Developmental and Comparative Immunology 146 (2023) 104735

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Phylogeny and expression of tetraspanin CD9 paralogues in rainbow trout (*Oncorhynchus mykiss*)

Carola E. Dehler [a], Pierre Boudinot [b], Bertrand Collet [b], SamuelA.M. Martin [a,*]

[a] Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen, UK
[b] Université Paris-Saclay, INRAE, UVSQ, Virologie et Immunologie Moléculaires, 78350, Jouy-en-Josas, France

ARTICLE INFO

Keywords:
Immune system
Salmonids
CD9
Whole genome duplication

ABSTRACT

CD9 is a member of the tetraspanin family, which is characterised by a unique domain structure and conserved motifs. In mammals, CD9 is found in tetraspanin-enriched microdomains (TEMs) on the surface of virtually every cell type. CD9 has a wide variety of roles, including functions within the immune system. Here we show the first in-depth analysis of the *cd9* gene family in salmonids, showing that this gene has expanded to six paralogues in three groups (*cd9a*, *cd9b*, *cd9c*) through whole genome duplication events. We suggest that through genome duplications, *cd9* has undergone subfunctionalisation in the paralogues and that *cd9c1* and *cd9c2* in particular are involved in antiviral responses in salmonid fish. We show that these paralogues are significantly upregulated in parallel to classic interferon-stimulated genes (ISGs) active in the antiviral response. Expression analysis of *cd9* may therefore become an interesting target to assess teleost responses to viruses.

1. Introduction

CD9 is a member of the tetraspanin family, which is characterised by conserved protein structure including four transmembrane domains and cysteine motifs (Horejsi and Vlcek, 1991). The crystal structure of human CD9 has been determined in humans showing typical tetraspanin conformation of 4 transmembrane regions (TM), a small extracellular loop (EC1), a large extracellular loop (EC2), a small intracellular loop (IC), as well as short intracellular C- and N-terminal ends (Horejsi and Vlcek, 1991; Umeda et al., 2019). Whilst TMs are highly conserved between different tetraspanin proteins, the extracellular loops and in particular EC2, show diverged sequences, which influence the range of protein binding partners (Umeda et al., 2020). Between tetraspanins, the mushroom-like structure of EC2 is highly variable in length and composition, which has been used to classify different tetraspanins (Huang et al., 2005).

Four major clades of tetraspanins have been identified, with the CD clade forming the largest family spanning 14 genes in vertebrates, including CD9 (Garcia-España et al., 2008). Due to their molecular similarities, *cd9*, *tspan2* and *cd81* are grouped into a lineage with a unique intron configuration and EC2 structure (Huang et al., 2010). This lineage is thought to have appeared in early vertebrate evolution and may have been derived from an ancient pre-vertebrate *tspan8*-like tetraspanin (Huang et al., 2005). The vase tunicate (*Ciona intestinalis*) has four tetraspanins closely related to the *cd9/tspan2/cd81* lineage, indicating that the ancestral genes to this lineage were present in deuterostomes before the emergence of vertebrates (Garcia-España et al., 2008). In the ancestral teleost, 80 ancient tetraspanin genes were proposed (Cao and Tan, 2018). In zebrafish, 44 of those are retained and 36 are lost (Cao and Tan, 2018). These genes were further duplicated and lost in subsequent emerging lineages (Cao and Tan, 2018).

Tetraspanin proteins are relatively small and highly abundant in every cell type and assemble with themselves and other transmembrane receptors to form tetraspanin-enriched microdomains (TEMs), which vary greatly in size between cell types (Yáñez-Mó et al., 2009). Within the tetraspanin web there are three types of associations: primary associations between tetraspanins and non-tetraspanin molecules, secondary associations between different tetraspanins and tertiary interactions that are indirect and cluster in the TEMs, enabling lateral dynamic organisation in the membrane and cross-talk with intracellular signalling and cytoskeletal structures (Levy and Shoham, 2005a). Different partnerships can be formed in different cell types with extracellular or intracellular domains of partner molecules (Levy and Shoham, 2005a). Tetraspanins are involved in a variety of molecular processes, including inter- and intracellular signalling, cell adhesion and membrane fusion (reviewed in Yáñez-Mó et al., 2009). TEMs on the

* Corresponding author. Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen, AB24 2TZ, UK.
E-mail address: sam.martin@abdn.ac.uk (SamuelA.M. Martin).

https://doi.org/10.1016/j.dci.2023.104735
Received 13 January 2023; Received in revised form 12 May 2023; Accepted 12 May 2023
Available online 13 May 2023
0145-305X/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

plasma membrane of cells are also highly involved in virus infection (reviewed in Florin and Lang, 2018; Hantak et al., 2019).

Mammalian CD9 has been found to associate with a wide variety of binding partners, including adhesion molecules (e.g., integrins), growth factors, signalling molecules and immune system molecules, reflecting the wide range of roles of this gene (reviewed in Reyes et al., 2018).

In mammals, *cd9* has been proposed as marker gene for a suite of cell types relevant to the immune system, including dendritic cells (DCs) (Unternaehrer et al., 2007), primitive hematopoietic stem cells (Karlsson et al., 2013), monocytes (Zilber et al., 2005) and marginal zone B cells (Won and Kearney, 2002). CD9 has several important functions in immunity, through its involvement in leukocyte differentiation, activation pathways and antigen presentation by MHC class II (reviewed in Brosseau et al., 2018). CD9 plays a role in T-cell activation in humans and mice (Reyes et al., 2015, 2018): it induces T cell co-stimulation in a CD28-independent manner and cooperates with CD3 in T cell activation during responses (Tai et al., 1996; Kobayashi et al., 2004). CD9 is also a key regulator of inflammation and is particular important for the secretion of IL10 by regulatory B cells (Ha et al., 2005). CD9 is also expressed in human macrophages and is transiently downregulated when exposed to LPS (Suzuki et al., 2009). CD9 is also a biomarker for exosomes secreted by antigen-presenting cells (APCs) (Hulsmans and Holvoet, 2013), with potential effects on antiviral innate immune responses (reviewed in Kouwaki et al., 2017).

TEM platforms including CD9 are attractive to hijacking by viruses as they locally increase the concentration of specific receptors, thereby improving infectivity of a cell (Yáñez-Mó et al., 2009). For example, in MERS-CoV infections, CD9 is required to bring primary viral receptor DPP4 and triggering TMPRSS2 in close proximity to allow virus fusion and infection (Earnest et al., 2017). Similarly, depletion of CD9 reduces influenza A virus (IAV) entry into mammalian cells (Earnest et al., 2015), and feline immunodeficiency virus and canine distemper virus (CDV) also use CD9 to infect target cells (Monk and Partridge, 2012).

Although CD9 and its role during viral infections is well studied in mammals, less is known about the *cd9* gene family and its functions in fish. The most complete studies have been in zebrafish (*Danio rerio*) and a previous study suggested to name the *cd9/tspan/cd81* lineage as *cd9* lineage and identified three *cd9* paralogues (*cd9a, cd9b, cd9c*), two *cd81* paralogues (*cd81*a, *cd81*b) and two *tspan2*TSPAN2 paralogues (*tspan2*a, *tspan2*b) (Briolat et al., 2014). This paralogue expansion is consistent with the basal teleost-specific whole genome duplication event (Ts3R) 316-226 million years ago (Ma) (Hurley et al., 2007). Whilst *cd9b* was found to be crucial to egg fertilisation and egg production in zebrafish (Greaves et al., 2021), *cd9c* was found to be highly induced in zebrafish infected with either infectious hematopoietic necrosis virus (IHNV) or Chikungunya virus (CHIKV) in parallel to upregulation of a suite of classic interferon stimulated genes (ISGs) (Briolat et al., 2014; Levraud et al., 2019). In Japanese flounder (*Paralichthys olivaceus*) three *cd9* paralogues (*cd9.1, cd9.2* and *cd9.3*) were found, with *cd9.1* and *cd9.3* being induced by bacterial and viral pathogens in a tissue-specific manner (He et al., 2021). The red stingray (*Dasyatis akajei*) is reported to have a single *cd9* gene with high expression in immune associated organs (Zhu et al., 2006). Furthermore, *cd9* in Arctic lamprey (*Lethenteron camtschaticum*) was significantly upregulated after LPS stimulations, also suggesting immune involvement of this gene (Wu et al., 2012).

The first salmonid *cd9* was identified by Fujiki et al. (2002) following gene enrichment by suppression subtractive hybridisation experiments in the head kidney RNA of Atlantic salmon exposed to IHNV. The Atlantic salmon sequence was then used to identify a *cd9* homolog in rainbow trout (Fujiki et al., 2002). A second salmonid *cd9* was characterised in rainbow trout by Castro et al. (2015) with 34.4% amino acid identity shared with the first *cd9* paralogue. The second *cd9* paralogue was proposed to play a role in B cell mediated immunity after infection with pathogens and vaccination (Castro et al., 2015). However, these studies are unlikely to reflect the complete *cd9* paralogue repertoire of salmonids due to the additional round of whole genome duplication 106 Ma (Ss4R; 95% Bayesian credibility interval 89–125 Ma) (Gundappa et al., 2021) and the whole genomes published for salmonids give us the opportunity to define the *cd9* genes in depth.

Here we identify 6 *cd9* paralogues in rainbow trout. We present for each paralog the gene structure and protein model, conserved motifs, gene synteny and phylogenetic analysis. Furthermore, we assess the potential of these paralogues to be involved in antiviral responses through *in silico* promoter analysis of immune-relevant transcription factor motifs and gene expression analysis.

2. Materials and methods

2.1. Identification of zebrafish and rainbow trout cd9 paralogues

Zebrafish *cd9a* (ENSDARG00000005842) and *cd9b* (ENSDARG00000016691) were identified in the NCBI database and used to blast against the zebrafish genome (GRCz11, GCA_000002035.4) to ensure all paralogues were identified. Zebrafish *cd9c* was first suggested by Briolat et al. (2014) and is annotated as zgc:65811 in NCBI and Ensembl (ENSDARG00000100904), showing high percentage of sequence identity with the other *cd9* paralogues and suggest confidence as a third *cd9* paralogue.

Each of the zebrafish *cd9* paralogue protein sequences were used to blast against the latest rainbow trout genome annotation (Genome Omyk_2.0, GCA_025558465.1) on NCBI which is cross-referenced with the latest Ensembl annotations. This process was repeated for other salmonid species with a published whole genome, namely Atlantic salmon (*Salmo salar;* Ssal_v3.1, GCA_905237065.2), Chinook salmon (*Oncorhynchus tshawytscha;* Otsh_v2.0, GCA_018296145.1), coho salmon (*Oncorhynchus kisutch;* Okis_V2, GCA_002021735.2), brown trout (*Salmo trutta;* SalTru1.1, GCA_901001165.1) and arctic char (*Salvelinus alpinus;* ASM291031v1, GCA_002910315.1), to cross-reference and validate the paralogue sequences identified in the rainbow trout genome. Nomenclature of the identified CD9 paralogues was assigned based on the sequence alignment, gene and protein structure, gene synteny and phylogenetic placement.

2.2. Intron and exon structure of rainbow trout cd9 paralogues

Intron and exon structures of the zebrafish and salmonid paralogue genes were compared as an additional measure of making sure that they all belong to *cd9* family. Intron and exon lengths were taken of the longest isoform of each paralogue from NCBI, which is based on RNAseq data. Intron and exon phases were taken from the corresponding sequences on Ensembl.

2.3. Protein structure and conserved motifs of rainbow trout cd9 paralogues

Protein structure and identification of conserved motifs was performed by TMHMM server (v.2.0) (http://www.cbs.dtu.dk/services/TMHMM/). Conserved "CCG" and single cysteine residues, typical of tetraspanins, were identified in the nucleotide sequences by manual inspection and N-glycosylation motifs were identified using NetNGlyc (v.1.0) (https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0).

2.4. Gene synteny of cd9 paralogues

Gene synteny was analysed with Genomicus (database version 109.01; https://www.genomicus.bio.ens.psl.eu/genomicus-109.01/cgi-bin/search.pl). Ensembl access numbers of the cd9 paralogues were used as input. In cases were Ensembl accession numbers were not yet included in the current Genomicus release (i.e. RapidEnsembl accession numbers) or had no Ensembl accession number, gene synteny was taken from NCBI gene tracks.

2.5. Phylogenetic analysis

Due to the evolutionary and molecular closeness of *cd9* with *cd81* and *tspan2* we included protein sequences for all genes of this lineage. As an outgroup, we also included four ancient tetraspanin genes from the vase tunicate (*Ciona intestinalis*) as literature suggests these as the closest related genes to the vertebrate *cd9/cd81/tspan2* (Garcia-España et al., 2008). Accession numbers for these genes were taken from Garcia-España et al. (2008) and adapted as needed where the accession numbers were outdated.

In terms of species, we included representatives of each non-fish vertebrate group: mouse (*Mus musculus;* mammal), Western clawed frog (*Xenopus tropicalis;* amphibian), green anole (*Anolis carolinensis;* reptile) and chicken (*Gallus gallus,* avian), West African lungfish (*Protopterus annectens*) as representative of a "living fossil" in fish, elephant shark (*Callorhinchus milii*) and great white shark (*Carcharodon carcharias*) as a representatives of cartilaginous fish, spotted gar (*Lepisosteus oculatus*) as representative of fish prior Ts3R, zebrafish, three-spined stickleback (*Gasterosteus aculeatus*), Atlantic cod (*Gadus morhua*) and Japanese flounder (*Paralichthys olivaceus*) as representatives of species after teleost specific whole genome duplication, northern pike (*Esox lucius*) as closest non-salmonid relative to salmonids, lake whitefish (*Coregonus clupeaformis*) as sister-group to the genera *Salmo* and *Oncorhynchus*, as well as rainbow trout and Atlantic salmon as our target species.

Furthermore, we constructed a separate fish specific *cd9* gene tree based on species selected across the whole phylogenetic diversity of this vertebrate group. The accession numbers for all species used for the *cd9/cd81/tspan2* and fish *cd9* tree can be found in supplementary material S1.

Protein sequences were transferred to MEGA11 and aligned with Clustal W. Evolutionary history analysis was performed using the Maximum Likelihood method and Jones-Taylor-Thronton (JTT) matrix-based model. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join (NJ) and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with superior log likelihood value. The resulting bootstrapped phylogenetic tree (n = 500) was manually annotated for clarity.

2.6. Promoter analysis of cd9 rainbow trout and atlantic salmon paralogues

To determine potential transcription factor binding sites for the *cd9* genes 2000 bp upstream of the transcription start site (TSS) of each paralogue were retrieved and analysed using TFBIND (https://tfbind.hgc.jp/) based on the TRANSFAC R.3.4 database.

For interferon stimulated response element (ISRE) sites we used both the database's consensus motif (5′ strand: "CAG TTTC WC TTTY CC") and the previously published salmonid-specific consensus motif (5′ strand: "DS TTTC $N_{1\text{-}2}$ TTTC H") (Castro et al., 2008) to assess the specificity of proposed motif sites and use a threshold of 0.8 confidence score output by TFBIND.

2.7. Expression analysis of salmonid cd9 paralogues in RNAseq datasets

To investigate the expression of *cd9* paralogues in immune-challenged salmonids we selected three RNAseq datasets that were focussed on antiviral activity in immune organs or in cell lines. These data sets were (Huang et al., 2010: accession for RNAseq experiment GSE176399; https://www.ncbi.nlm.nih.gov/gds) which were obtained from rainbow trout intestine and spleen RNA after 48 h infection with infectious hematopoietic necrosis virus (IHNV) compared to control. A further data set from our own lab involving two embryonic Chinook salmon cell lines (CHSE) stimulated with recombinant type I interferon (IFNA2) was used (Dehler et al., 2019). CHSE-EC is a transgenic cell line stably expressing eGFP and Cas9 produced previously and used as control in the IFNA2 stimulation experiment (Dehler et al., 2016). CHSE-GS2 is a gene edited mutant of CHSE-EC with a CRISPR/Cas9 knock-out of the *stat2* gene, which is involved in the antiviral interferon response cascade (Dehler et al., 2019). The raw data of this project is deposited in NCBI under BioProject 495492 (accession: SRX5803291-SRX5803302). Raw count data of the *cd9* paralogues identified in all datasets were used to produce graphs and determine statistical significance with one-way ANOVAs in R.

3. Results

3.1. Rainbow trout cd9 paralogues

In total we identified 6 *cd9* paralogues that can be split into three clades (*cd9a*, *cd9b*, *cd9c*) based on the zebrafish annotation and sequence similarity. There is strong evidence for two copies (Ss4R paralogues) for each clade in rainbow trout (*cd9a1* and *cd9a2, cd9b1* and *cd9b2, cd9c1* and *cd9c2*). We also compared our *cd9* paralogues to the previously identified *cd9* paralogues in rainbow trout (Fujiki et al., 2002; Castro et al., 2015). We found that the paralogue identified in Fujiki et al. (2002) shows highest sequence identity to our *cd9c2*, whilst the paralogue identified in Castro et al. (2015) is closest in sequence identity to the CD9b clade. The NCBI and Ensembl accession numbers can be found in supplementary material S1.

Notably, rainbow trout *cd9b1* and *cd9b2* are on unplaced scaffolds, not associated with a chromosome. This is also the case for *cd9b1* homologs in other salmonids (data not shown). Homologs for *cd9b2* could only be found in rainbow trout and coho salmon and a truncated sequence in Atlantic salmon (160 amino acids, 6 exons), which are all on unplaced scaffolds as well. No homologs to *cd9b2* could be found in brown trout, Chinook salmon or Arctic char.

3.2. Intron/exon structure

We summarised the intron and exon structure of the rainbow trout *cd9* paralogues (Fig. 1), which showed high conservation of eight exons persevered among the *cd9* paralogues. Whilst *cd9a* and *cd9b* paralogues almost have identical intron/exon structures (apart from *cd9a1* exon 6 being 3 exons shorter than the other paralogues), *cd9c* paralogues have a shorter exon 1 and longer exon 8.

Almost all introns of *cd9*, *cd81* and *tspan2* paralogues of rainbow trout were found to be phase 0 (no disruption of codons), apart from intron 2, which is phase 1 (codon disruption between first and second base). This can be expressed as intron string "0100000". We compared this with the intron strings of other vertebrate species used for phylogenetic analysis and found the same intron string in elephant shark, pike, zebrafish and mouse *cd9* paralogues. Interestingly, gar *cd9* has an intron string of "0120000", indicating a codon disruption between the second and third nucleotide in intron 3. In the vase tunicate, *citspan1* and *citspan15* have the same intron string as the *cd9/cd81/tspan2* lineage in rainbow trout, elephant shark, pike, zebrafish and mouse. *citspan7* has the same configuration as gar *cd9* and *citspan2* has a unique intron string of "0000000" (no codon disruptions) that was not found in any other sequences analysed here.

3.3. Protein structure and conserved motifs

We identified the different structural domains in CD9 paralogue sequences. The annotated CD9 paralogue protein alignment and the general protein structure is shown in Fig. 2 A and B respectively. The corresponding amino acid lengths of each structural element in Fig. 2B for each paralogue is shown in Supplementary Table S2. All transmembrane (TM) lengths are conserved with 23 amino acids among the paralogues, apart from TM2, which is only 20 amino acids in CD9c1 and CD9c2. The N-terminal length is highly conserved among all paralogues,

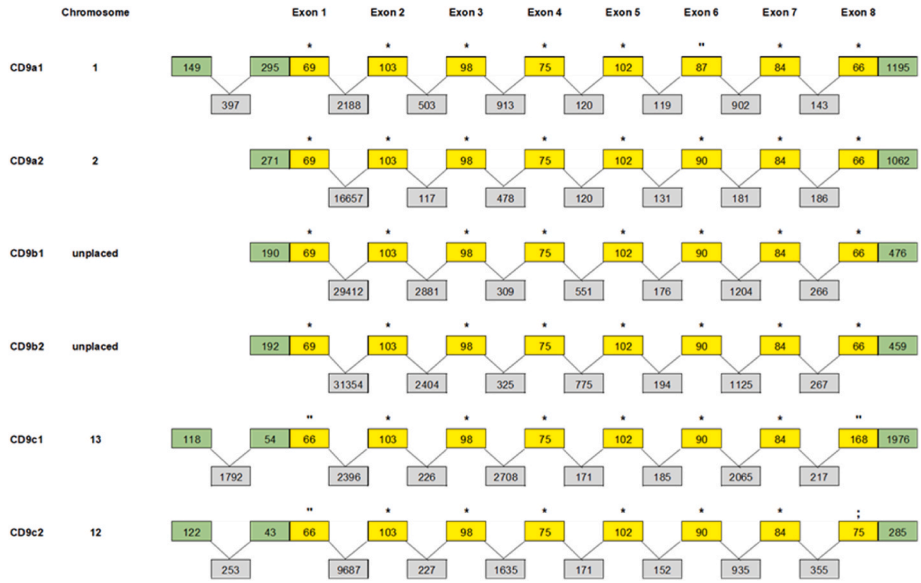

Fig. 1. Intron/Exon structure of rainbow trout (*Oncorhynchus mykiss*) CD9 paralogues. Green boxes: 5′ and 3′ untranslated (UTR) regions, gray boxes: introns, yellow boxes: exons. Icons above exons mark how the lengths are conserved among the paralogues. All accession numbers are in the S1 table. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

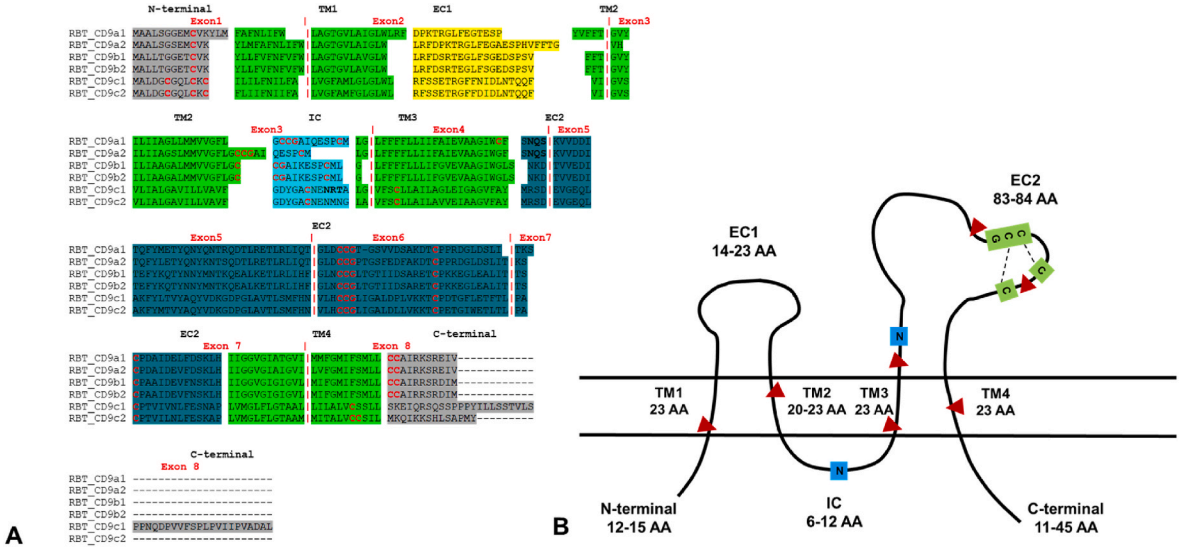

Fig. 2. Protein alignment of rainbow trout (*Oncorhynchus mykiss*) CD9 paralogues and the consolidated protein structure of the paralogues. A. Alignment of CD9 paralogues, annotated with protein structure elements and conserved motifs. Gray = N-terminal and C-terminal, green = transmembrane regions (TM), yellow = small extracellular loop (EC1), blue = intracellular loop (IC), teal = large extracellular loop (EC2). N-glycosylation sites are in bold. Conserved cysteine motifs are in red and bold. Exon boundaries are denoted with bold red lines. B. Consolidated protein structure of CD9 paralogues. Red triangles = exon/exon junctions. Blue "N" = sites of N-glycosylation, Green highlighted = conserved cysteine motifs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

except for CD9a1. The C-terminal lengths is highly conserved among CD9a and CD9b paralogues but is longer in CD9c1 and CD9c2 (45 and 14 amino acids, respectively). The small extracellular loop (EC1) is highly conserved in CD9b and CD9c paralogues but is shorter in CD9a1 (14 amino acids) and longer in CD9a2 (23 amino acids). The large extracellular loop (EC2) is conserved among CD9a1, CD9b1 and CD9b2 (83 amino acids) and CD9a2, CD9c1 and CD9c2 (84 amino acids). The intracellular loop (IL) is conserved in CD9a1, CD9c1 and CD9c2 but shorter in CD9a2 and CD9b paralogues (6 and 11 amino acids, respectively).

The highly conserved cysteine motif "CCG" could be identified in all rainbow trout CD9 paralogues, as well as two additional conserved cysteines (C). An additional CCG motif is present in CD9a paralogues at the junction of TM2 and IL, but it is unclear what function this has. In the EC2, all paralogues had 4 C residues (including the CCG). Additional C residues were also identified in N-terminal, IC, TM2, 3 and 4 and C-terminal regions and appear paralogue specific. As these cysteine residues are thought to be sites of palmitoylation that contribute to binding between different tetraspanin proteins, the difference in C motifs may lead to different preferential binding partners for the different paralogues.

N-glycosylation sites were identified in some rainbow trout CD9 paralogues in different structural areas. CD9a paralogues have a N-glycosylation site in the EC2 but is not predicted to be functional in

CD9a1. CD9a2 has a second predicted N-glycosylation site in EC2 but this is predicted to be non-functional. CD9c1 has a predicted functional N-glycosylation site in IC. No N-glycosylation sites were found in CD9b1, CD9b2 and CD9c2. N-glycosylation sites are summarised in supplementary material S3.

3.4. Chromosomal regions and gene synteny

We found patterns of gene synteny overlap between non-fish vertebrates and fish, as well as among different groups of fish (Fig. 3 and Supplementary Fig. S4 [*cd9a*], S5 [*cd9b*] and S6 [*cd9c*]). Among species with only a single copy of cd9, synteny is extremely conserved upstream of cd9, from mammal to elephant shark. Robustly shared genes across non-fish vertebrates, non-teleost cd9 and *cd9a* and *cd9b* clades of teleosts include *neurotrophin 3* (*ntf3*), *anoctamin 2* (*ano2)* and *von willebrand factor* (*vwf*). Interestingly, these genes are not present in the synteny of *cd9c* of any fish species examined.

Within the *cd9a* clade, there is a block of genes found downstream of cd9a in zebrafish and three-spined stickleback that is conserved in a rearranged position upstream of cd9a paralogues in pike and salmonids. Within the examined teleost species, there appear to be three groupings based on synteny: group 1 including zebrafish, common carp, channel catfish and Mexican tetra, group2 including all non-salmonid teleosts and group 3 including northern pike and salmonids. CD9 of elephant shark and *cd9a* of non-teleost fish also share gene synteny among each other. Interestingly, sterlet appears to have a duplication of cd9a, supported by the almost identical synteny of the two paralogues.

The *cd9b* clade has a similar group pattern as cd9a, however, no extensive gene rearrangement is evident between group1/2 and group 3. Gene synteny is more conserved between cd9b homologs of group 2 and group 3, especially downstream of cd9b. It should be noted that due to the cd9b paralogues of salmonid species not being assigned to a chromosome, the synteny information around these genes is limited.

Lastly, the cd9c clade synteny is very unique and shares almost no genes with single-copy cd9/cd9a or cd9b genes regardless of species. Within the *cd9c* clade, however, gene synteny is most stably conserved among teleost species, including salmonids. CD9c homologs were also found in non-teleost species, such as great white shark, thorny skate spotted gar and gray bichir. Synteny was almost completely conserved between great white shark and thorny skate, whilst gray bichir shares some gene synteny with group 1 teleosts. Interestingly, among teleosts, medaka appears to have undergone a thorough gene synteny rearrangement. Unfortunately, extensive comparisons with zebrafish *cd9c* could not be performed, due to a lack of synteny annotation of this gene in zebrafish. Zebrafish *cd9c* gene (zgc:65811; ENSDARG00000100904) was on the unplaced scaffold KN149855.1 (8111–10,664), with only one other gene [ENSDARG00000099947] encoding a GPCR, which is also found in the *cd9c* synteny of other teleosts, including salmonids.

There was a surprisingly great overlap of synteny between tetrapods and fish from distantly related families, which further increased between closely related species such as pike and rainbow trout. Interestingly, there is more notably consistency in synteny between the *cd9a* and *cd9b* clades, whereas the *cd9c* clade synteny is more divergent.

3.5. Evolutionary relationship between the paralogues by phylogeny

Due to the close relationship of *cd9* with *cd81* and *tspan2*, referred to as a lineage, we present a phylogenetic tree including all three genes (Fig. 4). In addition to the rainbow trout and Atlantic salmon sequences, we selected sequences from species as representatives for relevant evolutionary events, these included the spotted gar, great white shark and elephant shark as these species did not undergo teleost specific duplication, zebrafish, three-spined stickleback, Atlantic cod and Japanese flounder, representing teleosts withTs3R duplication, northern pike as sister taxa of salmonids that did not undergo Ss4R duplication and lake whitefish, a sister lineage to the *Salar* and *Oncorhynchus* genera. Mouse, chicken, frog, green anole and western clawed frog were chosen as tetrapod vertebrate group representatives and finally vase tunicate as a nonvertebrate chordate that contains ancestral tetraspanin-related sequences of this lineage. An extended *cd9* phylogenetic tree including fish species across their evolutionary diversity can be found in the supplementary material S7.

The vase tunicate genes, which represent the ancestral tetraspanins act as an outgroup and branch distantly from the vertebrate *cd9/cd81/tspan2* lineage. Within the vertebrate cluster, two subtrees are evident, with *cd9* in one group and *cd81* and *tspan2* in the second. Elephant shark had one copy each of *cd9, cd81* and *tspan2*, however, in other non-teleosts such as great white shark, thorny skate, spotted gar and gray bichir cd9a and cd9c homologs were identified, additionally to single copies of cd81 and tspan2. The cd9a homologs for non-teleost fish were

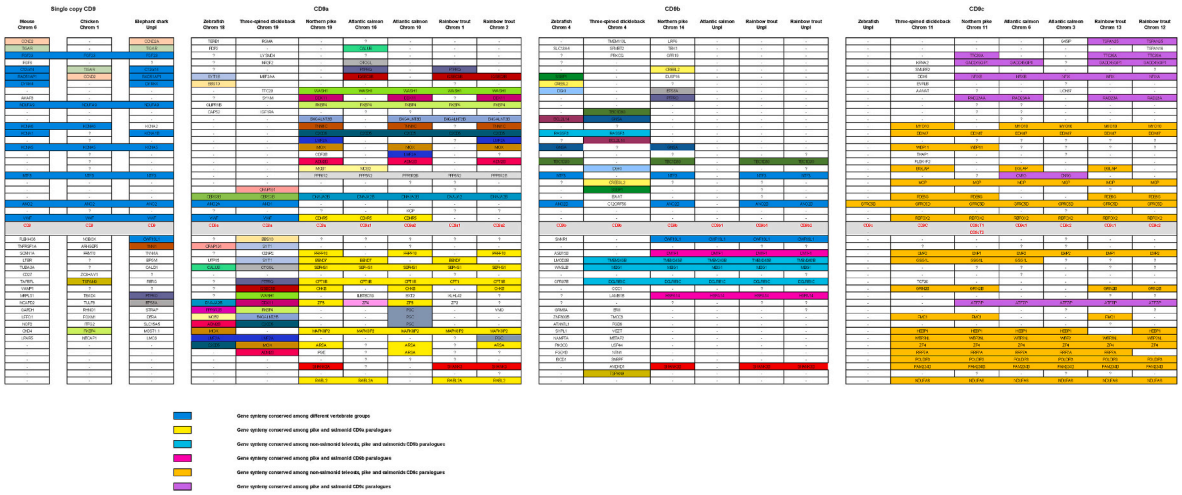

Fig. 3. Gene synteny of mouse (*Mus musculus*), chicken (*Gallus gallus*), elephant shark (*Callorhinchus milii*), zebrafish (*Danio rerio*), three-spined stickleback (*Gasterosteus aculeatus*), northern pike (*Esox lucius*), Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) CD9 paralogues as analysed with Genomicus or NCBI gene tracks. Dashes ("-") represent a missing gene and question marks ("?") indicate an unannotated gene. The colour figure legend reflects the main gene synteny conservation groups. Additional colours in the figure show gene synteny shared between species, however, at different positions of the chromosome. An extended version of the gene synteny analysis can be found in the supplementary material S4 (*cd9a*), S5 (*cd9b*) and S6 (*cd9c*). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

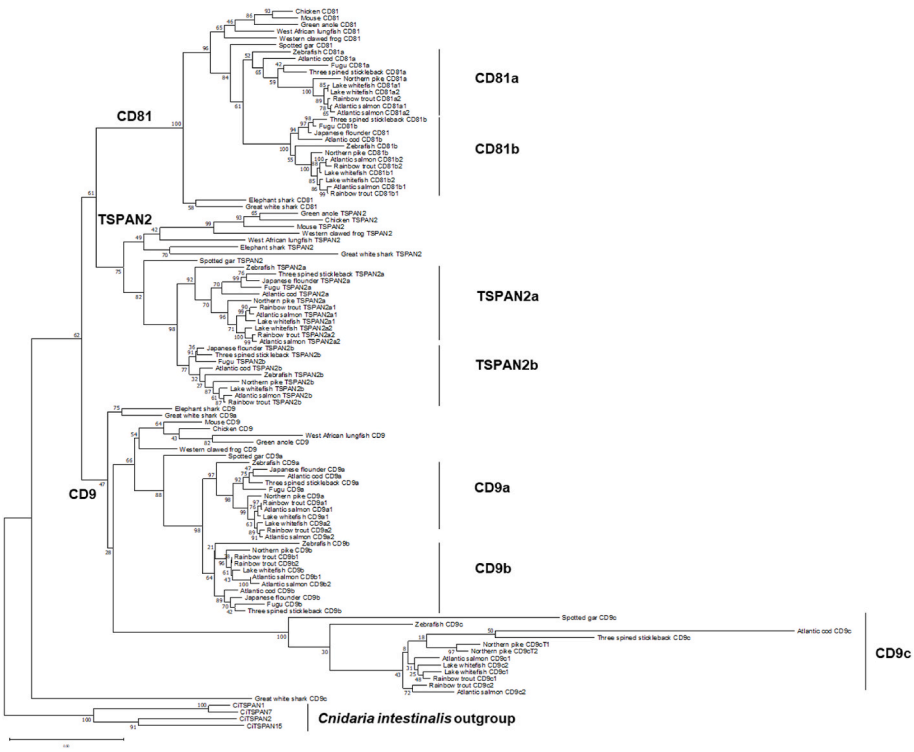

Fig. 4. Phylogenetic tree with the highest log likelihood of the cd9/tspan2/cd81 lineage as inferred by the Maximum likelihood method and John-Taylor-Thornton (JTT) matrix-based model (Mega11). The numbers next to the branches show the percentage of trees in which the associated taxa are clustered together (bootstrapped n = 500). The length of each branch reflects the evolutionary distance between taxa. Accession numbers of the species presented can be found in Supplementary Table S1. An extended phylogenetic tree of fish *cd9* homologs can be found in supplementary material S7.

placed in two distinct areas: Chondrichthyes homologs clustered with lamprey cd9-like proteins (which could not be assigned confidently to any cd9 clade, apart from one cd9c) on a subtree opposite of the cd9c branch, whilst non-teleosts sterlet, gray bichir and spotted gar clustered together as an outgroup to the cd9a/cd9b subtree. Apart from the sea lamprey cd9c, all other non-teleost cd9c proteins clustered together within the cd9c subtree but separately from the teleost cd9c proteins.

In teleosts, we found two copies each of *cd81* and *tspan2* plus 3 copies of *cd9*. This was also found in the salmonid sister taxa, northern pike, which diverged from the salmonid linage prior to the Ss4r whole genome duplication. In northern pike, guppy, European eel, common carp and gray bichir we identify two or three *cd9c* copies, closely linked on the same chromosome, which may be result of a tandem duplication in these species and are named *cd9cT*.

The *cd9* subtree shows the greatest complexity of the cluster, *cd9a* and *cd9b* genes form a distinct subtree, separated from the *cd9c* genes. In tetrapods and West African lungfish only one *cd9* gene is present, which is placed on a separate branch opposite to the *cd9a/b* branch. Gar has two cd9 paralogues: *cd9a* and *cd9c*, which sit on a separate branch outside the teleost *cd9a* and *cd9c* subtrees, respectively. Zebrafish, northern pike and other teleosts are found to have a member of *cd9a*, *b* and *c* with northern pike, guppy, European eel and common carp containing the additional *cd9c* tandem duplications as described earlier. Interestingly, *cd9c* subtree is distant from *cd9a* and *cd9b* with the largest distances from root to tip and may indicate a fish-specific event independent of the whole genome duplication and then underwent duplication in salmonids.

The *cd81* subtree is very consistent with the history of genome duplication events and vertebrate evolution, with one copy in tetrapods and non-teleost fish, two copies in teleosts and 4 copies in salmonids.

The *tspan2* subtree shows a similar story in copy numbers, with the exception of salmonids showing duplication in *tspan2a,* but only one gene in *tspan2b*, which could be due to loss of the duplication.

Accession numbers for all species and genes used for the phylogenetic trees (Fig. 4 and S7) can be found in Supplementary Table S1.

3.6. Interferon stimulated motifs in the cd9 gene promoters

To further understand and infer control of gene expression, we examine the proximal *cd9* promoter for transcription factor binding sites, with an emphasis on those driving interferon responses defined by ISRE sites (Fig. 5A and B).

We identified two potential ISRE sites in *cd9c1* and one in *cd9c2* promoters of both rainbow trout and Atlantic salmon, which aligned well with both consensus sequences and were above the 0.8 score confidence threshold. These ISRE sites were between 105 and 140 base pairs (bp) (*cd9c1*) and 130 bp (*cd9c2*) upstream of the transcription start site (TSS) for rainbow trout and 14 and 52 bp downstream (*cd9c1*) and 68 bp upstream (*cd9c2*) of the TSS site in Atlantic salmon. CD9c1 and *cd9c2* have an intron in the 5′UTR between the TSS and ATG start site, which vary in size from 1190 bp (*cd9c1*) to 253 bp (cd9c2) for rainbow trout and 1273 bp (*cd9c1*) to 231 bp (*cd9c2*) in Atlantic salmon. In rainbow trout, *cd9a1* also has an intron in the 5′ UTR. TATA boxes (consensus sequence 5′ TATA[A/T]A[A/T]) were only found in *cd9a1* and *cd9a2*. Promoter analysis of *cd9b1* and *cd9b2* was not feasible for neither rainbow trout nor Atlantic salmon, due to long sequence repeats in this area.

3.7. cd9 gene expression in salmonid cells at steady state and after viral infection

cd9 genes in rainbow trout were examined in both spleen and intestine following IHNV infection using an RNAseq data set generated by Huang et al. (2010) (Fig. 6A–F). In both tissues *cd9c1* and *cd9c2* were significantly increased in expression following the viral infection. Of these two paralogues, *cd9c1* had the higher basal expression level in intestine and spleen. The *cd9a* and *cd9b* paralogues did not show a significant increase in expression following the viral infection. In spleen there was a significant decrease in expression of *cd9a2*. No expression of *cd9b1* or *b2* was observed in spleen indicating tissue-specific basal expression of *cd9* paralogues.

Further examination of expression of *cd9* paralogues involved

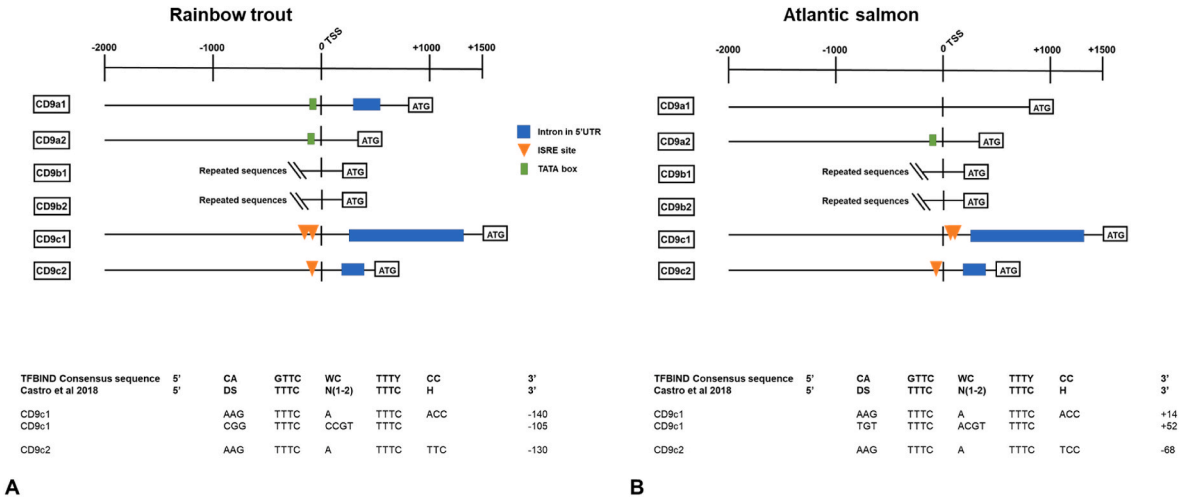

Fig. 5. Promoter and 5′ untranslated region (UTR) structure and interferon-stimulated response elements (ISRE) in CD9 paralogues of A) rainbow trout (*Oncorhynchus mykiss*) and B) Atlantic salmon (*Salmo salar*) 2000 bp upstream from the transcription start site (TSS). IUPAC codes: W = A or T, Y=C or T, D = A, G or T, S=C or G, N = any nucleotide, H = A, C or T.

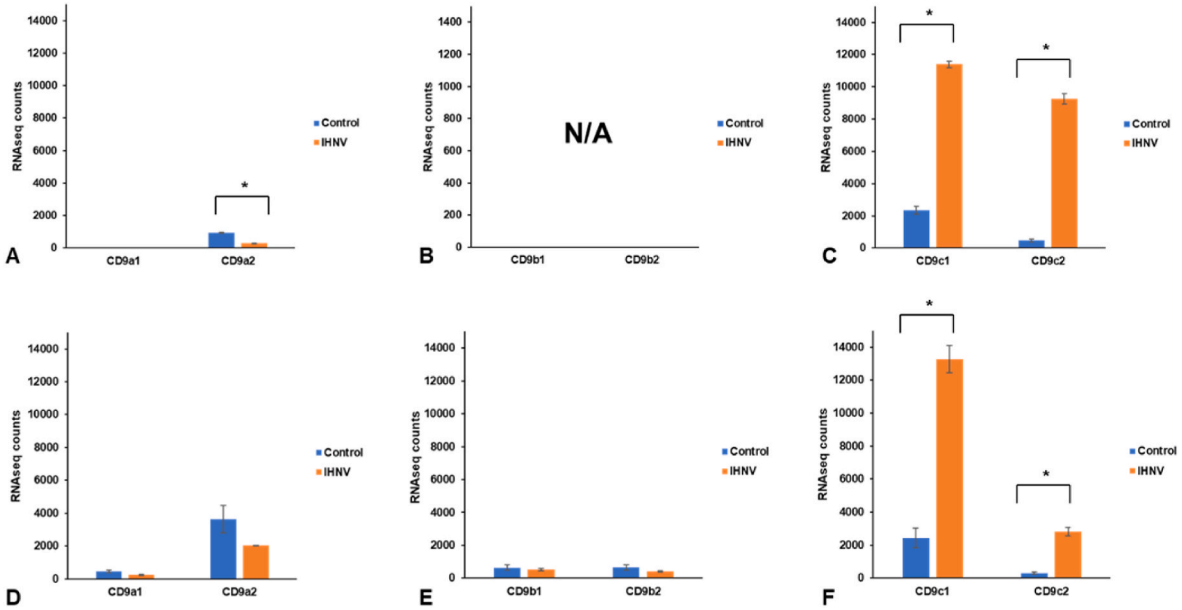

Fig. 6. CD9 paralogue expression (based on RNAseq count data) in the spleen (A = CD9a, B=CD9b, C=CD9c) and intestine (D = CD9a, E = CD9b, F=CD9c) of control (blue) and IHNV infected (orange) rainbow trout (*Oncorhynchus mykiss*) based on GEO RNAseq datasets GSE174274 and GSE176396, respectively (Huang et al., 2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

interrogating RNAseq data from a cell line with STAT2-knockout (CHSE-GS2) that are non-responsive to type I interferon (IFNA2) stimulation (Dehler et al., 2019) in comparison to the parent cell line (CHSE-EC) (Fig. 7A–F). *cd9c1* and *cd9c2* were the only paralogues to show significant increase in expression following IFNA2 stimulation in the parental cell type CHSE-EC, whereas in the CHSE-GS2 no changes were observed, suggesting these paralogues are interferon responsive genes. The *cd9c1* paralogue was expressed at a higher basal level than *cd9c2* in both cell lines. For *cd9a1*, expression was extremely low and only detected in the CHSE-EC unstimulated cells, whereas *cd9a2* was the highest expressed of the *cd9* genes. *cd9a2* gene expression was significantly decreased in expression in the CHSE-EC cells following IFNA2 stimulation. *cd9b1* expression was not assessed as this gene was could not be found in the current genome annotation and may not be present in chinook salmon. *cd9b2*, however, was expressed in both CHSE-EC and -GS2, with a small but significant decrease in expression following IFNA2 stimulation in both cell types. The response of the top four significantly upregulated genes in relation to *c9c1* and *cd9c2* expression in IFNA2 stimulated CHSE-EC cells is shown in Supplementary Fig. S8.

4. Discussion

CD9 proteins belonging to the tetraspanins are recognised as playing a major role in many cellular processes including immune function. Here we present the first in depth analysis of *cd9* paralogues in pike and salmonids where we identified four and six *cd9* copies, respectively, that can be defined into 3 clades: *cd9a*, *cd9b* and *cd9c, as in zebrafish*. Gene and protein structures are consistent with results from mammalian and teleost model species. Whilst gene synteny was well conserved between pairs within a clade only a few markers were conserved between neighbourhood of different *cd9* and between zebrafish and salmonids. There was also conservation in gene synteny to *cd9* paralogues in

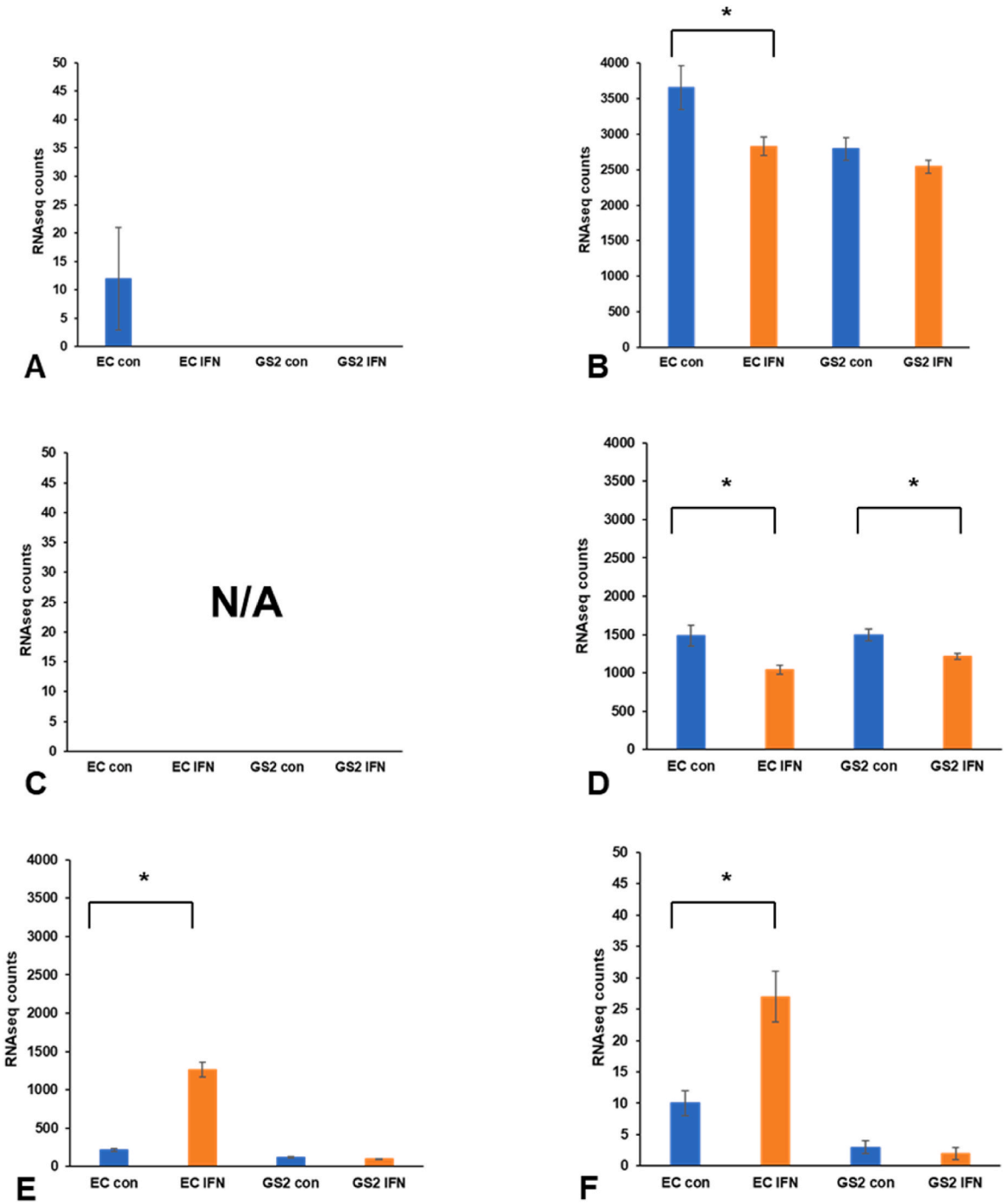

Fig. 7. RNAseq expression data (count data based on transcripts per million) of STAT2 knock-out experiment comparing "wild-type" CHSE-EC to STAT2-KO CHSE-GS2 cells in unstimulated and IFN-type I (IFNA2) stimulated condition. A = CD9a1, B=CD9a2, C=CD9b1, D = CD9b2, E = CD9c1, F=CD9c2. Blue bars reflect control groups, orange bars reflect IFNA2-stimulated groups. Asterixis identify significant differences between experimental groups. N/A marks CD9 paralo

northern pike, a non-salmonid sister group to rainbow trout and markers in common between zebrafish and rainbow trout *cd9*. Our phylogenetic tree tracks the evolution of *cd9* gene expansions through whole-genome duplication events. Promoter analysis identified interferon related elements, especially in *cd9c1* and *cd9c2*, which may point to a role for these genes in antiviral interferon activity. Gene expression analysis confirmed the *cd9c1* and *c2*, but not the other paralogs, as responsive to both an IHNV infection and to type I IFN in a model salmonid cell line.

4.1. Rainbow trout cd9 has 6 paralogues, which have a conserved gene and protein structure

Through bioinformatic analysis of the rainbow trout genome, we suggest six *cd9* paralogues, that can be split in three paralogue groups containing a pair of genes each. Many duplicates that appeared due to WGD events are eventually lost, but those that are retained may have developed novel functions and expression patterns (Parey et al., 2020). Immediately following a duplication event, duplicated genes are thought to be identical and therefore redundant (Glasauer and

Neuhauss, 2014). However, these redundant genes provide raw material for innovation with one of the duplicates able to develop a new function, termed neofunctionalization (Glasauer and Neuhauss, 2014). Alternatively, the ancestral gene functions may be divided between the duplicates, termed subfunctionalisation (Postlethwait et al., 2004). In vertebrates, silencing of one copy, termed non-functionalisation, or eventual loss of one copy are thought to be the most common outcomes of duplicate gene evolution (Glasauer and Neuhauss, 2014). In zebrafish and salmonids, however, duplicate retention rates are 20% (from the 3R) and 50% (from the 4R) respectively, suggesting ongoing re-diploidisation events in these fish with the retention rates correlating to the time-spans elapsed since the respective duplication events (Postlethwait et al., 2004; Glasauer and Neuhauss, 2014). Here we will discuss the evidence for salmonid CD9 duplications and the functional fate of the individual paralogues.

Intron and exon structure was highly conserved between rainbow trout *cd9* paralogues and other vertebrates (data not shown). *cd9a* and *cd9b* clades show more closely conserved exon lengths compared to *cd9c*, suggesting these genes are more diverged with potentially different rate of evolutionary change.

We also investigated the intron phases of rainbow trout *cd9* paralogues in comparison to the homologs in other species used for phylogenetic analysis as introns are thought of as ancient elements and their positions are usually conserved through evolution (Roy and Gilbert, 2005). Changes in intron phases are thought to be rare events, suggesting that most introns once inserted, remain at their position and retain their phase for long evolutionary times (Rogozin et al., 2000; Ruvinsky and Watson, 2007). Here we found that whilst intron positions were conserved among all species investigated, there were differences in intron phase patterns.

Protein sequence analysis showed that all six of the rainbow trout *cd9* paralogues conform to the typical tetraspanin structure of four TMs, EC1, IC and EC2 as identified in the crystal structure of human *cd9* (Umeda et al., 2019). The TMs have been shown to be highly conserved between different tetraspanin proteins, possibly due to their role in stabilising tetraspanins during biosynthesis and assembly and maintenance of the tetraspanin web (Levy and Shoham, 2005b; Umeda et al., 2020). This is consistent with our findings of a consistent TM length of 23 amino acids, except TM2 in CD9c1 and CD9c2 which only had a length of 20 amino acids. The EC2 is of particular interest as this domain is most important for facilitating protein-protein interactions and differences in this domain between tetraspanin paralogues, and most likely between *cd9* paralogues, will dictate different repertoires of binding partners and functional heterogeneity (Levy and Shoham, 2005b). Although we found that the length of EC2 is fairly conserved between *cd9* paralogues (between 83 and 84 amino acids), the protein sequences of this domain appear to vary significantly, which may influence the binding partner repertoire. It has been suggested that three major evolution events could account for the variation of EC2 between tetraspanin proteins: 1) co-evolution of major partners, 2) minor modification of partner repertoire and 3) duplication and subsequent alteration of partner repertoire (Huang et al., 2005). The cytoplasmic regions (terminal ends and IC) link the CD9 protein to the cytoskeleton and signalling molecules (Levy and Shoham, 2005b). We found the IC variable between the CD9 paralogues (6–12 amino acids). Furthermore, CD9a1 has a longer N-terminal and CD9c1 and CD9c2 have a longer C-terminal than other CD9 paralogues. This suggests differences in signalling with intracellular molecules and needs further investigation.

Apart from the conserved overall protein structure, one of the main features of tetraspanins is a distinct CCG amino-acid motif in the EC2, which is central for the formation of disulphide bridges with other conserved cysteine residues in this domain (Levy and Shoham, 2005b). We identified this motif in all of the rainbow trout *cd9* paralogues, further supporting these genes as true tetraspanins. Tetraspanins also contain 4, 6 or 8 conserved cysteine residues in the EC, as well as other conserved polar residues in the IC and TMs (Kovalenko et al., 2005). These conserved juxtamembrane cysteine residues are palmitoylated and play a central role in the formation of tetraspanin-tetraspanin interactions in the TEMs and the state of palmitoylation influences subcellular distribution and association of tetraspanins with their partner proteins (Charrin et al., 2002; Levy and Shoham, 2005a). 10 cysteine residues were identified in human *cd9* (Umeda et al., 2019). This was also found in most rainbow trout *cd9* paralogues, apart from *cd9a1* and *cd9c2*, which have 11 cysteine residues.

N-glycosylation sites are found in the EC2 in the majority of tetraspanins and sometimes are also present in the EC1 (Levy and Shoham, 2005a). However, little is known how post-translational modifications of N-glycosylation impact downstream functions of these genes (Termini and Gillette, 2017). In mammals, CD9 has one N-glycosylation site in EC1 (Boucheix et al., 1991), here we found that only CD9a2 had a functional N-glycosylation site predicted as functional in the EC2 and non-functional in CD9a1, finally CD9c1 had a functional N-glycosylation site predicted in the IC. A previous study also did not find a N-linked glycosylation sites in the EC1 of CD9 (our CD9c2) but suggested that this may not be needed for a fully functioning CD9 as it is also absent in cats (Fujiki et al., 2002) and are also absent in the red stingray CD9 (Zhu et al., 2006). To date no functional research into these N-glycosylation sites in CD9 has been performed.

4.2. Gene synteny analysis supports three clades for rainbow trout CD9 paralogues

Gene synteny analysis is proposed as good tool to resolve orthology and paralogy relationships of genes, additionally to phylogenetic analysis (Parey et al., 2020). Gene synteny evolves through independent mechanisms and is thought to be highly resilient to saturation at deep evolutionary times (Rokas and Holland, 2000; Parey et al., 2020). Here we found that gene synteny is poorly conserved between rainbow trout regions containing *cd9* paralogues and other vertebrates, even other teleosts as zebrafish. Only three genes were conserved between mouse and zebrafish and two genes between mouse, pike and rainbow trout. A block of seven genes was conserved between neighbourhood of zebrafish *cd9a* and pike *cd9a* and partly with rainbow trout *cd9a1* and *cd9a2* regions. This block, however, was present downstream of *cd9a* in zebrafish and upstream of *cd9a* in an inverted order in pike and rainbow trout. The best gene synteny conservation was found when comparing rainbow trout *cd9* paralogues to the paralogues of northern pike, a sister taxon to salmonids, which diverged 100–130 Mya (Rondeau et al., 2014). This suggests substantial reshuffling of gene order in the vicinity of *cd9* genes, which may be partly explained by WGD events. However, within the *cd9a, cd9b* and *cd9c* clades, gene pair synteny was highly conserved, suggesting origin of the duplicates from Ss4R WGD as opposed to small-scale duplication events.

4.3. cd9 paralogue expansion in rainbow trout is consistent with whole genome duplication events

Tetraspanin families, including the *cd9/cd81/tspan2* lineage, were produced by en bloc duplications as members of a family are separated in paraloguous genomic regions (Huang et al., 2005). The expansion of tetraspanin repertoire concurs with the whole-genome duplication events in vertebrates and specifically in fish (Huang et al., 2005; Abi-Rached et al., 2002). This is consistent with our findings of *cd9* paralogue expansion in salmonids, compared to the more basal spotted gar, which was not subjected to either Ts3R or Ss4R. The CD9c clade appears to be a special case as it appears to be present already in non-teleost fish, such as great white shark, gray bichir and spotted gar. Based on the position in the phylogenetic tree and synteny data of cd9c in these species, we suggest that this is an ancient gene that has undergone substantial modifications during fish evolution. This is consistent with a previous study, suggesting that the *cd9*-like branch in zebrafish (i.e. *cd9c*) has originated by independent small-scale

duplication (SSD) rather than teleost specific WGD (Huang et al., 2010). However, within the *cd9a*, *cd9b* and *cd9c* clades between salmonids and pike, gene pair synteny was highly conserved, suggesting origin of the duplicates from Ss4R WGD as opposed to small-scale duplication events. We also found that the duplications within the clades, i.e., *cd9a1* (chromosome 1) and *cd9a2* (chromosome 2), as well as *cd9c1* (chromosome 13) and *cd9c2* (chromosome 12) in rainbow trout are consistent with the evolutionary history of genome duplication and chromosomal rearrangement (Berthelot et al., 2014). Unfortunately, we cannot make any comments about the consistency for the *cd9b1* and *cd9b2* duplication, as both paralogues were not assigned to a chromosome.

We used a group of four ancestral tetraspanins (*citspan1, citspan2, citspan7 and citspan15*) as an outgroup to our *cd9/tspan2/cd81*, based on previous tetraspanin evolution analysis, which propose this group of genes as closest relation to the vertebrate cluster (Huang et al., 2010; Garcia-España et al., 2008; Garcia-España and DeSalle, 2009). The origin of the *cd9/tspan2/cd81* lineage is proposed to be in the chordate ancestor (525 Mya) (Garcia-España et al., 2008; Garcia-España and DeSalle, 2009). In invertebrates, the best homolog of *cd9* has been proposed to be from a family of *tspan8*-like genes of vase tunicate and amphioxus, suggesting the ancestor of the *cd9/tspan2/cd81* may have arisen in an ancient *tspan8*-like tetraspanin before the vertebrata radiation and that evolution of this lineage is associated with invention of new cell types and systems for brain and adaptive immune system (Huang et al., 2005).

Whilst in elephant shark, only one *cd9* paralogue was detected, in other "ancient" non-teleost fish species we found two *cd9* paralogues that clustered with either *cd9a* and *cd9c* homologs of teleost fish, albeit with some evolutionary distance as suggested by branch lengths in the phylogenetic tree and synteny conservation. We suggest therefore that the CD9b clade arose during the teleost-specific whole genome duplication from the duplication of the ancient *cd9a* gene, as no evidence of this clade could be found in non-teleost fish and due the closeness of gene synteny and phylogenetic placement of *cd9a* and *cd9b* paralogues. We found no evidence of duplication of the ancient *cd9c* in the Ts3R WGD, suggesting that this gene was either not successfully duplicated or the duplication was lost during rediploidisation. With the publication of whole annotated genomes of more non-teleost fish, followed by functional experiments, the evolutionary relationship of the *cd9* lineage will likely be more accurately resolved.

4.4. Transcription factor binding sites in cd9 paralogues support a role in antiviral immune system

Due to the suspected involvement of *cd9* paralogues in immune functions in fish, we investigated the promoter area of each paralogue for immune system relevant transcription factor binding sites. We found that predicted ISRE sites are consistent between rainbow trout and Atlantic salmon in terms of identified motifs and position. The differences in absolute position relative to the TSS between those two species may be due to the methods on how TSS are identified on NCBI, which use the longest identified transcript isoform to predict the TSS.

The vertebrate innate antiviral response is driven by type I interferon and is crucial for efficient defence against viral pathogens (Hambleton et al., 2013). In response to invading viruses, type I IFN signals through the Interferon stimulated gene factor 3 (ISGF3) complex which is formed by phosphorylation of transcription factor STAT2 by JAK kinases and oligomerisation with STAT1 and IRF-9 (Horvath et al., 1996; Blaszczyk et al., 2016). The ISGF3 complex in association with other nuclear transcription factors then binds to ISRE present in the promoters of interferon stimulated genes (ISGs), which mediate a rapid antiviral response (Horvath et al., 1996; Sadler and Williams, 2008; Blaszczyk et al., 2016). Here we find predicted ISRE sites in *cd9c1* and *cd9c2* within 150 bp upstream of the TSS. *cd9b1* and *cd9b2* promoter areas were defined by large repeats, therefore it was not possible to look for TF binding motifs with confidence. Interestingly *cd9b* paralogues in a previous publication were shown to have an important role in host/virus interactions (Castro et al., 2015), suggesting that better resolved promoter sequences for *cd9b* genes could confirm if their role in virus uptake is related to the interferon pathway. Alternatively, steady state level of *cd9b* might be important for antiviral defence even in absence of upregulation by type I IFN. Future comparative functional studies of the promoters of the six *cd9* paralogues will allow to demonstrate their IFN-dependent inducibility.

4.5. Gene expression suggests important roles of cd9c1 and cd9c2 in antiviral responses

Our analysis of RNAseq data of spleen and intestine of rainbow trout as well as CHSE cells suggests a tissue-specific expression of the different *cd9* paralogues. This was also found in Japanese flounder, in which three different *cd9* paralogues were identified (He et al., 2021). *cd9.1* and *cd9.2* showed high sequence identity, whilst *cd9.*3 had a low sequence identity (He et al., 2021). Both *cd9.1* and *cd9.3* expression increased after challenge with a viral pathogen, where *cd9.3* was more increased in mid and late stages of infection suggesting an important role of this gene in antiviral response (He et al., 2021). Interestingly, *cd9.3* shares closest sequence identity with rainbow trout *cd9b1/b2* (data not shown), supporting the findings of Castro et al. (2015). Furthermore, *cd9.1* and *cd9.3* showed protective roles in bacterial infections, with knockdown of *cd9.1* and *cd9.3* in FG (flounder gill) cells significantly weakened the ability of the cells to clear virus, whilst over-expression of these genes led to fewer bacteria in FG cells and in vivo fish (He et al., 2021).

Further evidence of the importance of *cd9* in antiviral immune responses in fish is provided by several previous studies. *cd9c* in zebrafish was found to be highly induced in response to virus infections in parallel to classic ISGs, similar to our findings in CHSE-EC cells (Briolat et al., 2014; Dehler et al., 2019). An immune system relevant role has also been suggested for *cd9* in red stingray (Zhu et al., 2006), Arctic lamprey (Wu et al., 2012) and sea lamprey (Uinuk-ool et al., 2002). *cd9* (closest sequence match to the *cd9b* clade presented here) in rainbow trout has shown to play an important role in immunity and was found to increase during rainbow trout development and increased significantly at first feeding, correlating with increased B lymphocyte activity (Castro et al., 2015). Furthermore, *cd9* was constitutively expressed in naïve B cells and decreased after stimulation with Viral Hemorrhagic Septicemia Virus (VHSV) or CpG, but not after stimulation with polyI:C or LPS (Castro et al., 2015), which is similar to the decrease we see in the intestine of IHNV-infected rainbow trout and in interferon-stimulated CHSE-EC cells. *cd9* was also increased in peritoneal cells and muscle of i.p. injection with VHSV or i.m. injection of VHSV vaccine, respectively but decreased in the gills of fish exposed to VHSV via bath infection (Castro et al., 2015). An unknown CD9 paralogue (but closest to our *cd9b* clade) was also identified in a study of the IgM^+ B cell surface proteome of Atlantic salmon (Peñaranda et al., 2019). In CHSE-EC cells, we could only detect expression of the *cd9b2* paralogue and show a small but significant downregulation after type I IFN stimulation. In addition, in Atlantic salmon heart *cd9* (paralogue type unknown) in innate-like B- and myeloid-cells were strongly induced in the infected with PMCV (piscine myocarditis virus), which coincided with a decreased antiviral response (Timmerhaus et al., 2011). Once virus levels plateaued, *cd9* levels decreased (Timmerhaus et al., 2011). These results are consistent with the findings in mammalian studies that *cd9* is expressed in cells of the immune system and play a role in virus pathogenesis (Levy and Shoham, 2005b; Yáñez-Mó et al., 2009). In mice, CD9 associate with FcγRs, an important receptor family for phagocytosis of macrophages, which can activate macrophages (Kaji et al., 2001). Murine peritoneal macrophages stimulated with IFNγ show a decrease of *cd9* expression (Wang et al., 2002). However, in STAT1 KO macrophages no reduction of *cd9* expression could be observed, suggesting that STAT1 pathway is required to reduce *cd9* in IFNγ activated macrophages (Wang et al., 2002). These findings are consistent with our observations of

STAT-related binding sites in the promoter of *cd9* paralogues. Additionally, we show that *cd9c1* and *cd9c2* are highly responsive to interferon stimulation in CHSE-EC cells, but not STAT2-KO cells CHSE-GS2, suggesting an association of these paralogues with the interferon pathway.

cd9 was found to be particularly involved in infections with enveloped viruses and their exit from infected cells, such as HIV, coronaviruses, influenza A viruses (IAVs), feline immunodeficiency virus and canine distemper virus (Fanaei et al., 2011; Gordon-Alonso et al., 2006; Earnest et al., 2015; Monk and Partridge 2012). Future studies using validation by genome editing in salmonid cells are needed to unravel the function of the different *cd9* paralogues with respect to viral specificity.

4.6. Expansion of cd9 paralogue repertoire in salmonids may have led to subfunctionalisation

In conclusion, we show by gene synteny and phylogenetic analysis that *cd9a* and *cd9b* clades are consistent duplication from the ancestor of fish before the Ts3R. The ancestor of the *cd9c* clade, however, may have appeared first after Ts3R (Huang et al., 2010). Promoter analysis and RNAseq gene expression analysis suggests a general antiviral role of *cd9c1* and *cd9c2*, but will need further validation through overexpression and knock-out experiments. Previous studies suggest *cd9b* clade to be relevant in B cell immunity (Castro et al., 2015; Peñaranda et al., 2019), which may suggest subfunctionalisation of *cd9* paralogue roles within the immune system. *cd9b* clade may also have a role in egg fertilisation as seen in zebrafish (Greaves et al., 2021). Our expression analysis does not provide any evidence that *cd9a* is involved in immunity. However, we found *cd9a2* as the highest expressed paralogue in unstimulated CHSE-EC cells, which could suggest a role in cell membrane health and maintenance, but further studies are needed to confirm this.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was funded by BBSRC project BB/R008973/1 and European Union's Horizon 2020 research and innovation program under grant agreement 817923 (AQUA-FAANG).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2023.104735.

References

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., Inoko, H., 2002. Evidence of en bloc duplication in vertebrate genomes. Nat. Genet. 31, 100–105. https://doi.org/10.1038/ng855.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Crollius, H.R., Guiguen, Y., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5, 3657. https://doi.org/10.1038/ncomms4657.

Blaszczyk, K., Nowicka, H., Kostyrko, K., Antonczyk, A., Wesoly, J., Bluyssen, H.A.R., 2016. The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. Cytokine Growth Factor Rev. 29, 71–81. https://doi.org/10.1016/j.cytogfr.2016.02.010.

Boucheix, C., Benoit, P., Frachet, P., Billard, M., Worthington, R.E., Gagnon, J., Uzan, G., 1991. Molecular cloning of the CD9 antigen. A new family of cell surface proteins. J. Biol. Chem. 266, 117–122. https://doi.org/10.1016/S0021-9258(18)52410-8.

Briolat, V., Jouneau, L., Carvalho, R., Palha, N., Langevin, C., Herbomel, P., Schwartz, O., Spaink, H.P., Levraud, J.-P., Boudinot, P., 2014. Contrasted innate responses to two viruses in zebrafish: insights into the ancestral repertoire of vertebrate IFN-stimulated genes. J. Immunol. 192, 4328–4341. https://doi.org/10.4049/jimmunol.1302611.

Brosseau, C., Colas, L., Magnan, A., Brouard, S., 2018. CD9 tetraspanin: a new pathway for the regulation of inflammation. Front. Immunol. 9, e2316 https://doi.org/10.3389/fimmu.2018.02316.

Cao, J., Tan, X., 2018. Comparative analysis of the tetraspanin gene family in six teleost fishes. Fish Shellfish Immunol. 82, 432–441. https://doi.org/10.1016/j.fsi.2018.08.048.

Castro, R., Martin, S.A.M., Bird, S., Lamas, J., Secombes, C.J., 2008. Characterisation of γ-interferon responsive promoters in fish. Mol. Immunol. 45, 3454–3462. https://doi.org/10.1016/j.molimm.2008.03.015.

Castro, R., Abós, B., González, L., Aquilino, C., Pignatelli, J., Tafalla, C., 2015. Molecular characterization of CD9 and CD63, two tetraspanin family members expressed in trout B lymphocytes. DCI 51, 116–125. https://doi.org/10.1016/j.dci.2015.03.002.

Charrin, S., Manie, S., Oualid, M., Billard, M., Boucheix, C., Rubinstein, E., 2002. Differential stability of tetraspanin/tetraspanin interactions: role of palmitoylation. FEBS Lett. 516, 139–144. https://doi.org/10.1016/s0014-5793(02)02522-x.

Dehler, C.E., Boudinot, P., Martin, S.A.M., Collet, B., 2016. Development of an efficient genome editing method by CRISPR/Cas9 in a fish cell line. Mar. Biotechnol. 18, 449–452. https://doi.org/10.1186/s12896-020-00626-x.

Dehler, C.E., Lester, K., Della Pelle, G., Jouneau, L., Houel, A., Collins, C., Dovgan, T., Machat, R., Zou, J., Boudinot, P., Martin, S.A.M., Collet, B., 2019. Viral resistance and IFN signalling in STAT2 knockout fish cells. J. Immunol. 203, 465–475. https://doi.org/10.4049/jimmunol.1801376.

Earnest, J.T., Hantak, M.P., Park, J.-E., Gallagher, T., 2015. Coronavirus and influenza virus proteolytic priming takes place in tetraspanin-enriched membrane microdomains. J. Virol. 89, 6093–6104. https://doi.org/10.1128/JVI.00543-15.

Earnest, J.T., Hantak, M.P., Li, K., McCray, P.B., Perlman, S., Gallagher, T., 2017. The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. PLoS Pathog. 13 https://doi.org/10.1371/journal.ppat.1006546.

Fanaei, M., Monk, P.N., Partridge, L.J., 2011. The role of tetraspanins in fusion. Biochem. Soc. Trans. 39, 524–528. https://doi.org/10.1042/BST0390524.

Florin, L., Lang, T., 2018. Tetraspanin assemblies in virus infection. Front. Immunol. 9, 1140. https://doi.org/10.3390/ijms222111609.

Fujiki, K., Gauley, J., Bols, N., Dixon, B., 2002. Cloning and characterisation of cDNA clones encoding CD9 from Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 54, 604–609. https://doi.org/10.1007/s00251-002-0506-0.

Garcia-España, A., Chung, P.-J., Sarkar, I.N., Stiner, E., Sun, T.-T., DeSalle, R., 2008. Appearance of new tetraspanin genes during vertebrate evolution. Genomics 91, 326–334. https://doi.org/10.1016/j.ygeno.2007.12.005.

Garcia-España, A., DeSalle, R., 2009. Intron sliding in tetraspanins. Commun. Integr. Biol. 2, 394–395. https://doi.org/10.4161/cib.2.5.8760.

Glasauer, S.M.K., Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genom. 289, 1045–1060. https://doi.org/10.1007/s00438-014-0889-2.

Gordon-Alonso, M., Yanez-Mo, M., Barreiro, O., Álvarez, S., Muñoz-Fernández, M.A., Valenzuela-Fernández, A., Sánchez-Madrid, F., 2006. Tetraspanins CD9 and CD81 modulate HIV-1 induced membrane fusion. J. Immunol. 15, 5129–5137. https://doi.org/10.4049/jimmunol.177.8.5129.

Greaves, S., Marsay, K.S., Monk, P.N., Roehl, H., Partridge, L.J., 2021. Tetraspanin Cd9b Plays a Role in Fertility in Zebrafish. bioRxiv. https://doi.org/10.1101/2021.11.25.469903.

Gundappa, M.K., To, T.-H., Grønvold, L., Martin, S.A.M., Lien, S., Geist, J., Hazlerigg, D., Sandve, S.R., Macqueen, D.J., 2021. Genome-wide reconstruction of rediploidisation following autopolyploidisation across one hundred million years of salmonid evolution. Mol. Biol. Evol. 39 https://doi.org/10.1093/molbev/msab310.

Ha, C.T., Waterhouse, R., Wessells, J., Wu, J.A., Dveksler, G.S., 2005. Binding of pregnancy-specific glycoprotein 17 to CD9 on macrophages induces secretion of IL-10, IL-6, PGE2 and TGF-beta1. J. Leukoc. Biol. 77, 948–957. https://doi.org/10.1189/jlb.0804453.

Hambleton, S., Goodbourn, S., Young, D.F., Dickinson, P., Mohamad, S.M.B., Valappil, M., McGovern, N., Cant, A.J., Hackett, S.J., Ghazal, P., Morgan, N.V., Randall, R.E., 2013. STAT2 deficiency and susceptibility to viral illness in humans. Proc. Natl. Acad. Sci. USA 110, 3053–3058. https://doi.org/10.1073/pnas.1220098110.

Hantak, M.P., Qing, E., Earnest, J.T., Gallagher, T., 2019. Tetraspanins: architects of viral entry and exit platforms. J. Virol. 93, e01429-17 https://doi.org/10.1128/JVI.01429-17.

He, J., Gu, H., Wang, W., Hu, Y., 2021. Two CD9 tetraspanin family members of Japanese flounder (*Paralichthys olivaceus*): characterisation and comparative analysis of the anti-infectious immune function. Vet. Res. 52, 28. https://doi.org/10.1186/s13567-021-00903-3.

Horejsi, V., Vlcek, C., 1991. Novel structurally distinct family of leucocyte surface glycoproteins including CD9, CD37, CD53 and CD63. FEBS Lett. 288, 1–4. https://doi.org/10.1016/0014-5793(91)80988-F.

Horvath, C.M., Stark, G.R., Kerr, I.M., Darnell Jr., J.E., 1996. Interactions between STAT and non-STAT proteins in the interferon-stimulated gene factor 3 transcription complex. Mol. Cell Biol. 16, 6957–6964. https://doi.org/10.1128/mcb.16.12.6957.

Huang, S., Yuan, S., Dong, M., Su, J., Yu, C., Shen, Y., Xie, X., Yu, Y., Yu, X., Chen, S., Zhang, S., Pontarotti, P., Xu, A., 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell-interactions from unicellular to multicellular organisms. Genomics 86, 674–684. https://doi.org/10.1016/j.ygeno.2005.08.004.

Huang, S., Tian, H., Chen, Z., Yu, T., Xu, A., 2010. The evolution of vertebrate tetraspanins: gene loss, retention, and massive positive selection after whole genome duplications. BMC Evol. Biol. 10, 306. https://doi.org/10.1186/1471-2148-10-306.

Hulsmans, M., Holvoet, P., 2013. MicroRNA-containing microvesicles regulating inflammation in association with atherosclerotic disease. Cardiovasc. Res. 100, 7–18. https://doi.org/10.1093/cvr/cvt161.

Hurley, I.A., Mueller, R.L., Dunn, K.A., Schmidt, E.J., Friedman, M., Ho, R.K., Prince, V. E., Yang, Z., Thomas, M.G., Coates, M.I., 2007. A new time-scale for ray-finned fish evolution. Proc. R. Soc. A B. https://doi.org/10.1098/rspb.2006.3749, 274489–498.

Kaji, K., Takeshita, S., Miyake, K., Takai, T., Kudo, A., 2001. Functional association of CD9 with the Fcγ receptors in macrophages. J. Immunol. 166, 3256–3265. https://doi.org/10.4049/jimmunol.166.5.3256.

Karlsson, G., Rörby, E., Pina, C., Soneji, S., Reckzeh, K., Miharada, K., Karlsson, C., Guo, Y., Fugazza, C., Gupta, R., Martens, J.H.A., Stunnenberg, H.G., Karlsson, S., Enver, T., 2013. The tetraspanin CD9 affords high-purity capture of all murine hematopoietic stem cells. Cell Rep. 4, 642–648. https://doi.org/10.1016/j.celrep.2013.07.020.

Kobayashi, H., Hosono, O., Iwata, S., Kawasaki, H., Kuwana, M., Tanaka, H., Dang, N.H., Morimoto, C., 2004. The tetraspanin CD9 is preferentially expressed on the human $CD4^{+}CD45RA^{+}$ naïve T cell population and is involved in T cell activation. Clin. Exp. Immunol. 137, 101–108. https://doi.org/10.1111/j.1365-2249.2004.02494.x.

Kouwaki, T., Okamoto, M., Tsukamoto, H., Fukushima, Y., Oshiumi, H., 2017. Extracellular vesicles deliver host and virus RNA and regulate innate immune response. Int. J. Mol. Sci. 18, 666. https://doi.org/10.3389/fimmu.2016.00335.

Kovalenko, O.V., Metcalf, D.G., DeGrado, W.F., Hemler, M.E., 2005. Structural organisation and interactions of transmembrane domains in tetraspanin proteins. BMC Struct. Biol. 5, 11. https://doi/10.1186/1472-6807-5-11.

Levraud, J.-P., Jouneau, L., Briolat, V., Laghi, V., Boudinot, P., 2019. IFN-stimulated genes in zebrafish and humans define an ancient arsenal of antiviral immunity. J. Immunol. 15, 3361–3373. https://doi.org/10.4049/jimmunol.1900804.

Levy, S., Shoham, T., 2005a. Protein-protein interactions in the tetraspanin web. Physiology 20, 218–224. https://doi.org/10.1152/physiol.00015.2005.

Levy, S., Shoham, T., 2005b. The tetraspanin web modulates immune-signalling complexes. Nature Reviews 5, 136–148. https://doi.org/10.1038/nri1548.

Monk, P.N., Partridge, L.J., 2012. Tetraspanins: gateways for infection. Infect. Disord.: Drug Targets 12, 4–17. https://doi.org/10.2174/187152612798994957.

Parey, E., Louis, A., Cabau, C., Guiguen, Y., Crollius, H.R., Berthelot, C., 2020. Synteny-guided resolution of gene trees clarifies the functional impact of whole-genome duplications. Mol. Biol. Evol. 37, 3324–3337. https://doi.org/10.1093/molbev/msaa149.

Peñaranda, M.M., Jensen, I., Tollersrud, L.G., Bruun, J.-A., Jørgensen, J.B., 2019. Profiling the Atlantic salmon IgM^{+} B cell surface proteome: novel information on teleost fish B cell protein repertoire and identification of potential B cell markers. Front. Immunol. 10, 37. https://doi.org/10.3389/fimmu.2019.00037.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., Yan, Y.-L., 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 20 https://doi.org/10.1016/j.tig.2004.08.001.

Reyes, R., Monjas, A., Yánez-Mó, M., Cardeñes, B., Morlino, G., Gilsanz, A., 2015. Different states of integrin LFA-1 aggregation are controlled through its association with tetraspanin CD9. Biochim. Biophys. Acta 1853, 2464–2480. https://doi.org/10.1016/j.bbamcr.2015.05.018.

Reyes, R., Cardeñes, B., Machado-Pineda, Y., Cabañas, C., 2018. Tetraspanin CD9: a key regulator of cell adhesion in the immune system. Front. Immunol. 9, 863. https://doi.org/10.3389/fimmu.2018.00863.

Rogozin, I.B., Lyons-Weiler, J., Koonin, E.V., 2000. Intron sliding in conserved gene families. Trends Genet. 16, 430–432. https://doi.org/10.1016/S0168-9525(00)02096-5.

Rokas, A., Holland, P., 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454–459. https://doi.org/10.1016/S0169-5347(00)01967-4.

Rondeau, E.B., Minkley, D.R., Leong, J.S., Messmer, A.M., Jantzen, J.R., von Schalburg, K.R., Lemon, C., Bird, N.H., Koop, B.F., 2014. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the neoteleostei. PLoS One 9 (7), e102089. https://doi.org/10.1371/journal.pone.0102089.

Roy, S.W., Gilbert, W., 2005. Rates of intron loss and gain: implications for early eukaryotic evolution. Proc. Natl. Acad. Sci. USA 102, 5773–5778. https://doi.org/10.1073/pnas.0500383102.

Ruvinsky, A., Watson, C., 2007. Intron phase patterns in genes: preservation and evolutionary changes. Open Evol. J. 1, 1–14. https://doi.org/10.2174/1874404400701010001.

Sadler, A.J., Williams, B.R., 2008. Interferon-inducible anti-viral effectors. Nat. Rev. Immunol. 8, 559–568. https://doi.org/10.1038/nri2314.

Suzuki, M., Tachibana, I., Takeda, Y., He, P., Minami, S., Iwasaki, T., Kida, H., Goya, S., Kijima, T., Yoshida, M., Kumagai, T., Osaki, T., Kawase, I., 2009. Tetraspanin CD9 negatively regulates lipopolysaccharide-induced macrophage activation and lung inflammation. J. Immunol. 182, 6485–6493. https://doi.org/10.4049/jimmunol.0802797.

Tai, X.G., Yashiro, Y., Abe, R., Toyooka, K., Wood, C.R., Morris, J., Long, A., Ono, S., Kobayashi, M., Hamaoka, T., Neben, S., Fujiwara, H., 1996. A role for CD9 molecules in T cell activation. J. Exp. Med. 184, 753–758. https://doi.org/10.1084/jem.184.2.753.

Termini, C.M., Gillette, J.M., 2017. Tetraspanins function as regulators of cellular signalling. Front. Cell Dev. Biol. 5, 34. https://doi.org/10.3389/fcell.2017.00034.

Timmerhaus, G., Krasnov, A., Nilsen, P., Alarcon, M., Afanasyev, S., Rode, M., 2011. Transcriptome profiling of immune responses to cardiomyopathy syndrome (CMS) in Atlantic salmon. BMC Genom. 12, 459. https://doi.org/10.1186/1471-2164-12-459.

Uinuk-ool, T., Mayer, W.E., Sato, A., Dongak, R., Cooper, M.D., Klein, J., 2002. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. Proc. Natl. Acad. Sci. USA 99, 22. https://doi.org/10.1073/pnas.21252769.

Umeda, R., Nishizawa, T., Osamu Nureki, O., 2019. Crystallisation of human tetraspanin protein CD9. Acta Crystallogr F Struct Biol Commun 75, 254–259. https://doi.org/10.1107/S2053230X1801840X.

Umeda, R., Satouh, Y., Takemoto, M., Nakada-Nakura, Y., Liu, K., Yokoyama, T., Shirouzu, M., Iwata, S., Nomura, N., Sato, K., Ikawa, M., Nishizawa, T., Nureki, O., 2020. Structural insights into tetraspanin CD9 function. Nat. Commun. 11, 1606. https://doi.org/10.1038/s41467-020-15459-7.

Unternaehrer, J.J., Chow, A., Pypaert, M., Inaba, K., Mellman, I., 2007. The tetraspanin CD9 mediates lateral association of MHC class II molecules on the dendritic cell surface. Proc. Natl. Acad. Sci. USA 104, 234–239. https://doi.org/10.1073/pnas.0609665104.

Wang, X.-Q., Evans, G.F., Alfaro, M.L., Zuckerman, S.H., 2002. Down-regulation of macrophage CD9 expression by IFNγ. Biochem. Biophys. Res. Commun. 290, 891–897. https://doi.org/10.1006/bbrc.2001.6293.

Won, W.J., Kearney, J.F., 2002. CD9 is a unique marker for marginal zone B cells, B1 cells and plasma cells in mice. J. Immunol. 168, 5605–5611. https://doi.org/10.4049/jimmunol.168.11.5605.

Wu, F., Su, P., Chen, L., Li, M., Liu, X., Li, Q., 2012. Cloning of Arctic lamprey (*Lethenteron camtschaticum*) cd9 with roles in the immune response. J. Fish. Biol. 81, 1147–1157. https://doi.org/10.1111/j.1095-8649.2012.03299.x.

Yáñez-Mó, M., Barreiro, O., Gordon-Alonso, M., Sala-Valdés, M., Sánchez-Madrid, F., 2009. Tetraspanin-enriched microdomains: a functional unit in cell plasma membranes. Trends Cell Biol. 19, 434–446. https://doi.org/10.1016/j.tcb.2009.06.004.

Zhu, J., Yan, K., Lu, L., Peng, C., Zhou, C., Chen, S., Xie, X., Dong, M., Xu, A., 2006. Molecular cloning and characterisation of CD9 cDNA from cartilaginous fish, red stingray, *Dasyatis akajei*. Mol. Immunol. 43, 1534–1540. https://doi.org/10.1016/j.molimm.2005.10.005.

Zilber, M.-T., Setterblad, N., Vasselon, T., Doliger, C., Charron, D., Mooney, N., Gelin, C., 2005. MHC class II/CD38/CD9: a lipid-raft-dependent signalling complex in human monocytes. Blood 106, 3074–3081. https://doi.org/10.1182/blood-2004-10-4094.