Journal Pre-proof

Expansion of fish CCL20_like chemokines by genome and local gene duplication: Characterisation and expression analysis of 10 CCL20_like chemokines in rainbow trout (*Oncorhynchus mykiss*)

Fuguo Liu, Tingyu Wang, Yehfang Hu, Guangming Tian, Christopher J. Secombes, Tiehui Wang

PII: S0145-305X(19)30378-7

DOI: https://doi.org/10.1016/j.dci.2019.103502

Reference: DCI 103502

To appear in: *Developmental and Comparative Immunology*

Received Date: 14 August 2019

Revised Date: 25 September 2019

Accepted Date: 25 September 2019

Please cite this article as: Liu, F., Wang, T., Hu, Y., Tian, G., Secombes, C.J., Wang, T., Expansion of fish CCL20_like chemokines by genome and local gene duplication: Characterisation and expression analysis of 10 CCL20_like chemokines in rainbow trout (*Oncorhynchus mykiss*), *Developmental and Comparative Immunology* (2019), doi: https://doi.org/10.1016/j.dci.2019.103502.

Expansion of fish CCL20_like chemokines by genome and local gene duplication: Characterisation and expression analysis of 10 CCL20_like chemokines in rainbow trout (*Oncorhynchus mykiss*)

Fuguo Liu[1], Tingyu Wang[1], Yehfang Hu[1], Guangming Tian[1,2], Christopher J. Secombes[1*], Tiehui Wang[1*]

[1]*Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen, AB24 2TZ, United Kingdom*

[2] School of Animal Science, Yangtze University, Jingzhou, 434020, P. R. China

***Corresponding authors:**

Dr. Tiehui Wang, E-mail: t.h.wang@abdn.ac.uk

Dr. Christopher J. Secombes, E-mail: c.secombes@abdn.ac.uk

ABSTRACT

Mammalian CCL20, or macrophage inflammatory protein-3α, can function as a homeostatic and inflammatory chemokine. In relation to the latter, it is responsible for the chemoattraction of lymphocytes and dendritic cells to mucosal immune sites under inflammatory and pathological conditions. CK1, CK8A and CK8B are rainbow trout (*Oncorhynchus mykiss*) CC chemokines that were reported previously to be phylogenetically related to mammalian CCL20. In the current study, an additional seven CCL20_L paralogues in rainbow trout are reported, that are divided into three subgroups and have been designated here as: CCL20_L1a (also referred to as CK1), CCL20_L1b1-2, CCL20_L2a (CK8A), CCL20_L2b (CK8B), CCL20_L3a, and CCL20_L3b1-4. Multiple CCL20_L genes were also identified in other salmonids that arose from both whole genome duplication and local gene duplication. Phylogenetic tree, homology and synteny analysis support that CCL20_L1-3 found in salmonids are also present in most of teleosts that arose from the 3R whole genome duplication and in some species, local gene duplication. Like mammalian CCL20, rainbow trout CCL20_L molecules possess a high positive net charge with a pI of 9.34-10.16, that is reported to be important for antimicrobial activity. Rainbow trout CCL20_L paralogues are differentially expressed and in general highly expressed in mucosal tissues, such as gills, thymus and intestine. The expression levels of rainbow trout CCL20_L paralogues are increased during development and following PAMP/cytokine stimulation. For example, in RTS-11 cells CCL20_L3b1 and CCL20_L3b2 are highly up-regulated by LPS, Poly I:C, recombinant(r) IFNa and rIL-1β. Trout CCL20_L paralogues are also increased after *Yersinia ruckeri* infection or Poly I:C stimulation *in vivo*, with CCL20_L3b1 and CCL20_L3b2 again highly up-regulated. Overall, this is the first report of the complete CCL20 chemokine subfamily in rainbow trout, and the analysis of their expression and modulation *in vitro* and *in vivo*. These results suggest that teleosts possess divergent CCL20_L molecules that may have important roles in anti-viral/ anti-bacterial defence and in mucosal immunity.

KEYWORDS: Rainbow trout, CCL20_L, chemokine, characterisation, expression, mucosal immune response

1. Introduction

Chemokines belong to a superfamily of small, secreted proteins that are important for regulating cell migration under both physiological and inflammatory conditions (Fu et al., 2017; Frick et al., 2016; Lally et al., 2003). In addition to this function in immunity, chemokines play important roles in normal development and growth, such as embryonic development, angiogenesis and organogenesis (Arenberg et al., 1997; Baoprasertkul et al., 2005; Zhou et al., 2018), and in tumour growth and metastasis (Ghadjar et al., 2009; Benkheil et al., 2018). According to the arrangement of cysteine residues in the N-terminal region, they can be divided into five subfamilies, CXC, CC, CX3C, XC and CX (fish specific) (Bacon et al., 2002; Lally et al., 2003; Nomiyama et al., 2008; Chen et al., 2013; Yuan et al., 2019). The CC and CXC subfamilies represent the two largest groups of chemokines.

The CC chemokines are the largest subfamily, with 28 members (CCL1-28) present in mammalian species. The number of CC chemokines has been expanded in teleosts, with 81 and 64 chemokine genes identified in the genomes of zebrafish and catfish, respectively (Bird and Tafalla, 2015; Nomiyama et al., 2008; Fu et al., 2017). Due to lineage-specific expansion and fast diversification, the orthologous relationship between fish and mammalian CC chemokines is difficult to establish. Seven groups of CC chemokines have been proposed in teleosts, the CCL20, CCL27/28, CCL19/21/25, CCL17/22, the macrophage inflammatory protein (MIP) and the monocyte chemotactic protein (MCP) groups, named according to their relationship with mammalian CC chemokines, and a fish-specific group, (Peatman and Liu, 2007; Laing and Secombes, 2004; Fu et al., 2017). The chemokine genes are present on chromosomes in a cluster (e.g. human MIP and MCP chemokines are grouped on Ch 17), a mini-cluster (e.g. human CCL19/21/27 on Ch 9), or singly/non-clustered (e.g. human CCL20, CCL25 and CCL28) (Nomiyama et al., 2010). Whilst the non-clustered and most of the mini-clustered chemokines are relatively conserved and are homeostatic (constitutively expressed) or dual-functional chemokines, the large number of clustered chemokines are often inflammatory (upregulated under inflammatory conditions) and show lineage-specific expansion (Nomiyama et al., 2008; Fu et al., 2017).

Mammalian CCL20 is a non-clustered dual-functional chemokine. Human CCL20, also known as macrophage inflammatory protein (MIP)-3α, exodus-1, and liver- and activation-regulated chemokine (LARC), was discovered independently by three research groups in 1997 (Rossi et al., 1997; Hromas et al., 1997; Hieshima et al., 1997). It is chemotactic for immature dendritic cells (DC), effector or memory CD4(+) T lymphocytes, and B lymphocytes that express CCR6, the only known receptor for CCL20 (Zhao et al., 2014). As a dual-functional chemokine, CCL20 is constitutively expressed by cells of mucosal related tissues, such as epithelial cells (keratinocytes, pulmonary epithelial cells, and intestinal epithelial cells) and by cells in some human organs such as lungs, lymph nodes, and appendix-associated lymphoid tissue, but not in spleen or bone marrow (Frick et al., 2016; Schmuth et al., 2002; Zhao et al. 2014). Under inflammatory conditions, CCL20 can be strongly induced in most

types of immune cells, including monocytes, macrophages, T lymphocytes (T helper 17 cells and CD8 T cells), DCs, neutrophils, eosinophils, mast cells, epithelial cells and melanocytes (Zhang et al, 2019; Zhao et al. 2014). Meanwhile, CCR6 is found on Th17 cells, memory CD4 and CD8 T cells, various B cell subtypes and other immune cells. Therefore, the CCL20/CCR6 axis is involved in the maintenance of immunological homeostasis, especially at mucosal sites and secondary lymphoid tissues, in inflammation and autoimmunity, and in regulation of effective humoral responses (Zhang et al, 2019; Zhao et al. 2014).

A single non-clustered CCL20 gene is present in mammals and birds, and multiple genes related to mammalian CCL20 have been reported in several teleost species (Nomiyama et al., 2010; Fu et al., 2017). Rainbow trout CC chemokine (CK)1 was the first chemokine sequence discovered in teleost fish (Dixon et al., 1998). It possesses 6 cysteine residues in contrast to the 4 conserved cysteine residues seen in mammalian CCL20 but nevertheless has been shown to be phylogenetically related to CCL20 (Laing and Secombes, 2004), as are trout CK8a and CK8b that do have 4 conserved cysteines. More recently, four CCL20 like genes have been discovered in the catfish genome (Fu et al., 2017). Although the amino acid sequences of fish CCL20 share poor identities with mammalian CCL20, their three-dimensional structure and tissue distribution are relatively well conserved (Lally et al., 2003; Fu et al., 2017; Mo et al., 2015; Leu et al., 2019). Despite the observation of an increased number of fish CCL20 genes, the contributions of lineage-specific whole genome duplications (WGD), e.g. additional third round (3R) WGD that have occurred at the base of the teleosts, and again in particular fish groups (e.g. 4R WGD in salmonids and cyprinids), and local gene duplications are not clear. Thus, investigation of the genomic location of CCL20 genes and comparison of the syntenic relationship with neighbouring genes will greatly help trace the origins of CCL20 genes in teleosts and higher vertebrates, and shed light on their modes of duplication and diversification.

In the present study, we identified 10 CCL20 like (CCL20_L) genes in the recently released rainbow trout genome, including the three genes (CK1, CK8a and CK8b) known before. These ten trout CCL20_L genes could be mapped to four chromosomes. Multiple CCL20_L genes have also been identified in other fish species, including other salmonids, Northern pike and spotted gar. Synteny and phylogenetic tree analysis clearly demonstrated that human and fish CCL20 genes arose from a common CCL20 ancestor before their divergence during vertebrate evolution. Teleost lineage-specific expansion via local gene duplications and WGD led to three groups (CCL20_L1-3) of divergent CCL20 genes in this fish group. We next investigated the tissue expression pattern of the ten trout CCL20 genes individually, and found that they were differentially expressed in tissues with high levels of expression found in mucosal related tissues. Lastly, we studied their modulation during developmental and after stimulation *in vivo* with Poly I:C, and *Yersinia ruckeri*, and *in vitro* with PAMPs and recombinant proinflammatory cytokines.

2. Materials and methods

2.1. Fish

Healthy rainbow trout were purchased from College Mill Trout Farm (Almondbank, Perthshire UK) and maintained in aerated fiberglass tanks supplied with a continuous flow of recirculating freshwater at 14 °C. Fish were fed twice daily on a commercial pellet diet (EWOS) and acclimatized for at least two weeks prior to use.

2.2. Gene identification and sequence analysis

Three CCL20_L genes (CK1, CK8a and CK8b) are known in rainbow trout (Dixon et al., 1998; Laing and Secombes, 2004). To identify additional CCL20_L genes in this species, we searched the recently released rainbow trout reference genome (GCF_002163495.1) using TBLASTN (Altschul et al., 1997) with the known trout CCL20_L genes as query, resulting in the identification of four genomic loci (Chromosomes (Ch) 8, 11, 15 and 28) that harbour CCL20_L genes. The 10 CCL20_L genes were then predicted as described previously (Wang et al., 2018; 2019a). Primers (supplementary **Table S1**) were subsequently designed in the predicted 5'- and 3'-UTRs for PCR cloning of the complete coding region of each predicted CCL20_L gene except CCL20_L1a (CK1). CCL20_L2a (CK8a) and CCL20_L2b (CK8b) were also re-cloned as they were previously predicted from expressed sequence tags (EST) with an incomplete coding region (Laing and Secombes, 2004). The general cloning and sequence analysis was as described previously (Wang et al., 2018; 2019a). The nucleotide sequences generated were assembled and analysed with the AlignIR programme (LI-COR, Inc.). The gene organization was predicted using the Splign program at NCBI. Protein translation was performed using Virtual Ribosome-version 2.0 and signal peptides predicted using the program SignalP4.1 Server. Identity and similarity analysis were performed using the matrix BLOSUM62 within the MatGAT program (Campanella et al., 2003), with a gap open penalty of 10 and gap extension penalty of 1. Multiple aa alignment was performed using Cluster Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and the conserved amino acids were shaded using BoxShade program (https://embnet.vital-it.ch/software/BOX_form.html). Disulfide bonding and cysteine connectivity were predicted using the Disulfide program (Ceroni et al., 2006). The Genomicus (database version 96.01, Muffato et al., 2010) was used to analyse the synteny of the CCL20 loci. Phylogenetic trees were constructed by the Neighbour-joining method using the MEGA7.0 program on full-length aa alignments and bootstrapped 10,000 times. The evolutionary distances were computed using the JTT matrix-based method with all ambiguous positions removed for each sequence pair.

2.3. Tissue distribution of expression of rainbow trout CCL20_L paralogues

Seventeen tissues (blood, thymus, gills, scales, skin, muscle, tail fins, adipose fin, brain, adipose tissue, spleen, liver, heart, intestine, gonad, head kidney (HK) and caudal kidney) of six healthy rainbow trout were collected and processed as described previously (Wang et al., 2011a). The RNA extraction, cDNA synthesis and real-time PCR of CCL20_L paralogues were also as described previously (et al., 2011a). In order to ensure that no genomic DNA could be amplified under the real-time PCR conditions used, at least one primer was designed to cross an intron. The details of the primers for qPCR analysis of CCL20_L paralogues are listed in Table 1. A common reference containing equal molar amounts of PCR products of each gene, including the house keeping gene elongation factor-1α (EF-1α), were prepared and used to directly compare the expression level of different CCL20_L paralogues. The expression level of each CCL20_L paralogue was calculated as arbitrary units after normalization against the expression level of EF-1α.

2.4. Ontogeny of the expression of CCL20_L paralogues

The ontogeny of the expression of CCL20_L paralogues was examined to investigate whether the expression of CCL20_L is correlated to immune capacity in early life. The archived samples from different developmental stages from a previous experiment have been described in detail in Wang et al. (2010). Briefly, eyed eggs, immediate post-hatch fry, pre-first feeding fry (at the stage of full disappearance of the yolk sac), and fry 3 weeks following first feeding were sampled and cDNA prepared. Six samples for each developmental stage were prepared and applied to real-time PCR analysis.

2.5. Modulation of the expression of trout CCL20_L paralogues in vivo by Poly I:C

The rainbow trout was stimulated *in vivo* by the viral dsRNA mimic polyinosinic:polycytidylic acid (Poly I:C, Sigma) as described previously (Wang et al., 2019b). Briefly, 12 fish (~100 g) were injected intraperitoneally (ip) with 1 mg Poly I:C in 0.2 ml PBS and another group of 12 fish were given the same amount of PBS as control. Six trout from each group were killed at 6 h and 24 h post injection, and spleen, HK, gills and intestine were collected for gene expression analysis as described previously (Wang et al., 2011a). The results were expressed as arbitrary units (AU) after normalisation to EF-1α, with the average expression level in control fish in each tissue at 6 h set as 1.

2.6. Modulation of the expression of trout CCL20_L paralogues by Yersinia ruckeri challenge

The bacterium *Yersinia ruckeri* is the causative agent of enteric red mouth disease (ERM), that has caused significant economic losses in salmonid aquaculture (Harun et al., 2011). A pathogenic strain (MT3072) of *Y. ruckeri* was obtained from Marine Scotland Science, Marine Laboratory, Aberdeen, UK. The bacteria were collected in PBS from a one-day old tryptic soy agar (TSA, Fluka BioChemika) plate, washed twice with PBS and resuspended in PBS containing 15% glycerol. Aliquots of bacteria were stored at −80 °C and titrated on TSA plates just prior to use. Two groups of fish were ip injected with 0.5 ml of *Y. ruckeri* at a concentration of 1 x 10^6 cfu/ml or injected with the

same volume of PBS as control. Four fish from each group were killed at day 1 and day 2, and gills, spleen, HK and intestine were collected for gene expression analysis as above. The results were expressed as AU after being normalised to EF-1α, with the average expression level in control fish in each tissue at day 1 set as 1.

2.7. Modulation of the expression of trout CCL20_L paralogues in RTS-11 cells by PAMPs and inflammatory cytokines

The rainbow trout monocyte/macrophage cell line RTS-11 (Ganassin and Bols, 1998) was maintained at 20 °C in Leibovitz L-15 medium (Invitrogen, UK) containing 30% foetal calf serum (FCS; Labtech international, UK), 100 U/ml penicillin and 100 mg/ml streptomycin. For experiments, the cells were collected by centrifugation at 200 x *g* for 5 min, washed once with L-15 containing 0.5% FCS, diluted in L-15 containing 10% FCS to a concentration of 1 x 10^6 cells/ml, and seeded at 2 ml/well into 12-well cell culture plates. RTS-11 cells were cultured overnight before stimulation.

Overnight cultures of RTS-11 cells were stimulated with pathogen-associated molecular patterns (PAMPs) and proinflammatory cytokines respectively. For PAMPs stimulation, the bacterial cell wall component lipopolysaccharide (LPS, from *E. coli* strain 055:B5, Sigma) at 25 μg/ml or Poly I:C at 50 μg/ml were added to the cells. For cytokine stimulation, IFNγ (20 ng/ml) (Wang et al., 2011b), IFNa (25 ng/ml, Wang et al., 2019b), IL-1β (25 ng/ml) (Hong et al., 2001), IL-6 (100 ng/ml) (Costa et al., 2011) or TNFα (50 ng/ml) (Hong et al., 2013) were added to the cells. All the cytokines were produced in *E. coli*, purified under denaturing conditions with stringent washing buffer to remove LPS contamination, refolded and re-purified under native conditions, and have been used recently (Wang et al., 2019b). For all treatments, medium alone was added as the control. The treatments were terminated by dissolving the cells in TRI reagent at 4 h, 8 h and 24 h post-stimulation. Total RNA exaction and gene expression were conducted as described above and expressed as a fold change calculated as the average expression level of stimulated samples divided by that of time-matched controls.

2.8. Statistical analysis

The data were analysed statistically using the SPSS Statistics package 24 (SPSS Inc., Chicago, Illinois). The analysis of real-time PCR data was as described previously (Wang et al., 2011a). Briefly, real-time quantitative PCR measurements were scaled, with the lowest expression level set as 1, and log2 transformed. One-way analysis of variance (ANOVA) and the LSD post hoc test were used to analyse the expression data from the same tissue after *in vivo* stimulation, or at the same time points after *in vitro* stimulation, with $P \leq 0.05$ between treatment and control groups considered significant. Additionally, a paired-samples T test was used to statistically analyse the expression of different CCL20_L paralogues in the same tissue, with $P \leq 0.05$ considered significant.

3. Results

3.1. Sequence analysis of CCL20_L paralogues in rainbow trout

10 trout CCL20_L genes have been predicted on four chromosomes. Except CK1 (CCL20_L1a), 9 of the 10 predicted CCL20_L genes have been sequence confirmed by cDNA cloning and named as CCL20_L1b1-2, CCL20_L2a (CK8A), CCL20_L2b (CK8B), CCL20_L3a and CCL20_L3b1-4, respectively, according to their homology and chromosomal location (**Tables 2-3, Supplementary Figs. S1-S9**). Each trout CCL20_L cDNA sequence has a complete open reading frame that encodes for 98-112 aa, with a signal peptide of 24-26 aa and a mature peptide of 74-79 aa. The mature peptide of each trout CCL20_L is 8.92-10.04 kDa with a basic pI of 9.34-10.16 (**Table 2**). A multiple alignment of all trout CCL20_L molecules and human CCL20 suggested general conservation including the four conserved cysteine residues that form two pairs of intra-molecular disulfide bonds (C1-C3 and C2-C4) (**Fig. 1**). However, marked differences do exist between trout CCL20_L1-3. First, trout CCL20_L1 molecules are 6-7 aa shorter before the N-terminal CC motif. Second, there is a four-basic aa (RRRR) motif at the N-terminal of trout CCL20_L3a, L3b3 and L3b4 that may function as a nuclear localisation signal (NLS) (Valgardsdottir et al., 2001). And third, trout CCL20_L1 (a/b1/b2) molecules have six conserved cysteine residues with the extra two cysteine residues potentially forming an additional disulfide bond. The six conserved cysteine residues were found in all CCL20_L1 molecules from other teleosts (**Fig. S10**).

3.2. Sequence analysis of CCL20_L paralogues in other salmonids and Northern pike

10 CCL20_L genes have also been identified or predicted on four homologous chromosomes in chinook salmon as seen in rainbow trout (**Table 3, Figs. S11-13**). However, only 9, 7 and 5 CCL20_L genes have been identified or predicted (**Figs. S14-16**) in the current genomes of coho salmon, Atlantic salmon and Arctic charr, respectively (**Table 3**). Furthermore, four CCL20_L genes, CCL20_L1-2 and CCL20_L3a/b, are present in the genome of Northern pike, the closest 3R relative of salmonids that has a sequenced genome (i.e. without the salmonid 4R WGD) (**Table 3**). The naming of salmonid and pike CCL20_L is in agreement with the phylogenetic tree analysis, genomic organisation and homology. An unrooted neighbour-joining phylogenetic tree revealed three groups of salmonid and pike CCL20_L (1-3) (**Fig. 2**). The pike CCL20_L1 and CCL20_L2 grouped outside of their salmonid counterparts, with salmonid CCL20_L1a/1b and 2a/2b forming independent clades within each group, a classical topology of paralogues that have arisen from the salmonid 4R WGD (Macqueen and Johnston, 2014). This topology is in agreement with the homology analysis that shows salmonid CCL20_L1a/1b and 2a/2b share higher aa identities between orthologues than paralogues (**Table 4**), with their genomic organisation and that each salmonid CCL20_L1a/1b and 2a/2b pair resides on homologous chromosomes (**Fig. S17**). Taken as a whole, these results suggest

that salmonid CCL20_L1a/1b and 2a/2b are 4R WGD paralogues that arose from ancestral genes that gave rise to CCL20_L1 and CCL20_L2 in Northern pike.

Curiously, salmonid CCL20_L3a, L3b1-2 and L3b3-4 formed three independent clades with pike CCL20_L3a/b closer to salmonid CCL20_L3b1-2, and salmonid CCL20_L3b3-4 closer to CCL20_L3a in this group (**Fig. 2**). This tree topology was in agreement with homology analysis that the salmonid CCL20_L3a, L3b1-2 and L3b3-4 molecules shared high aa identities within each group, and that salmonid CCL20_L3a shared higher aa identities to CCL20_L3b3-4 (57.5-69.5%) than to CCL20_Lb1-2 (42.6-48.6%), and pike CC20_L3a/b shared higher aa identities to salmonid CCL20_Lb1-2 (52.3-60.4%) than to salmonid CCL20_L3a and L3b3-4 (43.9-47.2%) (**Table 4**). However, the salmonid CCL20_L3b genes are closely linked at the same genomic locus at a site homologous to the salmonid CCL20_L3a locus that is on a different chromosome, as seen in rainbow trout, chinook salmon and Atlantic salmon (**Fig. S17**). This suggests that salmonid CCL20_L3a and L3b are indeed 4R WGD paralogues with salmonid CCL20_L3b further expanded by local gene duplication. Interestingly, only two CCL20_L3b genes have been found in Atlantic salmon vs four in rainbow trout and chinook salmon, with Atlantic salmon CCL20_L3b1 and L3b3 sharing similar high aa identities to trout CCL20_L3b1-2 and 3b3-4, respectively (**Table 4**). Similarly, only a single CCL20_L3b3 was found in Arctic charr that shares similar high aa identities to trout CCL20_L3b3-4. Although a complete CCL20_L3 repertoire in all salmonids needs to be confirmed once more advanced genome sequence is available, current data suggests that the CCL20_L3b gene was duplicated after the 4R WGD into the CCL20_L3b1/2 and Lb3/4 ancestral genes, which are preserved in some lineage as seen in Atlantic salmon, with subsequent duplication leading to four CCL20_L3b genes, as seen in rainbow trout and chinook salmon.

3.3. Evolutionary analysis of teleost CCL20_L genes

To have an insight into the evolution of teleost CCL20_L genes, we analysed CCL20_L genes in other teleost species in relation to genes in a holostean fish (spotted gar) and humans. Four CCL20_L genes (CCL20-1 to CCL20-4) are present at a single genomic locus in gar (**Table 3**), a 2R fish that represents an ancestral state since these fish have not undergone the teleost 3R WGD. Multiple (2-4) CCL20_L genes have been identified in other teleost (3R) fish species including zebrafish, catfish, Mexican tetra, Asian swamp eel, Japanese flounder, medaka and fugu (**Table 3**). These CCL20_L genes in 3R teleosts are present on two different chromosomes, as seen in zebrafish, pike, swamp eel and flounder (**Table 3**). Synteny analysis revealed significant conservation of the spotted gar CCL20_L locus to three CCL20_L loci on two chromosomes in 3R Northern pike and zebrafish, and to six CCL20_L loci on four chromosomes in 4R rainbow trout (**Fig. 3**). Furthermore, syntenic conservation of genes was also observed between the human CCL20 locus, the gar CCL20_L locus and teleost CCL20_L2 loci (**Fig. 3**). Taken as a whole, teleost CCL20_L genes are orthologues of

mammalian CCL20 and the expansion of CCL20_L genes in 3R/4R teleosts can be attributed to WGDs in addition to local gene duplications.

The aa sequences encoded by the teleost CCL20_L genes have relatively high identity to mammalian CCL3, CCL4 and CCL17, in addition to CCL20. For example, rainbow trout CCL20_L genes share the highest average aa identity of 32.2% to human CCL20, and somewhat lower average identities to human CCL4 (29.0%), CCL3 (28.4%) and CCL17 (28.2%), but relatively low (20.2-27.9%) identities to other human CC chemokines (**Table S2**). Therefore, a phylogenetic tree was generated based on a multiple alignment of aa sequences of teleost and gar CCL20_L molecules and selected CCL20, CCL3, CCL4 and CCL17 sequences from other vertebrates. All teleost CCL20_L molecules grouped with CCL20 genes from other vertebrate groups with high bootstrapping support (82%) and were separated from the CCL17, CCL3 and CCL4 groups (**Fig. 4**), further supporting the orthologue relationship of the teleost and gar CCL20_L molecules with CCL20 molecules in other vertebrates. In the CCL20/CCL20_L clade, independent groups of tetrapod CCL20, fish CCL20_L1 and CCL20_L2 were well supported with bootstrap values of 89-99%. The rest of the teleost and holostean CCL20_L molecules formed the CCL20_L3 group that was more divergent and showed lineage/species expansion/diversification as seen in spotted gar, Mexican tetra, zebrafish, pike and salmonids (**Fig. 4**). Moreover, teleost CCL20_L molecules were present in each of the three CCL20_L groups in most of the fish species analysed, including salmonids, pike, zebrafish, catfish, flounder, swamp eel and Mexican tetra.

Overall, the analysis above suggests that teleost CCL20_L genes arose from a single CCL20 ancestor in an early 2R vertebrate that led to tetrapod CCL20 gene, and duplicated locally in the ancestor of the ray-finned fish lineage (2R) (**Fig. 5**). Further local gene duplication in Holosteans likely led to the 4 CCL20_L genes seen in spotted gar. The two duplicated genes (CCL20_L1/2 and L3/4) were further duplicated by the 3R WGD to generate four CCL20_L genes (CCL20_L1-4) with CCL20_L4 subsequently lost in teleosts. Lineage/species specific local gene duplication/deletion led to different copies of CCL20_L genes as seen in Northern pike (two CCL20_L3 genes) and catfish (two CCL20_L1 genes known as CCL20b.1-2, Fu et al., 2017). The three CCL20_L genes (loci) on two chromosomes in 3R fish were further duplicated by the salmonid 4R WGD that would have generated 6 CCL20_L loci (CCL20_L1a, L1b, L2a, L2b, L3a and L3b) on four chromosomes, with CCL20_L1b and L3b duplicated locally before salmonid speciation. Additional local gene duplication of CCL20_L2b1/2 and Lb3/4 likely led to the 10 CCL20_L genes found in rainbow trout (**Fig. 5**).

3.4. Gene organisation analysis of CCL20_L genes in teleosts

Unlike most CC chemokine genes that possess three coding exons, human and other mammalian CCL20 genes possess four coding exons (Nomiyama et al., 2010). Fish CCL20_L genes also possess

a four coding exon/three intron structure with conserved intron phases (I, II and II), as seen in rainbow trout, Northern pike, spotted gar and humans (**Fig. 6**). The first exon encodes the signal peptide and the other exons encode the mature peptide. The third exon is well conserved with 78 bp in all genes except pike CCL20_L1 that has 81 bp. The second exon was more divergent between the three groups of CCL20_L genes. CCL20_L1 genes possess a smaller exon 2 (103 bp) compared to CCL20_L2 (115-118 bp) and CCL20_L3 (100-115 bp) (**Fig. 6**), leading to fewer aa before the first two cysteine residues (**Fig. 1**).

3.5. Tissue distribution of the expression of trout CCL20_L paralogues

The expression of CCL20_L paralogues was examined in seventeen tissues using gene specific primers and serially diluted references, and expressed as arbitrary units (AU) relative to EF-1α expression multiplied by 1,000,000 (**Fig. 7**). The expression of CCL20_L genes were detectable in all tissues examined, with average expression levels and ratios of paralogue expression, plus statistical analysis detailed in **Table S3**. AU < 100 was considered low level expression, 100 < AU <1,000 medium level expression and AU >1,000 high level expression. CCL20_L1a was highly expressed in caudal kidney, followed by medium levels in liver, heart and brain, and low levels in all other tissues (**Fig. 7A**). CCL20_L1b1/2 were highly expressed in gills, with medium level in thymus, and low levels in other tissues with the exception of medium level expression of CCL20_L1b2 in caudal kidney, adipose fin and brain (**Fig. 7A**). The CCL20_L1 paralogues were differentially expressed in most of the tissues with CCL20_L1a higher in caudal kidney, liver, heart, intestine, adipose tissue and gonad, but lower in mucosal tissues (gills, thymus, adipose fin and scales) (**Table S3**).

CCL20_L2a was expressed highly in gills and intestine, at medium levels in thymus, adipose fin, tail fins, scales, and caudal kidney, and at low levels in other issues (**Fig. 7B**). CCL20_L2b was highly expressed in mucosal tissues, including gills, thymus, adipose fin, tail fins and scales, and caudal kidney, at medium levels in skin, adipose tissue, spleen, brain and muscle, and at low levels in other tissues (**Fig. 7B**). The expression of CCL20_L2b was higher than CCL20_L2a in all tissues except in muscle, gonad and liver (**Table S3**).

CCL20_L3a was highly expressed in mucosal tissues such as gills, tail fins, adipose fin, intestine and thymus, but at low to medium levels in other tissues (**Fig. 7C**). CCL20_L3b1-4 were expressed highly in gills for all molecules, and in thymus for CCL20_L3b1 and in muscle for CCL20_L3b4. Other tissues expressed only low to medium levels of CCL20_L3b1-4 (**Fig. 7C**). The CCL20_L3 paralogues were differentially expressed in most of the tissues with L3a expressed more highly than L3b paralogues in these tissues except in spleen, head kidney and gills (**Table S3**).

In summary, the ten CCL20_L genes were differentially expressed with each tissue preferentially expressing a set of CCL20_L genes. The majority of the genes were highly expressed in mucosal

tissues, with gills having the largest expression levels of all genes except CCL20_L1a that was highest in caudal kidney.

3.6. The expression of trout CCL20_L genes during early developmental stages

Next, we investigated the expression of trout CCL20_L genes during early developmental stages, including eyed-eggs, immediate post-hatch fry, pre-first feeding fry and fry 3 weeks after first feeding, which is reported to be a critical period when the fish encounter potential pathogens from the environment and food (Wang et al., 2010). All genes showed an increase from eyed-eggs to post-hatch fry, although the expression level of CCL20_L3b2 and b4 was very low (**Fig. 8**). With CCL20_L1b1, L1b2, L2a, L2b, L3a, L3b1 and L3b2 this increased further at the pre-feeding stage, and further again in post-feeding fry for CCL20_L1b1. Whilst most of the genes were maintained at these levels, in the case of CCL20_L3b1 the level decreased in the post-feeding fry.

3.7. Modulation of the expression of trout CCL20_L genes in vivo by Poly I:C

The expression of trout CCL20_L genes in gills, spleen, HK and intestine after ip injection with Poly I:C was next investigated. A transient up-regulation at 6 h was observed for CCL20_L1a in spleen, CCL20_L1b2 in gills, and CCL20_L3b1 and L3b2 in spleen, HK and intestine. At 24 h only decreases in transcript level were seen, as with CCL20_L1b1, L1b2, L2a, L2b, L3b1 and L3b2 in spleen, L1b2, L2a, L3b1 and L3b2 in HK, and L1b2, L2b and L3b3 in intestine (**Fig. 9**). In summary, a transient induction of CCL20_L (L1a, L1b2, L3b1 and L3b2) was seen at 6 h post injection, with an inhibition of many CCL20_L genes observed at 24 h after Poly I:C stimulation.

3.8. Modulation of the expression of trout CCL20_L genes by Yersinia ruckeri challenge

Yersinia ruckeri induced the expression of CCL20_L1a in intestine at day 1 and in gills and spleen at day 2. CCL20_L3b1 was also induced in intestine at day 1, in gills from day 1 to day 2, and in spleen at day 2. Lastly, CCL20_L3b2 was also increased, in gills, HK and intestine from day 1 to day 2, and in spleen at day 2. Several genes were inhibited by *Y. ruckeri* infection. In gills this included CCL20_L2a at day 1 and day 2, L3a at day 2, and L3b3 at days 1 and 2. In spleen CCL20_L1b1 and L1b2 at day 1, and L2a and L2b at days 1 and 2 were all downregulated. In HK CCL20_L2a was inhibited at days 1 and 2, and L3b3 at day 2, whilst in intestine CCL20_L1b2, L2a, L2b, L3a and L3b3 were all decreased at day 2. The expression of CCL20_L3b4 was low and refractory. In summary, *Y. ruckeri* infection induced the expression of CCL20_L1a, L3b1 and L3b2 in multiple tissues, especially in mucosal tissues (gills and intestine), but an inhibitory effect was seen on other CCL20_L genes.

3.9. Modulation of trout CCL20_L gene expression in RTS-11 cells by PAMPs

The modulation of trout CCL20_L genes were then investigated in the macrophage cell line RTS-11 using Poly I:C and LPS, classical viral and bacterial PAMPs. Poly I:C could significantly induce the

up-regulation of CCL20_L3b1 and L3b2 from 4 h to 24 h, and that of CCL20_L2a at 8 h (**Fig. 11**). LPS could significantly induce the expression of CCL20_L3b1 and L3b2 from 4 h to 24 h, and that of CCL20_L3b3 from 8 h to 24 h. Both Poly I:C and LPS had little effect on the expression of CCL20_L1a, L1b1, L1b2, L2b and L3a (**Fig. 11**). However, the expression was inhibited at 24 h for CCL20_L2b by LPS and CCL20_L3b3 by Poly I:C. The expression level of CCL20_L3b4 in RTS-11 cells was low and refractory. In summary, both Poly I:C and LPS are good inducers of trout CCL20_L3b1-2 gene expression, with Poly I:C inducing a transient effect (peaked at 4 h) and LPS a longer lasting effect (peaked at later time points).

3.10. Modulation of trout CCL20_L genes by proinflammatory cytokines in RTS-11 cells

As the expression of trout CCL20_L genes can be modulated by PAMPs in RTS-11 cells, we finally examined the modulatory effects of pro-inflammatory cytokines that may also be induced by PAMP stimulation. All the recombinant cytokines used were produced in a similar way, are bioactive and show cytokine-specific effects in terms of modulation of gene expression, as seen previously (Wang et al., 2019b) and in this study. With regards to LPS contamination, the cytokines have been tested for their ability to upregulate known LPS-inducible genes in RTS-11 cells (**Fig. S18**). At least two of the six genes studied are refractory to treatment with any of the recombinant cytokines used, except IL-1β that is known to induce all the genes in RTS-11 cell. These results suggest that the effects observed in the present study are not due to LPS contamination or exogenous protein effects. IFNγ down-regulated the expression of CCL20_L2a at 8 h, L3b1 from 4 h to 8 h, L3b2 at 4 h, and L3b3 from 8 h to 24 h, but had no effect on CCL20_L1a, L1b1, L1b2, L2b and L3a (**Fig. 12**). IFNa induced the expression of CCL20_L2a and L2b at 8 h, CCL20_L3b1-2 from 4 h to 24 h, down-regulated the expression of CCL20_L3b3 at 24 h, had no effect on the expression of CCL20_L1a, L1b1-2 and L3a (**Fig. 12**). IL-1β induced the expression of CCL20_L2a and L3b3 from 4 h to 8 h, CCL20_L1b1 and L3a at 8 h, CCL20_L3b1-2 from 4 h to 24 h, and down-regulated the expression of CCL20_L2b at 24 h, but had no effect on CCL20_L1a and L1b2 (**Fig. 12**). IL-6 down-regulated the expression of CCL20_L2b and 3b2 at 24 h, CCL20_L3b1 at 4 h and 24 h, CCL20_L3b3 at 8 h, weakly up-regulated the expression of CCL20_L2a at 8 h, but had no effect on CCL20_L1a, L1b1-2 and L3a. Lastly, TNFα down-regulated the expression of CCL20_L1a at 4 h and CCL20_L2b at 24 h, had no effect on the other trout CCL20_L genes (**Fig 12**). CCL20_L3b4 expression was again low and refractory. In summary, IL-1β and type I IFNa are good inducers of CCL20_L expression in RTS-11 cells, especially for CCL20_L3b1 and L3b2, but other cytokines (type II IFNγ, IL-6 and TNFα) exhibited mainly inhibitory effects.

4. Discussion

Mammalian CCL20 is a single copy non-clustered dual-functional chemokine involved in the maintenance of immunological homeostasis and effective immune responses. By interrogation of salmonid genomes, we have identified and characterised ten expressed CCL20_L genes (three previously known) in rainbow trout, the largest number of CCL20_L genes known in a species to date. We then confirmed up to 10 CCL20_L genes exist in other salmonids and that multiple genes are commonly present in other teleosts and in spotted gar (holostean). The fish genes can be classified into three groups (CCL20_L1-3) and their expansion is the result of both gene and whole genome duplication during evolution. The ten trout CCL20_L genes are differentially expressed in tissues from healthy fish, with distinct expression patterns between and within groups. Their expression is increased during early developmental stages and differentially upregulated *in vivo* and *in vitro* after stimulation with PAMPs and cytokines, suggesting divergent roles of CCL20_L genes in anti-microbial defence in fish.

4.1. The evolution of CCL20_L genes in teleosts

Only a single CCL20 gene is present in tetrapods (Nomiyama et al., 2010) but up to 10 CCL20_L genes are present in salmonids and multiple genes are common in other teleosts and gar (holostean). Phylogenetic tree and homology analysis suggests three divergent groups of CCL20_L genes in teleosts, with most of the 3R teleosts having a gene belonging to each of CCL20_L1-3. Our gene synteny analysis suggests that these three groups of CCL20_L genes were generated by the 3R WGD from two CCL20_L ancestral genes that had been duplicated locally in a 2R actinopterygian ancestor, from an ancestral CCL20 gene that led to tetrapod CCL20 as seen in humans. One of the four daughter genes/loci (i.e. CCL20_L4) on the two duplicated chromosomes was likely lost early in 3R teleosts, and an insertion between CCL20_L2 and L3 happened that produced two CCL20_L loci on the same chromosome. Local gene duplications seem to have occurred in a lineage/species-specific manner, leading to three groups of CCL20_L genes residing in three loci on two chromosomes in teleosts. This model does not necessarily imply that the four CCL20_L genes of spotted gar were generated from the same two ancestral genes. Whilst gar CCL20-1 shares high aa identity and groups with one of the teleost subgroups (CCL20_L2) with high bootstrap support, the gar CCL20-2-4 genes are divergent and their grouping with teleost CCL20_L3 was not well supported. In addition, the relative transcriptional direction of the four gar genes is the opposite of that of CCL20_L2 and L3, suggesting that multiple rounds of local gene duplication/deletion with subsequent diversification might have occurred in this species.

The 4R WGD in salmonids appears to have generated the six CCL20_L loci on four chromosomes. It seems that all the six CCL20_L loci have been preserved in all extant salmonids analysed in this study. Again, independent local gene duplications are apparent, that happened before and after

salmonid speciation. Two CCL20_L1b (1-2) are present in rainbow trout, Atlantic, chinook and coho salmon that sit together and share high aa identities, representing a local gene duplication that occurred before salmonid speciation. Four CCL20_L3b (1-4) are present in rainbow trout, chinook salmon and likely also in coho salmon (genus Oncorhynchus) but only two have been found in Atlantic salmon (genus Salmo) and only one in Arctic charr (genus Salvelinus) in the current genome version of this species. All the CCL20_L3b paralogues are closely linked. The salmonid CCL20_L3b1/b2, and Lb3/b4 pairs share high aa identities within each pair (e.g. 79-98% aa identities) but only a medium level of aa identity between the two pairs (eg.50-53% aa identities). Hence the two Atlantic CCL20_L3b each belong to a single pair according to aa identity and phylogenetic tree analysis. These data suggest that a local gene duplication occurred in this locus before salmonid speciation to generate the CCL20_L3b1/2 and Lb3/4 ancestors. This state was retained in Salmo (and possibly in Salvelinus species), but further lineage-specific local gene duplications generated CCL20_L3b1-4 in Oncorhynchus species. Taken together, CCL20_L genes have been expanded in teleosts due to whole genome duplications and/or local gene duplications in a lineage/species-specific manner.

4.2. Indicators of functional diversification of teleost CCL20_L genes

Gene and genome duplications increase the genetic material that might then contribute to the increase of genomic and phenotypic complexity of an organism during evolution (Studer et al., 2010; Braasch and Salzburger, 2009; Escalera-Fanjul et al., 2019). Although the most common fate is the deletion of one of the duplicates by non-functionalization (or pseudogenization) to restore the initial state of a single copy, both duplicated copies can be preserved by sub- and/or neo-functionalization resulting in the generation of new or specialized functions (Kassahn et al., 2009; Glasauer and Neuhauss, 2014). There are many features of fish CCL20_L genes that may indicate functional diversification.

The CCL20_L genes of 3R teleost fish can be divided into three groups, with each group having different numbers of genes in different species/lineages. Although teleost CCL20_L molecules share high aa identities within each group from different species, the aa identities between groups are low between the three fish groups, and between mammalian CCL20 and fish CCL20_L, an indication of functional diversification of the three CCL20_L groups (Canestro et al., 2013). CCL20_L3 represent the largest and most divergent fish CCL20_L group as seen in the phylogenetic trees with up to 5 genes in salmonids and 2 in Northern pike. All salmonid CCL20_L3 genes were generated via the 4R salmonid WGD with the CCL20_L3b gene expanded by local gene duplications post the WGD. Therefore, salmonid CCL20_L3b genes should share similar identities to CCL20_L3a if the selection pressure was neutral. Interestingly, salmonid CCL20_L3a share apparent higher aa identities to CCL20_L3b3/4 than to CCL20_L3b1/2 molecules, suggesting asymmetric molecular evolution of CCL20_L3b genes with CCL20_L3b1/2 under adaptive selection with a higher evolutionary rate (Steinke et al., 2006).

Tetrapod CCL20 and fish CCL20_L all share a four exon/three intron gene organisation with conserved intron phases. The second exon of salmonid CCL20_L1 is smaller leading to only 2 aa at the N-terminal before the first two cysteine residues, compared to 8-9 aa for CCL20_L2-3. The N-terminal of a chemokine interacts with the binding pocket of the chemokine receptor leading to ligand recognition and receptor activation (Riutta bet Al., 2018). The different lengths of the N-terminal between the three types of CCL20_L molecules in fish may suggest differences in the binding affinities (and signalling) for their CCR6 receptors, with two copies known in 3R teleosts and 4 copies in salmonids (Mo et al., 2015; Grimholt et al. 2015; unpublished analysis of salmonid genomes). Another surprise is the presence of a potential polybasic NLS (the RRRR motif) only at the N-terminal of salmonid CCL20_L3a, L3b3 and L3b4. NLS in cytokines and cytokine receptors may be important for their function, as demonstrated in mammalian IFN-γ (Subramaniam et al., 2001) and CXCR4 (Bao et al., 2019). The NLS in salmonid CCL20_L3 molecules might be an adaptive evolutionary innovation that needs further investigation.

The majority of chemokines, including CC chemokines, possess four conserved cysteine residues that form two disulfide bonds to stabilise their structure and are important for their function (Legendre et al., 2013). An additional disulfide bond is found in three human CC chemokines that has a structural impact as seen with human CCL1 (aka. I-309, Keizer et al., 2000). An additional pair of cysteine residues are present in CCL20_L1, that is conserved in all 3R teleosts but missing in tetrapod CCL20 and fish CCL20_L2-3, suggesting another adaptive evolutionary innovation in 3R teleosts.

Functional diversification can also be reflected in differential gene expression and modulation that will be discussed later.

4.3. Driving force of teleost CCL20_L gene expansion

It has been proposed that the mechanisms of duplication have effects on the fate of the duplicated genes (Canestro et al., 2013). Hence, genes that are successfully duplicated by small-scale (local) duplication are usually not retained after WGD, and vice versa, due to dosage balance whereby perturbation of the relative ratios of genes in stoichiometric balance is deleterious (Rice and McLysaght, 2017). The dosage balance theory predicts that genes with products that interact with or are in a complex network could be preferentially co-retained as the ratios remaining unchanged, as seen in the fish TNFα/TNFR system where both the ligand and its receptors are present in two copies in 3R teleosts and four copies in 4R salmonids (Hong et al., 2013; 2019). This might also be the case with the expansion of CCL20_L and the receptor CCR6, with all paralogues duplicated after WGDs co-retained in 3R and 4R teleost species. However, teleost CCL20_L genes were also expanded by local gene duplication that may be attributed to the intrinsic adaptive value of cells carrying local duplicates (Rodrigo and Fares, 2018). This adaptive value might be related to the direct antimicrobial activities of CCL20 molecules. For example, human CCL20 possesses antibacterial activity (against

Escherichia coli and *Staphylococcus aureus*) of greater potency than human β-defensin-1 and -2 (Hoover et al., 2002), and may also be involved in anti-viral defence (Lee and Korner, 2017). Although there is no clear sequence similarity between β-defensins and CCL20, they do share similarities in terms of the abundance of cationic residues (basic pI), the presence of disulfide bonds (3 in β-defensins and 2-3 in CCL20_L) and crystal structure, providing molecular and structural properties of shared functions between CCL20 and β-defensins in binding to CCR6 and bactericidal activity (Hoover et al., 2002). The antimicrobial activity of fish CCL20_L molecules remains to be demonstrated. The further increase of CCL20_L1b and L3b by local gene duplications may enhance innate immunity to pathogens in anadromous salmonids that will likely face distinct pathogen repertoires when moving between fresh and salt water.

4.4. Potential role of teleost CCL20_L in mucosal immunity

All the 10 CCL20_L genes are differentially expressed in tissues from healthy fish. Their expression was differentially modulated *in vitro* by PAMPs and proinflammatory cytokines, and *in vivo* by Poly I:C stimulation and bacterial infection in a tissue specific manner, further suggesting functional diversification. Mammalian CCL20 is expressed predominately in mucosal related tissues and liver. The expression of trout CCL20_L2 and L3 paralogues is high in mucosal tissues including gills, thymus and intestine but low in liver. In contrast, the expression of CCL20_L1a in liver is high as reported by Lally et al. (2003) and is higher than any of the other CCL20_L paralogues, but low in mucosal tissues as seen in the present study. These expression patterns may suggest partitioning (subfunctionalisation) of the expression domain of trout CCL20_L genes to allow optimal function. The high level expression of the majority of CCL20_L genes in mucosal tissues suggests that they may play important roles in mucosal immunity in fish.

In agreement with the moderate expression of mammalian CCL20 in lymphoid tissues such as thymus and lymph nodes, and absence in spleen and bone marrow, the expression of all trout CCL20_L genes are low in central immune tissues/organs, including blood, HK and spleen. Interestingly, the expression of several trout CCL20_L genes, including CCL20_L1a, L1b2, L2a/b and L3b3, are 1-3 orders higher in caudal kidney than in HK. Indeed, the highest expression level of CCL20_L1a is found in caudal kidney. The kidney is an indispensable organ in vertebrates that regulates metabolic waste and fluid balance. In mammals the kidney consists of two bean-shaped organs found below the ribcage on either side of the spinal column. Fish kidney is a flattened structure located along the dorsal inner body wall and can be divided into HK and caudal kidney in salmonids. Fish HK is an unique organ analogous to mammalian bone marrow, that is a haematopoietic-lymphoid organ with no/limited renal function (Geven and Klaren, 2017). This site is dedicated to hormone production (from the interrenal and chromaffin cells) and haematopoiesis, including B and T cell development (Zapata et al., 2006). How the lymphocytes produced in the HK exit through the caudal kidney that is intertwined with blood filtering nephrons is not clear. Trout CCL20_L1a is a known chemoattractant

582 to peripheral blood leucocytes (Lally et al., 2003), and is highly expressed in caudal kidney at a level
583 > 2,000-fold higher than in HK. This may create a CCL20_L chemokine gradient from HK to caudal
584 kidney that guides the lymphocytes produced in the HK to exit, as another possible adaptive
585 innovation specific to ray-finned fish.

586

587 ***4.5. Potential role of teleost CCL20_L genes in anti-viral and anti-bacterial defence***

588 The expression of CCL20_L1a, L3b1 and L3b2 is highly induced *in vivo* by Poly I:C stimulation and
589 by *Y. ruckeri* infection, with the last two also highly induced by Poly I:C and LPS in the macrophage-
590 like cell line RTS-11, suggesting that these CCL20_L genes may have a role in anti-viral and anti-
591 bacterial defence. In line with CCL20 induction by IL-1β in mammals (Chabaud et al., 2001), IL-1β is
592 also a potent inducer of the expression of CCL20_L3b1-2 genes in rainbow trout. Type I IFNa, but
593 not type II IFNγ, was found to be another inducer of CCL20_L expression in RTS-11 cells. As both
594 IL-1β and IFNa are induced by Poly I:C and LPS in rainbow trout (Wang et al., 2019b), the Poly I:C
595 and bacterial infection induced CCL20_L gene expression may be mediated by these proinflammatory
596 cytokines.

597 Interestingly, there is a marked difference in terms of expression level in different tissues and
598 inducibility by Poly I:C stimulation/bacterial infection *in vivo* and PAMP/proinflammatory cytokine
599 stimulation *in vitro*. Several genes, e.g. CCL20_L1b2 and L2b, are highly expressed during
600 developmental stages and constitutively in tissues but were not be induced by stimulation *in vivo* and
601 *in vitro*, at least for the molecules used here. In contrast, the transcript level of other genes, e.g.
602 CCL20_L3b1, 3Lb2 and L1a, is not particularly high but is inducible *in vivo* and *in vitro*. Mammalian
603 CCL20 is known to be a dual-functional chemokine that is highly expressed constitutively in some
604 tissues and can be induced by inflammatory conditions. The distinct expression pattern and
605 inducibility of trout CCL20_L genes may suggest a sub-functionalisation of the homeostatic and
606 inflammatory properties of mammalian CCL20.

607 ***4.6. Conclusions***

608 In addition to the three known CCL20_L in trout (CK1, CK8a, CK8b)), seven further CCL20_L genes
609 have been identified in the genome of rainbow trout. Most of the studied teleost fish species possess 3
610 groups of CCL20_L genes residing at three loci on two chromosomes, while in salmonids there are
611 six CCL20_L loci on four chromosomes. Synteny analysis of these loci demonstrated that the
612 CCL20_L genes in trout arose from both whole genome duplication and local gene duplication. Trout
613 CCL20_L genes except CCL20_ L1a are generally highly expressed in mucosal related tissues, which
614 suggest that they may be involved in mucosal immunity. CCL20_L1a is highly expressed in caudal
615 kidney but low in HK, suggesting a role in directing lymphocyte trafficking of cells produced in the
616 HK. CCL20_L3b1 and CCL20_L3b2 were the most responsive to *in vivo* and *in vitro* stimulation,

with CCL20_L1a also induced relatively highly after injection with Poly I:C or infection with *Y. ruckeri*. These paralogues likely have an important role during viral and bacterial defence. The fish CCL20_L3 molecules are particularly novel, divergent and currently understudied.

Acknowledgements

Fuguo Liu was supported by a Newton International Fellowship funded by the Academy of Medical Sciences, UK (AMS, NIF004\1036). Tingyu Wang was supported by the Ministry of Science and Technology, Republic of China (Taiwan) (MOST 107-2917-I-564-019). YH was supported by a PhD Studentship from the Ministry of Education, Republic of China (Taiwan). Guangming Tian was supported financially by the State Scholarship Fund organised by the China Scholarship Council (201808420042).

628 **References.**

629 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997.
630 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic
631 Acids Res. 25: 3389-402.

632 Aquilino, C., Granja, A.G., Castro, R., Wang, T., Abos, B., Parra, D., Secombes, C.J., Tafalla, C.,
633 2016. Rainbow trout CK9, a CCL25-like ancient chemokine that attracts and regulates B cells and
634 macrophages, the main antigen presenting cells in fish. Oncotarget 7: 17547-64.

635 Arenberg, D.A., Polverini, P.J., Kunkel, S.L., Shanafelt, A., Hesselgesser, J., Horuk, R., Strieter,
636 R.M., 1997. The role of CXC chemokines in the regulation of angiogenesis in non‐small cell lung
637 cancer. J. Leukocyte Biol. 62: 554-562.

638 Arockiaraj, J., Bhatt, P., Kumaresan, V., Dhayanithi, N.B., Arshad, A., Harikrishnan, R., Arasu,
639 M.V., Al-Dhabi, N.A., 2015. Fish chemokines 14, 20 and 25: A comparative statement on
640 computational analysis and mRNA regulation upon pathogenic infection. Fish Shellfish Immunol.
641 47: 221-30.

642 Bacon, K., Baggiolini, M., Broxmeyer, H., Horuk, R., Lindley, I., Mantovani, A., et al., 2002.
643 Chemokine/chemokine receptor nomenclature. J. interferon & cytokine Res. 22: 1067-1068.

644 Bao, Y., Wang, Z., Liu, B., Lu, X., Xiong, Y., Shi, J., Li, P., Chen, J., Zhang, Z., Chen, M., Wang,
645 L., Wu, Z., 2019. A feed-forward loop between nuclear translocation of CXCR4 and HIF-1α
646 promotes renal cell carcinoma metastasis. Oncogene 38: 881-895.

647 Baoprasertkul, P., He, C., Peatman, E., Zhang, S., Li, P., Liu, Z., 2005. Constitutive expression of
648 three novel catfish CXC chemokines: homeostatic chemokines in teleost fish. Mol. Immunol. 42:
649 1355-1366.

650 Benkheil, M., Haele, M.V., Roskams, T., Laporte, M., Noppen, S., Abbasi, K., Delang, L., Neyts, J.,
651 Liekens S., 2018. CCL20, a direct-acting pro-angiogenic chemokine induced by hepatitis C virus
652 (HCV): Potential role in HCV-related liver cancer. Experimental Cell Research 372: 168-177.

653 Bird, S., Tafalla, C., 2015. Teleost chemokines and their receptors. Biology 4: 756-784.

654 Braasch, I., Salzburger, W., 2009. In ovo omnia: diversification by duplication in fish and other
655 vertebrates. J. Biol. 8: 25.

656 Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: an application that generates
657 similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4: 29.

658 Cañestro, C., Albalat, R., Irimia, M., Garcia-Fernàndez, J., 2013. Impact of gene gains, losses and
659 duplication modes on the origin and diversification of vertebrates. Semin. Cell Dev. Biol. 24: 83-94.

660 Ceroni, A., Passerini, A., Vullo, A., Frasconi, P., 2006. DISULFIND: a disulfide bonding state and
661 cysteine connectivity prediction server. Nucleic Acids Res. 34: 177-181.

662 Chen, J., Xu, Q., Wang, T., Collet, B., Corripio-Miyar, Y., Bird, S., Xie, P., Nie, P., Secombes, C.J.,
663 Zou. J., 2013. Phylogenetic analysis of vertebrate CXC chemokines reveals novel lineage specific
664 groups in teleost fish. Dev Comp Immunol. 41: 137-152.

665 Costa, M.M., Maehr, T., Diaz-Rosales, P., Secombes, C.J., Wang T., 2011. Bioactivity studies of
666 rainbow trout (*Oncorhynchus mykiss*) interleukin-6: effects on macrophage growth and antimicrobial
667 peptide gene expression. Mol. Immunol. 48: 1903-1916.

668 Dixon, B., Shum, B., Adams, E.J., Magor, K.E., Hedric, R.P., Muir, D.G., Parham, P., 1998. CK-1, a
669 putative chemokine of rainbow trout (*Oncorhynchus mykiss*). Immunological Reviews 166: 341-348.

670 Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L., González, A., 2019. Whole-Genome Duplication
671 and Yeast's Fruitful Way of Life. Trends Genet. 35: 42-54.

672 Frick, V.O., Rubie, C., Keilholz, U., Ghadjar, P., 2016. Chemokine/chemokine receptor pair
673 CCL20/CCR6 in human colorectal malignancy: An overview. World J. Gastroenterol 22: 833-841.

674 Fu, Q., Yang, Y., Li, C., Zeng, Q., Zhou, T., Li, N., Li, Y., Wang, X., Liu, S., Li, D., Liu, Z., 2017.
675 The chemokinome superfamily: II. The 64 CC chemokines in channel catfish and their involvement
676 in disease and hypoxia responses. Dev. Comp. Immunol. 73, 97-108.

677 Ganassin, R.C., Bols, N.C., 1998. Development of a monocyte/macrophage-like cell line, RTS-11,
678 from rainbow trout spleen. Fish Shellfish Immunol. 8: 457-476.

679 Geven, E.J.W., Klaren, P.H.M., 2017. The teleost head kidney: Integrating thyroid and immune
680 signalling. Dev. Comp. Immunol. 66: 73-83.

681 Ghadjar, P., Rubie, C., Aebersold, D.M. Keilholz, U., 2009. The chemokine CCL20 and its receptor
682 CCR6 in human malignancy with focus on colorectal cancer. Int. J. Cancer 125: 741-745.

683 Glasauer, S.M., Neuhauss, S.C., 2014. Whole-genome duplication in teleost fishes and its
684 evolutionary consequences. Mol. Genet. Genomics 289: 1045-60.

685 Grimholt, U., Hauge, H., Hauge, A.G., Leong, J., Koop, B.F., 2015. Chemokine receptors in Atlantic
686 salmon. Dev. Comp. Immunol. 49: 79-95.

687 Harun, N.O., Wang, T., Secombes, C.J., 2011. Gene expression profiling in naïve and vaccinated
688 rainbow trout after *Yersinia ruckeri* infection: insights into the mechanisms of protection seen in
689 vaccinated fish. Vaccine 26: 4388-4399.

690 Hieshima, K., Imai, T., Opdenakker, G., Van-Damme, J., Kusuda, J., Tei, H., Sakaki, Y., Takatsuki,
691 K., Miura, R., Yoshie, O., Nomiyama, H., 1997. Molecular cloning of a novel human CC chemokine
692 liver and activation-regulated chemokine (LARC) expressed in liver. Chemotactic activity for

693 lymphocytes and gene localization on chromosome 2. J. Biol. Chem. 272: 5846-5853.

694 Hong, S., Li, R., Xu, Q., Secombes, C.J., Wang, T., 2013. Two types of TNF-α exist in teleost fish:
695 phylogeny, expression, and bioactivity analysis of type-II TNF-α3 in rainbow trout *Oncorhynchus*
696 *mykiss*. J. Immunol. 191: 5959-5972.

697 Hong, S., Zou, J., Crampe, M., Peddie, S., Scapigliati, G., Bols, N., Cunningham, C., Secombes,
698 C.J., 2011. The production and bioactivity of rainbow trout (*Oncorhynchus mykiss*) recombinant IL-
699 1 beta. Vet. Immunol. Immunopathol. 81: 1-14.

700 Hong, S., Wang, T.Y., Secombes, C.J., Wang, T., 2019. Different origins of paralogues of salmonid
701 TNFR1 and TNFR2: Characterisation and expression analysis of four TNF receptor genes in
702 rainbow trout Oncorhynchus mykiss. Dev Comp Immunol. 99: 103403.

703 Hoover, D.M., Boulegue, C., Yang, D., Oppenheim, J.J., Tucker, K., Lu, W., Lubkowski, J., 2002.
704 The structure of human macrophage inflammatory protein-3alpha /CCL20. Linking antimicrobial
705 and CC chemokine receptor-6-binding activities with human beta-defensins. J. Biol. Chem. 277:
706 37647-54.

707 Hromas, R., Gray, P.W., Chantry, D., Godiska, R., Krathwohl, M., Fife, K., Bell, G.I., Takeda, J.,
708 Aronica, S., Gordon, M., Cooper, S., Broxmeyer, H.E., Klemsz, M.J., 1997. Cloning and
709 characterization of exodus, a novel beta-chemokine. Blood 89: 3315-3322.

710 Kassahn, K.S., Dang, V.T., Wilkins, S.J., Perkins, A.C., Ragan, M.A., 2009. Evolution of gene
711 function and regulatory control after whole-genome duplication: comparative analyses in vertebrates.
712 Genome Res. 19: 1404-18.

713 Keizer, D.W., Crump, M.P., Lee, T.W., Slupsky, C.M., Clark-Lewis, I., Sykes, B.D., 2000. Human
714 CC chemokine I-309, structural consequences of the additional disulfide bond. Biochemistry. 39:
715 6053-9.

716 Laing, K.J., Secombes, C.J., 2004. Trout CC chemokines: comparison of their sequences and
717 expression patterns. Mol. Immunol. 41: 793-808.

718 Lally, J., AI-Anouti, F., Bols, N., Dixon, B., 2003. The functional characterisation of CK-1, a
719 putative CC chemokine from rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol. 15,
720 411-424.

721 Lee, A.Y.S., Körner, H., 2019. The CCR6-CCL20 axis in humoral immunity and T-B cell
722 immunobiology. Immunobiology 224: 449-454.

723 Legendre, B., Tokarski, C., Chang, Y., De Freitas Caires, N., Lortat-Jacob, H., Nadaï, P.D., Rolando,
724 C., Duez, C., Tsicopoulos, A., Lassalle, P., 2013. The disulfide bond between cysteine 10 and
725 cysteine 34 is required for CCL18 activity. Cytokine 64: 463-70.

Leu, J.H., Tsai, C.H., Tsai, J.M., Yang, C.H., Hsueh, C.Y., Chou, H.Y., 2019. Identification and expression analysis of 19 CC chemokine genes in orange-spotted grouper (*Epinephelus coioides*). Fish Shellfish Immunol. 97: 1-10.

Macqueen, D. J., Johnston, I.A., 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. R. Soc. B. 281: 20132881.

Mo, Z.Q., Chen, R.A., Li, Y.W., Huang, X.Z., Li, A.X., Luo, X.C., Dan, X.M., 2015. Characterization and expression analysis of two novel CCR6 chemokine receptors and their three potential ligands CCL20Ls of grouper (*Epinephelus coioides*) post *Cryptocaryon irritans* infection. Fish Shellfish Immunol. 47: 280-288.

Muffato, M., Louis, A., Poisnel, C.E., Roest Crollius, H., 2010. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics 26: 1119-21.

Nelson, R.T., Boyd, J., Gladue, R.P., Paradis, T., Thomas, R., Cunningham, A.C., Lira, P., Brissette, W.H., Hayes, L., Hames, L.M., Neote, K.S., McColl, S.R., 2001. Genomic organization of the CC chemokine mip-3alpha/CCL20/larc/exodus/SCYA20, showing gene structure, splice variants, and chromosome localization. Genomics 73: 28-37.

Nomiyama, H., Hieshima, K., Osada, N., Kato-Unoki, Y., Otsuka-Ono, K., Takegawa, S., Izawa, T., Yoshizawa, A., Kikuchi, Y., Tanase, S., Miura, R., Kusuda, J., Nakao, M., Yoshie, O., 2008. Extensive expansion and diversification of the chemokine gene family in zebrafish: Identification of a novel chemokine subfamily CX. BMC Genomics 9: 222.

Nomiyama, H., Osada, N., Yoshie, O., 2010. The evolution of mammalian chemokine genes. Cytokine & Growth Factor Reviews 21: 253-262.

Peatman, E., Liu, Z., 2007. Evolution of CC chemokines in teleost fish: a case study in gene duplication and implications for immune diversity. Immunogenetics 59: 613-623.

Rice, A.M., McLysaght, A., 2017. Dosage-sensitive genes in evolution and disease. BMC Biol. 15: 78.

Riutta, S.J., Larsen, O., Getschman, A.E., Rosenkilde, M.M., Hwang, S.T., Volkman, B.F., 2018. Mutational analysis of CCL20 reveals flexibility of N-terminal amino acid composition and length. J. Leukoc. Biol. 104: 423-434.

Rodrigo, G., Fares, M.A., 2018. Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach. Elife 7: e29739.

Rossi, D.L., Vicari, A.P., Franz-Bacon, K., McClanahan, T.K., Zlotnik, A., 1997. Identification through bioinformatics of two new macrophage proinflammatory human chemokines: MIP-3alpha

759 and MIP-3beta. J. Immunol. 158: 1033-1036.

760 Schmuth, M., Neyer, S., Rainer, C., Grassegger, A., Fritsch, P., Romani, N., Heufler, C., 2002.
761 Expression of the C-C chemokine MIP-3α/CCL20 in human epidermis with impaired permeability
762 barrier function. Experimental Dermatology 11: 135-142.

763 Steinke, D., Salzburger, W., Braasch, I., Meyer, A., 2006. Many genes in fish have species-specific
764 asymmetric rates of molecular evolution. BMC Genomics 7: 20.

765 Studer, R.A., Robinson-Rechavi, M., 2010. Large-scale analysis of orthologs and paralogs under
766 covarion-like and constant-but-different models of amino acid evolution. Mol. Biol. Evol. 27: 2618-
767 27.

768 Subramaniam, P.S., Green, M.M., Larkin, J., Torres, B.A., Johnson, H.M., 2001. Nuclear
769 translocation of IFN-gamma is an intrinsic requirement for its biologic activity and can be driven by
770 a heterologous nuclear localization sequence. J. Interferon Cytokine Res. 21: 951-9.

771 Valgardsdottir, R., Brede, G., Eide, L.G., Frengen, E., Prydz, H., 2001. Cloning and characterization
772 of MDDX28, a putative DEAD-box helicase with mitochondrial and nuclear localization. The
773 Journal of Biological Chemistry, 276: 32056-32063.

774 Wang, B., Wangkahart, E., Secombes, C.J., Wang, T., 2019a. Insights into the Evolution of the
775 Suppressors of Cytokine Signaling (SOCS) Gene Family in Vertebrates. Mol. Biol. Evol. 36: 393-
776 411.

777 Wang, T., Diaz-Rosales, P., Costa, M.M., Campbell, S., Snow, M., Collet, B., Martin, S.A.,
778 Secombes, C.J., 2011a. Functional characterization of a nonmammalian IL-21: rainbow trout
779 *Oncorhynchus mykiss* IL-21 upregulates the expression of the Th cell signature cytokines IFN-
780 gamma, IL-10, and IL-22. J. Immunol. 186: 708-21.

781 Wang, T., Hu, Y.F., Wangkahart, E., Liu, F.G., Wang, A., Zahran, E., Maisey, K.R., Liu, M., Xu,
782 Q.Q., Imarail, M., Secombes, C.J., 2018. Interleukin (IL)-2 is a key regulator of T helper 1 and T
783 helper 2 cytokine expression in fish: functional characterization of two dvergent IL2 paralogs in
784 Salmonids. Frontiers in Immunology, 9: 1683.

785 Wang, T., Johansson, P., Abós, B., Holt, A., Tafalla, C., Jiang, Y., Wang, A., Xu, Q., Qi, Z., Huang,
786 W., Costa, M.M., Diaz-Rosales, P., Holland, J.W., Secombes, C.J., 2016. First in-depth analysis of
787 the novel Th2-type cytokines in salmonid fish reveals distinct patterns of expression and modulation
788 but overlapping bioactivities. Oncotarget 7: 10917-46.

789 Wang, T., Huang, W., Costa, M.M., Martin, S.A., Secombes, C.J., 2011b. Two copies of the genes
790 encoding the subunits of putative interleukin (IL)-4/IL-13 receptors, IL-4Rα, IL-13Rα1 and IL-
791 13Rα2, have been identified in rainbow trout (*Oncorhynchus mykiss*) and have complex patterns of

792 expression and modulation. Immunogenetics 63: 235-53.

793 Wang, T., Monte M.M., Huang, W.S., Boudinot, P., Martin, S.M., Secombes, C.J., 2010.
794 Identification of two FoxP3 genes in rainbow trout (*Oncorhynchus mykiss*) with differential
795 induction patterns. Mol. Immunol. 47, 2563-2574.

796 Wang, T.Y., Liu, F.G., Tian, G.M., Secombes, C.J., Wang T., 2019b. Lineage/species-specific
797 expansion of the Mx gene family in teleosts: differential expression and modulation of nine Mx
798 genes in rainbow trout *Oncorhynchus mykiss*. Fish Shellfish Immunol. 90: 413-430.

799 Yang, D., Chen, Q., Hoover, D.M., Staley, P., Tucker, K.D., Lubkowski, J., Oppenheim, J.J., 2003.
800 Many chemokines including CCL20/MIP-3alpha display antimicrobial activity. J. Leukoc. Biol. 74:
801 448-55.

802 Yuan, H., Li, Y., Han, P., Tian, G., Zhang, Z., Guo, H., Xu, Q., Wang, T., 2019. Identification and
803 characterization of three CXC chemokines in Asian swamp eel 3 (*Monopterus albus*) uncovers a
804 third CXCL11_like group in fish. Dev. Comp. Immunol. 101: 103454.

805 Zapata, A., Diez, B., Cejalvo, T., Gutiérrez-de Frías, C, Cortés, A., 2006. Ontogeny of the immune
806 system of fish. Fish Shellfish Immunol. 20: 126-36.

807 Zhang, L., Kang, Y.L., Chen, S.J., Wang, L., Jiang, M., Xiang, L.H., 2019. Circulating CCL20: A
808 potential biomarker for vitiligo together with the number of Th1/17 cells. Journal of Dermatological
809 Sciences 93: 92-100.

810 Zhao, L., Xia, J., Wang, X., Xu, F., 2014. Transcriptional regulation of CCL20 expression. Microbes
811 Infect. 16: 864-70.

812 Zhou, T., Li, N., Jin, Y., Zeng, Q., Prabowo, W., Liu, Y., Tian, C., Bao, L., Liu, S., Yuan, Z., Fu, Q.,
813 Gao, S., Gao, D., Dunham, R., Shubin, N.H., Liu, Z., 2018. Chemokine C-C motif ligand 33 is a key
814 regulator of teleost fish barbel development. Proc. Natl. Acad. Sci. USA 115: 5018-5027.

815

816

817

818

819 **Figure legend**

820 **Fig. 1. Amino acid sequence multiple alignment of rainbow trout CCL20_L molecules.** The
821 multiple alignment was generated using the Clustal Omega program, and conserved amino acid
822 residues were shaded using BOXSHADE (version 3.21). Human CCL20 was included in the
823 alignment for comparison. The signal peptides are highlighted in green, and cysteine residues in the
824 mature peptide are in red. The cysteine residues that potentially form an intramolecular disulfide bond
825 are linked above the alignment.

826 **Fig. 2. Phylogenetic tree analysis of salmonid and Northern pike CCL20_L molecules.** The
827 phylogenetic tree was constructed using a multiple alignment of salmonid and pike CCL20_L aa
828 sequences using the neighbour-joining (NJ) method within the MEGA 7.0 program. The JTT matrix-
829 based method with pairwise deletion option was chosen to compute the evolutionary distance. The
830 percentage bootstrap value (>50%) is shown next to the branches based on 10,000 replicates. The
831 accession number for each sequence is given after the common species and molecular names. The
832 bootstrap value at the root of each group is highlighted in red and the groupings are indicated on the
833 right. The sequences from trout and pike are in red and green, respectively, and other salmonid
834 sequences predicted in this study are in blue.

835 **Fig. 3. Synteny analysis of CCL20_L loci in selected teleosts, in comparison to the holostean gar**
836 **and human CCL20 locus.** The synteny was predicted using the Genomicus program (database
837 version 96.01 and trout 01.01) or information extracted from recently released reference genomes at
838 NCBI. The accession numbers are NC_023192 (spotted gar), NC_035087, NC_035091, NC_035084
839 and NC_035104 (rainbow trout); NC_025970 and NC_025988 (Northern pike); NC_007113 and
840 NC_007135 (zebrafish); and NC_000002 (human).

841 **Fig. 4. Phylogenetic tree analysis of teleost and holostean CCL20_L molecules, and CCL20,**
842 **CCL3, CCL4 and CCL17 of selected tetrapod groups**. The phylogenetic tree was constructed using
843 a multiple alignment of teleost and holostean CCL20_L, with CCL20, CCL3, CCL4 and CCL17 from
844 selected tetrapod groups, since these other CC chemokines share high aa sequence identities to
845 rainbow trout CCL20_L molecules. A neighbour-joining tree was constructed within the MEGA7.0
846 program using the JTT matrix-based method with pairwise deletion option. The percentage bootstrap
847 values (>50%) are shown next to the branches and based on 10,000 replicates. The accession number
848 for each sequence is given after the common species and molecular name. The bootstrap value at the
849 root of each group is highlighted in red and the groupings are indicated on the right. The sequences
850 from spotted gar and Northern pike are in red and green, respectively. The salmonid branches were
851 condensed and shown in blue.

852 **Fig. 5. A model of teleost CCL20_L gene evolution.** A single CCL20 gene was present in the early
853 2R vertebrates that was duplicated in ancestral actinopterygian fish. The two duplicated genes were

further duplicated by 3R and 4R WGDs with subsequent gene loss and local gene duplications. Lineage/species specific local gene duplication /deletion led to different copies of CCL20_L genes in extant teleosts.

Fig. 6. Gene organisation of CCL20_L molecules in rainbow trout, Northern pike, spotted gar and human. The black and white boxes represent coding and non-coding regions of exons and lines between boxes represent introns. The size (bp) of each exon is in the box, and the intron phase is indicated above the line. The genomic information is given in Table 2 (trout CCL20_L genes) and Table 3 (pike and gar CCL20_L genes).

Fig. 7. The expression level of trout CCL20_L genes in different tissues. The expression level of CCL20_L1a and CCL20_L1b1-2 (A), CCL20_L2a and CCL20_L2b (B), and CCL20_L3a and CCL20_L3b1-4 (C) was investigated by RT-qPCR in seventeen tissues from six fish. The transcript level was calculated according to a serial dilution of references that contained equal amounts of each gene and was normalized against the expression level of EF-1α. The data are presented as mean + SEM (n=6).

Fig. 8. The expression of trout CCL20_L genes during early developmental stages. cDNA samples were prepared from eyed-eggs, immediately post-hatch, pre-first feeding fry or fry 3 weeks after first feeding. Six independent samples from each developmental stage were selected for RT-qPCR analysis. Data were analysed by one-way ANOVA and presented as mean + SEM (n=6). Different letters above the bars indicate significant differences ($p \leqslant 0.05$).

Fig. 9. Modulation of the expression of trout CCL20_L genes *in vivo* by Poly I:C. Rainbow trout were injected ip with 0.2 ml of PBS containing 1 mg Poly I:C or 0.2 ml PBS as control. The gills, spleen, HK and intestine were collected at 6 h and 24 h post stimulation. The expression of trout CCL20_L genes was quantified by RT-qPCR. The data are presented as mean (+SEM) relative expression, where the expression level of each gene in each tissue of control fish at 6 h was defined as 1 (n=6). Different letters above the bars indicate significant differences in the same tissue ($p \leqslant 0.05$, one way-ANOVA). A bar over a tissue indicates no differences were detected.

Fig. 10. Modulation of the expression of trout CCL20_L genes by *Yersinia ruckeri* challenge. Rainbow trout were injected ip with 0.5 ml of *Y. ruckeri* at a concentration of 1×10^6 cfu/ml or 0.5 ml PBS as control. The gills, spleen, HK and intestine were collected at day 1 and day 2 post injection. The gene expression was quantified by RT-qPCR and normalised to the expression level of EF-1 α in each sample. The data are presented as the mean (+SEM) relative expression, where the expression level of each gene in each tissue of control fish at day 1 was defined as 1 (n=4). Different letters above the bars indicate significant differences (($p \leqslant 0.05$, one way-ANOVA). A bar over a tissue indicates no differences were detected.

888 **Fig. 11. Modulation of the expression of trout CCL20_L genes in RTS-11 cells by Poly I:C and**
889 **LPS.** RTS-11 cells were cultured overnight and then stimulated with Poly I:C (50 μg/ml), LPS (25
890 μg/ml) or medium as control for 4 h, 8 h and 24 h. The treatments were terminated by dissolving cells
891 in TRI reagent and the expression of trout CCL20_L genes was analysed by RT-qPCR. The data are
892 presented as mean (+SEM) fold change calculated as the expression level of stimulated samples
893 divided by that of time matched controls. One-way ANOVA and LSD post hoc tests were used to
894 analyse the expression data (* $p \leqslant 0.05$; **$p \leqslant 0.01$ and *** $p \leqslant 0.001$).

895 **Fig. 12. Modulation of the expression of trout CCL20_L genes in RTS-11 cells by pro-**
896 **inflammatory cytokines.** RTS-11 cells were cultured overnight and then stimulated with IFNγ (20
897 ng/ml), IFNa (25 ng/ml), IL-1β (25 ng/ml), IL-6 (100 ng/ml), TNFα (50 ng/ml), or medium as control
898 for 4 h, 8 h and 24 h. The treatments were terminated by dissolving cells in TRI reagent and the
899 expression of trout CCL20_L genes was analysed by RT-qPCR. The data are presented as mean
900 (+SEM) fold change calculated as the expression level of stimulated samples divided by that of time
901 matched controls. One-way ANOVA and LSD post hoc tests were used to analyse the expression data
902 (* $p \leqslant 0.05$; **$p \leqslant 0.01$ and *** $p \leqslant 0.001$).

903

904 **Table legend**

905

906 **Table 1. Primers used for real-time RT-PCR gene expression analysis of trout CCL20_L**
907 **paralogues.**

908 **Table 2. Summary of CCL20_L genes in rainbow trout.** The location and transcriptional direction
909 (F=forward, R=reverse) in the genome, and the accession numbers of genomic and cDNA sequences
910 of each gene are shown. The numbers of aa of the full-length protein, signal peptide and mature
911 peptide of the deduced aa sequence, and the pI and molecular weight (MW) of the mature peptide of
912 each gene are also shown.

913 **Table 3. Summary of CCL20_L genes from other fish species analysed in this study.** The location
914 and transcriptional direction (F=forward, R=reverse) in the genome, linkage group (LG) or scaffold,
915 and the accession numbers of cDNA and protein sequences of each gene are shown.

916 **Table 4. Comparison of aa sequence identities of CCL20_L molecules from rainbow trout and**
917 **from other salmonids and Northern pike.** The highest levels of aa identities to trout CCL20_L
918 molecules are shown in red and the second highest levels of identities among salmonids are shown in
919 blue.

920

Table 1. Primers used for real-time RT-PCR gene expression analysis of trout CCL20_L paralogues.

Gene	Forward (5' to 3')	Reverse (5' to 3')	Size
CCL20_L1a	CCACCACACAATCAGCGGACT	CGGCACACACAAACTTGTTCCTTAGAT	153
CCL20_L1b1	CCCCTCCACACAATCAGCAGATT	GCTTGAGCAGACAGCGTTGTACCT	203
CCL20_L1b2	CCCCTCCACACAATCAGCACATT	CTTGAGCAGGCAGCGTTGTACTC	202
CCL20_L2a	TCAACCTCAGCTGCATACGGTCTTA	TTCCTCACCCATTCGTCCTTAACTG	194
CCL20_L2b	CAACCTCAGCAGCATATGGTCCTC	CCTCACCCACTCGTCCTTAATGC	191
CCL20_L3a	CTTCTGCAGCTAAACGACCACGTC	GGCCTGTGAGGTCTTGATGTGAGA	245
CCL20_L3b1	GGTTTCTGCAGCTAAACAAGGTTTTCC	GCGAACAGCTTTCATAACCCAGC	202
CCL20_L3b2	GGTTTCTGCAGCTAAACAAGGTTTTCC	CAGCTTGTGAACAACTTTCATAACCCATT	208
CCL20_L3b3	CAGCAGCTCATCACATCACAGCA	TTTCCCATTTTTAGTGTGGAATATGATGG	262
CCL20_L3b4	GCTTCTGCAGCTAAACAAGCACGTC	CTGATGTGAACCGCCGTGC	233
EF-1α	CAAGGATATCCGTCGTGGCA	ACAGCGAAACGACCAAGAGG	327

Table 2. Summary of CCL20_L genes in rainbow trout. The location and transcriptional direction (F=forward, R=reverse) in the genome, and the accession numbers of genomic and cDNA sequences of each gene are shown. The numbers of aa of the full-length protein, signal peptide and mature peptide of the deduced aa sequence, and the pI and molecular weight (MW) of the mature peptide of each gene are also shown.

Gene	Genomic DNA [a]	Location	cDNA [b]	Orientation	Full length (aa)	Signal peptide/mature peptide (aa)	pI [c]	MW (kDa) [c]
CCL20_L1a	NC_035087 (Ch 11)	13,571,908-13,574,179	NM_001124254	F	100	26/74	9.34	9.05
CCL20_L1b1	NC_035091 (Ch 15)	11,142,715-11,140,091	MK986833	R	100	26/74	10.16	9.00
CCL20_L1b2	NC_035091 (Ch 15)	11,128,165-11,126,127	MK986834	R	98	24/74	9.96	9.02
CCL20_L2a	NC_035104 (Ch 28)	8,666,274-8,668,801	MK986835	R	112	24/88	9.39	10.04
CCL20_L2b	NC_035084 (Ch 8)	71,497,933-71,499,566	MK986836	F	102	24/78	9.46	9.00
CCL20_L3a	NC_035104 (Ch 28)	3,493,418-3,494,594	MK986837	F	105	26/79	10.07	8.86
CCL20_L3b1	NC_035084 (Ch 8)	74,979,211-74,980,774	MK986838	R	104	26/78	9.88	8.87
CCL20_L3b2	NC_035084 (Ch 8)	74,998,822-75,000,451	MK986839	R	104	26/78	9.87	8.90
CCL20_L3b3	NC_035084 (Ch 8)	75,009,667-75,010,564	MK986840	R	105	26/79	9.87	8.90
CCL20_L3b4	NC_035084 (Ch 8)	75,027,950-75,029,419	MK986841	R	105	26/79	9.96	8.82

[a] accession number of genomic DNA and chromosome (Ch) location
[b] accession number of cDNA
[c] pI and molecular weight of mature peptide

Table 3. Summary of CCL20_L genes from other fish species analyses in this study. The location and transcription direction (F=forward, R=reverse) on chromosome (Ch), linkage group (LG) or scaffold, and the accession numbers of cDNA and protein sequences of each gene are shown.

Gene	Chromosome	Location	Orientation	cDNA	Protein
Atlantic CCL20_L1a	Ch 19	66,965,509-66,975,716	R	XM_014159437	XP_014014912
Atlantic CCL20_L1b1	Ch 29	26,869,132-26,873,675	R	EG861420	
Atlantic CCL20_L1b2	Ch 29	26,377,437-26,395,612	R	EG865207	
Atlantic CCL20_L2a	Ch 3	36,707,891-36,714,434	F	XM_014191476	XP_014046951
Atlantic CCL20_L3a	Ch 3	43,703,189-43,710,324	F	XM_014192103	XP_014047578
Atlantic CCL20_L3b1	Ch 14	43,274,391-43,283,224	F	XM_014141244	XP_013996719
Atlantic CCL20_L3b3	Ch 14	43,261,929-43,270,154	F	XM_014141243	XP_013996718
Chinook CCL20_L1a	Ch 16	9,850,166-9,850,932	F	XM_024376526	XP_024232294
Chinook CCL20_L1b1	Ch 29	9,627,158-9,647,897	R	Predicted	
Chinook CCL20_L1b2	Ch 29	9,622,416-9,629,681	R	Predicted	
Chinook CCL20_L2a	Ch 28	8,967,696-8,969,882	R	XM_024391149	XP_024246917
Chinook CCL20_L2b	Ch 10	53,329,771-53,332,021	F	XM_024436478	XP_024292246
Chinook CCL20_L3a	Ch 28	2,337,830-2,339,345	R	XM_024390821	XP_024246589
Chinook CCL20_L3b1	Ch 10	56,639,430-56,642,246	R	XM_024436482	XP_024292250
Chinook CCL20_L3b2	Ch 10	56,621,123-56,624,348	R	XM_024436481	XP_024292249
Chinook CCL20_L3b3	Ch 10	56,674,999-56,677,552	R	XM_024436483	XP_024292251
Chinook CCL20_L3b4	Ch 10	56,657,884-56,662,928	R	Predicted	
Coho CCL20_L1a	LG 18	11,327,923-11,328,623	F	XM_020508624	XP_020364213
Coho CCL20_L1b1	LG 11	5,192,144-5,210,327	R	XM_020494039	XP_020349628
Coho CCL20_L1b2	LG 11	5,180,229-5,198,416	R	Predicted	
Coho CCL20_L2a	LG 27	6,568,085-6,574,388	R	XM_020463089	XP_020318678
Coho CCL20_L2b	LG 30	33,510,411-33,515,418	F	XM_020467471	XP_020323060
Coho CCL20_L3a	Scaffold 15709	95-2,366	R	XM_020479071	XP_020334660
Coho CCL20_L3b1	LG 24	14,372,742-14,378,054	F	XM_020459179	XP_020314768
Coho CCL20_L3b3	Scaffold 07847	3,792-9,351	R	XM_020477418	XP_020333007
Coho CCL20_L3b4	Scaffold 07847	9,987-18,004	R	Predicted	
Charr CCL20_L1a	LG 14	31,629,440-31,631,711	F	Predicted	
Charr CCL20_L2a	LG 19	8,419,893-8,424,609	R	XM_024009270	XP_023865038
Charr CCL20_L2b	LG 32	33,540,830-33,542,954	F	XM_023977823	XP_023833591
Charr CCL20_L3a	LG 33	8,100,306-8,102,511	R	XM_023979313	XP_023835081
Charr CCL20_L3b	Scaffold 6733	6,456-9,302	F	XM_024145097	XP_024000865
Pike CCL20-1	LG 21	17,967,668-17,979,891	F	XM_010890050	XP_010888352
Pike CCL20-2	LG 03	14,656,161-14,665,248	R	NM_001303712	NP_001290641
Pike CCL20-3a	LG 03	12,632,880-12,637,423	F	NM_001303695	NP_001290624
Pike CCL20-3b	LG 03	12,637,867-12,643,914	F	NM_001303676	NP_001290605
Gar CCL20-1	LG 14	13,048,845-13,059,212	R	XM_015361083	XP_015216569
Gar CCL20-2	LG 14	13,057,489-13,060,465	R	XM_01536116	XP_015216602
Gar CCL20-3	LG 14	13,071,958-13,079,221	F	XM_015361063	XP_015216549
Gar CCL20-4	LG 14	13,079,252-13,087,339	F	XM_015361062	XP_015216548
Zebrafish CCL20-1	Ch 24	26,001,300-26,013,395	F	NM_001113595	NP_001107067
Zebrafish CCL20-2	Ch 2	40,177,801-40,186,888	F	XM_005171348	XP_005171405
Zebrafish CCL20-3	Ch 2	45,188,850-45,196,113	F	NM_001136254	NP_001129726
Catfish CCL20a.1	Ch 7	7,202,821-7,208,868	F	XM_017470960	XP_017326449
Catfish CCL20	Ch 7	7,199,797-7,211,892	F	XM_017470960	XP_017326449
Catfish CCL20a.3	Scaffold 00585	83,016-87,559	F	XM_017461884	XP_017317373
Catfish CCL20b.1	Ch 23	9,215,269-9,227,366	F	XM_017453030	XP_017308519
Catfish CCL20b.2	Ch 23	9,219,892-9,231,987	F	XM_017453031	XP_017308520
Tetra CCL20-1	Ch 3	17,750,681-17,759,768	R	XM_007252332	XP_007252394
Tetra CCL20-2	Ch 16	9,108,011-9,137,066	F	XM_022677669	XP_022533390
Tetra CCL20-3	Scaffold 4423	287,457-297,760	R	XM_007241461	XP_007241523
Swamp eel CCL20-1	Scaffold 3.1	10,809,887-10,818,974	R	XM_0205995661	XP_020455222
Swamp eel CCL20-2	Scaffold 172.1	1,188,548-1,192,179	F	XM_020616001	XP_020471657
Swamp eel CCL20-3	Scaffold 172.1	43,528-47,159	F	XM_020616254	XP_020471910
Flounder CCL20-1	Scaffold 209	768,054-771,685	F	XM_020103201	XP_019958760
Flounder CCL20-2	Superscaffold 3	10,491,841-10,497,024	R	XM_020106948	XP_019962507
Flounder CCL20-3	Superscaffold 3	11,636,240-11,639,871	R	XM_020107025	XP_019962584
Medaka CCL20-1a	Ch 20	25,814,839-25,820,486	R	XM_023950372	XP_023806140
Medaka CCL20-1b	Ch 20	25,797,568-25,808,863	F	XM_023950385	XP_023806153
Medaka CCL20-3	Ch 17	23,158,631-23,165,895	R	XM_004079325	XP_004079373
Fugu CCL20-2	Ch 22	8243057-8247600	F	XM_011616507	XP_011614809
Fugu CCL20-3	Ch 22	7507868-7512827	F	XM_003975613	XP_003975662

Table 4. Comparison of aa sequence identities of CCL20_L molecules from rainbow trout and from other salmonids and Northern pike. The highest levels of aa identities to trout CCL20_L molecules are shown in red and the second highest levels of identities among salmonids are shown in blue.

		CCL20_L1			CCL20_L2		CCL20_L3				
	Trout	CCL20_L1a	CCL20_L1b1	CCL20_L1b2	CCL20_L2a	CCL20_L2b	CCL20_L3a	CCL20_L3b1	CCL20_L3b2	CCL20_L3b3	CCL20_L3b4
	Trout CCL20_L1a	100.0	74.0	73.0	22.9	25.5	30.3	29.0	29.9	30.6	30.6
	Atlantic CCL20_L1a	94.0	88.0	81.0	23.7	27.4	31.2	29.9	30.8	30.3	31.5
	Chinook CCL20_L1a	96.0	88.0	81.0	23.5	26.4	30.3	29.9	30.8	30.6	30.6
	Coho CCL20_L1a	95.0	87.0	80.0	23.5	25.5	29.4	29.9	30.8	30.6	30.6
	Charr CCL20_L1a	83.0	78.0	73.5	23.5	26.9	27.8	26.7	27.6	28.3	28.3
	Trout CCL20_L1b1	74.0	100.0	85.0	25.4	30.9	34.9	32.7	33.6	33.6	33.6
	Trout CCL20_L1b2	73.0	85.0	100.0	24.1	26.9	33.3	29.2	30.2	29.0	29.0
	Atlantic CCL20_L1b1	83.0	94.0	89.8	24.1	26.9	35.2	29.2	30.2	32.4	31.5
	Atlantic CCL20_L1b2	78.0	88.0	93.9	24.3	26.9	33.3	29.2	30.2	30.6	30.6
	Chinook CCL20_L1b1	77.2	85.1	77.2	21.3	24.8	32.7	29.4	31.2	29.7	29.7
	Chinook CCL20_L1b2	60.0	70.0	76.5	19.5	19.3	26.9	24.5	23.4	23.9	23.7
	Coho CCL20_L1b1	85.0	98.0	89.0	24.6	30.0	35.8	32.7	33.6	34.3	34.3
CCL20_L1	Coho CCL20_L1b2	75.5	86.3	89.2	25.4	25.0	34.2	31.2	32.1	33.6	33.6
	Trout CCL20_L2a	22.9	25.4	24.1	100.0	66.1	31.3	27.8	27.8	33.0	33.0
	Atlantic CCL20_L2a	23.5	25.2	23.0	89.4	69.9	32.8	26.5	26.5	31.0	30.2
	Chinook CCL20_L2a	21.5	24.0	23.6	79.5	58.3	28.5	23.3	23.3	29.2	29.2
	Coho CCL20_L2a	22.5	22.5	23.6	81.1	58.3	28.5	23.3	23.3	29.2	29.2
	Charr CCL20_L2a	21.9	24.6	22.1	82.1	67.0	29.6	22.4	23.3	27.8	27.0
	Trout CCL20_L2b	25.5	30.9	26.9	66.1	100.0	33.3	27.4	27.4	29.5	28.6
	Chinook CCL20_L2b	25.5	30.9	26.9	67.9	97.1	34.3	27.4	27.4	30.5	29.5
	Coho CCL20_L2b	25.5	30.9	26.9	67.9	97.1	34.3	27.4	27.4	30.5	29.5
CCL20_L2	Charr CCL20_L2b	26.4	31.8	27.8	66.1	93.1	34.3	27.4	27.4	32.4	31.4
	Trout CCL20_L3a	30.3	34.9	33.3	31.3	33.3	100.0	46.7	46.7	65.7	64.8
	Atlantic CCL20_L3a	31.2	35.8	35.2	33.0	34.3	94.3	48.6	48.6	69.5	68.6
	Chinook CCL20_L3a	30.3	34.9	33.3	31.3	33.3	97.1	48.6	48.6	68.6	67.6
	Coho CCL20_L3a	30.3	34.9	33.3	31.3	33.3	97.1	48.6	48.6	68.6	67.6
	Charr CCL20_L3a	31.4	33.9	32.5	34.2	30.8	79.2	42.6	42.6	58.3	57.5
	Trout CCL20_L3b1	29.0	32.7	29.2	27.8	27.4	46.7	100.0	90.4	51.4	51.4
	Trout CCL20_L3b2	29.9	33.6	30.2	27.8	27.4	46.7	90.4	100.0	51.4	51.4
	Trout CCL20_L3b3	30.6	33.6	29.0	33.0	29.5	65.7	51.4	51.4	100.0	98.1
	Trout CCL20_L3b4	30.6	33.6	29.0	33.0	28.6	64.8	51.4	51.4	98.1	100.0
	Atlantic CCL20_L3b1	29.0	31.8	30.2	30.7	28.8	44.9	79.8	81.7	51.9	51.9
	Atlantic CCL20_L3b3	28.7	33.3	31.8	33.9	29.5	66.7	53.3	53.3	89.5	88.6
	Chinook CCL20_L3b1	23.4	27.1	25.5	28.7	31.1	46.7	78.8	82.7	49.5	49.5
	Chinook CCL20_L3b2	23.4	27.1	25.5	28.7	31.1	46.7	78.8	82.7	49.5	49.5
	Chinook CCL20_L3b3	31.5	33.6	29.0	32.2	28.6	66.7	49.5	49.5	97.1	95.2
	Chinook CCL20_L3b4	31.5	33.6	29.0	32.2	28.6	66.7	49.5	49.5	97.1	95.2
	Coho CCL20_L3b1	30.6	32.7	30.2	28.7	28.3	46.7	90.4	93.3	51.4	51.4
	Coho CCL20_L3b3	31.5	35.5	33.6	33.0	30.5	67.6	49.5	49.5	94.3	92.4
	Coho CCL20_L3b4	31.5	35.5	33.6	33.0	30.5	67.6	49.5	49.5	94.3	92.4
CCL20_L3	Charr CCL20_L3b	28.7	33.9	28.4	33.0	30.5	67.6	51.4	51.4	87.6	85.7
	Pike CCL20_L1	63.5	62.5	56.7	26.7	26.4	26.8	27.5	27.5	29.7	29.7
	Pike CCL20_L2	23.8	22.0	25.4	53.2	50.8	27.1	23.1	23.1	26.2	25.4
	Pike CCL20_L3a	25.0	29.0	28.0	26.3	25.7	43.9	60.4	58.5	46.7	45.8
Pike	Pike CCL20_L3b	27.9	29.4	30.4	26.5	27.4	47.2	56.1	52.3	47.2	46.3

Figure 1

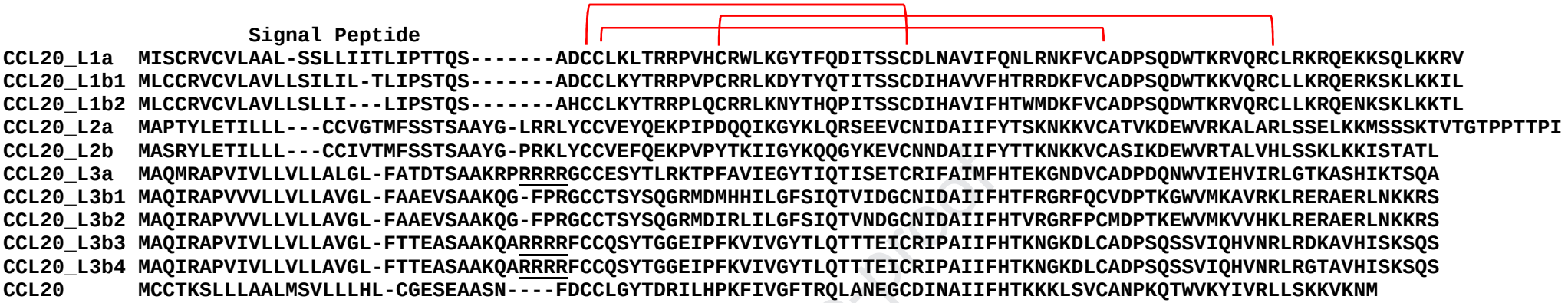


Figure 10

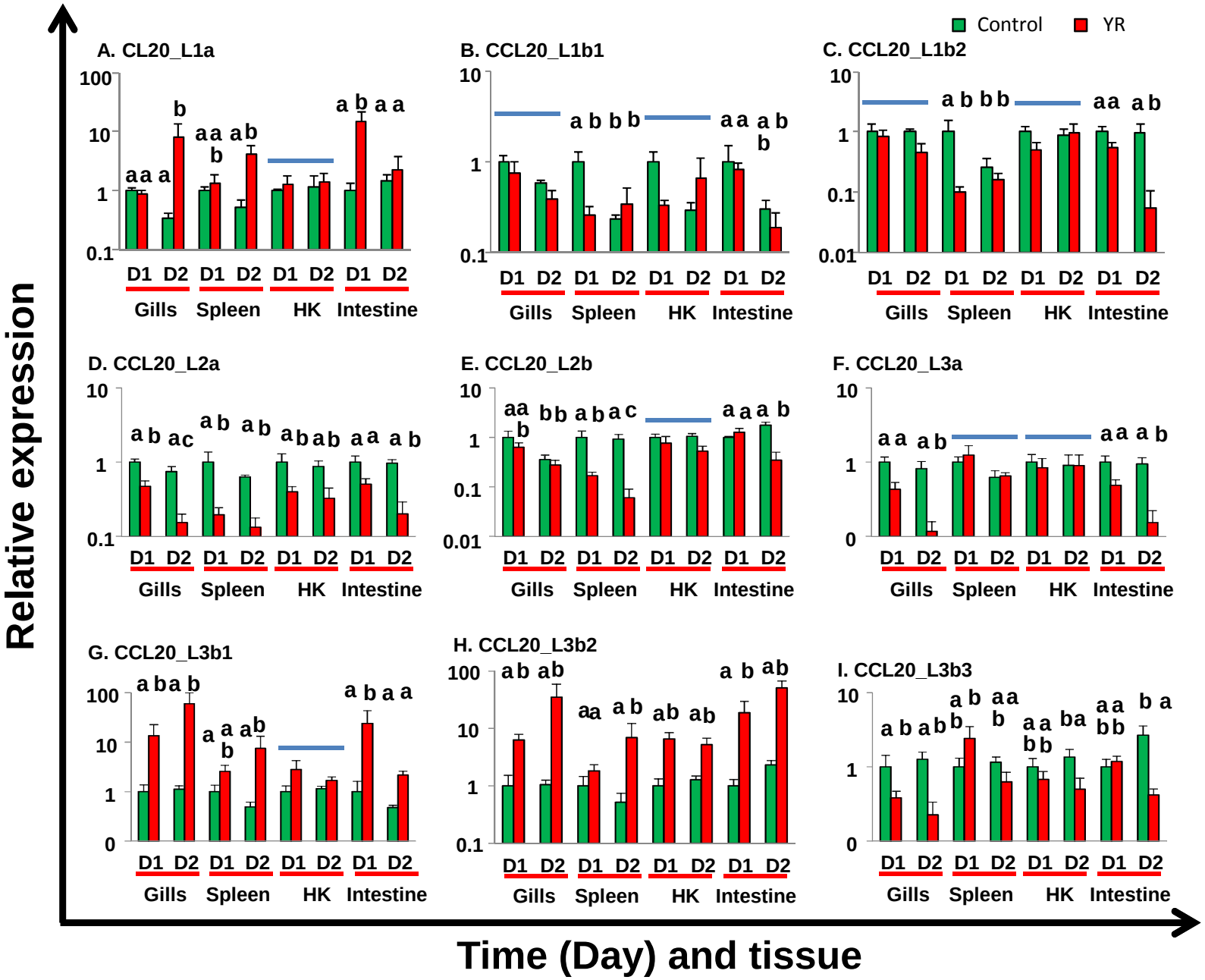

Control
YR
A. CL20_L1a
B. CCL20_L1b1
C. CCL20_L1b2
D. CCL20_L2a
E. CCL20_L2b
F. CCL20_L3a
G. CCL20_L3b1
H. CCL20_L3b2
I. CCL20_L3b3
D1 D2
Gills
Spleen
HK
Intestine
Relative expression
Time (Day) and tissue

Figure 11

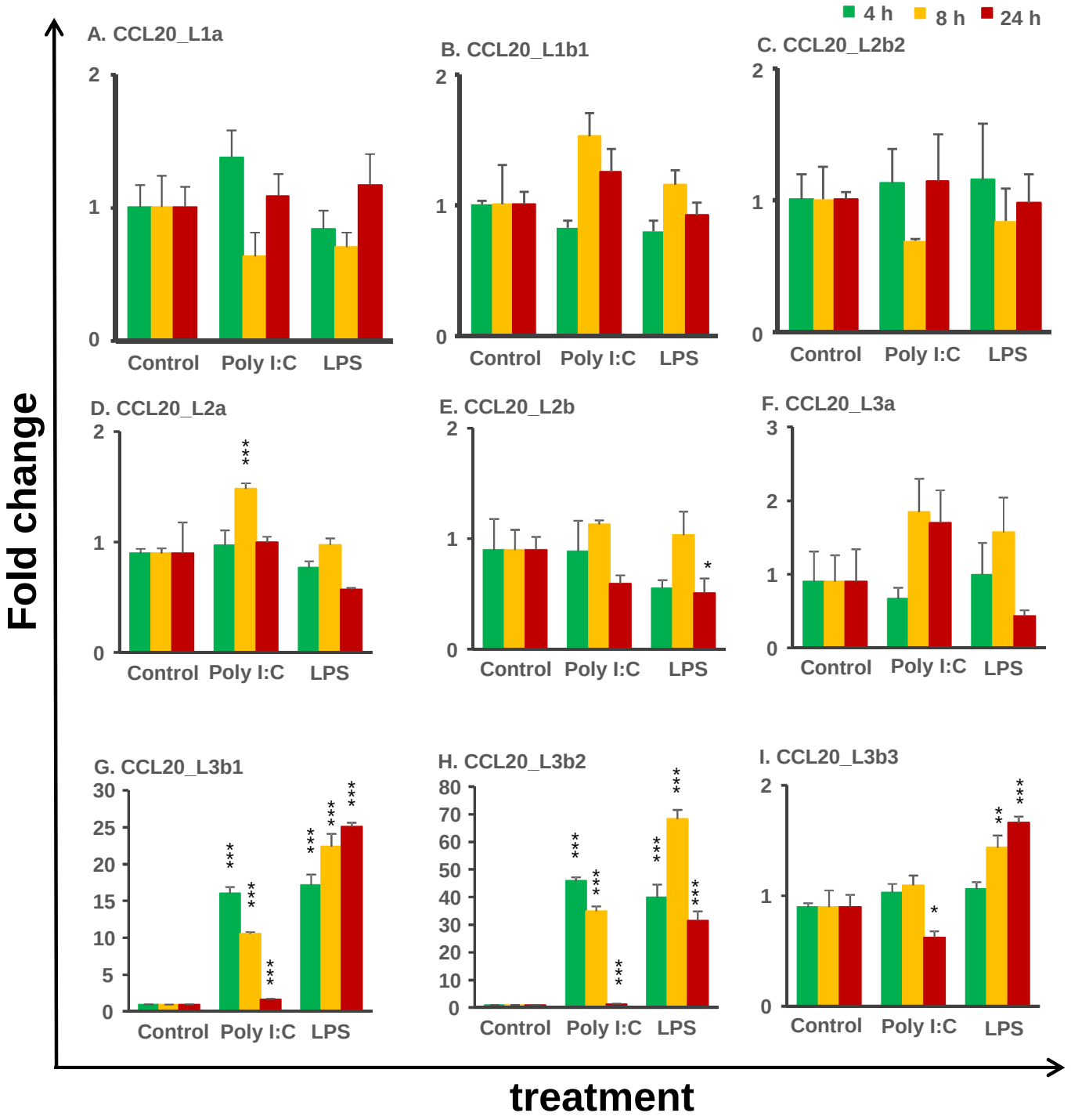

4 h
8 h
24 h
A. CCL20_L1a
B. CCL20_L1b1
C. CCL20_L2b2
D. CCL20_L2a
E. CCL20_L2b
F. CCL20_L3a
G. CCL20_L3b1
H. CCL20_L3b2
I. CCL20_L3b3
Control
Poly I:C
LPS
Fold change
treatment

Figure 12

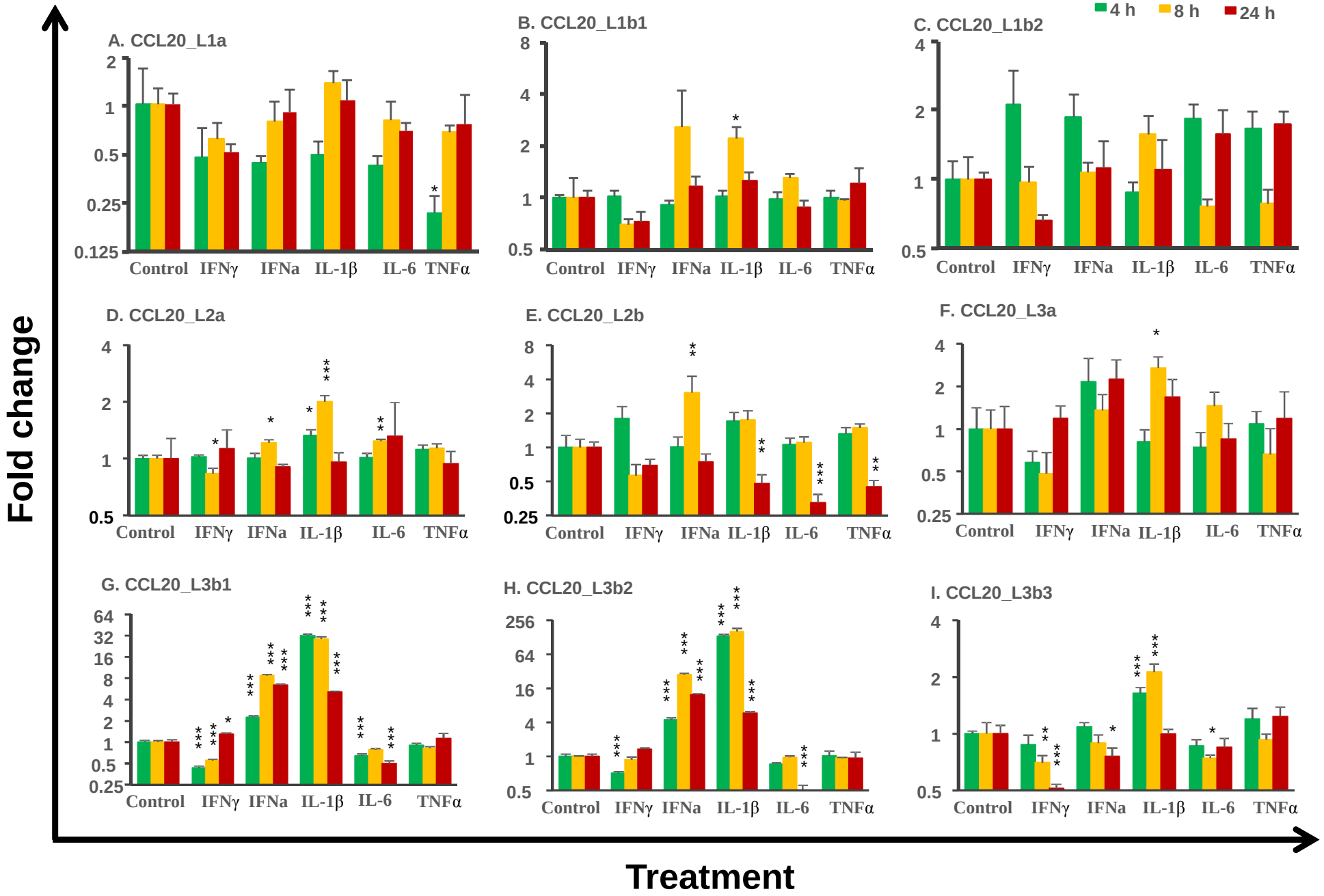

4 h
8 h
24 h
A. CCL20_L1a
B. CCL20_L1b1
C. CCL20_L1b2
D. CCL20_L2a
E. CCL20_L2b
F. CCL20_L3a
G. CCL20_L3b1
H. CCL20_L3b2
I. CCL20_L3b3
Control
IFNγ
IFNa
IL-1β
IL-6
TNFα
Fold change
Treatment

Figure 2

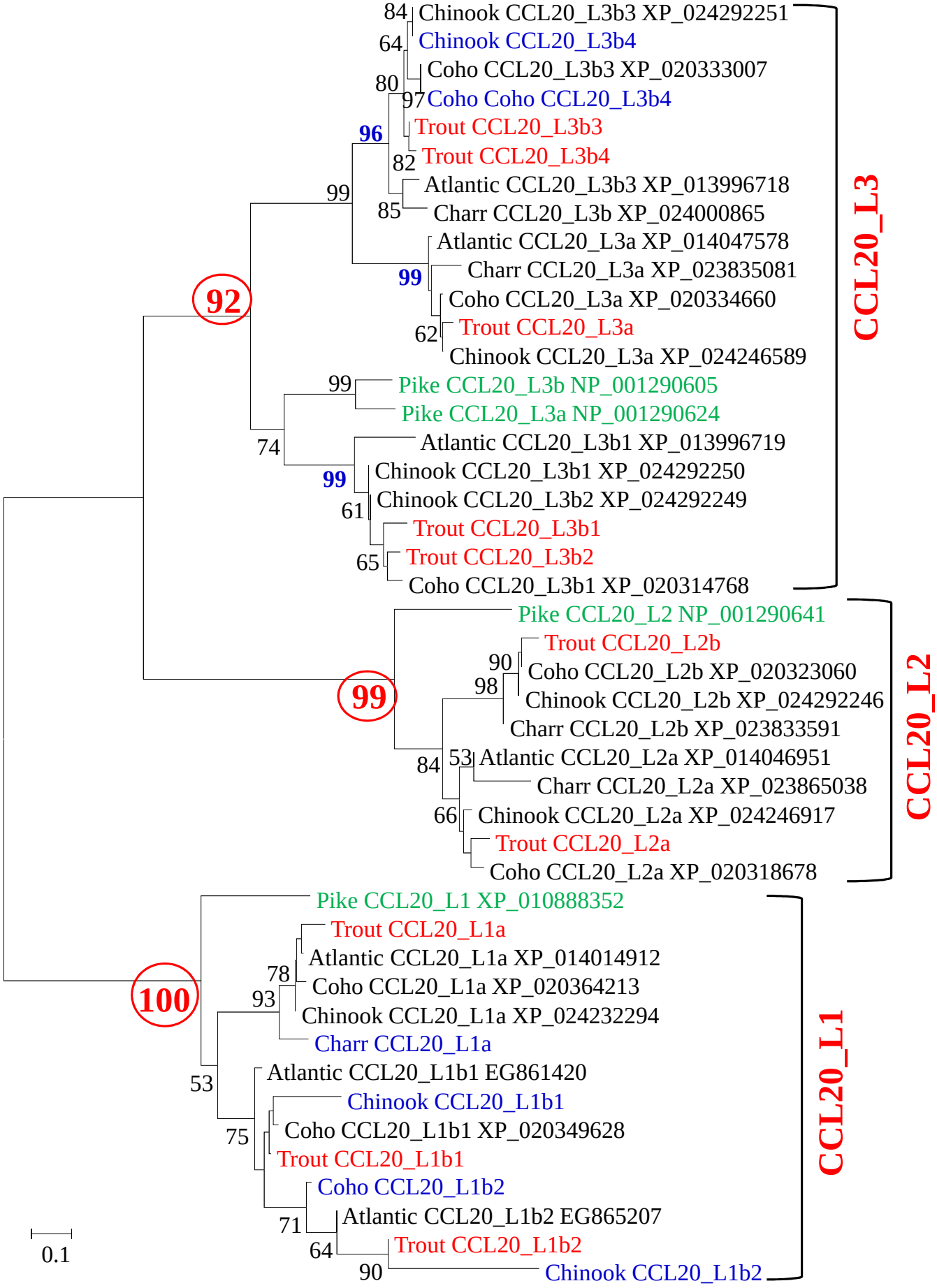

Chinook CCL20_L3b3 XP_024292251
Chinook CCL20_L3b4
Coho CCL20_L3b3 XP_020333007
Coho Coho CCL20_L3b4
Trout CCL20_L3b3
Trout CCL20_L3b4
Atlantic CCL20_L3b3 XP_013996718
Charr CCL20_L3b XP_024000865
Atlantic CCL20_L3a XP_014047578
Charr CCL20_L3a XP_023835081
Coho CCL20_L3a XP_020334660
Trout CCL20_L3a
Chinook CCL20_L3a XP_024246589
Pike CCL20_L3b NP_001290605
Pike CCL20_L3a NP_001290624
Atlantic CCL20_L3b1 XP_013996719
Chinook CCL20_L3b1 XP_024292250
Chinook CCL20_L3b2 XP_024292249
Trout CCL20_L3b1
Trout CCL20_L3b2
Coho CCL20_L3b1 XP_020314768
CCL20_L3
Pike CCL20_L2 NP_001290641
Trout CCL20_L2b
Coho CCL20_L2b XP_020323060
Chinook CCL20_L2b XP_024292246
Charr CCL20_L2b XP_023833591
Atlantic CCL20_L2a XP_014046951
Charr CCL20_L2a XP_023865038
Chinook CCL20_L2a XP_024246917
Trout CCL20_L2a
Coho CCL20_L2a XP_020318678
CCL20_L2
Pike CCL20_L1 XP_010888352
Trout CCL20_L1a
Atlantic CCL20_L1a XP_014014912
Coho CCL20_L1a XP_020364213
Chinook CCL20_L1a XP_024232294
Charr CCL20_L1a
Atlantic CCL20_L1b1 EG861420
Chinook CCL20_L1b1
Coho CCL20_L1b1 XP_020349628
Trout CCL20_L1b1
Coho CCL20_L1b2
Atlantic CCL20_L1b2 EG865207
Trout CCL20_L1b2
Chinook CCL20_L1b2
CCL20_L1
0.1

Figure 3

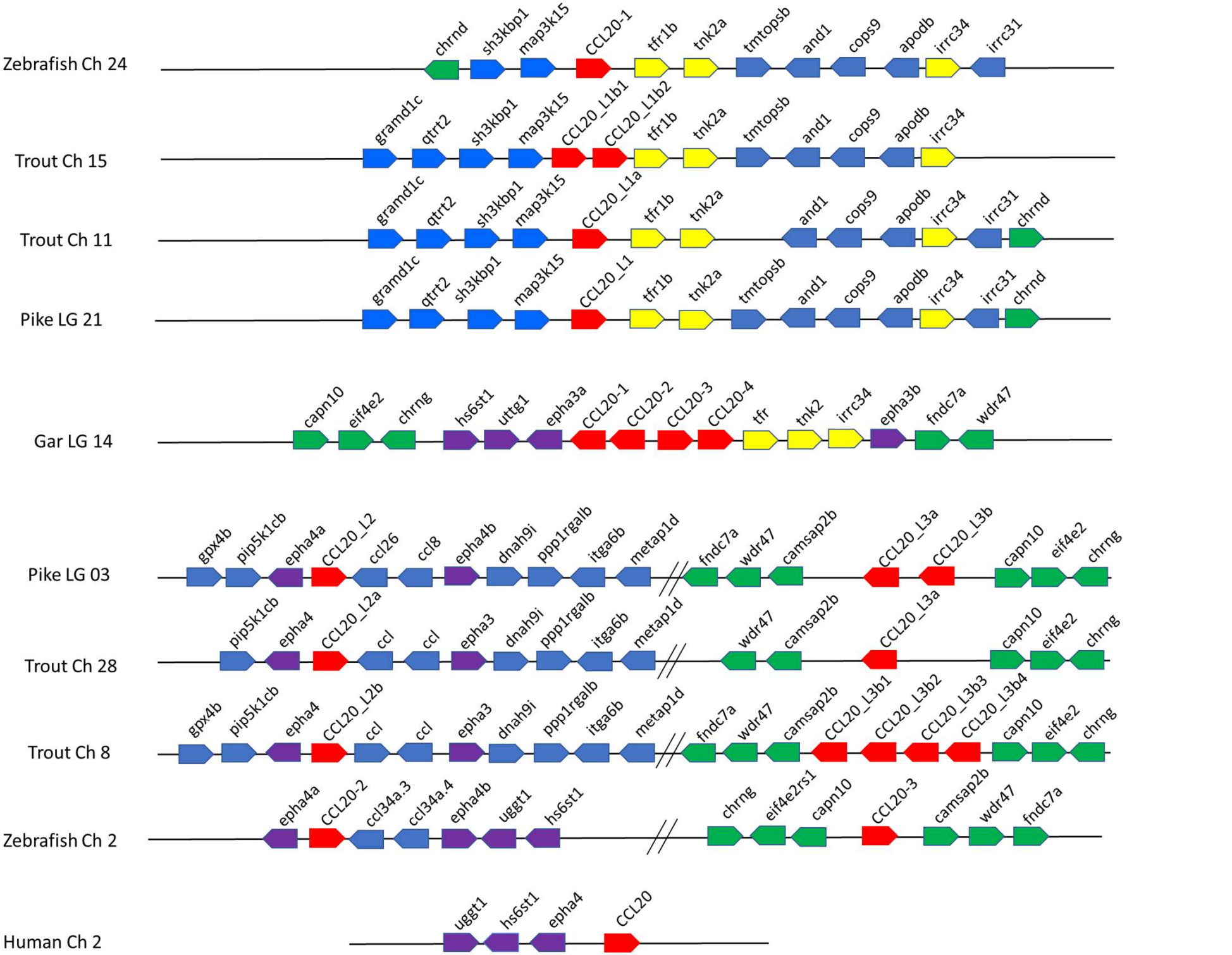

Zebrafish Ch 24
chrnd
sh3kbp1
map3k15
CCL20-1
tfr1b
tnk2a
tmtopsb
and1
cops9
apodb
irrc34
irrc31
Trout Ch 15
gramd1c
qtrt2
sh3kbp1
map3k15
CCL20_L1b1
CCL20_L1b2
tfr1b
tnk2a
tmtopsb
and1
cops9
apodb
irrc34
Trout Ch 11
gramd1c
qtrt2
sh3kbp1
map3k15
CCL20_L1a
tfr1b
tnk2a
and1
cops9
apodb
irrc34
irrc31
chrnd
Pike LG 21
gramd1c
qtrt2
sh3kbp1
map3k15
CCL20_L1
tfr1b
tnk2a
tmtopsb
and1
cops9
apodb
irrc34
irrc31
chrnd
Gar LG 14
capn10
eif4e2
chrng
hs6st1
uttg1
epha3a
CCL20-1
CCL20-2
CCL20-3
CCL20-4
tfr
tnk2
irrc34
epha3b
fndc7a
wdr47
Pike LG 03
gpx4b
pip5k1cb
epha4a
CCL20_L2
ccl26
ccl8
epha4b
dnah9i
ppp1rgalb
itga6b
metap1d
fndc7a
wdr47
camsap2b
CCL20_L3a
CCL20_L3b
capn10
eif4e2
chrng
Trout Ch 28
pip5k1cb
epha4
CCL20_L2a
ccl
ccl
epha3
dnah9i
ppp1rgalb
itga6b
metap1d
wdr47
camsap2b
CCL20_L3a
capn10
eif4e2
chrng
Trout Ch 8
gpx4b
pip5k1cb
epha4
CCL20_L2b
ccl
ccl
epha3
dnah9i
ppp1rgalb
itga6b
metap1d
fndc7a
wdr47
camsap2b
CCL20_L3b1
CCL20_L3b2
CCL20_L3b3
CCL20_L3b4
capn10
eif4e2
chrng
Zebrafish Ch 2
epha4a
CCL20-2
ccl34a.3
ccl34a.4
epha4b
uggt1
hs6st1
chrng
eif4e2rs1
capn10
CCL20-3
camsap2b
wdr47
fndc7a
Human Ch 2
uggt1
hs6st1
epha4
CCL20

Figure 4

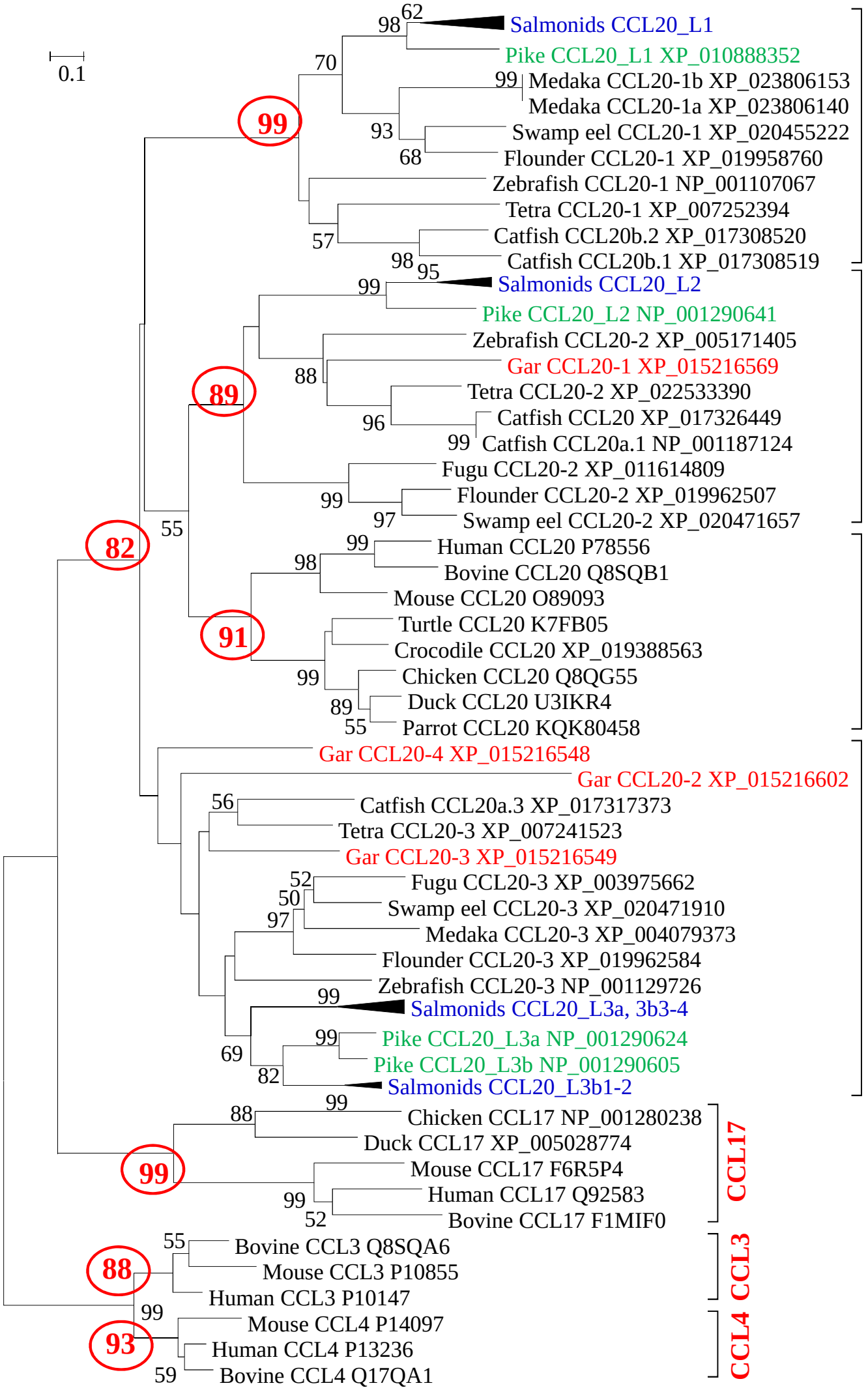

0.1
Salmonids CCL20_L1
Pike CCL20_L1 XP_010888352
Medaka CCL20-1b XP_023806153
Medaka CCL20-1a XP_023806140
Swamp eel CCL20-1 XP_020455222
Flounder CCL20-1 XP_019958760
Zebrafish CCL20-1 NP_001107067
Tetra CCL20-1 XP_007252394
Catfish CCL20b.2 XP_017308520
Catfish CCL20b.1 XP_017308519
Salmonids CCL20_L2
Pike CCL20_L2 NP_001290641
Zebrafish CCL20-2 XP_005171405
Gar CCL20-1 XP_015216569
Tetra CCL20-2 XP_022533390
Catfish CCL20 XP_017326449
Catfish CCL20a.1 NP_001187124
Fugu CCL20-2 XP_011614809
Flounder CCL20-2 XP_019962507
Swamp eel CCL20-2 XP_020471657
Human CCL20 P78556
Bovine CCL20 Q8SQB1
Mouse CCL20 O89093
Turtle CCL20 K7FB05
Crocodile CCL20 XP_019388563
Chicken CCL20 Q8QG55
Duck CCL20 U3IKR4
Parrot CCL20 KQK80458
Gar CCL20-4 XP_015216548
Gar CCL20-2 XP_015216602
Catfish CCL20a.3 XP_017317373
Tetra CCL20-3 XP_007241523
Gar CCL20-3 XP_015216549
Fugu CCL20-3 XP_003975662
Swamp eel CCL20-3 XP_020471910
Medaka CCL20-3 XP_004079373
Flounder CCL20-3 XP_019962584
Zebrafish CCL20-3 NP_001129726
Salmonids CCL20_L3a, 3b3-4
Pike CCL20_L3a NP_001290624
Pike CCL20_L3b NP_001290605
Salmonids CCL20_L3b1-2
Chicken CCL17 NP_001280238
Duck CCL17 XP_005028774
Mouse CCL17 F6R5P4
Human CCL17 Q92583
Bovine CCL17 F1MIF0
Bovine CCL3 Q8SQA6
Mouse CCL3 P10855
Human CCL3 P10147
Mouse CCL4 P14097
Human CCL4 P13236
Bovine CCL4 Q17QA1
Fish CCL20_L1
Fish CCL20_L2
Tetrapod CCL20
Fish CCL20_L3
CCL17
CCL3
CCL4
99
89
82
91
99
88
93

Figure 5

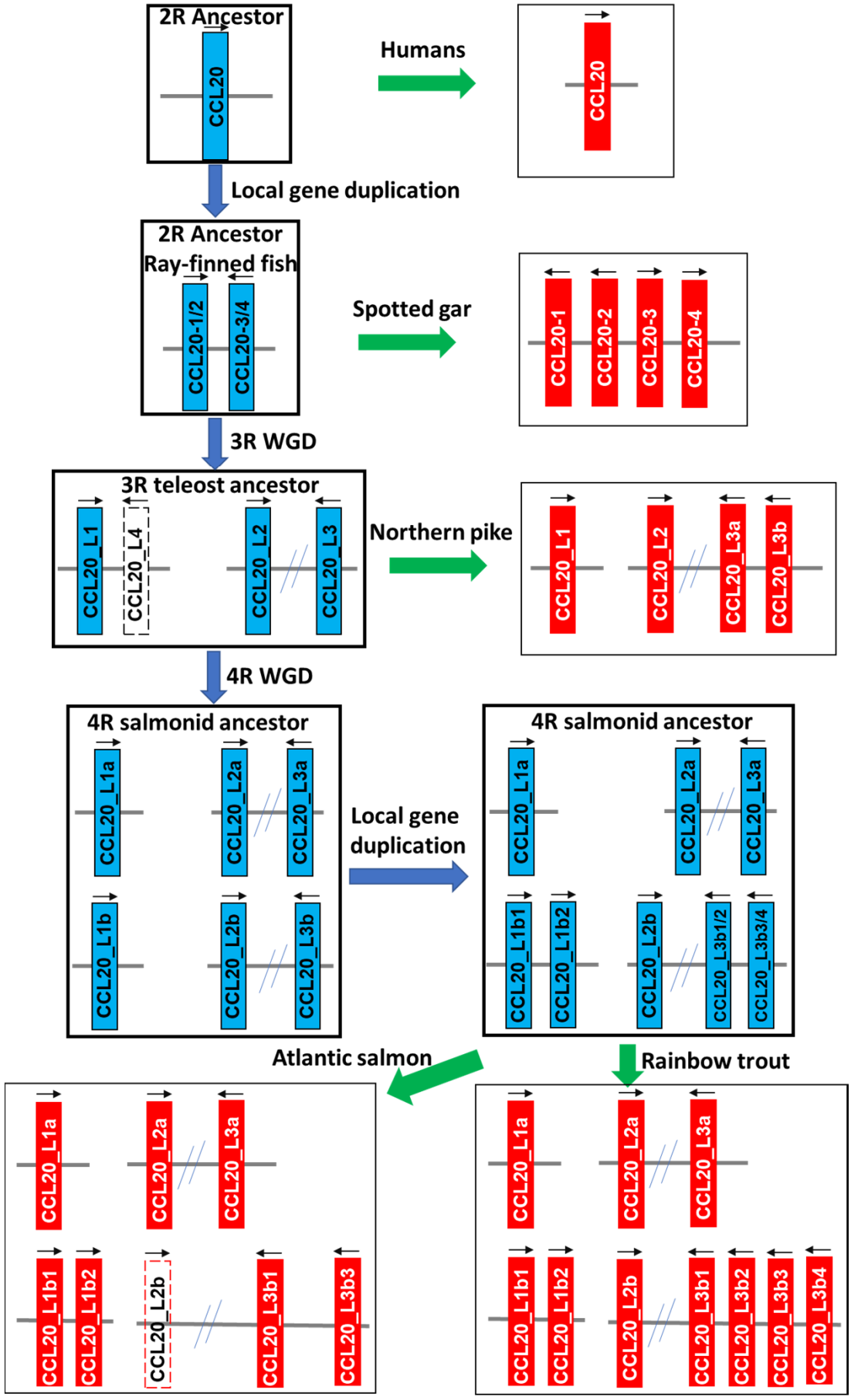
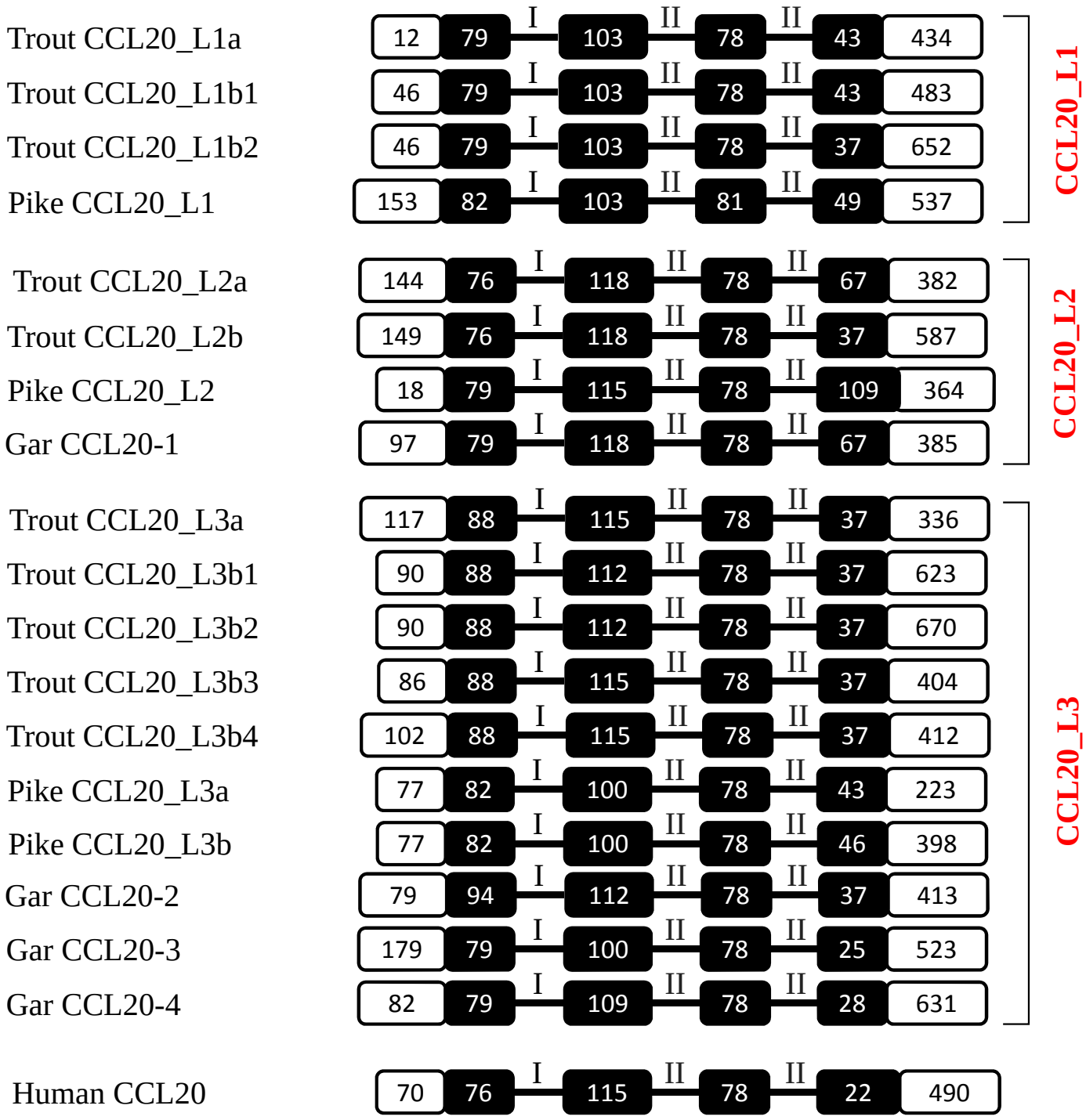

2R Ancestor
CCL20
Humans
CCL20
Local gene duplication
2R Ancestor
Ray-finned fish
CCL20-1/2
CCL20-3/4
Spotted gar
CCL20-1
CCL20-2
CCL20-3
CCL20-4
3R WGD
3R teleost ancestor
CCL20_L1
CCL20_L4
CCL20_L2
CCL20_L3
Northern pike
CCL20_L1
CCL20_L2
CCL20_L3a
CCL20_L3b
4R WGD
4R salmonid ancestor
CCL20_L1a
CCL20_L2a
CCL20_L3a
CCL20_L1b
CCL20_L2b
CCL20_L3b
Local gene duplication
4R salmonid ancestor
CCL20_L1a
CCL20_L2a
CCL20_L3a
CCL20_L1b1
CCL20_L1b2
CCL20_L2b
CCL20_L3b1/2
CCL20_L3b3/4
Atlantic salmon
Rainbow trout
CCL20_L1a
CCL20_L2a
CCL20_L3a
CCL20_L1b1
CCL20_L1b2
CCL20_L2b
CCL20_L3b1
CCL20_L3b3
CCL20_L1a
CCL20_L2a
CCL20_L3a
CCL20_L1b1
CCL20_L1b2
CCL20_L2b
CCL20_L3b1
CCL20_L3b2
CCL20_L3b3
CCL20_L3b4

Figure 6


Trout CCL20_L1a
12 79 I 103 II 78 II 43 434
Trout CCL20_L1b1
46 79 I 103 II 78 II 43 483
Trout CCL20_L1b2
46 79 I 103 II 78 II 37 652
Pike CCL20_L1
153 82 I 103 II 81 II 49 537
CCL20_L1
Trout CCL20_L2a
144 76 I 118 II 78 II 67 382
Trout CCL20_L2b
149 76 I 118 II 78 II 37 587
Pike CCL20_L2
18 79 I 115 II 78 II 109 364
Gar CCL20-1
97 79 I 118 II 78 II 67 385
CCL20_L2
Trout CCL20_L3a
117 88 I 115 II 78 II 37 336
Trout CCL20_L3b1
90 88 I 112 II 78 II 37 623
Trout CCL20_L3b2
90 88 I 112 II 78 II 37 670
Trout CCL20_L3b3
86 88 I 115 II 78 II 37 404
Trout CCL20_L3b4
102 88 I 115 II 78 II 37 412
Pike CCL20_L3a
77 82 I 100 II 78 II 43 223
Pike CCL20_L3b
77 82 I 100 II 78 II 46 398
Gar CCL20-2
79 94 I 112 II 78 II 37 413
Gar CCL20-3
179 79 I 100 II 78 II 25 523
Gar CCL20-4
82 79 I 109 II 78 II 28 631
CCL20_L3
Human CCL20
70 76 I 115 II 78 II 22 490

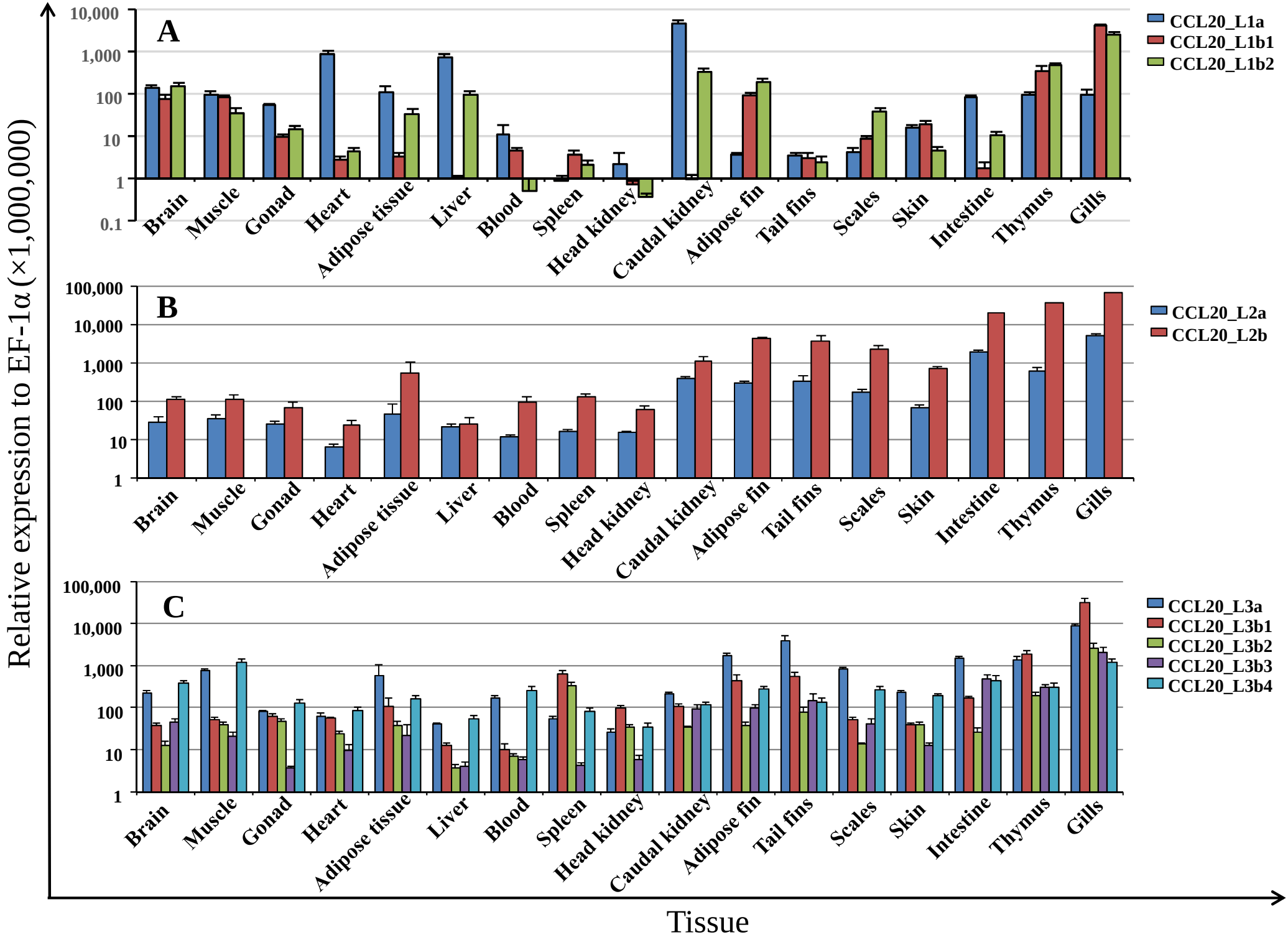

A
B
C
Relative expression to EF-1α (×1,000,000)
Tissue
10,000
1,000
100
10
1
0.1
100,000
CCL20_L1a
CCL20_L1b1
CCL20_L1b2
CCL20_L2a
CCL20_L2b
CCL20_L3a
CCL20_L3b1
CCL20_L3b2
CCL20_L3b3
CCL20_L3b4
Brain
Muscle
Gonad
Heart
Adipose tissue
Liver
Blood
Spleen
Head kidney
Caudal kidney
Adipose fin
Tail fins
Scales
Skin
Intestine
Thymus
Gills

Figure 8

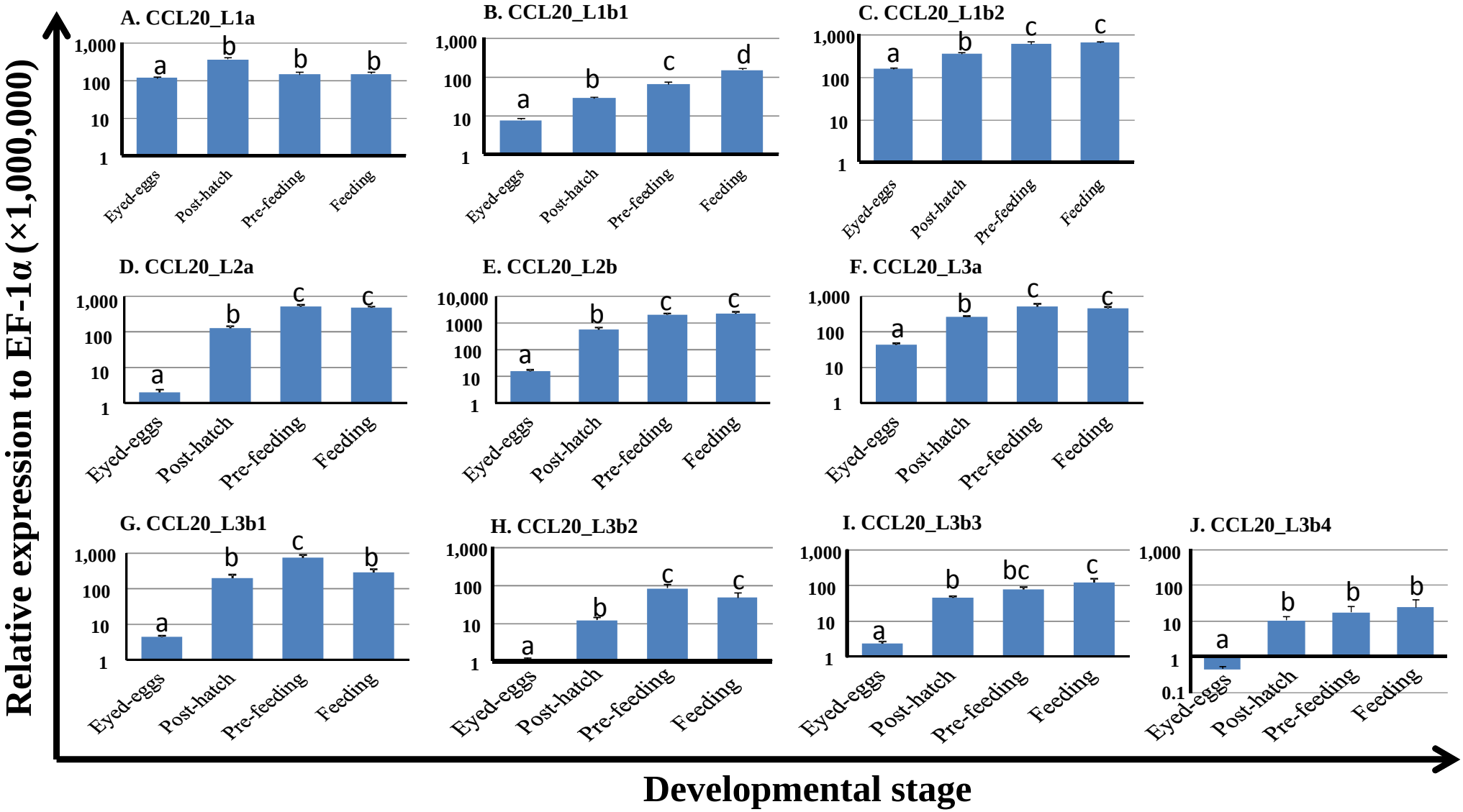
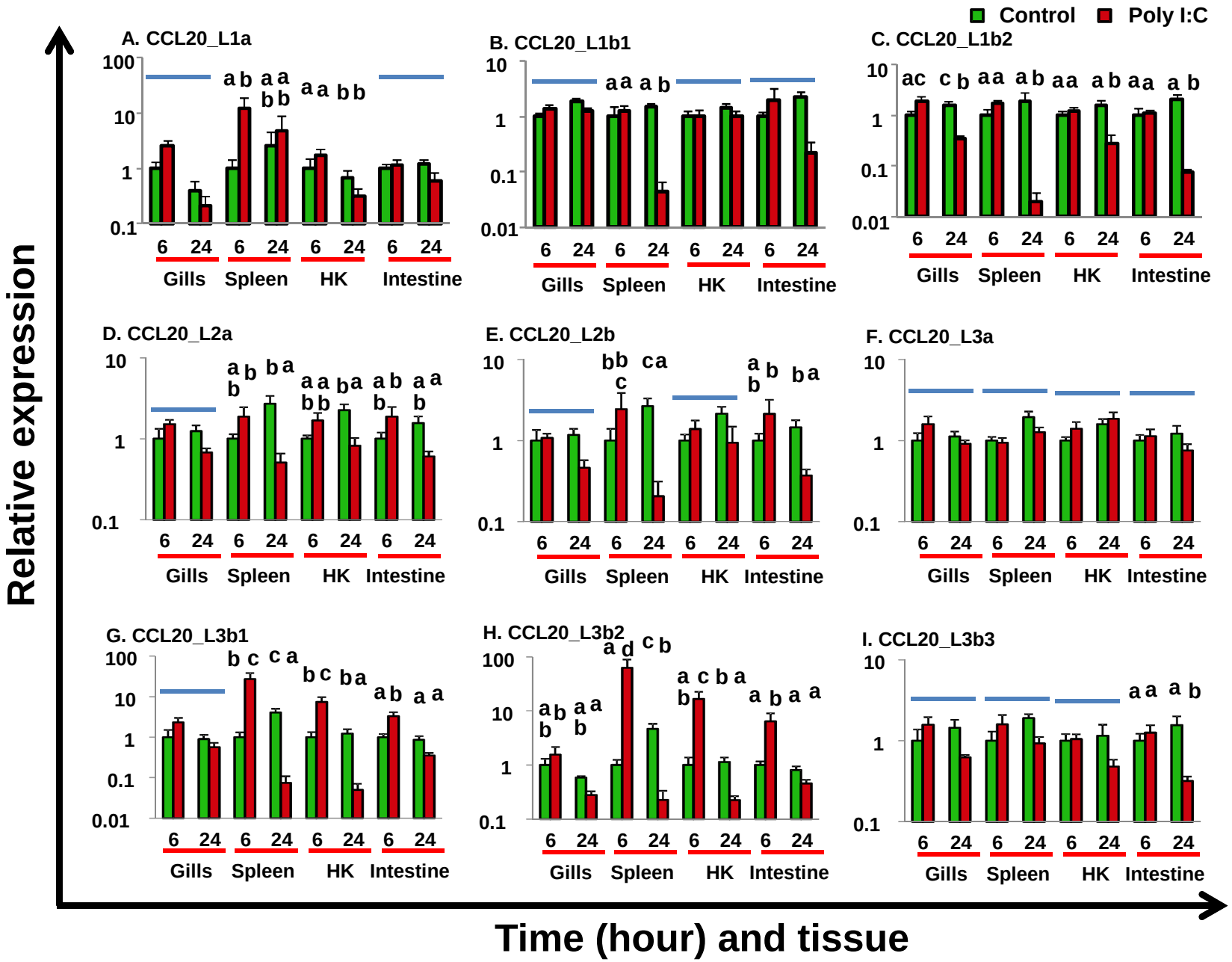

A. CCL20_L1a
B. CCL20_L1b1
C. CCL20_L1b2
D. CCL20_L2a
E. CCL20_L2b
F. CCL20_L3a
G. CCL20_L3b1
H. CCL20_L3b2
I. CCL20_L3b3
J. CCL20_L3b4
Eyed-eggs
Post-hatch
Pre-feeding
Feeding
Relative expression to EF-1α (×1,000,000)
Developmental stage

Figure 9


Control
Poly I:C
A. CCL20_L1a
B. CCL20_L1b1
C. CCL20_L1b2
D. CCL20_L2a
E. CCL20_L2b
F. CCL20_L3a
G. CCL20_L3b1
H. CCL20_L3b2
I. CCL20_L3b3
Relative expression
Time (hour) and tissue
Gills
Spleen
HK
Intestine
6
24

Highlights

Ten CCL20_L genes are present in rainbow trout that can be classified into three groups CCL20_L1-3.

CCL20_L1-3 gene are present in most teleosts, with lineage/species-specific expansion.

The expansion of CCL20_L genes was via whole genome duplication and local gene duplication.

Trout CCL20_L genes are differentially expressed and in general highly expressed in mucosal tissues.

Trout CCL20_L genes are increased during development and following PAMP/cytokine stimulation.

Trout CCL20_L genes are also increased after *Yersinia ruckeri* infection or Poly I:C stimulation *in vivo.*