teleost IL-4/13A and IL-4/13B loci: location, intron–exon organization, GATA-3 binding, and mRNA instability motifs
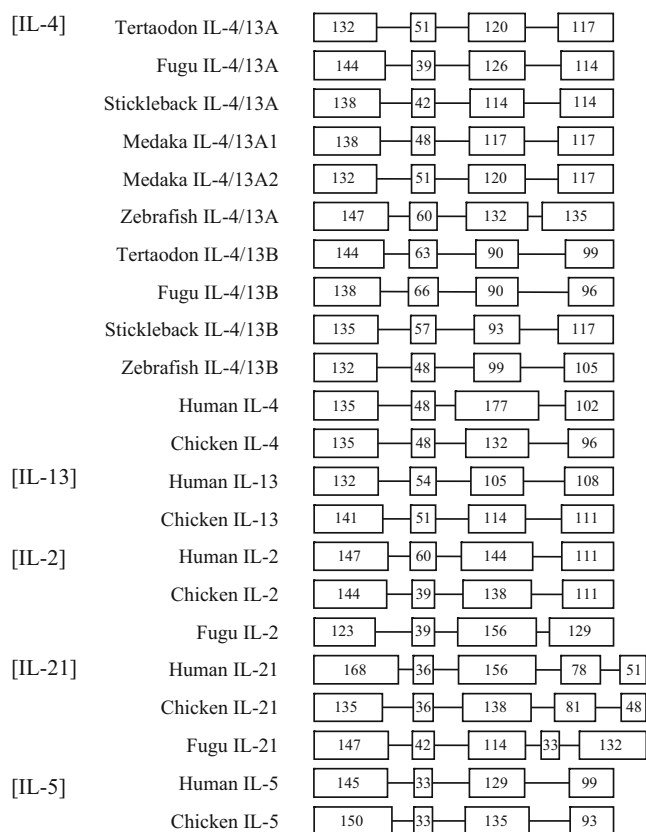
Except for medaka, a single IL-4/13 lineage gene was detected in each teleost block 11. MeChr14 block 11 harbors two similar *IL-4/13* genes, *IL-4/13A1* and *IL-4/13A2*, whereas no *IL-4/13B* gene was found in proximity of *KIF3A* on meChr10. However, the Ensembl database information on the respective stretch of meChr10 is incomplete (many residues are denoted as "n") and future improvement of the sequence information may reveal the presence of *IL-4/13B*. The teleost *IL-4/13A* genes are located downstream of both *RAD50* and *POU4F3*, in the same orientation as *RAD50*, and the teleost *IL-4/13B* genes are located downstream of *KIF3A* and upstream of *ATRX* or

DRD1, in an opposite orientation of *KIF3A*; the positioning of teleost *IL-4/13* relative to *RAD50* and *KIF3A* agrees with the tetrapod situation (e.g., Avery et al. 2004). The four-exon organization of the teleost *IL-4/13* genes is typical for the short-helix type I cytokine family (Fig. 3), with exons 2, 3, and 4 each starting at codon position 1 (e.g., Fig. 4). For zebrafish *IL-4/13A*, but for none of the other teleost *IL-4/13* genes, ESTs could be retrieved from the NCBI database (accessions CO922126, CK702857, EH475380, EH44 9554, EB991904, CK674382). The zebrafish *IL-4/13A* and *IL-4/13B* ORFs were confirmed by cDNA analysis. Because of some difficulty in predicting the last exons of the medaka *IL-4/13A* genes, also an exon 2–4 part of medaka *IL-4/13A2* was analyzed from cDNA. However, most teleost *IL-4/13* genes discussed in the present article were only predicted from genomic database sequences, and sometimes the exon borders are discussable (not shown).

The *IL-4/13A* sequence reported by Li et al. (2007) for *T. nigroviridis* is substantially different from that in the Ensembl *T. nigroviridis* database with more than 25-nucleotide substitutions and a six-nucleotide length difference between the ORFs. Because the Ensembl Tetraodon *IL-4/13A* sequence encodes a highly conserved cysteine, which is absent in the molecule predicted by Li and

coworkers, the present article uses the Ensembl Tetraodon sequence information for comparison with other fishes. The cDNA analysis by Li et al. (2007) was used, however, for predicting exon borders of pufferfish *IL-4/13A*.

The pufferfishes *T. nigroviridis* and *F. rubripes* separated 18–30 MYA, with an overall genetic divergence rate about twice as fast as observed between human and mouse (Jaillon et al. 2004; http://www.genoscope.cns.fr/). Tetraodon *IL-4/13A* and Fugu *IL-4/13A* are situated in identical gene settings, and so are Tetraodon *IL-4/13B* and Fugu *IL-4/13B* (not shown). Comparison between the pufferfish *IL-4/13A* sequences (Fig. 4) or *IL-4/13B* sequences (supplementary Fig. 2) indicates that relative to the poor sequence conservation of the ORFs (*IL-4/13A*=60% nucleotide identity when excluding gaps; *IL-4/13B*=61%), fragments upstream and downstream of the ORFs are well conserved (*IL-4/13A* upstream stretch depicted in Fig. 4= 73%; *IL-4/13A* downstream stretch=72%; *IL-4/13B* upstream stretch depicted in supplementary Fig. 2: 74%; *IL-4/13B* downstream stretch=77%). Better conservation of regulatory sequences than of coding sequences was also reported for cytokine genes in tetrapods (Avery et al. 2004). Between the pufferfishes and the other discussed teleost species, which separated longer ago, or between teleost *IL-4/13A* and *IL-4/13B*, the overall sequence similarities of the noncoding regions are very minor, but most teleost *IL-4/13* genes share a TATA box preceded by a GATA-3 binding motif in the presumable promoter region and multiple messenger RNA (mRNA) instability motifs (ATTTA) in the immediate downstream noncoding region (Fig. 4; supplementary Figs. 2 and 3). GATA-3 is a critical enhancing transcription factor for Th2 differentiation (Ansel et al. 2006; Ho and Pai 2007). The importance of the TATA box and GATA-3 binding motif for the teleost *IL-4/13A* genes is underlined by their conservation at very similar sites (Fig. 4 and supplementary Fig. 3), except in medaka *IL-4/13A1*. The closest GATA-3 binding motif upstream from the medaka *IL-4/13A1* ORF is only found at 0.9 kb, whereas these sites also tend to be further upstream from the *IL-4/13B* ORFs: 2.4 kb from Fugu *IL-4/13B*, 0.3 kb from Tetraodon *IL-4/13B*, 1.2 kb from the stickleback *IL-4/13B*, and 1.0 kb from zebrafish *IL-4/13B* (not shown). Overall, the conservation pattern suggests that GATA-3 has a more important direct enhancing function for *IL-4/13A* than for *IL-4/13B*. It also suggests that medaka *IL-4/13A1* has lost the original mode of *IL-4/13A* regulation. Whereas all teleost *IL-4/13A* have perfect TATA boxes, the proximal promoter regions of both Fugu and Tetraodon *IL-4/13B* lack such motifs (supplementary Fig. 2). This, combined with the several ESTs found for zebrafish *IL-4/13A* whereas none for zebrafish *IL-4/13B* (see above), suggests that generally the expression of teleost *IL-4/13A* genes is higher than that of teleost *IL-4/13B*.

		Exon 1	Intron/Exon 2	Exon 3	Exon 4	Exon 5
[IL-4]	Tertaodon IL-4/13A	132	51	120	117	
	Fugu IL-4/13A	144	39	126	114	
	Stickleback IL-4/13A	138	42	114	114	
	Medaka IL-4/13A1	138	48	117	117	
	Medaka IL-4/13A2	132	51	120	117	
	Zebrafish IL-4/13A	147	60	132	135	
	Tertaodon IL-4/13B	144	63	90	99	
	Fugu IL-4/13B	138	66	90	96	
	Stickleback IL-4/13B	135	57	93	117	
	Zebrafish IL-4/13B	132	48	99	105	
	Human IL-4	135	48	177	102	
	Chicken IL-4	135	48	132	96	
[IL-13]	Human IL-13	132	54	105	108	
	Chicken IL-13	141	51	114	111	
[IL-2]	Human IL-2	147	60	144	111	
	Chicken IL-2	144	39	138	111	
	Fugu IL-2	123	39	156	129	
[IL-21]	Human IL-21	168	36	156	78	51
	Chicken IL-21	135	36	138	81	48
	Fugu IL-21	147	42	114	33	132
[IL-5]	Human IL-5	145	33	129	99	
	Chicken IL-5	150	33	135	93	

Fig. 3 Intron–exon organization of *IL-4/13* lineage genes and several other short-helix type I cytokine genes. *Blocks* indicate the coding parts of exons. *Numbers* indicate length. Introns are not drawn to scale
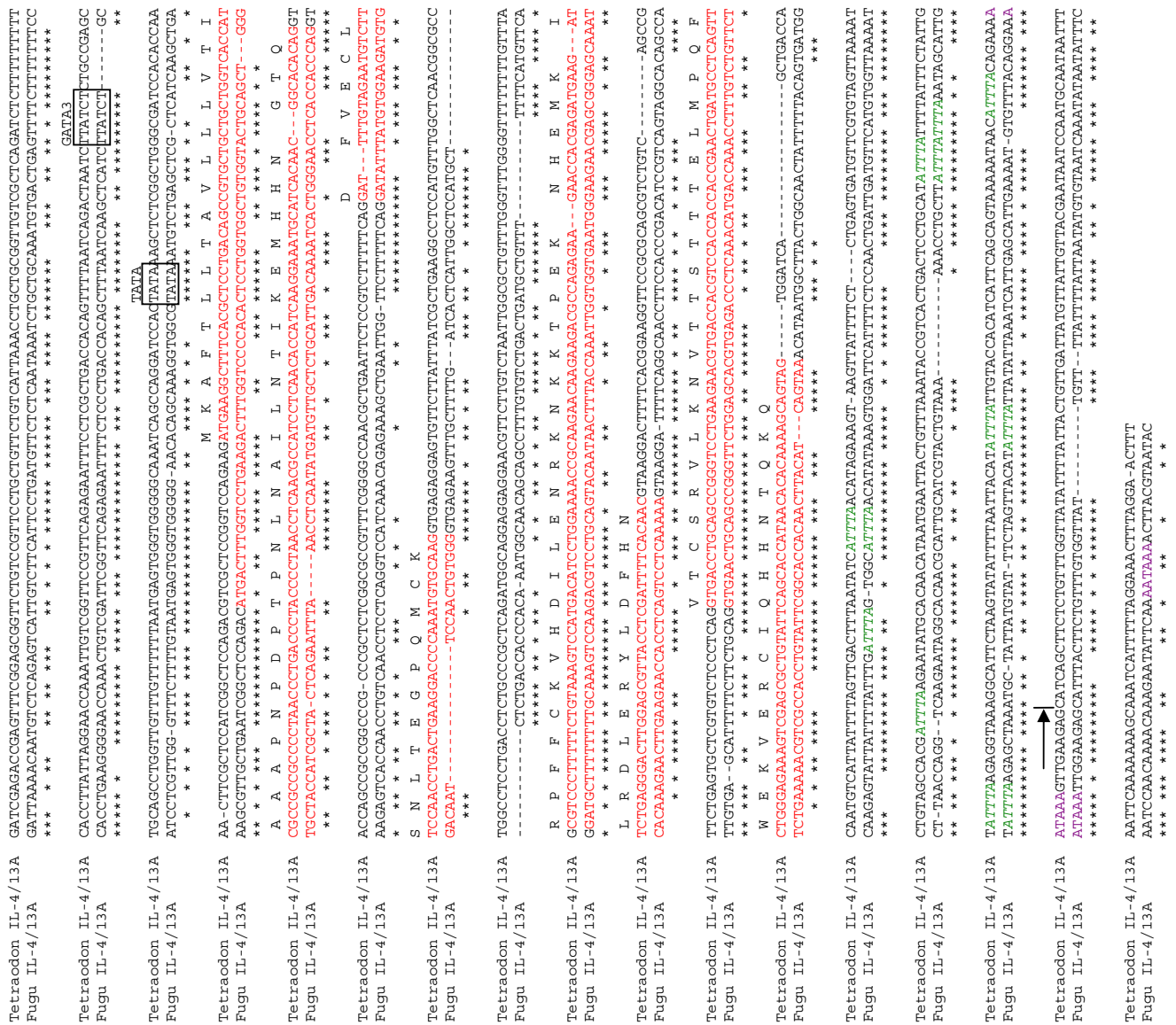
Fig. 4 Alignment of Tetraodon and Fugu *IL-4/13A* genomic sequences. Amino acids of the Tetraodon sequence are listed *above* first nucleotides of codons. TATA boxes and GATA-3 binding motifs are *boxed*. *Red*, open reading frames; *green*, mRNA destabilization motif; *purple*, poly-adenylation motif. *Arrow*, from this end the poly-A tail was attached (Li et al. 2007)

Mammalian and chicken *IL-4* and *IL-13* promoter regions also have TATA boxes preceded by a GATA-3 motif, but only for mammalian *IL-13* and not mammalian *IL-4*, this GATA-3 motif was found to be important (Ranganath et al. 1998; Lavenu-Bombled et al. 2002; Avery et al. 2004; Ansel et al. 2006). GATA-3, however, does have an important enhancing effect on mammalian *IL-4* by inducing epigenetic chromatin changes by binding elsewhere in the region (for an extensive review on transcription regulation in the extended *IL-4* region, see Ansel et al. 2006).

Li et al. (2007) determined the full-length 3′ untranslated region of Tetraodon *IL-4/13A* and already reported the many mRNA instability motifs (ATTTA; Sachs, 1993), which are also found in Fugu *IL-4/13A* (Fig. 4) and the other teleost *IL-4/13* genes (supplementary Fig. 2 and data not shown). Short half-life induced by such motifs improves regulatory control (Sachs 1993; Umland et al. 1998).

Hopefully, we detected all teleost IL-4/13 lineage genes, but with the short exons, the low similarities, and the low expression levels, short-chain type I cytokine genes are truly hard to find.

Other short-chain type I cytokine genes

Detection of teleost IL-3 family members was anticipated because of the detection of zebrafish *IL-3Rβc* (common

component of the receptors for IL-3, IL-5, and GM-CSF; Liongue and Ward 2007), but none was reported so far. We did not search very intensively for this family and found only a single possible gene situated in zeChr14 block 11 in a reverse orientation to *IL-4/13B*. This zebrafish gene, which was not determined by cDNA analysis but predicted from genomic sequence by computer software (GenBank accession XP_001332104; no annotation), has an intron–exon organization agreeing with other short-helix type I cytokines and encodes a 161-amino acid protein with a probable leader sequence and some motifs reminiscent of IL-3 family members (not shown). For synteny reasons (Fig. 2), this gene was tentatively designated *IL-5* (Fig. 1), but further investigation is needed for proper identification.
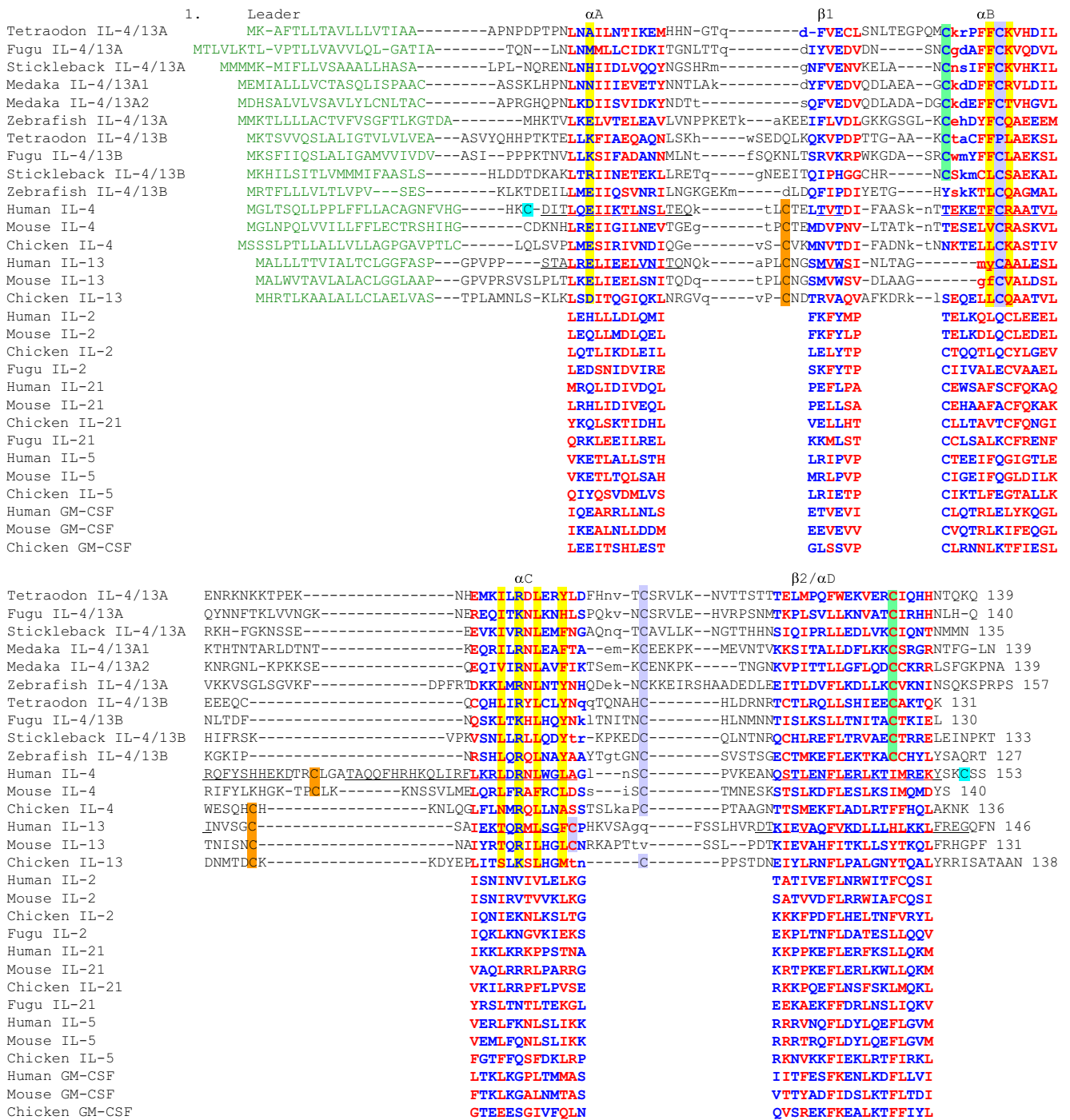
Tetraodon *IL-15* has been mapped to chromosome 18 (Fang et al. 2006), which agrees with the presence of other block 3 homologues (supplementary Table 1) and the reported C1 identity of this Tetraodon chromosome (Jaillon et al. 2004). A fish-specific copy of the IL-15 lineage, *IL-15X*, did not map to chromosomes with a major C1 or C2 identity (Fang et al. 2006). We did not search for teleost TSLP or IL-9.

Deduced teleost IL-4/13 molecules share important characteristics with tetrapod IL-4 and IL-13 while distinguished by a unique cysteine pair

Figure 5 shows the deduced amino acid sequences of teleost IL-4/13A and IL-4/13B in alignment with full-length tetrapod IL-4 and IL-13 and fragments of a few other IL-2 family members (IL-2 and IL-21) and a few IL-3 family members (IL-5 and GM-CSF). The five fragments represent the consensus structural framework within the short-chain type I cytokines as distinguished by Rozwarski et al. (1994) based on overlapping segments of superimposed IL-2, IL-4, IL-5, GM-CSF, and M-CSF structures. These distinguished fragments basically consist of (major parts of) the α-helix A (αA), β-strand 1 (β1), αB, αC, and the tightly connected β2 plus αD. We resolved to this alignment method because, even between subfamily members, there is an enormous divergence in sequence and variation in helix lengths and positions of disulfide bridges (e.g., Fig. 5; Rozwarski et al. 1994; Eisenmesser et al. 2001; Moy et al. 2001; Avery et al. 2004), making full-length alignment virtually impossible. Within the framework (Fig. 5 blue and red residues), Rozwarski et al. distinguished residues that form the inner core of the proteins (Fig. 5, red residues). They determined that most of the inner core residues were hydrophobic but that 23% of them had a side chain with hydrogen-bonding potential, which mostly folded out of the core but in some cases was involved in buried hydrogen-bonding interactions. We added the IL-13 and IL-21 sequences to Fig. 5 based on structural alignments by

Eisenmesser et al. (2001) and Bondensgaard et al. (2007), respectively. The αB helix of mammalian IL-13 is shorter than found in most type I cytokines (Eisenmesser et al. 2001; Moy et al. 2001), whereas chicken IL-13 αB probably has a more usual length (Fig. 5; Avery et al. 2004). The Fig. 5 sequences between the framework fragments predominantly form loop structures, and we aligned them based on sequence similarities and exon mapping (Fig. 5 small font relates to exon borders); like for other short-helix type I cytokines, the IL-4/13 lineage leader+αA, β1, αB+αC, and β2+αD domains are mostly encoded by exons 1, 2, 3, and 4, respectively (Rozwarski et al. 1994). Cysteines forming disulfide bridges in human IL-4 and IL-13 molecules (Walter et al. 1992; Wlodawer et al. 1992; Powers et al. 1993; Eisenmesser et al. 2001; Moy et al. 2001) or expected in the other depicted IL-4/13 lineage molecules are depicted with identical color shading. Other than these cysteines, positions with residues characteristic for the IL-4/13 lineage are depicted in Fig. 5 with yellow shading. A highly conserved cysteine pair in the IL-4/13 lineage, only modified in mammalian IL-13, consists of a cysteine in αB and a cysteine in the CD loop, which is encoded by the beginning of exon 4 (gray pair); a typical setting of the cysteine in αB, found in many IL-4/13 lineage molecules, is the preceding by an aromatic residue and/or the following by a basic residue. The yellow-shaded glutamic acid in αA and arginine in helix αC were shown to be the important sites for binding of human IL-4 to the IL-4Rα receptor chain (Wang et al. 1997; Hage et al. 1999), and other tetrapod IL-4/13 lineage molecules also have acidic and basic residues at these positions, respectively (Fig. 5; Eisenmesser et al. 2001; Moy et al. 2001; Avery et al. 2004). The huIL-4 αA glutamic acid forms part of a hydrogen network including IL-4Rα tyrosine and serine residues (Hage et al. 1999), thus exchange for other groups with hydrogen-bonding potential as observed in most teleost IL-4/13 molecules does not necessarily interfere with the binding mode; however, the hydrophobic residues at this position in Tetraodon and Fugu IL-4/13A do not agree. HuIL-4 αC arginine (Fig. 5, yellow) forms an ion pair with an aspartic acid on IL-4Rα (Hage et al. 1999), and the positive charges of the teleost IL-4/13 residues at this position agree with a similar binding mode. According to Fig. 5, the ancestral setting of this αC arginine probably was (L/I)XRXLXX(Y/F).

Two cysteine pairs collectively distinguish tetrapod IL-4 and IL-13 from the teleost IL-4/13 molecules. An ABloop–BCloop pair is only found in the tetrapod molecules (Fig. 5, orange), whereas an ABloop–αD pair (Fig. 5, green) is only found in the teleost molecules. The simplest explanation based on genomic positions (paragraphs above) and positioning of these disulfide pairs is that a single ancestral IL-4/13 gene duplicated independently in the tetrapod and

```
                1.      Leader                                              αA                        β1                      αB
Tetraodon IL-4/13A      MK-AFTLLTAVLLLVTIAA--------APNPDPTPNLNAILNTIKEMHHN-GTq--------d-FVECLSNLTEGPQMCkrPFFCKVHDIL
Fugu IL-4/13A           MTLVLKTL-VPTLLVAVVLQL-GATIA-------TQN--LNLNMMLLCIDKITGNLTTq--------dIYVEDVDN-----SNCgdAFFCKVQDVL
Stickleback IL-4/13A    MMMMK-MIFLLVSAAALLHASA-------LPL-NQRENLNHIIDLVQQYNGSHRm---------gNFVENVKELA----NCnsIFFCKVHKIL
Medaka IL-4/13A1        MEMIALLLVCTASQLISPAAC--------ASSKLHPNLNNIIIEVETYNNTLAk--------dYFVEDVQDLAEA--GCkdDFFCRVLDIL
Medaka IL-4/13A2        MDHSALVLVSAVLYLCNLTAC-------APRGHQPNLKDIISVIDKYNDTt-----------sQFVEDVQDLADA-DGCkdEFFCTVHGVL
Zebrafish IL-4/13A      MMKTLLLLACTVFVSGFTLKGTDA--------MHKTVLKELVTELEAVLVNPPKETk---aKEEIFLVDLGKKGSGL-KCehDYFCQAEEEM
Tetraodon IL-4/13B      MKTSVVQSLALIGTVLVLVEA---ASVYQHHPTKTELLKFIAEAQQNLSKh-----wSEDQLKQKVPDPTTG-AA--KCtaCFFPLAEKSL
Fugu IL-4/13B           MKSFIIQSLALIGAMVVIVDV--ASI--PPPKTNVLLKSIFADANNMLNt-----fSQKNLTSRVKRPWKGDA--SRCwmYFFCLAEKSL
Stickleback IL-4/13B    MKHILSITLVMMMIFAASLS-------HLDDTDKAKLTRIINETEKLLRETq-----gNEEITQIPHGGCHR-----CSkmCLCSAEKAL
Zebrafish IL-4/13B      MRTFLLLVLTLVPV---SES--------KLKTDEILLMEIIQSVNRILNGKGEKm-----dLDQFIPDIYETG----HYskKTLCQAGMAL
Human IL-4              MGLTSQLLPPLFFLLACAGNFVHG-----HKC-DITLQEIIKTLNSLTEQk-------tLCTELTVTDI-FAASk-nTTEKETFCRAATVL
Mouse IL-4              MGLNPQLVVILLFFLECTRSHIHG-------CDKNHLREIIGILNEVTGEg-------tPCTEMDVPNV-LTATk-nTTESELVCRASKVL
Chicken IL-4           MSSSLPTLLALLVLLAGPGAVPTLC------LQLSVPLMESIRIVNDIQGe-------vS-CVKMNVTDI-FADNk-tNNKTELLCKASTIV
Human IL-13            MALLLTTVIALTCLGGFASP---GPVPP----STALRELIEELVNITQNQk-----aPLCNGSMVWSI-NLTAG------myCAALESL
Mouse IL-13           MALWVTAVLALACLGGGLAAP---GPVPRSVSLPLTLKELIEELSNITQDq------tPLCNGSMVWSV-DLAAG------gfCVALDSL
Chicken IL-13         MHRTLKAALALLCLAELVAS---TPLAMNLS-KLKLSDITQGIQKLNRGVq-----vP-CNDTRVAQVAFKDRk--lSEQELLCQAATVL
Human IL-2                                        LEHLLLDLQMI                     FKFYMP              TELKQLQCLEEEL
Mouse IL-2                                        LEQLLMDLQEL                     FKFYLP              TELKDLQCLEDEL
Chicken IL-2                                      LQTLIKDLEIL                     LELYTP              CTQQTLQCYLGEV
Fugu IL-2                                         LEDSNIDVIRE                     SKFYTP              CIIVALECVAAEL
Human IL-21                                       MRQLIDIVDQL                     PEFLPA              CEWSAFSCFQKAQ
Mouse IL-21                                       LRHLIDIVEQL                     PELLSA              CEHAAFACFQKAK
Chicken IL-21                                     YKQLSKTIDHL                     VELLHT              CLLTAVTCFQNGI
Fugu IL-21                                        QRKLEEILREL                     KKMLST              CCLSALKCFRENF
Human IL-5                                        VKETLALLSTH                     LRIPVP              CTEEIFQGIGTLE
Mouse IL-5                                        VKETLTQLSAH                     MRLPVP              CIGEIFQGLDILK
Chicken IL-5                                      QIYQSVDMLVS                     LRIETP              CIKTLFEGTALLK
Human GM-CSF                                      IQEARRLLNLS                     ETVEVI              CLQTRLELYKQGL
Mouse GM-CSF                                      IKEALNLLDDM                     EEVEVV              CVQTRLKIFEQGL
Chicken GM-CSF                                    LEEITSHLEST                     GLSSVP              CLRNNLKTFIESL

                                                           αC                              β2/αD
Tetraodon IL-4/13A      ENRKNKKTPEK----------------NHEMKILRDLERYLDFHnv-TCSRVLK--NVTTSTTTELMPQFWEKVERCIQHHNTQKQ 139
Fugu IL-4/13A           QYNNFTKLVVNGK---------------NEREQITKNLKNHLSPQkv-NCSRVLE--HVRPSNMTKPLSVLLKNVATCIRHHNLH-Q 140
Stickleback IL-4/13A    RKH-FGKNSSE---------------EEVKIVRNLEMFNGAQnq-TCAVLLK--NGTTHHNSIQIPRLLEDLVKCIQNTNMMN 135
Medaka IL-4/13A1        KTHTNTARLDTNT---------------KEQRILRNLEAFTA--em-KCEEKPK---MEVNTVKKSITALLDFLKKCSRGRNTFG-LN 139
Medaka IL-4/13A2        KNRGNL-KPKKSE---------------QEQIVIRNLAVFIKTSem-KCENKPK-----TNGNKVPITTLLGFLQDCCKRRLSFGKPNA 139
Zebrafish IL-4/13A      VKKVSGLSGVKF-------------DPFRTDKKLMRNLNTYNHQDek-NCKKEIRSHAADEDLEEITLDVFLKDLLKCVKNINSQKSPRPS 157
Tetraodon IL-4/13B      EEEQC----------------------CQHLIRYLCLYNqqTQNAHC--------HLDRNRTCTLRQLLSHIEECAKTQK 131
Fugu IL-4/13B           NLTDF-----------------------QSKLTKHLHQYNklTNITNC--------HLNMNNTISLKSLLTNITACTKIEL 130
Stickleback IL-4/13B    HIFRSK--------------------VPKVSNLLRLLQDYtr-KPKEDC--------QLNTNRQCHLREFLTRVAECTRRELEINPKT 133
Zebrafish IL-4/13B      KGKIP----------------------RSHLQRQLNAYAAYTgtGNC--------SVSTSGECTMKEFLEKTKACCHYLYSAQRT 127
Human IL-4              RQFYSHHEKDTRCLGATAQQFHRHKQLIRFLKRLDRNLWGLAGl---nSC--------PVKEANQSTLENFLERLKTIMREKYSKCSS 153
Mouse IL-4              RIFYLKHGK-TPCLK-------KNSSVLMELQRLFRAFRCLDSs---iSC--------TMNESKSTSLKDFLESLKSIMQMDYS 140
Chicken IL-4           WESQHCH-------------KNLQGLFLNMRQLLNASSTSLkaPC--------PTAAGNTTSMEKFLADLRTFFHQLAKNK 136
Human IL-13            INVSGC-------------------SAIEKTQRMLSGFCPHKVSAgq-----FSSLHVRDTKIEVAQFVKDLLLHLKKLFREGQFN 146
Mouse IL-13            TNISNC-------------------NAIYRTQRILHGLCNRKAPTtv------SSL--PDTKIEVAHFITKLLSYTKQLFRHGPF 131
Chicken IL-13          DNMTDCK------------------KDYEPLITSLKSLHGMtn------C---------PPSTDNEIYLRNFLPALGNYTQALYRRISATAAN 138
Human IL-2                                        ISNINVIVLELKG                   TATIVEFLNRWITFCQSI
Mouse IL-2                                        ISNIRVTVVKLKG                   SATVVDFLRRWIAFCQSI
Chicken IL-2                                      IQNIEKNLKSLTG                   KKKFPDFLHELTNFVRYL
Fugu IL-2                                         IQKLKNGVKIEKS                   EKPLTNFLDATESLLQQV
Human IL-21                                       IKKLKRKPPSTNA                   KKPPKEFLERFKSLLQKM
Mouse IL-21                                       VAQLRRRLPARRG                   KRTPKEFLERLKWLLQKM
Chicken IL-21                                     VKILRRPFLPVSE                   RKKPQEFLNSFSKLMQKL
Fugu IL-21                                        YRSLTNTLTEKGL                   EEKAEKFFDRLNSLIQKV
Human IL-5                                        VERLFKNLSLIKK                   RRRVNQFLDYLQEFLGVM
Mouse IL-5                                        VEMLFQNLSLIKK                   RRRTRQFLDYLQEFLGVM
Chicken IL-5                                      FGTFFQSFDKLRP                   RKNVKKFIEKLRTFIRKL
Human GM-CSF                                      LTKLKGPLTMMAS                   IITFESFKENLKDFLLVI
Mouse GM-CSF                                      FTKLKGALNMTAS                   VTTYADFIDSLKTFLTDI
Chicken GM-CSF                                    GTEEESGIVFQLN                   QVSREKFKEALKTFFIYL
```

Fig. 5 Alignment of IL-4/13 lineage molecules with structural framework fragments of several other short-chain type I cytokines. α-Helix and β-strand regions determined for human IL-4 (Powers et al. 1993) and human IL-13 (Eisenmesser et al. 2001) are *underlined*; the border between human IL-4 β2 and αD is between threonine and leucine. *Green font*, (predicted) leader sequence; *other colors* are explained in the main text

teleost lines into IL-4 and IL-13 versus IL-4/13A and IL-4/13B. However, the sequence motifs are probably too few to be conclusive on this mode of evolution.

Homology comparison of Zebrafish IL-4/13B with the NCBI protein database (blastp analysis) retrieved various tetrapod IL-4 sequences as top matches (top score 39.7 bits, identity ~27%). In contrast, the other teleost IL-4/13 sequences did not retrieve tetrapod IL-4 or IL-13 among their top matches and showed only minor scores with these proteins (<30 bits). We assume that the somewhat higher similarity of zebrafish IL-4/13B with tetrapod IL-4 is only the result of similar fluctuation around a sequence theme

and does not relate to their closer common ancestry within the IL-4/13 lineage.

Tertiary structural modeling of teleost IL-4/13 suggests the presence of four stable α-helices in "up-up-down-down" orientation and two disulfide bridges

When the various teleost IL-4/13 molecules were analyzed for the likelihood of tertiary structures by the 3D-PSSM software, many retrieved IL-4 and/or IL-13 sequences among the best matches (not shown), but only zebrafish IL-4/13 sequences retrieved them as a true top match. The top match of zebrafish IL-4/13B was human IL-4 with the highly significant PSMM score 4.25e−0.7. The top match of zebrafish IL-4/13A was human IL-13, with PSMM score 1.44. To investigate the likelihood of a disulfide bridge in ABloop–αD, which is unique for teleost IL-4/13 (Fig. 5, green), we continued 3D modeling of zebrafish IL-4/13A (since, as a teleost exception, this cysteine pair lacks in zebrafish IL-4/13B). To this end, zebrafish IL-4/13A and human IL-13 were aligned inspired by the 3D-PSSM result, which in addition to the expected helix regions αA, αB, αC, and αD, predicted an additional α-helix in the CD loop

(supplementary Fig. 4). The so aligned structure was investigated by Insight II 2000 software, and the initial predicted conformation was built to satisfy the requirement of existence of the extra helix. However, when forming the two predicted disulfide bridges in molecular dynamics simulation, the extra α-helix melted away (thick yellow stretch in Fig. 6), while αA, αB, αC, and αD were stably maintained (Fig. 6). The result shows that these four α-helices are energetically stable and suggests reliability of existence of the disulfide bridges C44–C120 (Fig. 5, green) and C50–C92 (Fig. 5, gray). The 3D-PSSM blast results for the various teleost IL-4/13 sequences and the Insight II 2000 modeling result for zebrafish IL-4/13A collectively support that teleost IL-4/13 molecules have a four α-helical "up-up-down-down" structure typical of short-chain type I cytokines and, in most cases, two disulfide bridges.

Conclusion

Teleost short-chain type I cytokine genes were found at genomic loci orthologous to tetrapod *IL-4* and *IL-13*. In contrast to the tandemly organized tetrapod *IL-4* and *IL-13*,

Fig. 6 3D modeling by Insight II 2000 software of zebrafish IL-4/13A (*left*) in comparison with human IL-13 (*right*). Stretches with α-helix potential according to 3D-PSSM software are indicated by *thick coloring* (as in supplementary Fig. 4); upon formation of the disulfide bridges Cys44-Cys120 and Cys50-Cys92 the α-helices A-D are stable but the thick yellow stretch does not form an α-helix. *Above*, top view from side with N termini of helices A and B and C termini of helices C and D. *Below*, side view with ascending helices A and B and descending helices C and D. *N-term.*, N terminus of protein; *C-term.*, C terminus of protein; *green* and *red*, sulfur and other atoms in cysteines forming disulfide bridges
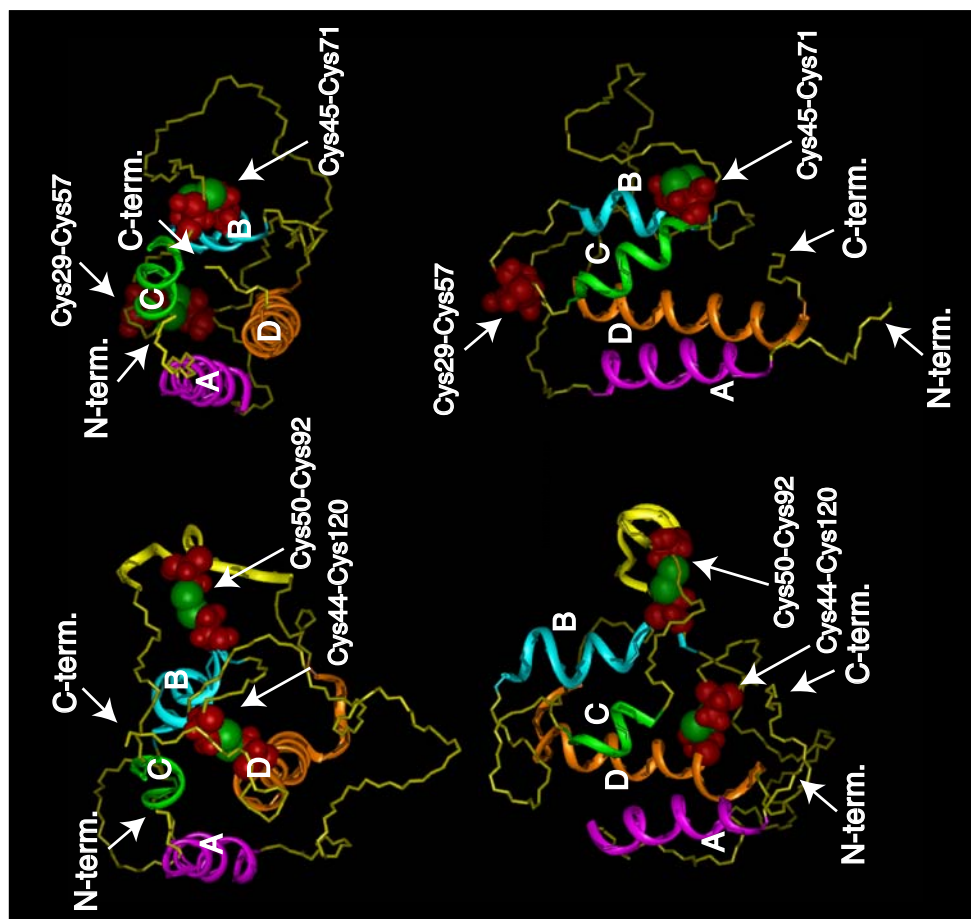
most teleost *IL-4/13* genes are organized as singletons. However, presumably due to the fish-specific WGD (FS3R WGD), teleost fishes acquired two IL-4/13 lineage loci as well, designated *IL-4/13A* and *IL-4/13B*. Peculiar is that the duplicated region including teleost *IL-4/13* was translocated to a chromosome with which it shares earlier paralogy, probably from 2R WGD. GATA-3 binding motifs in the promoter regions of teleost *IL-4/13* support that these genes encode Th2 cytokines indeed. The positioning of the GATA-3 binding motifs and the levels of expression inferred from the representation in the EST database suggest that the expression of teleost *IL-4/13A* versus teleost *IL-4/13B* somewhat resembles that of tetrapod *IL-13* versus tetrapod *IL-4* (the promoter GATA-3 binding motif is more important for *IL-13*, and this gene is expressed at higher levels than *IL-4*). The teleost IL-4/13 molecules share typical IL-4/13 lineage motifs determined from tetrapod IL-4 and IL-13, the lineage signature being most pronounced in zebrafish IL-4/13B. The teleost versus tetrapod IL-4/13 lineage molecules are distinguishable by a characteristic cysteine pair each. Probably, there is insufficient information to conclude if tetrapod *IL-4* and *IL-13* duplicated before or after fish and tetrapods separated. Future studies of teleost IL-4/13 should investigate if T lymphocytes in fish have a Th2 differentiation potential.

References

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714

Ansel KM, Djuretic I, Tanasa B, Rao A (2006) Regulation of Th2 differentiation and Il4 locus accessibility. Annu Rev Immunol 24:607–656

Avery S, Rothwell L, Degen WD, Schijns VE, Young J, Kaufman J, Kaiser P (2004) Characterization of the first nonmammalian T2 cytokine gene cluster: the cluster contains functional single-copy genes for IL-3, IL-4, IL-13, and GM-CSF, a gene for IL-5 that appears to be a pseudogene, and a gene encoding another cytokinelike transcript, KK34. J Interferon Cytokine Res 24:600–610

Bazan JF (1990) Structural design and molecular evolution of a cytokine receptor superfamily. Proc Natl Acad Sci USA 87:6934–6938

Bei JX, Suetake H, Araki K, Kikuchi K, Yoshiura Y, Lin HR, Suzuki Y (2006) Two interleukin (IL)-15 homologues in fish from two distinct origins. Mol Immunol 43:860–869

Bird S, Zou J, Kono T, Sakai M, Dijkstra JM, Secombes C (2005) Characterisation and expression analysis of interleukin 2 (IL-2) and IL-21 homologues in the Japanese pufferfish, *Fugu rubripes*, following their discovery by synteny. Immunogenetics 56:909–923

Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y (2006) The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol 7:R43

Bondensgaard K, Breinholt J, Madsen D, Omkvist DH, Kang L, Worsaae A, Becker P, Schiødt CB, Hjorth SA (2007) The existence of multiple conformers of interleukin-21 directs engineering of a superpotent analogue. J Biol Chem 282:23326–23336

Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol 23:1808–1816

Cao L, Ding X, Yu W, Yang X, Shen S, Yu L (2007) Phylogenetic and evolutionary analysis of the septin protein family in metazoan. FEBS Lett 581:5526–5532

Chomarat P, Banchereau J (1998) Interleukin-4 and interleukin-13: their similarities and discrepancies. Int Rev Immunol 17:1–52

de Vries JE (1998) The role of IL-13 and its receptor in allergy and inflammatory responses. J Allergy Clin Immunol 100:792–801

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:e314

Eisenmesser EZ, Horita DA, Altieri AS, Byrd RA (2001) Solution structure of interleukin-13 and insights into receptor engagement. J Mol Biol 310:231–241

Ekker M, Akimenko MA, Allende ML, Smith R, Drouin G, Langille RM, Weinberg ES, Westerfield M (1997) Relationships among msx gene structure and function in zebrafish and other vertebrates. Mol Biol Evol 14:1008–1022

Fang W, Xiang LX, Shao JZ, Wen Y, Chen SY (2006) Identification and characterization of an interleukin-15 homologue from *Tetraodon nigroviridis*. Comp Biochem Physiol B Biochem Mol Biol 143:335–343

Furuhashi Y, Iwata T, Sikorski P, Atkins E, Doi Y (2000) Structure and morphology of the aliphatic polyester poly-b-propiolactone in solution-grown chain-folded lamellar crystals. Macromolecules 33:9423–9431

Gamse JT, Kuan YS, Macurak M, Brösamle C, Thisse B, Thisse C, Halpern ME (2005) Directional asymmetry of the zebrafish epithalamus guides dorsoventral innervation of the midbrain target. Development 132:4869–4881

Gu X, Wang Y, Gu J (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 31:205–209

Hage T, Sebald W, Reinemer P (1999) Crystal structure of the interleukin-4/receptor alpha chain complex reveals a mosaic binding interface. Cell 97:271–281

Hillier LW, Miller W, Birney E et al (2004) Genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432:695–712

Ho IC, Pai SY (2007) GATA-3—not just for Th2 cells anymore. Cell Mol Immunol 4:15–29

Howard M, Farrar J, Hilfiker M, Johnson B, Takatsu K, Hamaoka T, Paul WE (1982) Identification of a T cell-derived b cell growth factor distinct from interleukin 2. J Exp Med 155:914–923

Huising MO, Kruiswijk CP, Flik G (2006) Phylogeny and evolution of class-I helical cytokines. J Endocrinol 189:1–25

Itoh N, Ornitz DM (2004) Evolution of the Fgf and Fgfr gene families. Trends Genet 20:563–569

Jaillon O, Aury JM, Brunet F, Petit JL et al (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957

Kasahara M, Naruse K, Sasaki S et al (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719

Kelley LA, MacCallum RM, Sternberg MJ (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299:499–520

Ko LJ, Engel JD (1993) DNA-binding specificities of the GATA transcription factor family. Mol Cell Biol 13:4011–4022

Lavenu-Bombled C, Trainor CD, Makeh I, Romeo PH, Max-Audit I (2002) Interleukin-13 gene expression is regulated by GATA-3 in T cells: role of a critical association of a GATA and two GATG motifs. J Biol Chem 277:18313–18321

Leonard WJ (2003) Type I cytokines and interferons and their receptors. In: Paul WE (ed) Fundamental Immunology. 5th edn. Lippincott Williams & Wilkins, Philadelphia, pp 701–747

Li JH, Shao JZ, Xiang LX, Wen Y (2007) Cloning, characterization and expression analysis of pufferfish interleukin-4 cDNA: the first evidence of Th2-type cytokine in fish. Mol Immunol 44:2078–2086

Liongue C, Ward AC (2007) Evolution of Class I cytokine receptors. BMC Evol Biol 7:120

Liu TX, Zhou Y, Kanki JP, Deng M, Rhodes J, Yang HW, Sheng XM, Zon LI, Look AT (2002) Evolutionary conservation of zebrafish linkage group 14 with frequently deleted regions of human chromosome 5 in myeloid malignancies. Proc Natl Acad Sci USA 99:6136–6141

McLysaght A, Hokamp K, Wolfe KH (2002) Extensive genomic duplication during early chordate evolution. Nat Genet 31:200–204

Minty A, Chalon P, Derocq JM, Dumont X, Guillemot JC, Kaghad M, Labit C, Leplatois P, Liauzun P, Miloux B et al (1993) Interleukin-13 is a new human lymphokine regulating inflammatory and immune responses. Nature 362:248–250

Morgan RO, Jenkins NA, Gilbert DJ, Copeland NG, Balsara BR, Testa JR, Fernandez MP (1999) Novel human and mouse annexin A10 are linked to the genome duplications during early chordate evolution. Genomics 60:40–49

Moy FJ, Diblasio E, Wilhelm J, Powers R (2001) Solution structure of human IL-13 and implication for receptor binding. J Mol Biol 310:219–230

Murata T, Obiri NI, Puri RK (1998) Structure of and signal transduction through interleukin-4 and interleukin-13 receptors (review). Int J Mol Med 1:551–557

Murphy KM, Reiner SL (2002) The lineage decisions of helper T cells. Nat Rev Immunol 2:933–944

Nakatani Y, Takeda H, Kohara Y, Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res 17:1254–1265

Nelms K, Keegan AD, Zamorano J, Ryan JJ, Paul WE (1999) The IL-4 receptor: signaling mechanisms and biologic functions. Annu Rev Immunol 17:701–738

Obiri NI, Debinski W, Leonard WJ, Puri RK (1995) Receptor for interleukin 13 Interaction with interleukin 4 by a mechanism that does not involve the common gamma chain shared by receptors for interleukins 2, 4, 7, 9, and 15. J Biol Chem 270:8797–8804

Ohno S (1970) Evolution by gene duplication. George Allen and Unwin, London

Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P (1998) Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 15:1145–1159

Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM (1993) The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 32:6744–6762

Rabie TS, Crooijmans RP, Morisson M, Andryszkiewicz J, van der Poel JJ, Vignal A, Groenen MA (2004) A radiation hybrid map of chicken Chromosome 4. Mamm Genome 15:560–569

Ranganath S, Ouyang W, Bhattarcharya D, Sha WC, Grupe A, Peltz G, Murphy KM (1998) GATA-3-dependent enhancer activity in IL-4 gene regulation. J Immunol 161:3822–3826

Read LR, Raynard SJ, Ruksc A, Baker MD (2004) Gene repeat expansion and contraction by spontaneous intrachromosomal homologous recombination in mammalian cells. Nucleic Acids Res 32:1184–1196

Robinson-Rechavi M, Boussau B, Laudet V (2004) Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. Mol Biol Evol 21:580–586

Rozwarski DA, Gronenborn AM, Clore GM, Bazan JF, Bohm A, Wlodawer A, Hatada M, Karplus PA (1994) Structural comparisons among the short-chain helical cytokines. Structure 2:159–173

Sachs AB (1993) Messenger RNA degradation in eukaryotes. Cell 74:413–421

Secombes CJ, Bird S, Hong S, Laing KJ, Zou J (2001) Phylogeny of vertebrate cytokines. Adv Exp Med Biol 484:89–94

Sémon M, Wolfe KH (2007) Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. Trends Genet 23:108–112

Stolte EH, van Kemenade BM, Savelkoul HF, Flik G (2006) Evolution of glucocorticoid receptors with different glucocorticoid sensitivity. J Endocrinol 190:17–28

Takizawa F, Araki K, Kobayashi I, Moritomo T, Ototake M, Nakanishi T (2008) Molecular cloning and expression analysis of T-bet in ginbuna crucian carp (Carassius auratus langsdorfii). Mol Immunol 46:127–136

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

Umland SP, Razac S, Shah H, Nahrebne DK, Egan RW, Billah MM (1998) Interleukin-5 mRNA stability in human T cells is regulated differently than interleukin-2, interleukin-3, interleukin-4, granulocyte/macrophage colony-stimulating factor, and interferon-gamma. Am J Respir Cell Mol Biol 18:631–642

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA 101:1638–1643

Walter MR, Cook WJ, Zhao BG, Cameron RP Jr, Ealick SE, Walter RL Jr, Reichert P, Nagabhushan TL, Trotta PP, Bugg CE (1992) Crystal structure of recombinant human interleukin-4. J Biol Chem 267:20371–20376

Wang T, Holland JW, Carrington A, Zou J, Secombes CJ (2007) Molecular and functional characterization of IL-15 in rainbow trout Oncorhynchus mykiss: a potent inducer of IFN-gamma expression in spleen leukocytes. J Immunol 179:1475–1488

Wang Y, Shen BJ, Sebald W (1997) A mixed-charge pair in human interleukin 4 dominates high-affinity interaction with the receptor alpha chain. Proc Natl Acad Sci USA 94:1657–1662

Wlodawer A, Pavlovsky A, Gustchina A (1992) Crystal structure of human recombinant interleukin-4 at 2.25 A resolution. FEBS Lett 309:59–64

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet 2:333–341

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

Wraith A, Tornsten A, Chardon P, Harbitz I, Chowdhary BP, Andersson L, Lundin LG, Larhammar D (2000) Evolution of the neuropeptide Y receptor family: gene and chromosome duplications deduced from the cloning and mapping of the five receptor subtype genes in pig. Genome Res 10:302–310

Wynn TA (2003) IL-13 effector functions. Annu Rev Immunol 21:425–456

Yoshiura Y, Kiryu I, Fujiwara A, Suetake H, Suzuki Y, Nakanishi T, Ototake M (2003) Identification and characterization of Fugu orthologues of mammalian interleukin-12 subunits. Immunogenetics 55:296–306

Zou J, Yoshiura Y, Dijkstra JM, Sakai M, Ototake M, Secombes C (2004) Identification of an interferon gamma homologue in Fugu, *Takifugu rubripes*. Fish Shellfish Immunol 17:403–409

Zou J, Carrington A, Collet B, Dijkstra JM, Yoshiura Y, Bols N, Secombes C (2005) Identification and bioactivities of IFN-gamma in rainbow trout *Oncorhynchus mykiss*: the first Th1-type cytokine characterized functionally in fish. J Immunol 175:2484–2494

Zurawski G, de Vries JE (1994) Interleukin 13, an interleukin 4-like cytokine that acts on monocytes and B cells, but not on T cells. Immunol Today 15:19–26