Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes

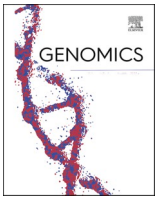
Thomas C Clark, Shahmir Naseer, Manu Kumar Gundappa, Audrey Laurent, Aline Perquis, Bertrand Collet, Daniel J Macqueen, Samuel A.M. Martin, Pierre Boudinot

Genomics 115 (2023) 110663

Contents lists available at ScienceDirect

Genomics

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



Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes

Thomas C. Clark [a,b,1], Shahmir Naseer [b,1], Manu Kumar Gundappa [c], Audrey Laurent [d], Aline Perquis [d], Bertrand Collet [a], Daniel J. Macqueen [c], Samuel A.M. Martin [b,*], Pierre Boudinot [a,*]

[a] Université Paris-Saclay, INRAE, UVSQ, VIM, Jouy-en-Josas 78350, France
[b] Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK
[c] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, UK
[d] INRAE, LPGP, Rennes, France

A B S T R A C T

Antiviral innate immunity is orchestrated by the interferon system, which appeared in ancestors of jawed vertebrates. Interferon upregulation induces hundreds of interferon-stimulated-genes (ISGs) with effector or regulatory functions. Here we investigated the evolutionary diversification of ISG responses through comparison of two salmonid fishes, accounting for the impact of sequential whole genome duplications ancestral to teleosts and salmonids. We analysed the transcriptomic response of the IFN pathway in the head kidney of rainbow trout and Atlantic salmon, which separated 25–30 Mya. We identified a large set of ISGs conserved in both species and cross-referenced them with zebrafish and human ISGs. In contrast, around one-third of salmonid ISG lacked orthologs in human, mouse, chicken or frog, and often between rainbow trout and Atlantic salmon, revealing a fast-evolving, lineage-specific arm of the antiviral response. This study also provides a key resource for in-depth functional analysis of ISGs in salmonids of commercial significance.

1. Introduction

Jawed vertebrates have a powerful innate antiviral response based mainly on specialized cytokines called interferons (IFN), and their subsequent activation of IFN stimulated genes (ISGs) that drive antiviral activity in cells. There are 3 types of IFN in mammals: type I, II, and III, with type I and III directly involved in antiviral defense. Almost all nucleated cells can produce type I IFN upon viral detection through stimulation of germline encoded pattern recognition receptors (PRRs). These receptors can be located on the cell surface or intracellularly and recognize motifs characteristic of viral nucleic acids or proteins [35]. Orthologs of type I and II, but not type III IFN, are present in fish [31]. Fish type I IFN genes differ in structure from mammalian orthologs with the latter lacking introns, but encoding proteins with an overall similar structure [21]. Receptors of fish type I IFN have conserved structure with mammalian type I IFN receptors IFNAR1 and IFNAR2 ([2,28], reviewed in [43]). Activation of these receptors induces a downstream signalling cascade through the conserved JAK/STAT pathway leading to the formation of the interferon-stimulated gene factor 3 (ISGF-3) complex by association of interferon regulatory factor (IRF)9 with STAT1 and STAT2 [6]. ISGF-3 then translocates to the nucleus, where it binds to motifs in the promoters of hundreds of ISGs with multitudes of effector and regulatory functions [41]. These functions include direct antiviral defense (MX, PKR), signalling resulting in positive or negative feedback loops (STATs, IRFs), immune cell activation and attraction (CXCL9–11), and many other basic cellular functions [40,42]. The function of many ISGs remains unknown even in mammals; in fish, and particularly salmonids, the functional complexity of the ISG repertoire is further increased due to historic whole genome duplication events (WGDs) and potential subsequent sub-functionalisation or functional specialization of the retained gene duplicates (paralogs).

The innate antiviral response is of paramount importance for fish health in aquaculture with the further expansion of aquaculture globally constrained by a number of viral diseases. In salmonids specifically, these include infectious salmon anaemia virus (ISAV) [1], piscine myocarditis virus (PMCV) [22], infectious pancreatic necrosis virus (IPNV) [13], salmonid alphavirus (SAV) [23], viral hemorrhagic septicemia virus (VHSV) [3] or infectious haematopoietic necrosis virus (IHNV) [14] and can often cause high mortality in fry or in adult fish. While various vaccines exist and are in use, many viral diseases cannot

* Corresponding authors.
E-mail addresses: sam.martin@abdn.ac.uk (S.A.M. Martin), pierre.boudinot@inrae.fr (P. Boudinot).
[1] Equal contribution.

https://doi.org/10.1016/j.ygeno.2023.110663
Received 13 April 2023; Received in revised form 24 May 2023; Accepted 31 May 2023
Available online 5 June 2023

be completely prevented by vaccination. Furthermore, as antiviral treatments are either not available or too expensive, genomic based selection for improved resistance to viral pathogens is a highly promising approach, underpinned by functional knowledge of fish genomes and immune function. A comprehensive characterization of the response of zebrafish larvae to viral infection and type I IFN stimulation recently led to the identification of a core list of ISGs conserved between zebrafish and human [10,27]. About 40% of the genes induced by the type I IFN response in zebrafish (137/337 genes) lacked any orthologs in human, potentially corresponding to ISGs restricted to ray-finned fishes.

Here, we investigated the diversity of ISGs in two salmonid species of global significance for aquaculture: Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). These species are paleotetraploid owing to a WGD event that occurred in the salmonid ancestor ~100 million years ago (Mya), termed the salmonid-specific whole genome duplication (ssWGD) [20,33]. While many paralogs produced from an earlier WGD in the common teleost ancestor ~300 Mya (teleost-specific, tsWGD) are conserved in modern teleost genomes (*e.g.* 20% in salmonids), a much higher proportion (55%) of paralog pairs have been retained following ssWGD, making salmonids an ideal system for studying gene evolution over successive WGD rounds [30]. Antiviral immune responses have been studied intensively in these species independently, including at the transcriptomic level ([5,19,24] and refs therein). A direct comparison of antiviral responses between these two species, which diverged around 30 million years ago, would provide novel insights into conserved or divergent mechanisms of innate immunity, including in relation to WGD events.

In this study, we compared the transcriptomic response of rainbow trout and Atlantic salmon to polyinosinic:polycytidylic acid (poly I:C), a synthetic analog of dsRNA, in both head kidney tissue (*in vivo*) and in purified head kidney leukocytes (*in vitro*). The head kidney is the primary haematopoietic site in fish, and also a secondary lymphoid organ with a high diversity of leukocyte subtypes. Poly I:C was used as a common, powerful inducer of type I IFN to investigate the response of these closely related species under identical experimental conditions. Transcriptome data were mapped to the most recent reference genomes to optimize inferences on paralogous gene expression and to ensure the relevance of our work as a reference for the fish immunology community.

2. Results

2.1. Overview of study

We compared head kidney samples stimulated by *in vivo* and *in vitro* by poly I:C, a synthetic analog of dsRNA, to controls with two objectives: (1) to compare the anti-viral response induced by a viral RNA mimic in two salmonid species separated by ~30 Mys of evolution [30], following the same stimulation and analysis methods. The rationale for using both *in vivo* and *in vitro* stimulations is that all cellular interactions are better reflected by the reaction of the tissue *in vivo*, while extracted leukocytes receive more potent exposure to the stimulant *in vitro*. This allowed a standardized comparison of the conservation of constitutive gene expression and inducibility by poly I:C between both species; and (2) to identify a comprehensive repertoire of rainbow trout and Atlantic salmon genes responsive to poly I:C (hence, type I IFN), with a detailed phylogenetic analysis inferring paralogy and orthology relationships for the most conserved and functionally relevant genes, including in relation to tsWGD and ssWGD events. We use "orthology" here in the frame of a purely phylogenetic definition, which does not necessarily imply that orthologous genes have the same function. The experiments were performed in young sexually immature adults, where adaptive immunity was fully in place with all immune cell types present.

2.2. Trout and salmon head kidney transcriptional response in vivo and in vitro

Strong transcriptional responses were induced by poly I:C *in vivo* and *in vitro* in both species. Individual samples were clearly separated by principal component analysis (PCA) into groups reflecting infection status and *in vivo* versus *in vitro* models (Fig. 1A and B). A large number of genes were up-regulated in both conditions (see intersect *in vivo / in vitro* Fig. 1C and D: 801 in trout, 658 in salmon). Many more genes were significantly modulated *in vivo* by poly I:C injection compared to *in vitro* stimulation in both species, and more genes were differentially expressed in rainbow trout than Atlantic salmon (Tables S1 and S2 for rainbow trout and Atlantic salmon, respectively). In rainbow trout, more genes were down-regulated than up-regulated *in vivo*, while the opposite was observed *in vitro*. For Atlantic salmon, more genes were up-regulated than down-regulated under both conditions. While the response appeared more extensive in rainbow trout than in Atlantic salmon, the ratios of numbers of responsive genes between experimental groups were similar in both species (compare Fig. 1C and D). Strikingly, large numbers of genes were differentially expressed *in vivo* only, in both species, highlighting the importance of intact cell-cell direct interactions, paracrine modulation by circulating factors, and possibly the contribution of non-leukocytic cells missing in the *in vitro* assay.

To compare the global response to poly I:C stimulation across species, we performed gene set enrichment for Gene Ontology identifiers of Atlantic salmon and rainbow trout based on official names (HGNC) of human orthologs of the differentially expressed genes. Several GO terms were commonly enriched in all four experimental groups (*i.e.*, *in vivo* and *in vitro*, for both species) (Fig. S1, Table S3). These comprised generic GO terms associated with immune or inflammatory response, or more specific terms related to IFN response, with *defense response to virus* and *type I interferon signalling pathway* being the most enriched terms in Atlantic salmon *in vivo* and *in vitro*. Interestingly, *stimulatory C-type lectin receptor signalling pathway* and *response to lipopolysaccharide* (explained by genes encoding rREL and many TNF superfamily factors), typically associated with responses to bacteria, were enriched in all conditions.

The GO term *positive regulation of interleukin-12 production* was enriched in rainbow trout and Atlantic salmon leukocytes stimulated *in vitro*, but not *in vivo*. GO enrichment for IL-12 production was based on a signature comprising *CD40, RIPK2, IRF8, IRF5, CCL19, TIRAP* and *TLR3*, but not *IL12* itself. In fact, *il12* genes were induced only in rainbow trout, not in Atlantic salmon, and this GO term enrichment should be considered with caution. The interpretation of induction patterns of *il12* related genes in salmonids is complex both due to the retention of several paralogs and because cytokines of the IL12 family are heterodimers. For example, IL12 is composed of 2 separate chains, P35 (encoded by IL12a) and P40 (encoded by IL12b). Rainbow trout and Atlantic salmon each have three *il12a-p35* genes (ENSOMYG00000046134; ENSOMYG00000037837; ENSOMYG00000043265; ENSSSAG00000003180; ENSSSAG00000072619; ENSSSAG00000065452), and five *il12b-p40* genes (ENSOMYG00000017650; ENSOMYG00000017163; ENSOMYG00000015114; ENSOMYG00000019564; ENSOMYG00000016603; ENSSSAG00000079839; ENSSSAG00000069633; ENSSSAG00000068948; ENSSSAG00000041830; ENSSSAG00000009655). To further complicate matters, P35 and P40 subunits can heterodimerise with EBI3 or P19 to form IL35 and IL23 all respectively. In contrast, *positive regulation of GTPase activity* and *peptidyl-tyrosine phosphorylation* (a key step in inflammatory/STAT pathway) were enriched in both species after *in vivo* stimulation only.

Other GO terms were enriched in rainbow trout specifically, after both *in vivo* and *in vitro* stimulation, notably *T-cell activation* and *Ag processing/presentation*. Several terms were enriched in only one experimental group. These comprise the induction of inflammatory IL1 and IL6 pathways, which were only found in rainbow trout cells after *in vitro* stimulation. In contrast, *positive regulation of smooth muscle cell proliferation*, *angiogenesis* and *response to mechanical stimulus* were only identified after *in vivo* stimulation in Atlantic salmon, implying a regeneration

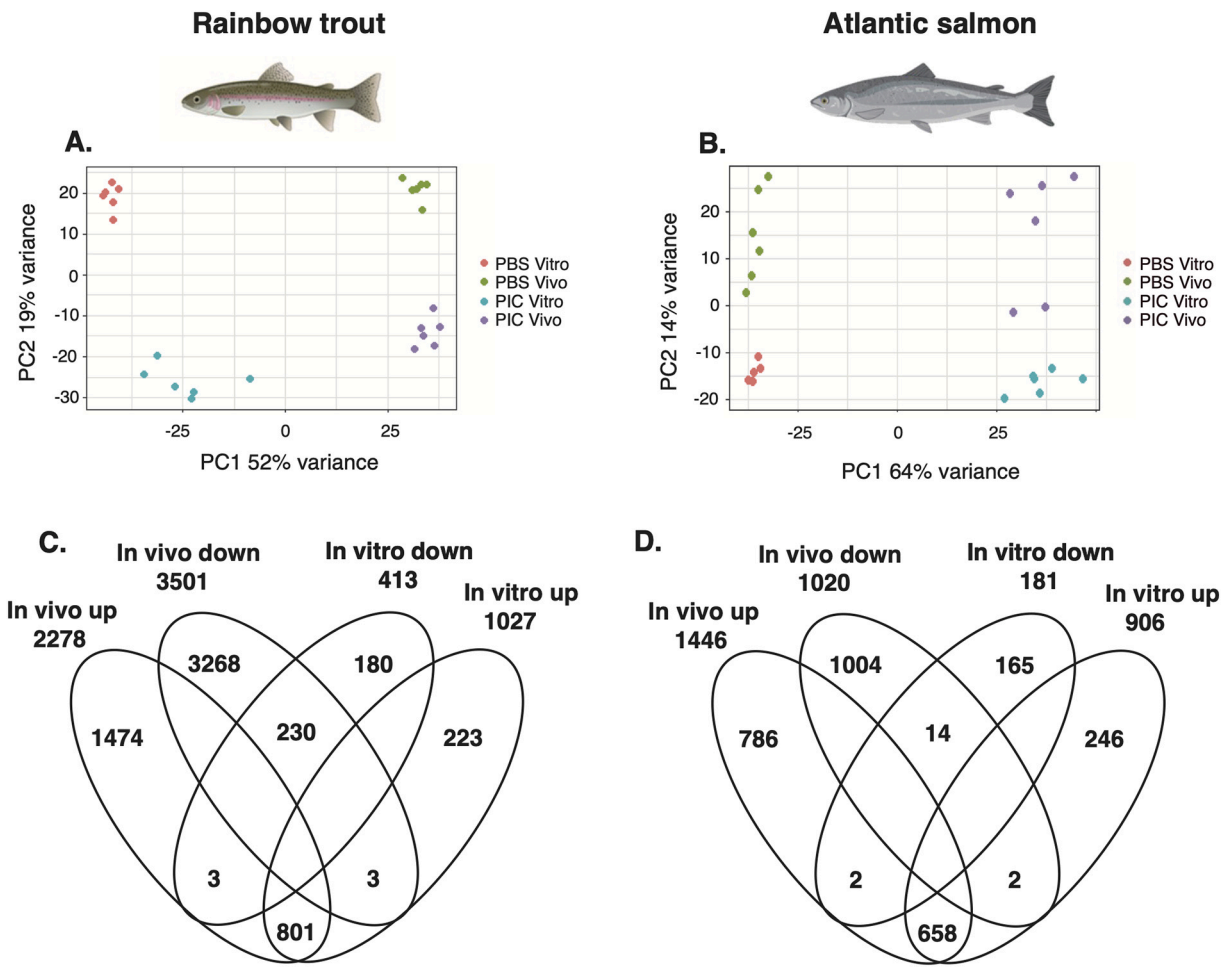


Fig. 1. Transcriptome response of rainbow trout head kidney to poly I:C stimulation.
Principal component analysis of expression levels of up- or down regulated genes after *in vitro* or *in vivo* poly I:C stimulation for 24 h in rainbow trout (A) and Atlantic salmon (B). PIC: poly I:C stimulated groups; PBS: control groups. Projection on the two first axis is shown (dimension 1: horizontal axis; dimension 2: vertical axis). Venn diagrams showing the number of rainbow trout (C) and Atlantic salmon (D) genes significantly up- or down- regulated (*adj* $p < 0.05$; Fold Change (FC) > 2 or < 0.5) by poly I:C *in vivo* or *in vitro* stimulation. The total numbers of genes are indicated for each condition, and full annotated lists are available in Tables S1 and S2.

response in head kidney upon infection.

Based on the global homology prediction used, we cross-referenced the complete list of genes up-regulated *in vivo* and/or *in vitro* between the two species. Strikingly, only 37% of Atlantic salmon up-regulated genes showed up-regulation in rainbow trout. An even lower proportion (25%) of rainbow trout up-regulated genes matched orthologs of poly I:C induced Atlantic salmon genes. Thus, a significant fraction of the transcriptional response to poly I:C stimulation was not shared between orthologous genes in the two salmonid species.

2.3. Induction of type I and type II IFN by poly I:C stimulation

Focusing on the antiviral response, a critical component was the large repertoire of type I IFN genes reported in rainbow trout and Atlantic salmon [31,47,51]. In salmonids more than twenty type I IFN genes are present belonging to six subtypes: IFNa, IFNb, IFNc, IFNd, IFNe and IFNf [31,38]. Within each subtype, rainbow trout and Atlantic salmon genes often diversified independently, and one-to-one orthologous relationships cannot be defined between individual genes [31]. While mapping RNAseq reads on a large array of short and highly similar genes is complicated, our results provide insights about the contribution of different subsets of IFN in these salmonid species. Of the type I IFNs, only genes belonging to the IFNA subtype were significantly modulated, while expression levels of the other type I IFNs were negligible (Table 1). Three rainbow trout genes and five Atlantic salmon *ifna* co-orthologs were up-regulated both *in vitro* and *in vivo* (ENSOMYG00000063614, ENSOMYG00000043922, ENSOMYG00000063439; ENSSSAG00000104134, ENSSSAG00000086201, ENSSSAG00000088069, ENSSSAG00000121777, ENSSSAG00000117268). One IFNA subtype gene from each species was significantly up-regulated solely *in vivo* (ENSOMYG00000043885 and ENSSSAG00000108534).

Regarding type II IFN, 3 *ifnγ* genes were found in rainbow trout and 6 in Atlantic salmon with variable expression patterns. Despite the extra numbers of Atlantic salmon genes in this family, only 2 for each species were found to be expressed, both of which were significantly upregulated *in vitro* while unchanged *in vivo*. However, the expression level after stimulation *in vitro* remained very low in this species. Rainbow trout *ifnγ* (ENSOMYG00000023065) was significantly upregulated *in vitro* and belonged to the same phylogenetic clade in a 1:3 relationship as the 2 significantly modulated Atlantic salmon *ifnγ* genes (ENSSSAG00000105299, ENSSSAG00000116696) (Table 1). The other significantly induced rainbow trout *ifnγ* gene (ENSOMYG00000061657) was phylogenetically related 1:1 to (ENSSSAG00000102271), however expression could not be detected for this Atlantic salmon gene (Table 1). The remaining *ifnγ* genes of both species were not expressed in our study.

Table 1
Significantly induced type I and II IFN and cytokines following poly I:C stimulation in rainbow trout and Atlantic salmon.

Family	Trout gene ID	Rainbow trout				Phylogenetic relationship	Atlantic salmon				
		In vivo expression[1,2]	*In vitro* expression	*In vivo* FC	*In vitro* FC		*In vivo* FC	*In vitro* FC	*In vivo* expression	*In vitro* expression	Salmon gene ID
IL10	ENSOMYG00000005081	0.04	1.52	[1]	[2]57.94	1:1		5.20	5.54	0.03	ENSSSAG00000107544
	ENSOMYG00000071035	0.02	0.05			1:1			0.19	1.17	ENSSSAG00000111615
IFNA	ENSOMYG00000063614	0.36	0.27	8.70	49.62		69.61	233.78	1.04	0.77	ENSSSAG00000104134
	ENSOMYG00000043885	0.08	0.11		9.30		6.55	8.78	0.67	0.45	ENSSSAG00000086201
	ENSOMYG00000043922	0.25	0.34	18.36	7.69		5.22	9.18	0.21	0.26	ENSSSAG00000088069
	ENSOMYG00000063439	0.38	0.41	20.31	9.80		35.68	10.76	0.58	0.45	ENSSSAG00000121777
	ENSOMYG00000041047	0.15	0.12					15.30	0.80	0.27	ENSSSAG00000108534
	ENSOMYG00000057496	0.01				many:many	6.09	10.30	0.94	0.50	ENSSSAG00000117268
	ENSOMYG00000042483										ENSSSAG00000087874
	ENSOMYG00000063541										ENSSSAG00000055115
											ENSSSAG00000120909
											ENSSSAG00000119937
IFNG	ENSOMYG00000023065	0.068	0.264		17.55			34.04	0.22	0.00	ENSSSAG00000105299
						1:3		28.26	0.39	0.01	ENSSSAG00000116696
											ENSSSAG00000091489
	ENSOMYG00000061657	0.077	0.845		26.63	1:1					ENSSSAG00000102271
	ENSOMYG00000051259										ENSSSAG00000112202
						1:2					ENSSSAG00000086488

[1] Only significant fold change (FC) values are shown.
[2] Expression values were obtained by normalising gene count against total raw counts and then the ratio to ELF1a multiplied by 1000, trout (ENSOMYG00000038328) and salmon (ENSSSAG00000077892) counts, blank cells = no counts for the gene.

2.4. Salmonid core type I IFN induced response genes (ISGs) conserved in human (Set 1 and Set 2)

2.4.1. Overlap of salmonid poly I:C induced genes with human orthologs

To appreciate the wider evolutionary conservation of the type I IFN response, we then assessed the proportion of genes induced by poly I:C in rainbow trout or Atlantic salmon, that had an ortholog in human. Based on Ensembl release 108 annotations, we searched for human orthologs of salmonid genes induced in at least one condition (*i.e.*, *in vivo* and/or *in vitro*). In rainbow trout, 65% of the up regulated genes (1541 out of 2504) had an ortholog in human. In Atlantic salmon, these proportions were lower, with 48% of the up-regulated genes (819 out of 1694) having an ortholog in human. These observations are consistent with our previous report that 60% of genes up-regulated by IFNφ1 in zebrafish larvae had a human ortholog [27].

2.4.2. Set 1: salmonid genes induced by poly I:C belonging to a core conserved ISG set

We previously identified 95 human ISGs (in 77 paralogy groups) sharing ortholog(s) with zebrafish ISGs [27], representing a core ISG set present in the last common ancestor of teleosts and tetrapods (LCATT) that lived ~450 Mya. This gene set comprised most of the key signalling components of the type I IFN system [27]. Rainbow trout and Atlantic salmon (co-)orthologous genes could be found for all these core ISGs (except for *gimap* and *pmaip*): in total, 325 genes in trout and 348 genes in Atlantic salmon (Table S4). Of all these salmonid genes, 303 and 301 rainbow trout genes (*in vitro* and *in vivo*, respectively), and 305 and 305 Atlantic salmon genes (*in vitro* and *in vivo*, respectively) were expressed in our datasets. Around 50% of these genes were differentially expressed after poly I:C stimulation *in vivo* and/or *in vitro* (Set 1, Fig. 2). Specifically, 160 and 144 genes were significantly modulated *in vitro,* in rainbow trout and Atlantic salmon, respectively, while 191 and 154 genes were significantly modulated *in vivo*, in rainbow trout and Atlantic salmon, respectively (Table S4). All members of some gene families were highly up-regulated (Table S4). For instance, all expressed members of the gene families: ADAR, DDX58, EPSTI1, IFI35, IRF (IRF7 and IRF9), ISG15, RSAD2, USP18, HELZ2, RNF213, PKR, and HERC5/6, displayed significant up-regulation across both conditions in both species. Further to this, only 8 ISGs out of 95 conserved in human and zebrafish (*CASP1, CASP7, FAM111, GIMAP, MS4A, PMAIP1, PTMA, SETX*) had no rainbow trout or Atlantic salmon up-regulated orthologs, confirming that the core ISG response is overall well conserved between human, zebrafish, rainbow trout and Atlantic salmon (Table S4), hence likely across Euteleostomi. Two gene families (AHNAK and RARRES3) had only significant down-regulated (*i.e.*, no induced) representatives in salmonids, while they were up-regulated in zebrafish larvae [27]. Detailed phylogenetic relationships between conserved ISG and their orthologs in rainbow trout and Atlantic salmon genes are provided in Table S4, based on our global homology prediction (see Methods) in relation to their expression pattern across analysed conditions. For example, 87 rainbow trout and Atlantic salmon genes induced by poly I:C were 1-to-1 orthologs, while 39 genes from either species had a 1:1 ortholog that was not significantly up-regulated after stimulation.

Collectively these observations indicate that most core ISGs defined by Levraud and co-authors [27] are conserved in salmonids, with relatively few complete losses, and that a consistent response to poly I:C occurs in most cases in both salmonid species.

To better understand evolution after the tsWGD and ssWGD events, we studied homology relationships between representatives of Set 1 genes in human, zebrafish (Superorder: Ostariophysi), northern pike (Superorder: Protacanthopterygii, Order: Esociformes), rainbow trout and Atlantic salmon (Superorder: Protacanthopterygii, Order: Salmoniformes) (Table S4). Inclusion of zebrafish allowed for a direct comparison with our previous work [27] while the inclusion of northern pike helped identify if salmonid Set 1 genes were retained from ssWGD, as this species belongs to the sister lineage to salmonids (Esociformes),

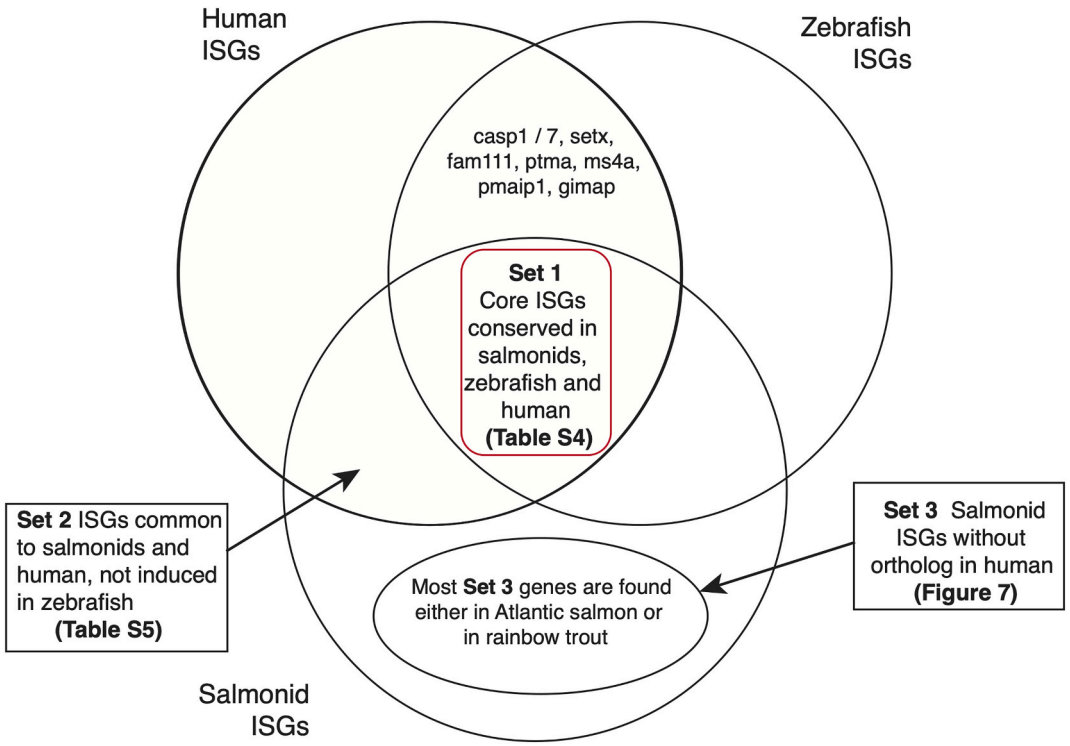


Fig. 2. Overview of the different gene sets involved in the type I IFN/poly I:C induced response according to their evolutionary conservation across different vertebrate clades.
Gene sets analysed in this study are in bold. Supplementary tables showing the list and annotations of these sets are mentioned.

which is closely related in terms of gene repertoires, but does not share ssWGD. However, local duplications or gene losses add an extra layer of complexity for these interpretations. Therefore, gene number and duplications were also assessed based on a general consensus of the other teleost and salmonid species contained within our homology prediction, to interpret retentions or losses following WGD.

As shown in Figs. 3, 6.6% of human core ISGs followed a 1 (in human) to 2 (in zebrafish or pike) to 4 (in salmonid species) evolutionary pattern, consistent with sequential retention of paralogs from tsWGD and then ssWGD, implying evolutionary pressure to maintain duplicated copies. These genes belonged to various families and had

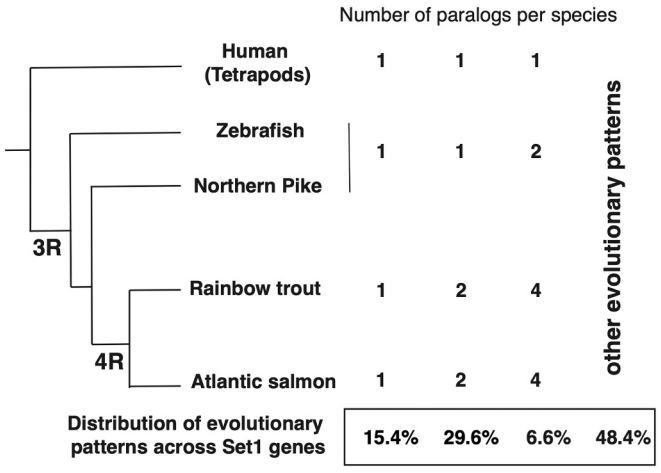


Fig. 3. Evolutionary patterns of core conserved ISG.
Phylogenetic relationships between human ISG and their conserved orthologs in zebrafish, pike, rainbow trout and Atlantic salmon were determined based on Ensembl database and Scorpio analysis (see Table S4). The % were computed based on the number of human ISGs corresponding to each pattern.

multiple functions. This included *cGAS*, a cytosolic DNA sensor involved in type I IFN induction, *socs1*, a key inhibitor of janus kinase (regulator of downstream type I IFN signalling), and *ncoa7* (a nuclear receptor coactivator). Using equivalent species designations, a 1:1:2 evolutionary pattern represents a higher fraction of Set 1 genes (29.6% of human genes), indicating paralog loss after tsWGD, followed by paralog retention following ssWGD (Fig. 3). The proportion of human core ISGs associated with a 1:1:1 evolutionary pattern by the same definition was 15.4% (Fig. 3), requiring sequential paralog loss after tsWGD and ssWGD. They comprised key genes of the jak/stat cascade (*stat2, stat6, nmi),* three RNA helicases (*helz2, rig I/ddx58, mda5/ifih1*), a key mitochondrial regulator of inflammation (*cmpk2*), a regulator of PKR (*epsti1*), and factors involved in ubiquitination (*rnf138, rnf180, uba7*). Evolutionary pattern assignations can be viewed in column C of Table S4.

Fold change distributions of salmonid Set 1 genes were significantly correlated between *in vitro* and *in vivo* stimulation (rainbow trout in Fig. 4A, Atlantic salmon in Fig. 4B), with notable differences observed for genes across a range of induction values, especially *in vitro*. We then correlated normalized expression levels between rainbow trout and Atlantic salmon Set 1 orthologs (defined in Table S4) after *in vitro* (Fig. 4C) and *in vivo* (Fig. 4D) stimulation. Many Set 1 genes showed highly correlated induction in both species, as illustrated for *helz2, rsad2, irf7, samd9, stat1* and *stat2*. However, there were important exceptions such as *isg12,* with *isg12–1* and *– 2* highly induced in rainbow trout after both *in vivo* and *in vitro* stimulation (Fig. 4A), but the single-copy *isg12* gene in Atlantic salmon does not show significant change of expression after stimulation (Table S2).

2.4.3. Set 2: salmonid genes induced by poly I:C with ISG orthologs in human, which are not induced by IFNφ1 or lost in zebrafish

Strikingly, 100 rainbow trout/Atlantic salmon genes which had no ortholog induced by IFNφ1 in zebrafish larvae [27], but have a human ISG ortholog, and were up-regulated by poly I:C in at least one of the groups of the present study (Table S5). The 47 human orthologs of these

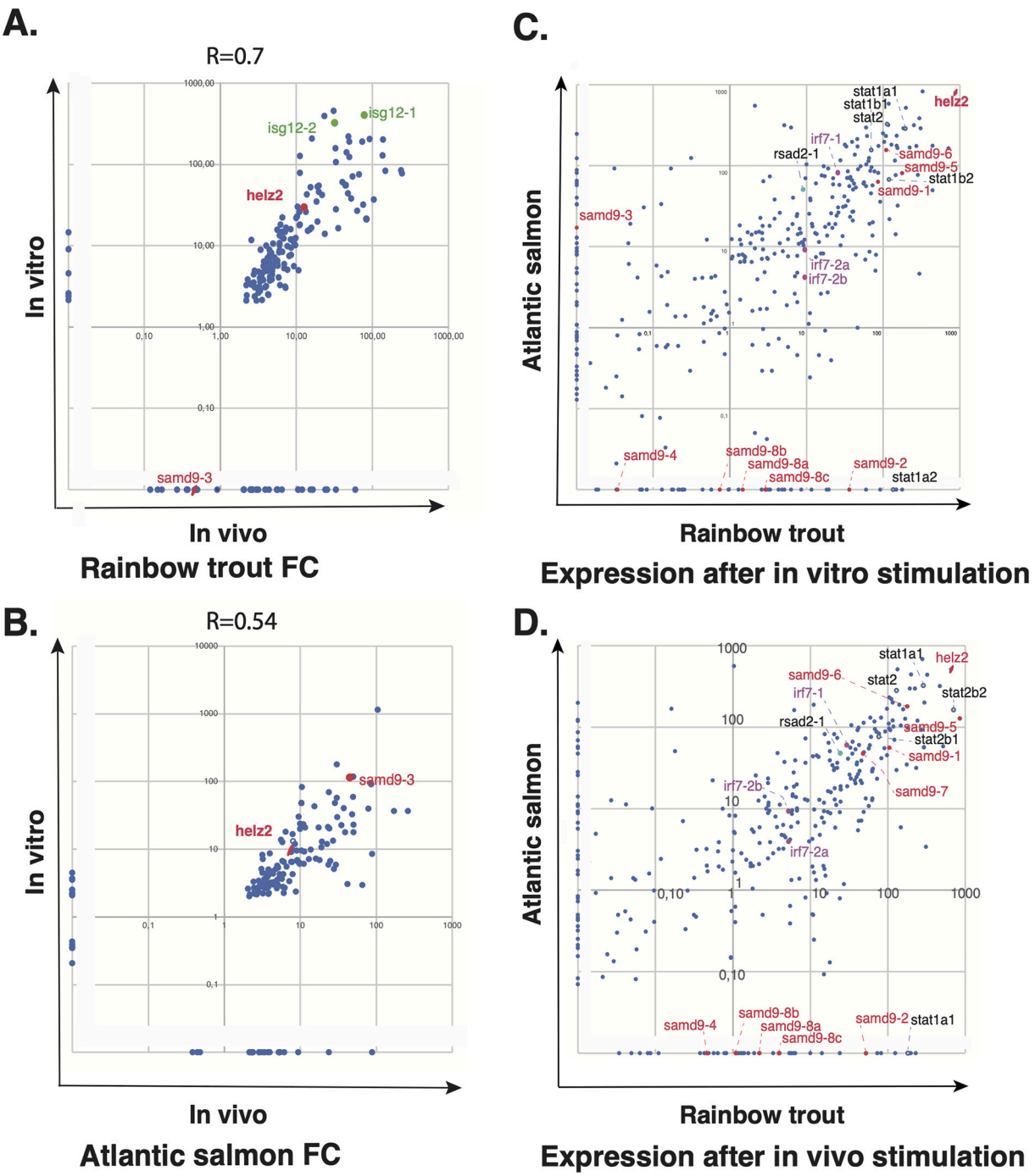


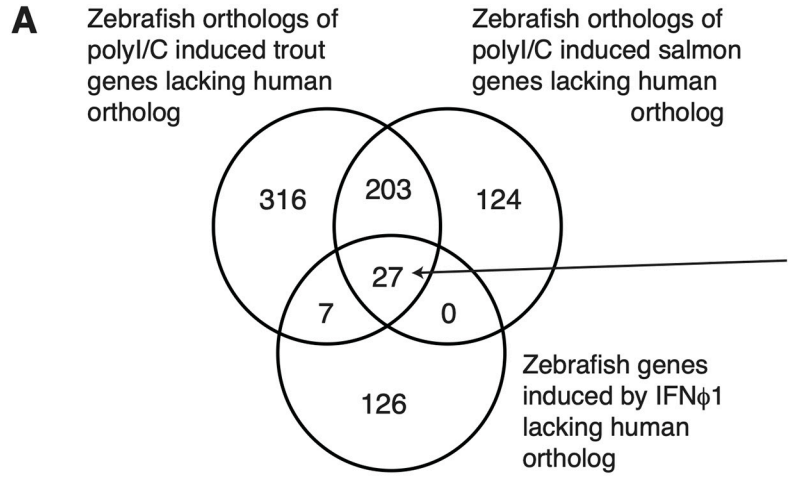
Fig. 4. Comparison of the core transcriptional response in Atlantic salmon and rainbow trout.
Comparison of fold change distribution after poly I:C stimulation of rainbow trout (A) and Atlantic salmon (B) head kidney *in vivo* and *in vitro*, limited to conserved ISGs (Set 1). Genes expressed only *in vivo* or only *in vitro* are represented on additional horizontal and vertical axes.
Comparison of normalized expression levels (see Table S4) of rainbow trout and Atlantic salmon (co)orthologs after *in vitro* (C) and *in vivo* (D) stimulation. The representation is limited to conserved ISGs (set 1), and "many to many" orthologs have been excluded as they cannot be easily represented in this graph. Genes expressed or present only in one species are represented on additional horizontal and vertical axes. Key genes commented in the text are indicated, with their gene family identified by different colours/symbols.

genes comprised genes encoding 10 transcription factors (*ATF3, BCL2L14, BCL3, JUNB, MAFB, ZCCHC2, IRF1, IRF2, LMO2, CSRNP1*), 16 enzymes (*ARG2, CD38, CYP1B1, GBP2, GMPR, NCF1, SPTLC2, ST3GAL4, USP42, RBCK1, SDS, RNF19B, DUSP5, MKL, RIN2, TNFAIP3*), 4 secreted factors (*SAA1, SERPINE1, ANGPTL1, TNFAIP6*), 2 lectins (*CD69* and *CLEC2*), a member of the IL1 family (*IL1RN*), the nucleic acid sensor *TLR7* and the adaptor *MYD88*, membrane receptors (*IFNGR1, MILR1, MICB, TNFSF10*), the immuno-proteasome *PSMB9,* the extracellular matrix component *EHD4,* the RNA binding protein *ELAVL4* and genes involved in signalling or cell metabolism (*FFAR2, FNDC3B, THEMIS2, CDKN1A, PLIN2*). These genes were not identified by Levraud and co-workers for several possible reasons: (1) a few genes are absent in cyprinids/zebrafish (*BCL2L14*; MAFB; SDS; IL1RN/37) (Table S5), (2) when orthologs were present, the lack of induction in our previous study could be explained by a wider range of induction by poly I:C, compared to zebrafish IFNφ1 and chikungunya virus; (3) while zebrafish larvae

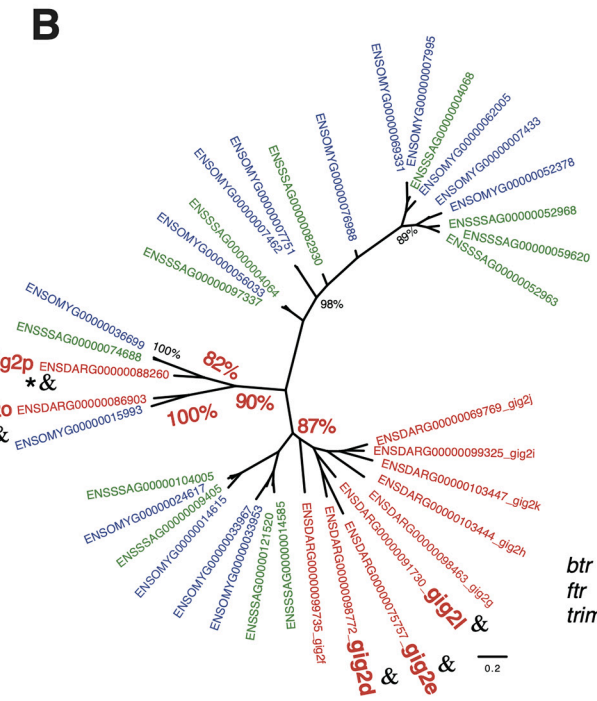
have no (or very few) lymphocytes, we analysed here mature populations of salmonid head kidney leukocytes, which likely express a wider set of ISGs. The TLR7/MYD88 pathway, presumably expressed in the fish counterparts of plasmacytoid dendritic cells is likely a good example [18].

2.4.4. Set 3: salmonid poly I:C stimulated genes lacking a human ortholog – characterized by large gene families induced by IFNφ1 in zebrafish

We next focused on salmonid ISGs/poly I:C induced genes that lacked an ortholog in human (Set 3, *i.e.,* 963 genes in rainbow trout, 875 in Atlantic salmon, see Fig. 4). As we previously described the repertoire of ISGs lacking a human ortholog in zebrafish [27], we asked if these genes were shared between salmonids and zebrafish. We first determined zebrafish orthologs for rainbow trout and Atlantic salmon Set 3 genes. As shown in Fig. 5A, only 27 Set 3 genes were previously observed to be IFNφ1 responsive in zebrafish. These genes comprised 2 tapasin-like sequences, a grass carp reovirus -induced gene (*gig*)2, 12 *trim* genes, a galectin as well as 11 unannotated genes. While these genes lacked a human ortholog, six had orthologs in other tetrapods in the Ensembl database, indicating that a subset of Set 3 genes were present in



Ensembl ID	gene name	Ortholog in tetrapods
ENSDARG00000030954	finTRIM	no
ENSDARG00000032043		
ENSDARG00000038646		
ENSDARG00000039852		
ENSDARG00000041220		
ENSDARG00000053293		
ENSDARG00000079537		
ENSDARG00000092961		
ENSDARG00000095032		
ENSDARG00000095345		
ENSDARG00000095678		
ENSDARG00000074360		
ENSDARG00000076483	tapasin-like	yes
ENSDARG00000097019		
ENSDARG00000088251	lgals17	no
ENSDARG00000088260	gig2p	yes
ENSDARG00000067718	unknown	no
ENSDARG00000073789		yes
ENSDARG00000078832		no
ENSDARG00000087575		no
ENSDARG00000089432		yes
ENSDARG00000092228		no
ENSDARG00000096562		no
ENSDARG00000098204		yes
ENSDARG00000100660		no
ENSDARG00000100968		no
ENSDARG00000101189		no



C. *btr, fintrim,* and *trim25* genes induced by poly I:C

	Rainbow trout				Atlantic salmon			
	Total nb of genes	In vivo + in vitro	In vivo only	In vitro only	Total nb of genes	In vivo + in vitro	In vivo only	In vitro only
btr	27	21	6	0	26	18	7	1
ftr	59	52	4	3	157	127	17	13
trim25	2	1	0	0	2	1	0	0

Fig. 5. Genes involved in the antiviral response of zebrafish, rainbow trout or Atlantic salmon, but lacking a human ortholog.
Venn diagrams showing the intersect between zebrafish, trout and salmon lists of poly I:C induced genes with no ortholog in human and no ortholog in chicken. (A) overlap between zebrafish type I IFN induced genes, zebrafish Ensembl orthologs of rainbow trout poly I:C induced genes and zebrafish Ensembl orthologs of Atlantic salmon poly I:C induced genes. (B) Maximum likelihood tree of protein sequences of *gig2* family members from zebrafish (red), rainbow trout (blue) and Atlantic salmon (green). All *gig2* zebrafish paralogs were included in the analysis; those induced by IFNφ1 in Levraud et al. [27] are marked with an asterisk "*", while those induced by viral infection in Balla et al. [4] are marked with an "&". All salmonid *gig2* paralogs included in this analysis are induced in rainbow trout and/or in Atlantic salmon. The evolutionary history was inferred by using the Maximum Likelihood (ML) method and JTT matrix-based model implemented in MEGA X [25]. Briefly, initial trees for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model based on the 38 amino acid sequences of GIG2 proteins, followed by the ML approach. The tree with the highest log likelihood (−7480.38) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. There were a total of 246 positions in the final dataset. 100 bootstrap replicates were performed. (C) Total numbers of *trim25, btr,* and *ftr* paralogs up-regulated by poly I:C in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the LCATT. As *gig2* and *trim* represent large gene subsets with members involved in the response to poly I:C/virus across fish clades, we analysed their contribution to the poly I:C induced response in rainbow trout and Atlantic salmon.

gig2 (for "grass carp reovirus (GCRV)-induced gene 2") genes were previously identified as a fish ISG with antiviral activity, and constitute a family encoding proteins weakly similar to poly(ADP-ribose) polymerases (PARPs) in agnathans, cartilaginous and bony fish, and amphibians but not amniote tetrapods [50]. In zebrafish larvae, only two family members (*gig2o* and *gig2p*) were induced by IFNφ1 [27]; additional zebrafish *gig2* genes *(gig2d, gig2e*, and *gig2l)* were upregulated by viral infection in [4], which could also be ISGs. A higher number of *gig2* genes were up-regulated by poly I:C in rainbow trout (15 genes) and Atlantic salmon (13 genes), both *in vivo* and *in vitro*. As illustrated by phylogenetic analysis (Fig. 5B), only ENSDARG00000088260 (*gig2p*) has a putative ortholog in rainbow trout and Atlantic salmon, while ENSDAG00000086903 (*gig2o*) apparently has an ortholog only in rainbow trout. In contrast, the other zebrafish *gig2* paralogs, including *gig2d, gig2e*, and *gig2l,* grouped together in a single clade, representing a sister group to a clade containing multiple *gig2* genes from Atlantic salmon and rainbow trout, while a separate paraphyletic grouping of salmonid *gig2* genes was identified with no apparent zebrafish orthologs. This branching pattern is typical of gene families that diversified extensively in different fish lineages. The large number of poly I:C induced *gig2* genes in salmonids was remarkable, with 13 and 14 genes in rainbow trout and Atlantic salmon respectively. In both salmonid species, the basal levels of expression *in vivo* and *in vitro* were generally similar, with a few exceptions (Table 2). Furthermore, fold changes were not higher *in vitro* than *in vivo*, in contrast to the general trend observed in this study where magnitude of response was generally greater *in vitro*. There was no clear correlation between *in vitro/in vivo* induction pathway, and orthologous relationships of trout and salmon *gig2* genes.

Tripartite motif proteins (TRIM) are defined by the combination of a RING finger motif, B Boxes and by a coiled-coil (CC) domain followed by various C-terminal domains [36] often with E3 protein ligase activity. *Trim* genes encode a number of important factors that restrict viral infections, such as TRIM5 that blocks HIV in non-human primates [46]. Three gene expansions of *trim* genes have been reported in teleosts, which are not present in other vertebrates [9]: (1) the fish novel trim or *fintrim (ftr)*, (2) the bloodthirsty-like trim (*btr*) related to *trim39* and (3) the *trim35* genes [8]. Our previous work in zebrafish showed that many *ftr* and *btr*, but not *trim 35*, can be classified as ISGs [27]. We found a similar situation in salmonids after poly I:C stimulation, but with higher numbers of genes (Fig. 5C): 27 and 26 genes and 59 and 157 genes were annotated as *btr* and *ftr* in rainbow trout and Atlantic salmon, respectively. These findings are in accordance with the extensive expansion of *trim* genes in salmonids, particularly in Atlantic salmon [48]. The precise annotation of such large and recent gene expansions is notoriously difficult, and the mapping of sequencing data represents an additional potential source of error. To test the annotation of these genes in rainbow trout and Atlantic salmon, we therefore checked the domain composition of the protein models predicted for poly I:C-induced *btr* and *ftr*; we then performed a phylogenetic analysis of the sequences containing a complete domain combination (Ring B box CC and B30.2 domains) and found a very high consistency between the curated zebrafish annotation [26] and the salmonid annotation (*btr* versus *trim25* versus *ftr*) available for rainbow trout and Atlantic salmon in Ensembl 108 (Fig. S2). Specifically, zebrafish, rainbow trout and Atlantic salmon TRIM25 (the closest relative of ftr and btr present in human; and a conserved ISG), BTR and FTR sequences grouped into well supported branches.

Thus, our data identify *gig2* and *ftr/btr* as multigenic families bringing important contributions to the transcriptome response of salmonid head kidney cells to poly I:C.

2.5. A large fraction of the response to poly I:C is restricted to each species

To assess the proportion of Set 3 genes which had orthologs in non-human tetrapods and identify the "fish-specific" fraction of responding genes, we then looked for orthologs in mouse, chicken and the frog *Xenopus tropicalis*. Fig. 6A indicates that a minor fraction of Set 3 (25% in rainbow trout, 20% in Atlantic salmon) was present in the LCATT, while being lost in human. The large majority of Set 3 genes (Set 3* genes – for rainbow trout: 721 out of 963 genes = 75%; for Atlantic salmon: 700 out of 875 genes = 80%) had no ortholog in these tetrapod species, and therefore likely represent fish-specific ISGs. We then analysed the phylogenetic relationships of Set 3* genes between rainbow trout and Atlantic salmon. A substantial proportion had no ortholog in the other

Table 2
Poly IC induction of gig2 family members. Correspondence between salmon and trout genes is based on phylogenetic relationships corrected by Scorpio.[1]

	Rainbow trout				Phylogenetic relationship	Atlantic salmon				
Trout gene ID	*In vivo* expression[2]	*In vitro* expression	*In vivo* FC	*In vitro* FC		*In vivo* FC	*In vitro* FC	*In vivo* expression	*In vitro* expression	Salmon gene ID
ENSOMYG00000007433	9.61	16.32	7.5[2]	9.8	1:1	6.2	5.0	35.85	49.25	ENSSSAG00000059620
ENSOMYG00000056033	0.04	0.06	13.9	8.7		21.1	84.5	46.90	41.27	ENSSSAG00000097337
					1:2	7.4	86.9	9.74	21.44	ENSSSAG00000004064
ENSOMYG00000007995	31.14	11.32	104.6	[12]N.S	2:2	4.6	5.4	0.33	0.68	ENSSSAG00000105375
ENSOMYG00000069331	0.04	0.09	29.9	N.S		7.8	86.7	15.74	29.72	ENSSSAG00000004068
ENSOMYG00000014615	1.67	1.18	2.3	N.S	2:2	4.2	2.1	2.98	1.58	ENSSSAG00000104005
ENSOMYG00000024617	2.11	2.17	N.S	2.2		2.9	N.S	3.22	2.42	ENSSSAG00000009405
ENSOMYG00000007462	1.01	2.11	181.4	45.0		6.5	4.3	18.54	17.59	ENSSSAG00000082930
ENSOMYG00000007751	0.34	0.55	79.9	102.2						
ENSOMYG00000076988	12.12	8.48	6.7	8.4	3:1					
					0:1	7.1	3.5	0.27	0.27	ENSSSAG00000074688
					0:1	7.6	4.6	16.15	13.62	ENSSSAG00000052963
					0:1	5.4	0:1	0.82	0.87	ENSSSAG00000052968
ENSOMYG00000033953	209.35	216.27	8.7	5.7	1:0					
ENSOMYG00000033967	0.84	2.20	22.5	4.8	1:0					
ENSOMYG00000015993	36.06	46.22	4.0	5.1	1:0					
ENSOMYG00000062005	15.79	30.25	57.9	99.3	1:0					
ENSOMYG00000052378	14.32	15.04	6.5	8.7		N.S	9.7	10.45	27.75	ENSSSAG00000014585
					1:2	N.S	7.3	0.13	0.36	ENSSSAG00000121520

[1] N.S = Non-significant, significant fold change (FC) values.

[2] Expression values were obtained by normalising gene count against total raw counts and then the ratio to ELF1a multiplied by 1000, trout (ENSOMYG00000038328) and salmon (ENSSSAG00000077892) counts.

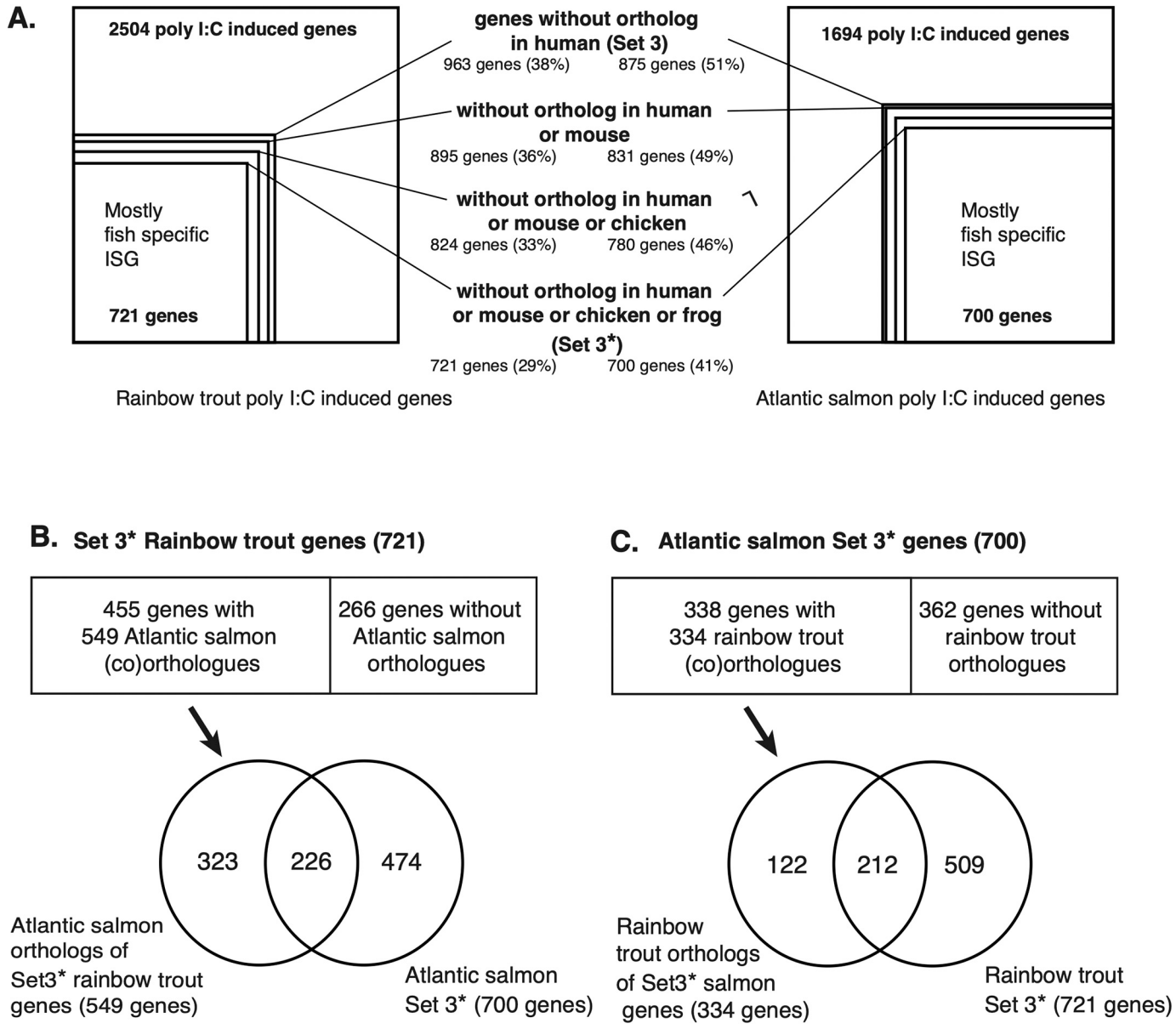


Fig. 6. Genes induced by poly I:C in rainbow trout and Atlantic salmon lacking a human ortholog: a species-specific response.
(A) Analysis of rainbow trout and Atlantic salmon Set3 genes orthology relationships with mouse, chicken, and tropical clawed frog. (B) and (C) Overlap between trout and salmon lists of Set3* genes without orthologs in human, mouse, chicken, and frog.

species (37% for rainbow trout, 51% for Atlantic salmon (Fig. 6B and C). Within the list of 549 Atlantic salmon orthologs of rainbow trout Set 3* genes, only 226 were present in Atlantic salmon Set 3* (representing 32% of this list) (Fig. 6B). Importantly, the other 323 salmon orthologs of rainbow trout Set 3* genes were not induced by poly I:C and had no orthologs in frog, chicken, mouse or human. Rainbow trout orthologs of Atlantic salmon Set 3* genes followed a similar pattern. Overall, the Set 3* fraction of poly I:C responding genes appears to be largely restricted to each fish species, even at close evolutionary timescales such as that separating Atlantic salmon and rainbow trout.

3. Discussion

We provide here a comprehensive repertoire of rainbow trout and Atlantic salmon genes responsive to poly I:C (hence, type I IFNs), with a detailed phylogenetic annotation, pointing to paralogous and orthologous relationships for the most conserved and functionally relevant genes. We characterized the type I IFN response across the two species in live animals and purified primary cultured leukocytes, which are classical models to study immune responses. Our data supports the existence of two distinct arms of antiviral immunity, with divergent evolutionary dynamics: a core conserved gene set shared widely across vertebrate clades and present in the LCATT, and a rapidly evolving get set that has diversified very rapidly among lineages.

3.1. While the core conserved poly I:C induced response is correlated in vivo and in vitro, the overall transcriptional responses show a limited overlap

Many more genes were differentially expressed after poly I:C stimulation *in vivo*, compared to *in vitro*, and those induced in both conditions represented only about a quarter of the complete list. This limited overlap was unexpected, but was observed independently in both species. While this may be partly due to the elimination of connective tissue and erythrocytes by the percoll gradient applied (see Methods), it also underscores the importance of cellular interactions in the tissue.

Importantly, it shows that *in vitro/ex vivo* stimulations cannot be considered representative of the *in vivo* transcriptome response to strong stimulants such as poly I:C, even when primary culture conditions are well established. However, highly induced genes were largely shared between both tested conditions. For example, 68 trout genes of the top 100 induced genes *in vivo* were also induced *in vitro*, of which 65 showed fold change values >10. Further, all top 100 genes induced *in vivo* were also induced *in vitro*, of which 93 showed fold change values >10. Among core conserved ISGs (Set 1), 151 genes were induced both *in vivo* and *in vitro* in rainbow trout (129 genes in Atlantic salmon), with overall good correlation of fold change values *in vivo* and *in vitro*, in both species. Focused analyses of the members of large multigenic families using specially designed sequencing strategies will determine whether or not subsets of paralogs are specialized in particular stimulation conditions.

GO analysis pointed to several defined pathways that were specifically triggered *in vivo* or *in vitro* only, but such results have to be taken with caution due to the complexity of paralog functional specialization, as for *il12*-related genes. Positive regulation of GTPase activity, and peptidyl-tyrosine phosphorylation that induces a key step in inflammatory/STAT pathway, were enriched only *in vivo*, likely in connection with a particular interaction between leukocyte subsets that was lost *in vitro*. More significant are probably the specific induction of IL6, IL10 and IFNγ *in vivo* only. Further experiments, including single cell transcriptome analyses, will be required to clarify which cell types are involved in these responses, and this should be investigated *in vivo*.

3.2. Different transcriptional responses to poly I:C in rainbow trout and Atlantic salmon

The response to poly I:C involved more genes in rainbow trout than in Atlantic salmon, while the numbers of coding genes are very similar in these species (48,326 in rainbow trout, 47,205 in Atlantic salmon; numbers from assemblies USDA_OmykA_1.1and Ssal_v3.1, Ensembl release 208). This difference was most pronounced *in vivo,* indicating it may require cell/cell interactions in an intact organ. Alternatively, it might also reflect a different impact of the aquatic environment during *in vivo* experiments with rainbow trout and Atlantic salmon, including temperature. Furthermore, a significant number of up-regulated genes in one species had no orthologs among the up-regulated genes of the other (>600 in rainbow trout). Given the close relationship of rainbow trout and Atlantic salmon genomes, this was unexpected, and likely reflects that virus-induced evolutionary pressures have led to a fast adaptation of regulatory sequences and transcriptional profiles.

Gene set enrichment identified two major functional modules involved in the response of only one species. First, GO terms linked to antigen (Ag) presentation and T cell responses were enriched only in rainbow trout. More research will be necessary to determine if type I IFN pathways trigger Ag presentation and cellular responses distinctly in these species, or if intrinsic differences of kinetics explain our observations. A second GO term linked to protein ADP ribosylation was also enriched in rainbow trout only, both *in vivo* and *in vitro*. ADP-ribosylation is promoted by poly-adenosine diphosphate-ribose polymerases (PARPs), which have various antiviral functions through direct modification of viral proteins or epigenetic regulation of IFN I, ISG and inflammation genes [16]. It is important to note that several paralogs of *parp9, parp12* and *parp14* were induced by poly I:C both in rainbow trout and Atlantic salmon. A direct functional study of *parp* paralogs will be required to understand how they participate to the response in each condition.

3.3. Core conserved ISGs have diversified in salmonids and show parallel responsiveness to poly I:C

There are 127 zebrafish orthologs to a starting set of 95 human ISGs, corresponding to 77 ancestral genes [27], for which we identified 325 rainbow trout and 348 Atlantic salmon orthologs, approximately half of which were up-regulated by poly I:C (*in vivo* and/or *in vitro*). Genes present as single copy in human, zebrafish, northern pike and salmonids, which likely evolved under a strong selection pressure to eliminate duplicates, comprised several key RIGI-like helicases involved in virus sensing. A single copy of these genes is present in the genomes of mammals and most other tetrapod species in the Ensembl Genome Browser; *rig I* has been lost by chicken [34], but is present in other bird species. Interestingly, a single copy of these three genes has been retained in the allotetraploid genome of *Xenopus laevis* (rigi.L ID: 108713708; mda5/ifih1.L, ID: 108701345; helz-2 ID: 108700953), all on the L subgenome. *Stat2* was another example that is present as a single copy across bony fish and tetrapod genomes, in contrast to *stat1* [7].

Almost all genes induced in rainbow trout or Atlantic salmon had (co)-orthologs induced in the other species. Among one-to-one orthologs in these two species, the response to poly I:C was well correlated. In contrast, the thirty-three genes which had no orthologs in the other species, constituted most of the highly contrasted responses between rainbow trout and Atlantic salmon. Among the exceptions, the case of *isg12* was particularly striking, as *isg12* is one of the most induced ISGs in human and zebrafish; in zebrafish larvae, members of this expanded family were among the most highly induced genes [27]. In rainbow trout, two *isg12* genes were highly induced both *in vivo* and *in vitro*; in contrast, the only gene found in Atlantic salmon was not significantly induced after poly I:C stimulation. As intraspecific variation may also occur, we cannot exclude that some of the contrasting responses between rainbow trout and Atlantic salmon do not reflect purely interspecific differences.

The core set of ISGs conserved in human, zebrafish and salmonids is complemented by about 100 rainbow trout and Atlantic salmon genes up-regulated by poly I:C, which possess an ISG ortholog between salmonids and humans, but not human and zebrafish (named “Set 2”). These conserved sets of ISG present in the LCATT could in the future be complemented by analyses of the response in other teleost species, as each fish group likely has, during evolutionary time, lost and expanded ISGs independently. While the other genes induced by poly I:C are largely species-specific, Sets 1 and 2 provides a comprehensive list of salmonid paralogs with their phylogenetic relationships and induction level during antiviral response, allowing a salmonid-based annotation of ISG paralogs for the first time.

3.4. Gig2, ftr and btr multigene families are not present in mammals but are involved in response to poly I:C both in salmonids and zebrafish

Salmonid genes induced by poly I:C and lacking a human ortholog (*i.e.*, Set 3) were mostly species-specific. However, a fraction of Set 3 genes had orthologs in rainbow trout, Atlantic salmon and zebrafish. Such genes comprised members of *gig2* and several class *IV trim* subsets, which have been subjected to considerable genomic expansion in several teleost species [48]. Based on zebrafish, *Tetraodon* and medaka, our previous data and others’ indicated that *ftr* and *btr trim* subsets are present in different fish families, in which they diversified independently and repeatedly [8,39]. Furthermore, zebrafish *fintrim* genes showed signatures of strong positive selection [49], as also reported for other antiviral *trim* groups such as mammalian *trim5* genes. Our work confirms that salmonids possess well defined and large groups of *trim25, btr* and *ftr*, with many genes responding to poly I:C and presumably type I IFN. Examining these expanded gene families will determine to what extent the large diversification observed in salmonids [48] occurred before or after the divergence of *Salmo* and *Oncorhynchus* and as such the evolution of these important antiviral genes.

4. Conclusions

In this work, we provided a comprehensive description of rainbow trout and Atlantic salmon gene families responsive to poly I:C in the

context of their phylogenetic relationships. This was made possible by the availability of high-quality reference genomes annotated in the frame of the AQUA-FAANG consortium (https://www.aqua-faang.eu/). Our work improves the annotation of the immune-responsive transcriptome of fishes, which is generally heavily reliant on functional knowledge available for mammals, built on assumptions that fish and mammalian orthologs have similar functions. In salmonids, the rich diversity of paralogs, which resulted from both tsWGD and ssWGD, often gets an inaccurate annotation based on a unique mouse or human counterpart. Our analysis of the sets of paralogs involved in virus-induced innate immunity, either conserved across vertebrates or fish-specific, provides a resource supporting in-depth genome-wide functional analysis in the future.

5. Materials and methods

5.1. Animal studies

Immature rainbow trout (~153 g) were raised in the freshwater fish facilities of Institut National de la Recherche en Agriculture et environnement (INRAE, Jouy en Josas, France). For rainbow trout, all fish experiments were carried out in accordance with the recommendations of the European Union guidelines for the handling of laboratory animals (http://ec.europa.eu/environment/chemicals/lab_animals/index_en.htm). The experimental protocols were approved by the INRAE Institutional Ethics Committee "Comethea" (permit license no. 15–60). Freshwater stage immature Atlantic salmon (~70 g) were obtained from the University of Stirling, Buckieburn hatchery and kept in the aquarium of the zoology building at the University of Aberdeen (UK). For Atlantic salmon, all procedures described were carried out in compliance with the Animals (Scientific Procedures) Act 1986 under UK Home Office license PPL number 70/8071 and approved by the ethics committee at the University of Aberdeen, UK. Despite the size difference between species, both were sexually immature at the time of sampling. All fish were healthy and monitored throughout the study.

5.2. Immunological stimulations in vivo

Poly I:C (Sigma #P1530) was diluted to 5 mg/ml in PBS. Before use, an aliquot was heated to 55 °C for 15 mins and then allowed to cool to room temperature for 20 mins. Fish ($n = 6$) were injected intraperitoneally (IP) with 100 μl of either 1× PBS or poly I:C (500 μg per fish). Following IP injections, PIT (passive integrated transponder) tagged rainbow trout were kept for 24 h in one 300 l tank supplied with recirculating dechlorinated water with a flow rate of 1000 L/h, at 10 °C, with a photoperiod of 10:14 light:dark. Atlantic salmon were kept for 24 h in one400 l tank with a flow rate of 1000 l/h in fresh water at 12 °C and a photoperiod of 12:12 light:dark. A computerised control system was used to monitor pH, ammonia concentration and oxygen levels over the duration of the stimulation in both species. Fish were euthanised by over exposure to anaesthetic (MS-222 [E10521 Merck, Sigma-Aldrich] at a final concentration of 50 mg/l and buffered to a pH of 7 for rainbow trout, and 2-phenoxyethanol [77,699 Merck, Sigma-Aldrich] at a concentration of 1.25 ml (neat) per l of water for Atlantic salmon), followed by destruction of the brain using a scalpel. Head kidney tissue was sampled and then either flash frozen in liquid nitrogen for rainbow trout or stored in RNA later for Atlantic salmon before storage at −80 °C.

5.3. Immunological stimulations in vitro

Six healthy unstimulated Atlantic salmon and rainbow trout from the same populations were euthanised as described above before aseptic extraction of the full head kidney tissue. The head kidney from each fish was placed in a falcon tube filled with 20 ml of extraction media (L15 [Gibco], 2% FBS and 0.02% EDTA) and kept on ice. Head kidney tissue was then placed onto a 100 μm cell strainer and cells were dissociated from tissue by gentle mashing through the strainer and pipetting with a further 20 ml of extraction media. To isolate leucocytes, the total cell suspension (20 ml) was layered onto a 51% percoll gradient (10 ml) and centrifuged for 30 mins at 400*g* and 4 °C with no brake on. Cells at the interface were carefully pipetted into a fresh 50 ml tube and washed twice with extraction media, re-pelleting at 400 g and 4 °C for 10 mins. Cell viability was assessed using Trypan blue staining and cell counts were done using a Malassez counting chamber before dilution to final counts of 1×10^6 cells per ml. Cells from each fish were then plated into 6 well plates (1 well per fish per treatment) at a density of 2×10^6 cells per well in 2 ml growth media (GMEM, 10% FBS and 0.1% pen/strep). One hundred micrograms poly I:C (50 μg per ml) was then added to the wells, while control cells were left untouched, and plates were then left in the incubator at 20 °C for 24 h in both species. Following cell stimulations, media was extracted from the stimulated ($n = 6$) and control ($n = 6$) wells and centrifuged at 500*g* for 5 mins in a 2 ml Eppendorf tube to collect cells in suspension. One ml of TRIzol (Thermo Fisher Scientific) was then added to each well and a cell scraper was used to detach remaining cells, before combining with the pelleted cells collected from supernatants. Samples were then frozen at -80 °C before further processing.

5.4. Sample processing and sequencing

For *in vivo*, head kidney tissue was homogenised with ceramic beads in a FastPrep-24 5 G tissuelyser in 1 ml of Trizol, while *in vitro* cells were already frozen in TRIzol (see above). Following homogenisation, total RNA was extracted from both *in vivo* and *in vitro* studies following the manufacturer instructions for a standard TRIzol total RNA extraction. Concentration and purity of the RNA was estimated using a Nanodrop 2000C Spectrophotometer, alongside further confirmation on an Agilent Bioanalyser 2100 (Agilent) to generate RNA integrity (RIN) values. Libraries were sequenced using the Illumina NovaSeq 6000 platform to generate around 30 million paired end 150 bp reads per sample. Raw reads were deposited in the European nucleotide archive (EBI) and are accessible in the FAANG data coordination centre under accession numbers (PRJEB50076 and PRJEB49984) for Atlantic salmon and rainbow trout, respectively.

5.5. Annotation of Ensembl rapid release genomes of Atlantic salmon and rainbow trout

Annotation of gene models predicted in the latest Atlantic salmon (GCA_905237065.2; [45]) and rainbow trout (GCA_013265735.3; [17]) genome assemblies downloaded from the Ensembl rapid release portal (prior to release of Ensembl 106, which provided full annotations for the same gene models) was performed using a blast-based approach to generate a list of best matching hits against various databases. Proteomes from each species were retrieved and filtered to retain only the longest isoform for each gene. Each filtered protein sequence was subjected to blastp using Diamond v2.0.9.147 [11] against the Atlantic salmon and rainbow trout proteomes downloaded from previously released genome assembly versions, available on Ensembl release 104 (GCA_000233375.4 for Atlantic salmon; GCA_002163495.1 for rainbow trout). All parameters were set to default except setting –max-target-seqs to 1 and –outfmt to 6. Further, all the query protein sequences were subjected to a blastp search against the human (GCA_000001405.28) and zebrafish (GCA_000002035.4) peptide sequences retrieved from Ensembl release 104. Finally, the current Ensembl annotation (release 108) was extracted using Biomart at the end of the project and added to Tables S1 and S2.

5.6. Mapping and analysis of transcriptome data

Sequencing data from both species were mapped against Atlantic salmon (GCA_905237065.2) [45] and rainbow trout

(GCA_013265735.3) [17] reference genomes. All steps of the analysis were done using the same nf-core pipeline RNA-seq version 3.6 [15] with default parameters. Therein, mapping of sequencing reads was carried out using STAR [12] and then RSEM [29] was used to quantify the mapped reads. The raw counts were then imported into R v4.1.1 before analysis using DESeq2 [32]. Count data was log transformed using the rlog function in DESeq2 to regularise the sequences in order to conduct pre-processing, quality checks and identify any potential outliers through graphical analysis such as PCA plots, clustering and heatmaps. Differentially expressed genes (DEGs) were identified using the Wald test in DESeq2 and were considered significant at a Benjamini–Hochberg False Discovery Rate (FDR) adjusted p-value <0.05 and a fold change >2 or < 0.5 representing a doubling or halving of gene expression, respectively. Analyses of rainbow trout and Atlantic salmon data were performed in parallel using an identical pipeline.

5.7. Gene ontology analysis

Gene ontology analysis was performed using DAVID (PMID: 35325185) with the human genome as a reference. Lists of official gene symbols corresponding to rainbow trout and Atlantic salmon differentially expressed genes were generated from Ensembl using Biomart, and visually curated. Up- and down-regulated genes were mapped on KEGG_Pathways and GO ontology (GOterm_BP, GOterm_CC and GO_term_MF). GO enrichment was performed by DAVID, and p values adjusted with Bonferroni or Benjamini correction. The same lists of official gene symbol were also analysed with Ingenuity Pathway Analysis (IPA, QIAGEN).

5.8. Phylogenetic analyses

Global phylogenetic relationships between paralogs and gene family members were analysed based on ([27] and references therein), and phylogenetic relationships (gene trees) available on Ensembl. For conserved ISGs, phylogenetic trees extracted from Ensembl (release 106, April 2022) were also verified using synteny-based corrected data produced by SCORPiOs [37]. SCORPiOs reconciles species trees and sequence-based gene trees while accounting for the local syntenic landscape of paralogs stemming from WGD events. Finally, compatibility with the current annotation of rainbow trout and Atlantic salmon genomes in Ensembl release 108 was checked.

Additional phylogenetic reconstruction was performed for targeted gene families (*gig2, fintrim, btr, trim25*). For each gene, the protein sequence of the longest isoform was used in the analysis. Ensembl gene ID are shown as tip labels of trees. The evolutionary history was inferred using MegaX [44] using the Maximum Likelihood (ML) method, with parameters indicated in figure legends (Figs. 5, S2). The trees were edited using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/Figtree/).

CRediT authorship contribution statement

Thomas Clark: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Shahmir Naseer:** Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **Manu Kumar Gundappa:** Methodology, Data curation, Writing – original draft. **Audrey Laurent:** Conceptualization, Methodology, Investigation, Data curation, Supervision, Writing – original draft, Writing – review & editing. **Aline Perquis:** Methodology, Investigation, Data curation, Supervision, Writing – original draft. **Bertrand Collet:** Conceptualization, Methodology, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Daniel J. Macqueen:** Conceptualization, Methodology, Investigation, Data curation, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing. **Samuel A.M. Martin:** Conceptualization, Methodology, Investigation, Data curation, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing. **Pierre Boudinot:** Conceptualization, Methodology, Investigation, Data curation, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data at the Attach file step.

Acknowledgements

This work was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 817923 (AQUAFAANG), by ANR (ANR-21-CE35-0019, LipofishVac), by the Eranet ICRAD-Nucnanofish (ANR_13001498 and BBSRC_ICRAD BB/V019902/1), by the BBSRC Institutional strategic programme award (BBS/E/D/20002174) and by institutional grants from INRAE. We thank Drs Hugues Roest Crollius, Camille Berthelot, Alexandra Louis and Elise Parey for sharing synteny-based corrected data produced by SCORPiOs, and Louis du Pasquier for insightful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110663.

References

[1] M. Aamelfot, O.B. Dale, K. Falk, Infectious salmon anaemia - pathogenesis and tropism, J. Fish Dis. 37 (2014) 291–307.
[2] D. Aggad, M. Mazel, P. Boudinot, K.E. Mogensen, O.J. Hamming, R. Hartmann, S. Kotenko, P. Herbomel, G. Lutfalla, J.P. Levraud, The two groups of zebrafish virus-induced interferons signal via distinct receptors with specific and shared chains, J. Immunol. 183 (2009) 3924–3931.
[3] L. Baillon, E. Mérour, J. Cabon, L. Louboutin, E. Vigouroux, A.L. Farias Alencar, L. Cuenca, Y. Blanchard, N.J. Olesen, V. Panzarin, T. Morin, M. Brémont, S. Biacchesi, The viral hemorrhagic septicemia virus (VHSV) markers of virulence in rainbow trout (*Oncorhynchus mykiss*), Front. Microbiol. 11 (2020), 574231.
[4] K.M. Balla, M.C. Rice, J.A. Gagnon, N.C. Elde, Linking virus discovery to immune responses visualized during zebrafish infections, Curr. Biol. 30 (2020) 2092–2103.
[5] J.W. Bledsoe, J. Ma, K. Cain, T.J. Bruce, A. Rawles, J. Abernathy, T. Welker, K. Overturf, Multi-tissue RNAseq reveals genetic and temporal differences in acute response to viral (IHNV) infection among three selected lines of rainbow trout with varying resistance, Fish Shellfish Immunol. 124 (2022) 343–361.
[6] A.R. Bluyssen, J.E. Durbin, D.E. Levy, ISGF3 gamma p48, a specificity switch for interferon activated transcription factors, Cytokine Growth Factor Rev. 7 (1996) 11–17.
[7] P. Boudinot, S. Bird, L. Du Pasquier, B. Collet, The repertoire of vertebrate STAT transcription factors: origin and variations in fish, Dev. Comp. Immunol. 116 (2021), 103929.
[8] P. Boudinot, L.M. van der Aa, L. Jouneau, L. Du Pasquier, P. Pontarotti, V. Briolat, A. Benmansour, J.P. Levraud, Origin and evolution of TRIM proteins: new insights from the complete TRIM repertoire of zebrafish and pufferfish, PLoS One 6 (2011), e22022.
[9] P. Boudinot, J. Zou, T. Ota, F. Buonocore, G. Scapigliati, A. Canapa, J. Cannon, G. Litman, J.D. Hansen, A tetrapod-like repertoire of innate immune receptors and effectors for coelacanths, J. Exp. Zool. B Mol. Dev. Evol. 322 (2014) 415–437.
[10] V. Briolat, L. Jouneau, R. Carvalho, N. Palha, C. Langevin, P. Herbomel, O. Schwartz, H.P. Spaink, J.P. Levraud, P. Boudinot, Contrasted innate responses to two viruses in zebrafish: insights into the ancestral repertoire of vertebrate IFN-stimulated genes, J. Immunol. 192 (2014) 4328–4341.
[11] B. Buchfink, K. Reuter, H.G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND, Nat. Methods 18 (2021) 366–368.
[12] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[13] C.P. Dopazo, The infectious pancreatic necrosis virus (IPNV) and its virulence determinants: what is known and what should be known, Pathogens 9 (2020) 94.
[14] K. Einer-Jensen, L. Delgado, E. Lorenzen, G. Bovo, Ø. Evensen, S. Lapatra, N. Lorenzen, Dual DNA vaccination of rainbow trout (*Oncorhynchus mykiss*) against

two different rhabdoviruses, VHSV and IHNV, induces specific divalent protection, Vaccine 27 (2009) 1248–1253.
[15] P.A. Ewels, A. Peltzer, S. Fillinger, H. Patel, J. Alneberg, A. Wilm, M.U. Garcia, P. Di Tommaso, S. Nahnsen, The nf-core framework for community-curated bioinformatics pipelines, Nat. Biotechnol. 38 (2020) 276–278.
[16] A.R. Fehr, S.A. Singh, C.M. Kerr, S. Mukai, H. Higashi, M. Aikawa, The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions, Genes Dev. 34 (2020) 341–359.
[17] G. Gao, S. Magadan, G.C. Waldbieser, R.C. Youngblood, P.A. Wheeler, B. E. Scheffler, G.H. Thorgaard, Y. Palti, A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout, G3 (Bethesda) 11 (2021) jkab052.
[18] M. Gilliet, W. Cao, Y.J. Liu, Plasmacytoid dendritic cells: sensing nucleic acids in viral infection and autoimmune diseases, Nat. Rev. Immunol. 8 (2008) 594–606.
[19] O. Gervais, A. Papadopoulou, R. Gratacap, B. Hillestad, A.E. Tinch, S.A.M. Martin, R.D. Houston, D. Robledo, Transcriptomic response to ISAV infection in the gills, head kidney and spleen of resistant and susceptible Atlantic salmon, BMC Genomics 23 (2022) 775.
[20] M.K. Gundappa, T.H. To, L. Gronvold, S.A.M. Martin, S. Lien, J. Geist, D. Hazlerigg, S.R. Sandve, D.J. Macqueen, Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of salmonid evolution, Mol. Biol. Evol. 39 (2022) msab310.
[21] O.J. Hamming, G. Lutfalla, J.P. Levraud, R. Hartmann, Crystal structure of zebrafish interferons I and II reveals conservation of type I interferon structure in vertebrates, J. Virol. 85 (2011) 8181–8187.
[22] O. Haugland, A. Mikalsen, P. Nilsen, K. Lindmo, B.J. Thu, T.M. Eliassen, N. Roos, M. Rode, O. Evensen, Cardiomyopathy syndrome of Atlantic salmon (*Salmo salar* L.) is caused by a double-stranded RNA virus of the totiviridae family, J. Virol. 85 (2011) 5275–5286.
[23] T.K. Herath, H.W. Ferguson, M.W. Weidmann, J.E. Bron, K.D. Thompson, A. Adams, K.F. Muir, R.H. Richards, Pathogenesis of experimental salmonid alphavirus infection in vivo: an ultrastructural insight, Vet. Res. 47 (2016) 7.
[24] A. Krasnov, L.H. Johansen, C. Karlsen, L. Sveen, E. Ytteborg, G. Timmerhaus, C. C. Lazado, S. Afanasyev, Transcriptome responses of Atlantic Salmon (*Salmo salar* L.) to viral and bacterial pathogens, inflammation, and stress, Front. Immunol. 12 (2021), 705601.
[25] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[26] C. Langevin, J.P. Levraud, P. Boudinot, Fish antiviral tripartite motif (TRIM) proteins, Fish Shellfish Immunol. 86 (2019) 724–733.
[27] J.-P. Levraud, L. Jouneau, V. Briolat, V. Laghi, P. Boudinot, Interferon-stimulated genes in zebrafish and human define an ancient arsenal of antiviral immunity, J. Immunol. 203 (2019) 3361–3373.
[28] J.P. Levraud, P. Boudinot, I. Colin, A. Benmansour, N. Peyrieras, P. Herbomel, G. Lutfalla, Identification of the zebrafish IFN receptor: implications for the origin of the vertebrate IFN system, J. Immunol. 178 (2007) 4385–4394.
[29] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinform. 12 (2011) 323.
[30] S. Lien, B.F. Koop, S.R. Sandve, J.R. Miller, M.P. Kent, T. Nome, T.R. Hvidsten, J. S. Leong, D.R. Minkley, A. Zimin, et al., The Atlantic salmon genome provides insights into rediploidization, Nature 533 (2016) 200–205.
[31] F. Liu, T. Wang, J. Petit, M. Forlenza, X. Chen, L. Chen, J. Zou, C.J. Secombes, Evolution of IFN subgroups in bony fish - 2. Analysis of subgroup appearance and expansion in teleost fish with a focus on salmonids, Fish Shellfish Immunol. 98 (2020) 564–573.
[32] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[33] D.J. Macqueen, I.A. Johnston, A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification, Proc. Biol. Sci. 281 (2014) 20132881.
[34] K.E. Magor, D. Miranzo Navarro, M.R. Barber, K. Petkau, X. Fleming-Canepa, G. A. Blyth, A.H. Blaine, Defense genes missing from the flight division, Dev. Comp. Immunol. 41 (2013) 377–388.
[35] F. McNab, K. Mayer-Barber, A. Sher, A. Wack, A. O'Garra, Type I interferons in infectious disease, Nat. Rev. Immunol. 15 (2015) 87–103.
[36] K. Ozato, D.M. Shin, T.H. Chang, H.C. Morse 3rd., TRIM family proteins and their emerging roles in innate immunity, Nat. Rev. Immunol. 8 (2008) 849–860.
[37] E. Parey, A. Louis, C. Cabau, Y. Guiguen, H. Roest Crollius, C. Berthelot, Synteny-guided resolution of gene trees clarifies the functional impact of whole-genome duplications, Mol. Biol. Evol. 37 (2020) 3324–3337.
[38] B. Robertsen, The role of type I interferons in innate and adaptive immunity against viruses in Atlantic salmon, Dev. Comp. Immunol. 80 (2018) 41–52.
[39] M. Sardiello, S. Cairo, B. Fontanella, A. Ballabio, G. Meroni, Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties, BMC Evol. Biol. 22 (2008) 1–22.
[40] J.W. Schoggins, D.A. MacDuff, N. Imanaka, M.D. Gainey, B. Shrestha, J.L. Eitson, K.B. Mar, R.B. Richardson, A.V. Ratushny, V. Litvak, et al., Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity, Nature 505 (2014) 691–695.
[41] J.W. Schoggins, S.J. Wilson, M. Panis, M.Y. Murphy, C.T. Jones, P. Bieniasz, C. M. Rice, A diverse range of gene products are effectors of the type I interferon antiviral response, Nature 472 (2011) 481–485.
[42] J.W. Schoggins, S.J. Wilson, M. Panis, M.Y. Murphy, C.T. Jones, P. Bieniasz, C. M. Rice, A diverse range of gene products are effectors of the type I interferon antiviral response, Nature 472 (2011) 481–485.
[43] C.J. Secombes, J. Zou, Evolution of interferons and interferon receptors, Front. Immunol. 8 (2017) 209.
[44] G. Stecher, K. Tamura, S. Kumar, Molecular evolutionary genetics analysis (MEGA) for macOS, Mol. Biol. Evol. 37 (2020) 1237–1239.
[45] K. Stenlokk, M. Saitou, L. Rud-Johansen, T. Nome, M. Moser, M. Arnyasi, M. Kent, N.J. Barson, S. Lien, The emergence of supergenes from inversions in Atlantic salmon, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 377 (2022) 20210195.
[46] M. Stremlau, C.M. Owens, M. Perron, M. Kiessling, P. Autissier, J. Sodroski, The cytoplasmic body component TRIM5α restricts HIV-1 infection in old world monkeys, Nature 427 (2004) 848–853.
[47] B. Sun, B. Robertsen, Z. Wang, B. Liu, Identification of an Atlantic salmon IFN multigene cluster encoding three IFN subtypes with very different expression properties, Dev. Comp. Immunol. 33 (2009) 547–558.
[48] J. Suurvali, C.J. Garroway, P. Boudinot, Recurrent expansions of B30.2-associated immune receptor families in fish, Immunogenetics 74 (2022) 129–147.
[49] L. van der Aa, J.P. Levraud, M. Yahmi, E. Lauret, V. Briolat, P. Herbomel, A. Benmansour, P. Boudinot, A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish, BMC Biol. 2009 (7) (2009) 7.
[50] Y.B. Zhang, T.K. Liu, J. Jiang, J. Shi, Y. Liu, S. Li, J.F. Gui, Identification of a novel Gig2 gene family specific to non-amniote vertebrates, PLoS One 8 (2013), e60588.
[51] J. Zou, B. Gorgoglione, N.G. Taylor, T. Summathed, P.T. Lee, A. Panigrahi, C. Genet, Y.M. Chen, T.Y. Chen, M. Ul Hassan, et al., Salmonids have an extraordinary complex type I IFN system: characterization of the IFN locus in rainbow trout *Oncorhynchus mykiss* reveals two novel IFN subgroups, J. Immunol. 193 (2014) 2273–2286.