



HAL
open science

Profiling the T Cell Receptor Alpha/Delta Locus in Salmonids

Eva Stina Isabella Edholm, Christopher Graham Fenton, Stanislas Mondot, Ruth Hracky Paulssen, Marie Paule Lefranc, Pierre Boudinot, Susana Magadan

► **To cite this version:**

Eva Stina Isabella Edholm, Christopher Graham Fenton, Stanislas Mondot, Ruth Hracky Paulssen, Marie Paule Lefranc, et al.. Profiling the T Cell Receptor Alpha/Delta Locus in Salmonids. *Frontiers in Immunology*, 2021, 12, 10.3389/fimmu.2021.753960 . hal-03434949

HAL Id: hal-03434949

<https://hal.science/hal-03434949v1>

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Profiling the T Cell Receptor Alpha/Delta Locus in Salmonids

Eva-Stina Edholm^{1*}, Christopher Graham Fenton², Stanislas Mondot^{3,4}, Ruth H. Paulssen², Marie-Paule Lefranc⁵, Pierre Boudinot⁶ and Susana Magadan^{7,8*}

¹ Faculty of Biosciences, Fisheries & Economics, Norwegian College of Fishery Science, University of Tromsø—The Arctic University of Norway, Tromsø, Norway, ² Clinical Bioinformatics Research Group, Genomics Support Centre Tromsø (GSCT), Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø - The Arctic University of Norway, Tromsø, Norway, ³ Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France, ⁴ GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, ⁵ IMGT®, The International ImMunoGeneTics Information System (IMGT), Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), CNRS, University of Montpellier, Montpellier Cedex, France, ⁶ Université Paris Saclay, INRAE, UVSQ, Virologie et Immunologie Moléculaires, Jouy-en-Josas, France, ⁷ Immunology Laboratory, Biomedical Research Center (CINBIO), University of Vigo, Vigo, Spain, ⁸ Galicia Sur Health Research Institute (IIS-GS), Hospital Alvaro Cunqueiro, Vigo, Spain

OPEN ACCESS

Edited by:

Robert David Miller,
University of New Mexico,
United States

Reviewed by:

M. Carla Piazzon,
Torre de la Sal Aquaculture Institute
(IATS), Spain
Lars Torkel Hellman,
Uppsala University, Sweden

*Correspondence:

Susana Magadan
smaga@uvigo.es
Eva-Stina Edholm
eva-stina.i.edholm@uit.no

Specialty section:

This article was submitted to
Comparative Immunology,
a section of the journal
Frontiers in Immunology

Received: 05 August 2021

Accepted: 28 September 2021

Published: 18 October 2021

Citation:

Edholm E-S, Fenton CG, Mondot S,
Paulssen RH, Lefranc M-P,
Boudinot P and Magadan S (2021)
Profiling the T Cell Receptor Alpha/
Delta Locus in Salmonids.
Front. Immunol. 12:753960.
doi: 10.3389/fimmu.2021.753960

In jawed vertebrates, two major T cell populations have been characterized. They are defined as α/β or γ/δ T cells, based on the expressed T cell receptor. Salmonids (family *Salmonidae*) include two key teleost species for aquaculture, rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) which constitute important models for fish immunology and important targets for vaccine development. The growing interest to decipher the dynamics of adaptive immune responses against pathogens or vaccines has resulted in recent efforts to sequence the immunoglobulin (IG) or antibodies and T cell receptor (TR) repertoire in these species. In this context, establishing a comprehensive and coherent locus annotation is the fundamental basis for the analysis of high-throughput repertoire sequencing data. We therefore decided to revisit the description and annotation of TRA/TRD locus in Atlantic salmon and two strains of rainbow trout (Swanson and Arlee) using the now available high-quality genome assemblies. Phylogenetic analysis of functional TRA/TRD V genes from these three genomes led to the definition of 25 subgroups shared by both species, some with particular feature. A total of 128 TRAJ genes were identified in *Salmo*, the majority with a close counterpart in *Oncorhynchus*. Analysis of expressed TRA repertoire indicates that most TRAV gene subgroups are expressed at mucosal and systemic level. The present work on TRA/TRD locus annotation along with the analysis of TRA repertoire sequencing data show the feasibility and advantages of a common salmonid TRA/TRD nomenclature that allows an accurate annotation and analysis of high-throughput sequencing results, across salmonid T cell subsets.

Keywords: T cell receptor, repertoire, salmonid fish, VDJ annotation, TRA/TRD locus, gene rearrangement, adaptive immunity

INTRODUCTION

The emergence of vertebrates was accompanied by morphological and functional innovations, such as the development of an adaptive immune system (1, 2). Vertebrate adaptive immunity relies on the clonal expression of somatically diversifying antigen receptors on lymphocytes, which generates diversity and antigen specific recognition far beyond those offered by the allelic variation or alternative splicing processes. In jawed vertebrates, including jawed fish, the adaptive immune components include B and T lymphocytes, which express specific antigen receptors, the immunoglobulins (IG) or antibodies and T cell receptors (TR), respectively (3–5).

TR can be formed by two distinct polypeptide chains, α/β or γ/δ , which distinguish the two major T cell populations across jawed vertebrates (4, 6–8). The germline structure of TR loci provides the basic pieces from which genes encoding these receptors are assembled to generate a vast repertoire of T cells bearing structurally diverse receptors (potentially $>10^{13}$ sequences in humans) for specific antigen recognition (4, 6). During T cell development, TR genes, called variable (V), diversity (D), and joining (J) undergo somatic rearrangement mediated by RAG1/2 and other enzymes to encode the different chains of the receptor. The V-(D)-J combinations made from the germline gene pool generate the wide diversity of TR chains required for antigen recognition.

TR locus organization has been analyzed in many mammalian species (4, 9–15) but the data available for teleost and lower ectothermic vertebrates are much more limited (9, 11, 16, 17). The TRB locus has a common genomic structure in several mammalian species, with different numbers of TRBV genes positioned upstream of tandem-aligned TRBD-J-C clusters, each composed of a single D (TRBD) gene, several J (TRBJ) genes, and one C (TRBC) gene (4, 10, 13, 18, 19). The TRG locus shows greater gene organization plasticity. Generally, the TRG genes comprise an array of TRGV genes linked to TRGJ and TRGC genes organized in J-C clusters (4, 9, 11). In all studied vertebrates including several teleost species, the TRA and TRD genes present a conserved linkage in a unique TRA/TRD locus (4, 15, 20–23). The TRA/TRD locus described in the amphibian *Xenopus laevis* is highly conserved among tetrapods (15, 24). The TRD locus is imbedded within the TRA locus with the following arrangement: TRA/TRD V-TRDD-TRDJ-TRDC-TRAJ-TRAC, with some V genes being used in the synthesis of both α and δ TR chains (9, 11, 21, 24). Furthermore, the presence of IGHV-related genes within the TRA/TRD locus, which are expressed exclusively with δ TR chains, has been also reported in different tetrapods, including *Xenopus* and the Chinese alligator (24, 25). The organization of the TRA/TRD locus in teleosts, based on data from zebrafish (*Danio rerio*) (17), Japanese pufferfish (*Takifugu rubripes*) (26), green pufferfish (*Tetraodon nigroviridis*) (20) and Atlantic salmon (*Salmon salar*) (23) is different. On the chromosome in the Watson orientation (FWD5'>FWD3'), the multiple TRA/TRD V genes are followed by the TRAC-TRAJ-TRDC-TRDJ-TRDD cluster, which is in inverted orientation to the TRA/TRD V genes. Defining the locus orientation, by the transcriptional orientation of the TRDC and TRAC genes, the teleost TRA/TDR locus is in reverse orientation (REV) on the

chromosome, with the TRDD genes at the 5' end of the locus and the TRA/TRD V genes translocated downstream at the 3' end of the locus in opposite orientation (IMGT index > Genomic orientation, <http://www.imgt.org/IMGTindex/genomicOrientation.php>). TRA and/or TRD genes have been identified in other teleosts, such as channel catfish (*Ictalurus punctatus*) (16), mandarin fish (*Siniperca chuatsi*) (27), sea bass (*Dicentrarchus labrax*) (28), common carp (*Cyprinus carpio*) (29), and flounder (*Parlichthys olivaceus*) (30), but the organization of genomic locus has yet to be characterized in these species.

Salmonids (family *Salmonidae*) include two key species for aquaculture, rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) which constitute important models for fish immunology and important targets for vaccine development. While TRA and TRB chain cDNAs were initially cloned in rainbow trout (31, 32), a first description of the organization of the TRA/TRD locus was reported from Atlantic salmon by Yazawa et al. (23). On *Salmo salar* chromosome 14 genomic sequence, a total of 292 TRA/TRD V genes (128 functional) were identified upstream of the TRAC-TRAJ-TRDC-TRDJ-TRDD cluster, which is in inverted orientation to TRA/TRD V genes. One hundred and twenty two TRAJ genes (113 functional), one TRDJ and 3 TRDD genes were also found in the locus (23).

In this study, we re-examined the genomic sequence of the TRA/TRD locus in Atlantic salmon and rainbow trout and revisited its annotation to establish a universal salmonid nomenclature based on IMGT rules and standards (4, 33). Recent progress in sequencing technologies has led to much improved versions of genome assemblies for these species. In rainbow trout, homozygous clonal lines from two different strains have been sequenced. The first reference genome, the Omyk_1.0 assembly, was from the line “Swanson”, which was derived from a semi-wild population from the Swanson River in the Kenai Peninsula of Alaska (34). The second assembly (released 2020, USDA_OmykA_1.1) was generated from the Arlee clonal line (35), which derives from the “Arlee” strain. This domesticated hatchery strain was used by the Montana Department of Fish, Wildlife and Parks and is thought to have originally been collected from Northern California like most farmed rainbow trout stocks that were imported to Europe (Gary Thorgaard, Personal Communication). Focusing on functional sequences, phylogenetic analysis of TRA/TRD V genes from these three genomes led to the definition of 25 subgroups shared by *Salmo* and *Oncorhynchus*, and an additional single gene subgroup present only in *Oncorhynchus*. Maps from Atlantic salmon and rainbow trout assemblies were established and compared, which will facilitate updates of the nomenclature based on future sequence data. In addition, this work on TRA/TRD locus allows accurate annotation and analysis of high-throughput sequencing repertoire data, across salmonid T cell subsets.

MATERIAL AND METHODS

Annotation of TRA/TRD Loci

Atlantic salmon (ICSASG_v2; GenBank GCA_000233375.4, isolate Sally, female, breed: double haploid), rainbow trout Arlee (USDA_OmykA_1.1; GenBank GCA_013265735.3,

isolate Arlee, male) and Swanson (Omyk_1.0; GenBank GCA_002163495.1, isolate Swanson, male) genome assemblies were accessed through the NCBI website (<https://www.ncbi.nlm.nih.gov/>). To identify the chromosomes that embrace the TRA/TRD locus, previously published salmonid TRAC and TRDC sequences (23, 31, 32, 36, 37) were used for BLASTn and tBLASTn searches. The TRA/TRD locus was located on chromosome 14, in Atlantic salmon (CM003292.1, NC_027313.1, 93901823 bp), and on chromosome 8 in rainbow trout Arlee (CM023226.2, NC_048572.1, 91622588 bp) and Swanson (CM007942.1, NC_035084.1, 83778284 bp). These chromosomes were selected for in depth analysis. TR genes were searched using the BLAST tool at Galaxy website (<https://usegalaxy.org>) and visual analysis using SnapGene software (from Insightful Science; available at snapgene.com). Previously published TRA and TRD constant sequences from rainbow trout and Atlantic salmon (23, 31, 32, 36, 37) were used as queries to identify the chromosomal regions containing TRAC and TRDC genes, intron splice signals were identified for all sequences and were used to determine the limits of the TRAC and TRDC coding exons. To annotate TR variable V, joining J, and diversity D genes, the recombinant signal (RS) sequences were identified, respectively V-RS, J-RS, and 5'D-RS and 3'D-RS. Splice signals were used to determine the limits of the coding nucleotide sequences for the V (L-PART1 donor splice and L-PART2 acceptor splice) and the J (J-REGION donor splice). The annotation and functionality of TRA/TRD genes were established according to IMGT rules and standards (4, 33). In silico analysis of gene function considered the following parameters: a) presence of appropriate splice sites, b) presence of RS sequence compatible with effective rearrangement, c) open reading frames which included conserved cysteine and tryptophan codons at positions 23 (1st-CYS), 41 (CONSERVED-TRP) and 104 (2nd-CYS) of TRA/TRD V genes (IMGT unique numbering system) (32). Briefly, a germline entity (TRA or TRD V, TRDD, TRDJ or TRAJ gene) was considered functional (F) if the coding region has an open reading frame, no defect in splicing site, or in RS and presents key conserved amino acids; as open reading frame (ORF) if the coding region has an open reading frame but there are defects in the splicing sites or RS sequences, and/or changes of conserved amino acids that lead to incorrect folding; or annotated as pseudogene (P) if the coding region has stop codon or frameshift mutations (4, 33).

Nomenclature Rules

Salmonid TRA/TRD V genes were named based on nucleotide similarity, phylogenetic analysis and positional information. We followed the same principles as in our recent work on salmonid IGH loci (38, 39). The proposed TRAV names are constituted as follows: first, TRAV, then the subgroup number, then a first dash and a number defining the gene rank in the locus (4, 33). For example, the name TRAV2-2 denotes a gene belonging to the subgroup 2, located at the reference rank 2 within the locus. This rank was determined by the locus 5'-3' transcriptional orientation of the TRDD-TRDJ-TRDC-TRAJ-TRAC cluster. Additional numeric qualifiers could be added when relevant to

indicate the position of a gene in a microcluster, as described in (39). Accordingly, we first classified and named the Atlantic salmon TRAV genes. We then mapped rainbow trout TRA/TRD V genes to the corresponding subgroups based on a threshold off 75% nucleotide identity and named them based on positional information. This name format was chosen as it provides flexibility to integrate additional genes, and/or describe gaps while keeping the nomenclature consistent across haplotypes or genomes updates (39). There is no criterion to differentiate TRAV and TRDV a priori. Our choice of naming genes "TRAV" was made for the sake of coherence and standardization (according to the IMGT Scientific chart rules) and has been validated by the IUIS Nomenclature Committee. The expression of genes by TR alpha chains is documented by our repertoire sequencing data, and does not preclude of the expression by TR delta chain. The TRAJ genes were named following the same schedule proposed for human TRA/TRD locus (4, 40, 41), with TRAJ1 being the closest TRAJ to TRAC gene. This annotation also fits with the previous description given in Atlantic salmon by Yazawa et al. (23). Chromosomal coordinates given in the TR gene annotation files (**Supplementary Data 1**) include the recombinant signal (RS) sequence and coding regions of TRAV, TRAJ, TRDJ and TRDD genes, and refer to the genomic region that contains the 3 exons of the TRAC and TRDC genes (4).

Phylogenetic Analysis

Phylogenetic analysis was performed on functional TRAV sequences from Atlantic salmon and rainbow trout. Amino acid sequences were aligned using ClustalW in the MEGA-X software (42). The phylogenetic trees were then constructed using the Maximum Likelihood method in MEGA X, corrected using the Poisson model, and bootstrapped 500 times.

Samples and RNA Extraction for Sequencing of Atlantic Salmon TRA Transcripts

Head kidney (HK), spleen and gill were collected from three non-vaccinated Atlantic salmon, parramorphs, strain NLA reared in hatchery at the Kårvika Aquaculture Research station, Troms, Norway and confirmed free of the salmon pathogens ISAV, SAV, PRV and IPNV by RT-qPCR. They were kept in running freshwater at 10°C, exposed to continuous light and fed with commercial dry feed. Tissues were aseptically harvested, kept in RNALater at -20°C until use and total RNA was extracted using the RNeasy Mini Kit (Qiagen). All animal handling were performed in accordance with the recommendations of the current animal welfare regulations: FOR-1996-01-15-23 (Norway).

DNA Library Preparation and Sequencing

Reverse transcription (RT) was carried out using the SMARTer RACE 5'/3' kit (Clontech). Briefly, first strand synthesis was performed following the manufacturers protocol using 0.5 µg of total RNA per reaction. Then, double-stranded cDNA was obtained with PCR assay carried out using 2.5 µL of diluted RT product according to the manufacturers protocol using a

TRAC reverse (GSP1) primer (5'-GATTACGCCAAGCTTGGCAAGCTGTGGTATTGCTTGAGTTC-3'). Touchdown PCR cycling conditions were performed as follows: five cycles at 94°C for 30 s and 72°C for 2 min; five cycles at 94°C for 30 s, 70°C for 30 s, and 72°C for 2 min; 25 cycles at 94°C for 30 s, 68°C for 30 s, and 72°C for 2 min; and a final step at 72°C for 10 min. The resulting PCR reaction was run on a 1,2% agarose/EtBr gel and the resulting ~500bp single band was extracted and gel purified using the NucleoSpin Gel and PCR Clean-Up Kit (Clontech).

The amount of input DNA for library preparation was quantified using a Qubit-fluorometer 3.0 (Invitrogen). Libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) using the standard protocol. Each library was then quality checked with the Agilent 2100 Bioanalyzer. Prior to pooling and sequencing, libraries were quantified through PCR using the KAPA Library Quantification Kits, according to the manufacturer's instructions. Libraries were pooled (4nM) and were diluted to 1.8 pM and subsequently paired-end sequenced with an Illumina Next Seq550 platform and with 300 cycle run chemistry.

Raw Data Processing

All forward and reverse illumina reads were merged using the flash software (43) together to produce extended fasta reads. Only extended reads greater than 150 bp were considered. Using the seqkit software (44) all reads were relabeled with their respective sample name. The reads were then concatenated into one large fasta file. A blast database was constructed from the resultant sample labeled extended fragment file with the makeblastdb command from the ncbi blast executables. The C region sequence, J region sequences and V region sequences were then blasted against the locally created blastdb database. Only sequences with a 97% identity were considered as valid C, J, or V hits. The R data.table package was used to load the blast result files for C, J, and V. Viable blast ids met the following criteria; a C region greater than 30 bp, J region with an identity greater than 99.9%, V region with an identity greater than 99.9% and a length greater than or equal to 50 bp. Remaining V and J blast hits were then sorted by expectancy and any duplicates were removed. The remaining valid blast ids contain a C region, J region, and V region meeting the above criteria. The extended and labeled fasta sequence could be retrieved by blast_ids. Each sequence was translated in all six reading frames to define productive and unproductive sequences. To account for variation with respect to library size samples were normalized using the rlog function of the DESeq2 package to transform the count data to the log2 scale.

RESULTS

Salmonids TRA/TRD Genomic Organization: Comparison Between Atlantic Salmon and Rainbow Trout The Atlantic Salmon TRA/TRD Locus

The TRA/TRD locus of Atlantic salmon was manually annotated using the current genome assembly [ICSASG_v2; GenBank

GCA_000233375.4], as well as cDNA data available in GenBank and sequences previously reported by Yazawa et al. (23). The TRA/TRD locus is located on chromosome 14, in forward (Watson) orientation, defined from the FWD5' (short arm) telomeric end to the FWD3' (long arm) telomeric end of the chromosome (**Figure 1A**). It spans 1,14 Megabases (Mb) and includes, from FWD5' to FWD3', 239 TRAV, one TRAC, 128 TRAJ, one TRDC, a single TRDJ and three TRDD (**Figure 1A**). Defining the *Salmo salar* TRA/TRD locus orientation, by the transcriptional orientation of the TRAC and TRDC genes, the locus is in reverse orientation (REV) on chromosome 14, with the TRDD genes at the 5' end of the locus and the TRAV genes translocated downstream at the 3' end of the locus in opposite orientation. Thus, the Atlantic salmon TRA/TRD locus comprises from 5' (chr FWD3') to 3' (chr FWD5'), first the TRDD-TRDJ-TRDC-TRAJ-TRAC cluster followed by the TRA/TRD V genes translocated downstream in 3' in opposite orientation.

On the chromosomal FWD5' side of the region containing V genes, a first block of 29 TRAV genes is in inverted orientation. It is followed by a region > 300 kilobases (Kb), between the TRAV10-8 and the TRAV12-7 genes, where we could not identify any TRAV genes and there are no gaps (**Figure 1A**). The other 210 TRAV genes are in forward orientation on the chromosome. In conclusion, we identified a total of 239 TRAV genes (**Supplementary Data 1**), while Yazawa et al. (23) reported 292 TRA/DV genes.

The TRAJ-TRAC cluster spans about 52 Kb. The exon-intron organization of the TRAC gene was confirmed, comprising 3 exons that encode a protein of 110 amino acids (AA). A total of 128 TRAJ genes are found on the genomic region between the TRAC and TRDC genes (**Figures 1A, 2A**), i.e. 6 more genes than previously identified (23). The TRD genes (TRDD, TRDJ and TRDC) correspond to those previously characterized by Yazawa et al. (23), with three TRDD genes, one TRDJ and one TRDC gene that comprises 3 exons.

The Rainbow Trout TRA/TRD Locus

The analysis of genome assemblies from isogenic fish from the Swanson (Omyk_1.0; GCA_002163495.1) and Arlee (USDA_OmykA_1.1; GCA_013265735.3) strains resulted in a first description of the TRA/TRD locus in rainbow trout. In both genome assemblies the TRA/TRD genes are located on chromosome 8, with a similar FWD5' to FWD3' configuration (TRA/TRD V followed by a TRAC-TRAJ-TRDC-TRDJ-TRDD cluster comprising genes in inverted orientation) as described for Atlantic salmon (**Figure 1B** and **Supplementary Data 2**). Defining the *Oncorhynchus mykiss* TRA/TRD locus orientation, by the transcriptional orientation of the TRAC and TRDC genes, the trout locus on chromosome 8 is, like that of salmon on chromosome 14, in reverse orientation (REV), with the TRDD genes at the 5' end of the locus and the TRAV genes translocated downstream at the 3' end of the locus in opposite orientation. Thus, the *Oncorhynchus mykiss* TRA/TRD locus comprises from 5' (chr FWD3') to 3' (chr FWD5'), first the cluster TRDD-TRDJ-TRDC-TRAJ-TRAC cluster followed by the TRA/TRD V genes translocated downstream in 3' in opposite orientation as described for Atlantic salmon.

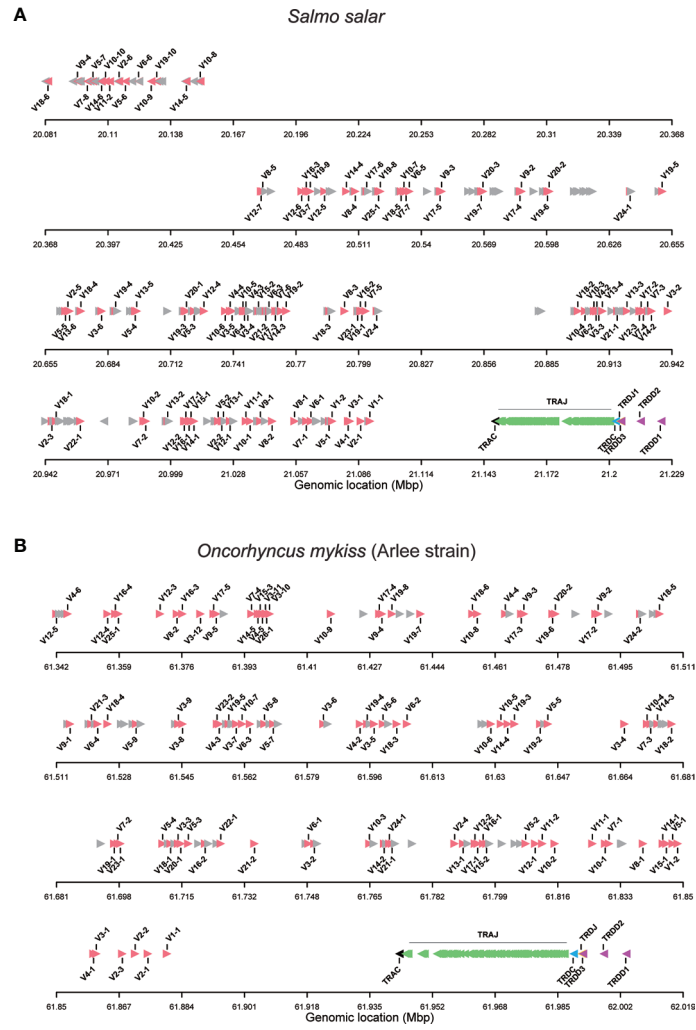


FIGURE 1 | Organization of TRA/TRD locus in salmonids. **(A)** Atlantic salmon (*Salmo salar*) and **(B)** rainbow trout (*Oncorhynchus mykiss*, Arlee strain). TRAV gene names are according to IMGT nomenclature (4, 33). Functional and ORF TRAV genes are in red. TRAV pseudogenes with frameshift(s) in the V-REGION are in grey and their name are not displayed. The arrow indicates the orientation. The symbols representing the genes are not to scale. Information about gaps (length and location) is available in **Supplementary Data 3**.

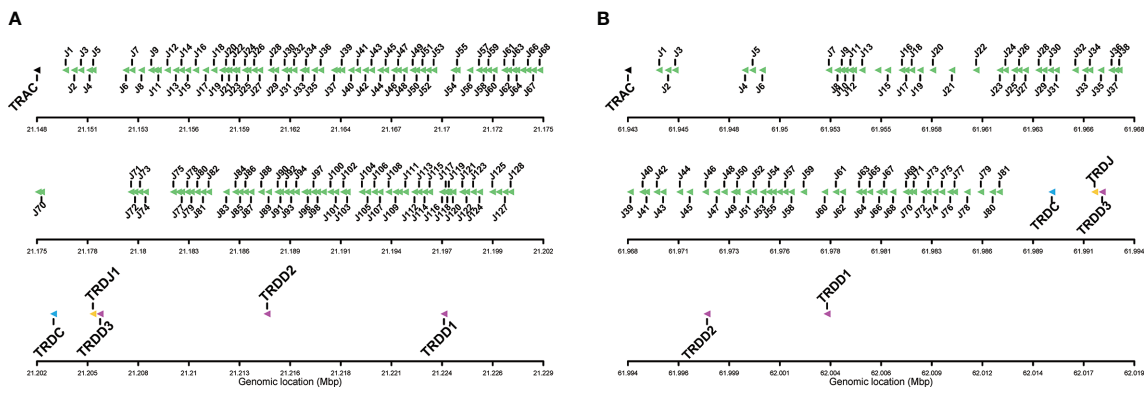


FIGURE 2 | Detailed view of the TRAJ cluster and the TRD genes, including TRDC, TRDJ and TRDD genes. **(A)** *Salmo salar* and **(B)** *Oncorhynchus mykiss*, Arlee line. The boxes representing the genes are not to scale.

Unlike what was identified in Atlantic salmon TRA/TRD locus, in rainbow trout all annotated TRAV genes are in the same opposite orientation (Figures 1A, B). The TRA/TRD locus spans a total length of approximately 825Kb in Omyk_1.0 genome (Swanson strain), compared to only 660Kb in the USDA_OmykA_1.1 genome (Arlee strain). Also, the number of annotated TRAV and TRAJ genes differs significantly between assemblies. Thus, 164 TRAV genes were found in Arlee for only 57 in Swanson (Supplementary Data 1), and 81 TRAJ genes were annotated in Arlee compared to 66 in Swanson (Supplementary Data 2 and Figure 2B). These discrepancies may be due to either problems in the genome assembly or real differences between the loci of the different rainbow trout strains. Further studies are required to elucidate this point. However, the presence of gaps in the Swanson TRA/TRD locus suggests that the assembly of this region may be incorrect (Figure 1B and Supplementary Data 2). TRAJ genes were classified in accordance with their position and their high nucleotide identity with their corresponding Atlantic salmon orthologues.

Salmonid TRAV Genes

A total of 239 TRAV genes were identified in the Atlantic salmon TRA/TRD locus. Of these, 104 TRAV are functional (F), 12 TRAV are ORF, and 123 TRAV are pseudogenes. The rainbow trout Arlee TRA/TRD locus comprises 163 TRAV genes: 93 TRAV are functional, 14 TRAV are ORF and 56 TRAV are

pseudogenes (Figure 1B and Supplementary Data 1), while in Swanson assembly only 57 TRAV genes could be annotated. Of these 25 TRAV are functional, 6 TRAV are ORF and 26 TRAV are pseudogenes (Supplementary Data 1 and Data 2).

Based on the percentage of identity between nucleotide sequences of the V-REGION (threshold 75%), Atlantic salmon F and ORF TRAV genes (116 TRAV sequences) could be classified into 25 TRAV gene subgroups (Table 1). All F and ORF TRAV genes identified in both lines of rainbow trout could be classified into the same 25 subgroups defined for the Atlantic salmon TRAV, except for one ORF TRAV gene (TRAV26) from the Arlee assembly, which represents a single gene subgroup (Table 1). To confirm the subgroup clustering deduced amino acid sequences of F and ORF TRAV genes were further analyzed. Both phylogenetic trees of functional TRAV genes (Figure 3 and Supplementary Data 4) and the pairwise distance matrix (Figure 4A and Supplementary Data 5) generated from F and ORF TRAV sequences support these 26 subgroups. Representative Atlantic salmon and rainbow trout (Arlee) amino acid sequences from each TRAV subgroup were aligned with sequences from the 75 zebrafish TRAV subgroups defined by Seeley et al. (17) and a phylogenetic tree was constructed using NJ method (Supplementary Data 6). We chose the zebrafish because it was the only other teleost species in which a comprehensive analysis of the TRA/DV repertoire had been reported and was available in IMGT, allowing to look for highly conserved genes.

TABLE 1 | Characterization of the TRAV subgroups (number of genes per functionality F and ORF and CDR-IMGT lengths) in Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) Arlee and Swanson strains.

IMGT TRAV sub-group	<i>Salmo salar</i> ¹	CDR-IMGT lengths	<i>Oncorhynchus mykiss</i> ¹ , Arlee	CDR-IMGT lengths	<i>Oncorhynchus mykiss</i> ¹ , Swanson	CDR-IMGT lengths
TRAV1	2F ²	[8.6.3]	2F	[8.6.3]		
TRAV2	6F	[7.6.3]	4F	[7.6.3]	3F	[7.6.3]
TRAV3	7F	[4 (5).0.3]	11F, 1ORF	[4 (5) (8).0.3]	1F	[4.0.3]
TRAV4	2F, 2ORF ³	[4.8.1 (2)]	5F, 1ORF	[4.7 (8).2 (3)]		
TRAV5	6F, 1ORF	[5.7.3]	9F	[5.7.2 (3)]	3ORF	[5.7.3]
TRAV6	5F, 1ORF	[4.8.2 (3)]	4F	[4.8.3]		
TRAV7	7F, 1ORF	[7 (8) (9).6.3]	3F, 1ORF	[7 (8).2 (6).3]	1F	[8.6.3]
TRAV8	4F, 1ORF	[7.10.3]	1F, 1ORF	[7.10.3]	1F	[7.10.3]
TRAV9	4F	[7.6.3]	5F	[7.6.3]	2F	[7.6.3]
TRAV10	8F, 2ORF	[7.6.3]	5F, 4ORF	[7.6.3 (10)]	2F, 3ORF	[6 (7).6.3]
TRAV11	2F	[7.6.3]	2F	[7.6.3]	1F	[7.6.3]
TRAV12	6F, 1ORF	[7 (8).6.3]	5F	[7.6.3]	2F	[7.6.3]
TRAV13	4F, 2ORF	[5.6.3]	1F	[5.6.3]		
TRAV14	4F, 1ORF	[4.3 (7).3]	5F	[4.7.3]	2F	[4.7.3]
TRAV15	2F	[5.7 (8).3]	2F, 1ORF	[5.8.3]	1F	[5.8.3]
TRAV16	3F	[6.10.3]	3F, 1ORF	[6.10.3 (4)]	2F	[6.10.3 (4)]
TRAV17	6F	[8.5.3]	5F	[8.5.3]	1F	[8.5.3]
TRAV18	6F	[7.4.2 (6)]	6F	[7.4.2 (5)]	1F	[7.4.5]
TRAV19	10F	[4.7.3]	6F, 2ORF	[4.7.3]	4F	[4.7.3]
TRAV20	3F	[6.3.3]	2F	[6.3.3]	1F	[6.3.3]
TRAV21	2F	[4.7.2]	3F	[7.4.2 (5)]		
TRAV22	1F	[7.3.11]	1ORF	[7.3.11]		
TRAV23	1F	[5.8.3]	2F	[5.8.3]		
TRAV24	1F	[5.2.3]	2F	[5.2.3]		
TRAV25	1F	[5.7.3]	1F	[5.8.3]		
TRAV26			1ORF			
Total	103F, 12 ORF		94F, 14 ORF		25F, 6ORF	

¹TRAV genes per subgroup ²F, functional; ³ORF, Open reading frame (as defined in the text).

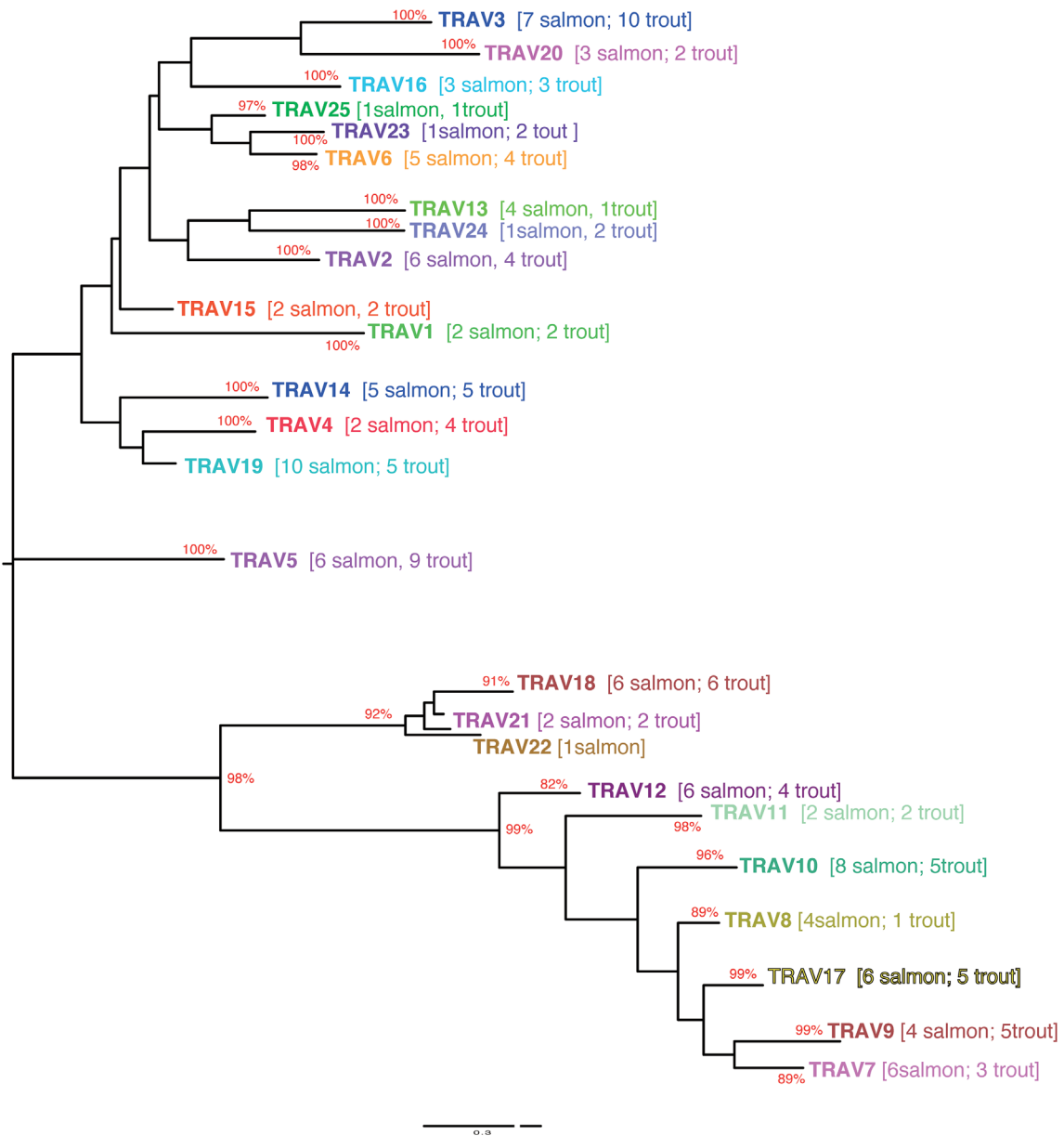


FIGURE 3 | Evolutionary tree of TRAV protein sequences from Atlantic salmon and rainbow trout. The tree was inferred using the Maximum Likelihood method and JTT matrix based model. A bootstrap of 500 replicates was used. The tree with highest likelihood is shown, and is drawn to scale, with branch lengths representing the number of substitutions per site. The percentage of trees in which the associated taxa cluster together is indicated when >80. Subgroups have been collapsed for the sake of readability. The number of Atlantic salmon and rainbow trout sequences in each subgroup is indicated between brackets (an extended tree is represented in **Supplementary Figure 4**). The analysis has been performed only on functional V sequences. Hence it comprises only 25 subgroups since subgroup 26 is only represented by an ORF sequence in the Arlee rainbow trout genome.

The structure of the tree which lacked stable branches with sequences from the three species, shows that subgroups defined in salmonids and in the zebrafish reflect completely different evolutionary pathways. This is in contrast to what is observed between different groups of mammals, with a number of TRAV subgroups conserved across primates, rodents, Cetartiodactils and carnivores (12). Understanding the evolution of TRA/DV genes would require an extensive phylogenetic work involving

sequences from multiple fish species, but this simple analysis indicates that there is apparently no highly conserved TRA/DV subsets between Salmonids and Cyprinids, two basal branches of teleosts. Comparing the Atlantic salmon and Arlee TRA/TRD locus, we did not find an obvious difference in the number of members belonging to each TRAV subgroup. In Atlantic salmon, four subgroups (TRAV22 to TRAV25) are represented by a single functional gene, while TRAV10 and TRAV19 subgroups

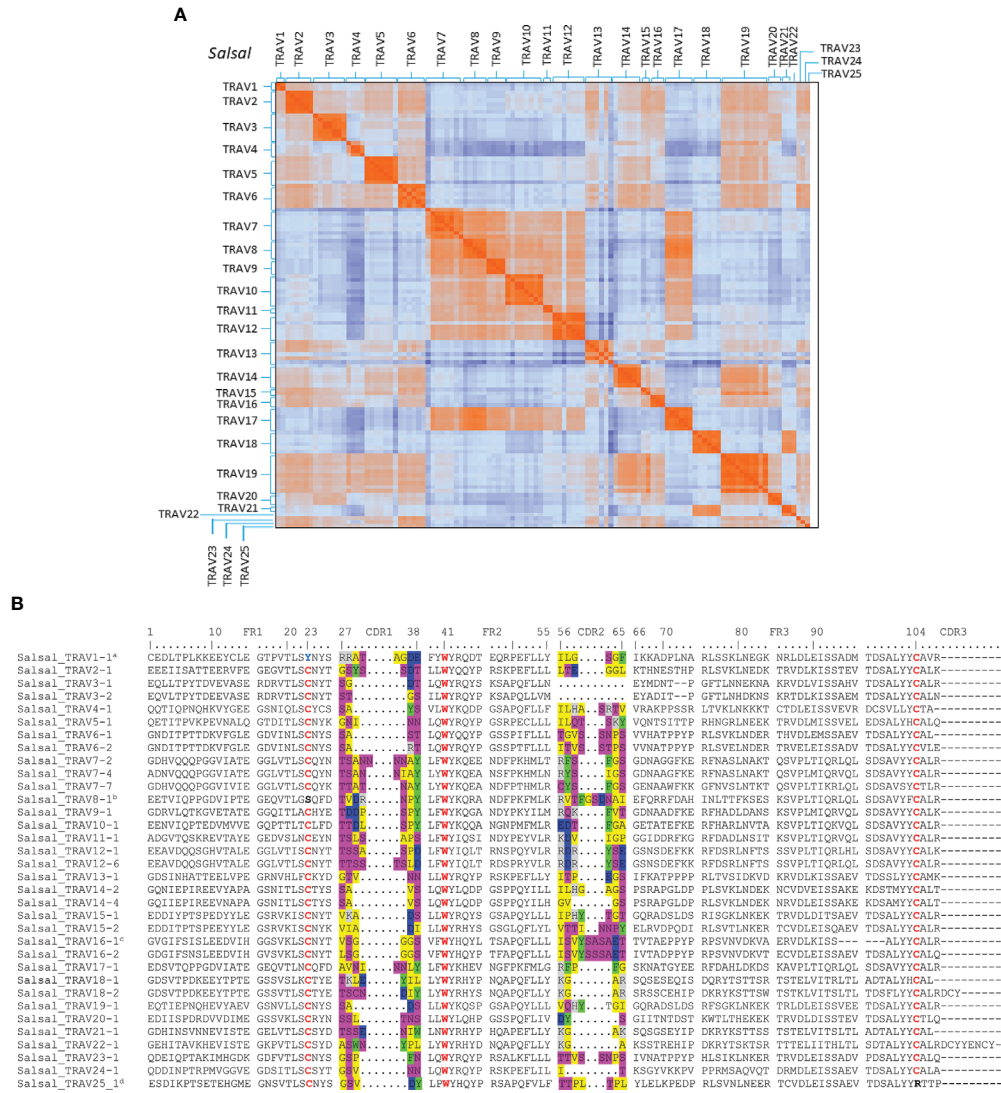


FIGURE 4 | Comparison of TRAV sequences from Atlantic salmon. **(A)** Heatmap representation of pairwise distance matrix of functional and ORF TRAV. Pairwise distances were estimated from amino acid alignments using the Poisson model. 103 functional TRAV and 12 ORF (TRAV4-2, TRAV4-4, TRAV5-7, TRAV6-3, TRAV7-1, TRAV8-5, TRAV10-3, TRAV10-5, TRAV12-3, TRAV13-4, TRAV13-6 and TRAV14-1) were included in the analysis. Color in each pixel indicate pairwise distance among and between each TRAV sequence. **(B)** Multiple Alignment of representative TRAV subgroups for Atlantic salmon ("Salsal" in the IMGT 6-letter abbreviation for genus and species). The CDR1 and CDR2 encoded amino acids are colored as follows: *gray* positively charged R groups, *blue* negatively charged R groups, *green* aromatic R groups, *pink* polar uncharged R groups, *yellow* nonpolar R groups, *red* conserved cysteines (1st-CYS 23 and 2nd-CYS 104) and conserved tryptophan 41. ^aConserved 1st-CYS 23 is replaced by a Tyrosine (Y). ^bConserved 1st-CYS 23 is replaced by a Serine (S). ^csix amino-acid deletion in FR3 compared to the other genes of the subgroup. ^dConserved 2nd-CYS 104 is replaced by a Arg.

each contain 10 genes. In rainbow trout, the 25 TRAV subgroups comprise one (TRAV1, TRAV13, TRAV22, TRAV25) to twelve (TRAV3) members. The most striking differences are found in the TRAV3 subgroup, with 7 TRAV genes in Atlantic salmon and 12 TRAV genes in Arlee, and the TRAV13 subgroup, with 6 members in Atlantic salmon and 1 in Arlee (Table 1). In Swanson genome assembly (Omyk_1.0), we only identified TRAV genes belonging to 16 subgroups (TRAV2, 3, 5, 7 to 12 and 14 to 20). All of them could be easily matched to a unique counterpart in Arlee (*i.e.*, with a nucleotide sequence >90%

similar). The distribution of TRAV genes in Arlee and Swanson was very similar, and functional genes were found in the same order. There was only one exception, the duplicated TRAV10-8D gene found in Swanson was lacking in the Arlee genome assembly (Figure 1B and Supplementary Data 2).

Multiple alignments of amino acid sequences of representative members of each TRAV subgroup are shown in Figure 4B and Supplementary Data 5. In most subgroups sequences show typical framework regions (FR-IMGT) and complementarity determining regions (CDR-IMGT) as well as

canonical conserved amino acids: cysteine 23 (1stCYS) in FR1-IMGT, tryptophan 41 (CONSERVED-TRP) in FR2-IMGT, hydrophobic AA 89 (generally Leu), and cysteine 104 (2ndCYS) in FR3-IMGT (33) (**Figure 4B** and **Supplementary Data 5**).

However, TRAV sequences show structural diversity both in AA composition and in length. CDR1-IMGT and CDR2-IMGT lengths range from 4 to 9 AA and 0 to 10 AA in length, respectively, and the germline CDR3-IMGT is two to eleven AA long. The diversity is also reflected in the FR3-IMGT, most of them are 38 AA long (**Figure 4**), but there are TRAV genes with 32 (Salsal TRAV16-1), 35 (Oncmyk TRAV3-6) or 36 (Salsal and Oncmyk TRAV3-1 and TRAV3-2) AA.

Atypical Salmonid TRAV Genes

Three subgroups - TRAV1, TRAV3 and TRAV22 - have unusual structural particularities as shown in **Figures 5** and **6** in comparison to a typical TRAV (TRAV19).

TRAV1 from salmon and trout consistently lack the cysteine 23 (1st-CYS), which is replaced by a tyrosine (**Figure 5A**). No other structural particularity was identified, as illustrated by the "IMGT Collier de Perles" which provides a standardized delimitation of the strands and loops of the V-DOMAIN (32) (45), or by 3D modelling based on crystallographic structures (**Figures 5B, C**).

Another peculiar structural feature was found in salmon and trout TRAV22: as illustrated by the "IMGT Collier de Perles" representation, these sequences comprise a long stretch of 11 codons between the 2nd-CYS 104 and the V-HEPTAMER of the V-RS, instead of 3 or 4, as in most salmonid TRAV genes (**Figures 5D, E**). Such germline-encoded CDR3 sequences might lead to particular function of TR, with possible involvement in invariant-like TR. **Figure 5F** shows that the CDR3 stretch encoded by the V gene indeed constitutes a large fraction of the loop. Remarkably, it may contain particular structural elements such as two Cys possibly involved in a S-S bridge and a di-glycine motif in the sequence presented in **Figure 5F**.

All members of TRAV3 completely lack CDR2, reminiscent of the trout TRBV6 sequences (47) (**Figure 6A**). This was observed in all twelve functional rainbow trout TRAV3 (11 in Arlee and 1 in Swanson, see **Table 1**), and the seven functional Atlantic salmon TRAV3 as well as in the other salmonids, and in species as diverse as goby, Atlantic cod, sea bass and barramundi perch. Again, this feature could not be retrieved in all teleost groups. Comparing TRAV3, TRAV19 and TRAV16 IMGT Collier de Perles and 3D models, **Figures 6B, C** illustrate the wide variation of the CDR2 length observed across salmonid TRAV domains, raising the issue of potential modes of interaction with major histocompatibility (MH) proteins.

Salmonid TRAJ Genes

A total of 128 TRAJ genes were identified in the Atlantic salmon genome assembly. The TRAJ genes were named following the same rule proposed for human TRA/TRD locus (40), classifying the TRAJ1 as the closest TRAJ to the TRAC gene. This annotation also corresponds to the description given in Atlantic salmon by Yazawa et al. (23) (**Figures 1A** and **2A**).

The nucleotide and deduced AA sequences of the 128 TRAJ genes are reported in **Supplementary Data 7**. One hundred and four TRAJ appeared to be functional genes, with canonical FGXG J-motif, without stop codons and flanked by the J-RS at the 5' end, and a donor splice at the 3' end. Interestingly, 15 TRAJ genes lacked the canonical FGXG motif and can be considered ORF. Nine pseudogenes were also identified in Atlantic salmon TRA locus, with the presence of stop codons. A lower number of TRAJ genes were found in rainbow trout. A total of 81 (73 F, 6 ORF, 2P) and 66 (62 F, 2 ORF and 2P) TRAJ genes were annotated in Arlee and Swanson, respectively (**Figure 2B**, **Supplementary Data 2** and **Data 7**).

When we compared the deduced amino acid sequences of TRAJ genes from both salmonid species, we could identify at least one counterpart, except for Arlee or Swanson TRAJ9, Arlee TRAJ61 and for eight Atlantic salmon TRAJ genes (**Supplementary Data 7**). It means that TRAJ genes are conserved in both species and many TRAJ genes in Atlantic salmon are duplicated.

Salmonid TRDJ and TRDD Genes

Only one TRDJ gene was found in salmonids (**Figures 1, 2** and **Supplementary Data 8**). The deduced amino acid sequence is highly conserved in Atlantic salmon (DKLTFGKAINLIVEP) and rainbow trout (AKLTFGKAINLIVEP). In both species the FGKA motif replaces the hallmark motif (FGXG) identified in higher vertebrates TR J regions.

The nucleotide sequences of the three TRDD genes identified in the genome assemblies of both species are reported in **Supplementary Data 8**. They consist of 8 nt (TRDD1), 13 nt (Atlantic salmon TRDD2), 14 nt (rainbow trout TRDD2), and 13 nt (TRDD3) sequence that can be read in its three coding frames without stop codon. The 5'D-RS and 3'D-RS that flank the D-REGION are well conserved.

Expression of Atlantic Salmon TRA Genes

The functionality and expression of TRA genes was investigated by deep sequencing of 5'-RACE products produced from spleen, head-kidney (HK) and gill of three unvaccinated presmolt Atlantic salmon. Prior to sequencing the fish were confirmed free of the salmon pathogens ISAV, SAV3, PRV and IPNV by RT-qPCR. The RACE was initiated from primers located in the TRAC gene and the total number of productive reads obtained from each sample were normalized to provide a quantitative assessment of the respective usage of TRAV, TRAJ genes and TRAV-TRAJ combinations in the different individuals and across representative tissues. Combining all datasets, the relative expression of all productive and unproductive sequences for all TRAV/TRAJ combinations are represented in **Figures 7A, B**, respectively. There was no obvious correlation between TRAV or TRAJ gene expression and the relative position in the locus. Further, TRAV genes located in either of the two TRAV blocks were used at comparable frequencies indicating that neither the transcriptional orientation nor the large gene-less region present in the locus significantly influence expression. Among the 14 inverted, functional TRAV all but one (TRAV6-6) where expressed, representing ~0.31 to 1.45% of total productive rearrangements.

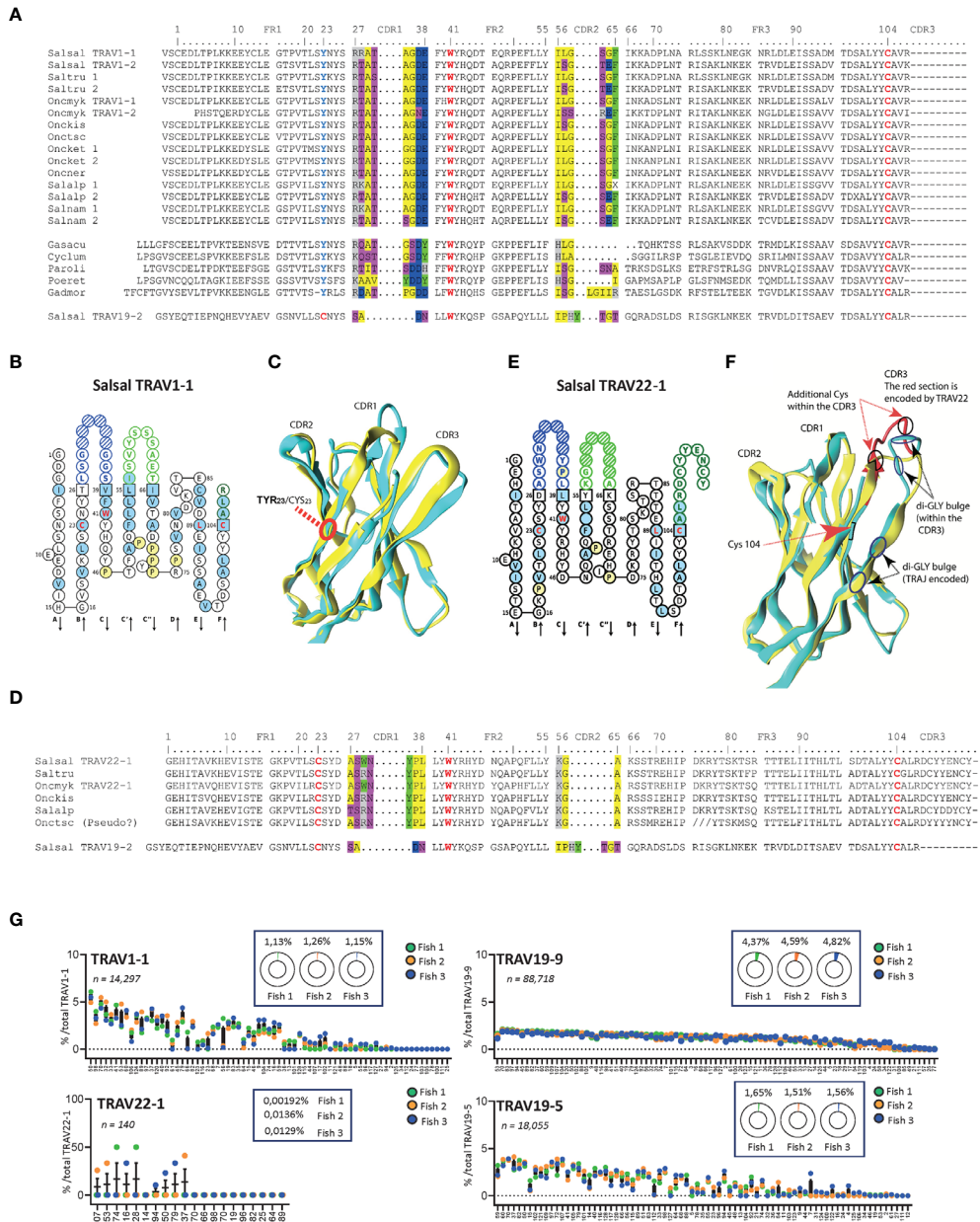


FIGURE 5 | Specific features of salmonid TRAV1 and TRAV22. **(A)** Multiple alignment of TRAV1 protein sequences from Atlantic salmon (Salsal) and rainbow trout (Oncmyk) with sequences presenting similar features from other Salmonids and other teleost. **(B)** IMGT Collier de Perles representation of Atlantic salmon TRAV1-1 based on the IMGT unique numbering for V-DOMAIN (45, 46). CDR1-IMGT is shown in dark blue and the CDR2-IMGT in light green and the germline CDR3-IMGT in dark green (33). Amino acids are shown in the one-letter abbreviation. Hatched circles correspond to missing positions according to the IMGT unique numbering (33, 46). **(C)** Salmon TRAV1 domain (cyan) aligned to human TR 2w20 alpha chain V domain. Position 23 is frame in red. **(D)** Multiple alignment of TRAV22 protein sequences from rainbow trout and Atlantic salmon with related sequences from other salmonids. **(E)** IMGT Collier de Perles of Atlantic salmon TRAV22-1. **(F)** Salmon TRAV22 domain (cyan) aligned to human TR Ipha chain V domain. Note the long CDR3. The section of CDR3 encoded by TRAV22 is in red. 2nd-CYS104, additional Cys within the CDR3 loop, and Gly amino acids in di-Gly motifs (one in the G strand and one within the CDR3) are highlighted. **(G)** Relative TRAJ usage shown as percentage of total reads containing a specific TRAJ gene among TRAV1-1, TRAV22, TRAV19-5 or TRAV19-9 containing productive rearrangements. Individual TRAJ genes are shown on the X-axis, average number of reads from each fish containing the indicated TRAV-TRAJ combination are shown (fish 1 is in green, fish 2 in orange and fish 5 in blue). The total number of sequences for each TRAV is indicated and the inlaid box shows the average representation of each TRAV per total number of productive rearrangements from each fish. (Atlantic salmon, *Salmo salar*: Salsal; rainbow trout, *Oncorhynchus mykiss*: Oncmyk; Brown trout, *Salmo trutta*: Saltru; Coho salmon, *Oncorhynchus kisutch*: Onckis; Chinook salmon, *Oncorhynchus tshawytscha* Onctsh; Chum salmon, *Oncorhynchus keta*: Oncket; Sockeye salmon, *Oncorhynchus nerka*: Oncner; Artic char, *Salvelinus alpinus*: Salalp; Lake trout, *Salvelinus namaycush*: Salnam; Three-spined stickleback, *Gasterosteus aculeatus*: Gasacu; Lumpfish, *Cyclopterus lumpus*: Cyclum; Olive flounder, *Paralichthys olivaceus*: Paroli; goby, *Gobiodon reticulatus*: Gobret; and Atlantic cod, *Gadus morhua*: Gadmor; guppy, *Poecilia reticulata* Poeret; sea bass, *Dicentrarchus labrax*: Diclab; barramundi perch, *Lates calcarifer*: Latcal).

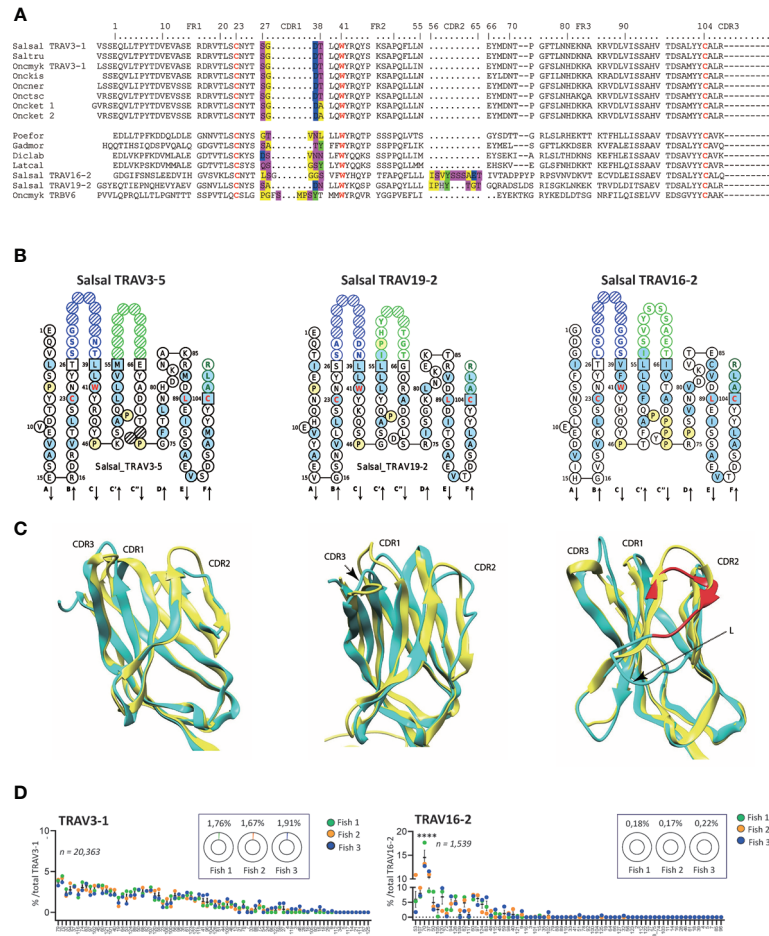


FIGURE 6 | Variations of CDR2-IMGT across salmonid TRAV. **(A)** Multiple alignment of selected TRAV (TRAV22, TRAV16, TRAV19) protein sequences from Atlantic salmon (Salsal) and rainbow trout (Oncmyk) with sequences presenting similar features from other Salmonids and other teleosts. **(B)** IMGT Collier de Perles of Atlantic salmon TRAV3-5, TRAV19-2, TRAV16-2, illustrating CDR2-IMGT length variation. **(C)** Salmon TRAV3, TRAV19, and TRAV16 (cyan) aligned to human TR B7 alpha chain V domain. Note the lack of CDR2 in salmon TRAV3. Salmon TRAV19 is shown for comparison of CDR2 modelling, and TRAV16 domain for its very long CDR2 region (in red) that induces an additional loop (L) in the model, which may not reflect the reality but illustrates the constraint on the domain. **(D)** Relative TRAJ usage shown as percentage of total reads containing a specific TRAJ gene for TRAV3-1 or TRAV19-9 containing productive rearrangements. Individual TRAJ genes are shown on the X-axis, average number of reads from each fish containing the indicated TRAV-TRAJ combination are shown (fish 1 is in green, fish 2 in orange and fish 5 in blue). The total number of sequences for each TRAV is indicated and the inlaid box shows the average representation of each TRAV per total number of productive rearrangements from each fish. Atlantic salmon, Salsal; rainbow trout, Oncmyk; Brown trout, Saltru; Coho, Oncnkis; Chinook salmon, Oncnsc; Chum salmon, Oncnet; Sockeye, Oncner; African molly, Poefor; Atlantic cod, Gadmor; barramundi perch, Latcal).

To determine if these proportions reflect selective processes or rather are the result of recombination probability of TRAV genes we compared the distribution of TRAV/TRAJ among productive and unproductive transcripts. As these unproductive rearrangements are independent of TR mediated selection processes, they should reflect probabilities of rearrangements. **Figures 7A, B** show a remarkable similarity of frequency heatmap, providing a global overview of rearrangement efficiency. In general, albeit not statistically significant the TRAJ genes located on the 5' side of the cluster (*i.e.*, those closest to TRAV genes) tend to be less expressed (**Figures 7C, D**). However, unlike the situation in many mammals (12, 48), where combinations of TRAV and TRAJ genes more proximal to each other tend to be favored and thus

represented at a higher frequency, no clear pattern of preferential combinations could be discerned here based solely on location.

While some inter-individual variation of expression of specific TRAV-J combinations was observed the overall gene usage, both for TRAV and TRAJ, was remarkably similar among individuals and across tissues (**Figure 7D**). The four most frequently expressed V genes TRAV2-1 (1.58, 1.70 and 1.58% in fish 1, 2 and 5 respectively), TRAV15-1 (1.70, 1.69 and 1.61%), TRAV19-8 (1.64, 1.63 and 1.69%) and TRAV19-9 (1.64, 1.67 and 1.73%) were shared among all three individuals. Similarly, a highly consistent pattern in terms of genes expressed at low frequencies was apparent across individual fish, with a few TRAV genes weakly represented within the dataset, including TRAV2-6 (0.31, 0.31 and 0.32%), TRAV6-1 (0.079, 0.079 and

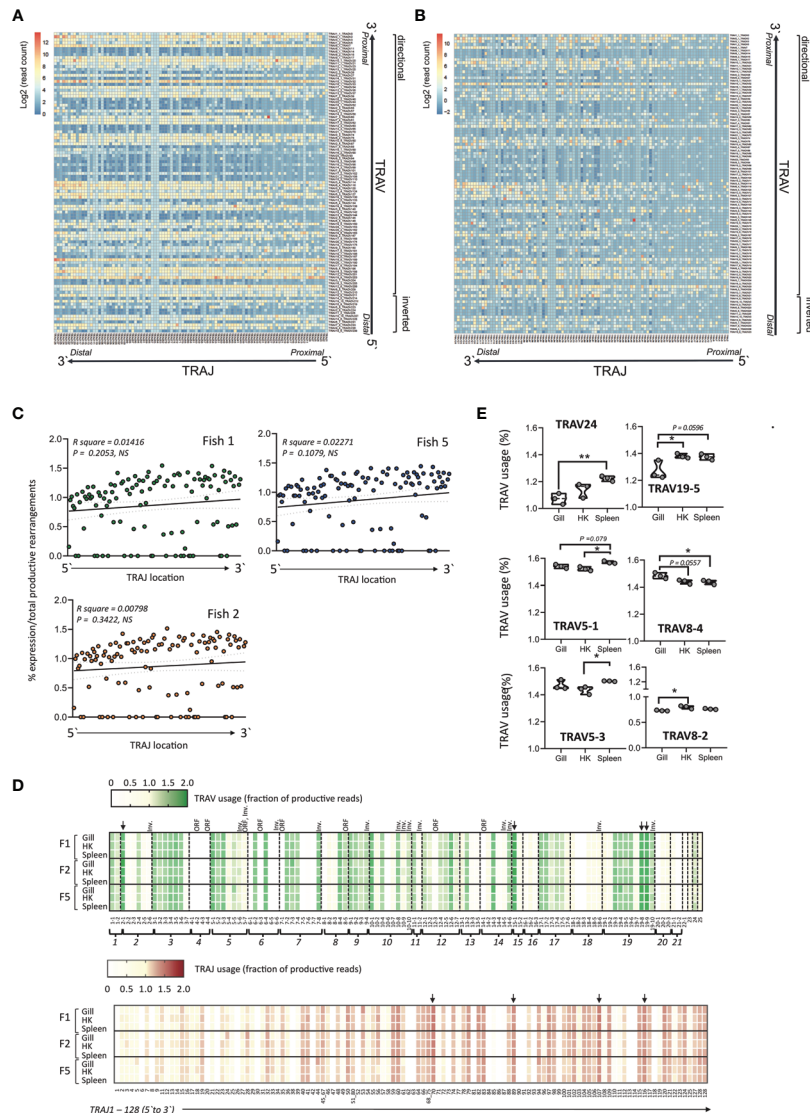


FIGURE 7 | Heat map of representative productive **(A)** or unproductive **(B)** TRAV- J combinations in Atlantic salmon. The average expression of TRAV-J combinations was computed from the TRA repertoire of the total number of productive sequences from gill, head kidney and spleen from three individual presmolt Atlantic salmon. The expression of each combination was expressed as total reads and the resulting matrix was normalized and represented as a heat map with a log₂ count scale. TRAJ and TRAV genes are displayed as each gene appear in the genome following a 5' to 3' direction and the transcriptional direction of the TRAV genes is indicated. **(C)** Scatterplot of the relative expression of TRAJ genes versus genome location showing no significant correlation. The most 5' TRAJ corresponds to TRAJ1 and is located closest to the TRAC gene. The average of normalized expression frequencies in gill, HK and from three fish is shown. **(D)** TRAV and TRAJ usage in Atlantic salmon. TRAV and TRAJ usage was determined from deep sequencing datasets of C alpha (TRAC) primed 5'-RACE generated from gill, head kidney (HK) and spleen of three unvaccinated pre-smolt Atlantic salmon. TRAV and TRAJ usage in each individual and tissue was expressed as the percentage of total productive TRA rearrangements from normalized data sets. Inv. = inverted orientation of TRAV genes, ORF = open reading frame genes (TRAV4-2, TRAV4-4, TRAV5-7, TRAV6-3, TRAV7-1, TRAV12-3 and TRAV14-1). Downward pointing arrows indicating the four most frequently used TRAV (TRAV2-1, TRAV15-1, TRAV19-8 and TRAV19-9) and TRAJ genes (TRAJ70, TRAJ89, TRAJ107 and TRAJ116) common to all three fish. **(E)** Tissue preferential TRAV usage. The relative expression of specific TRAV genes in gill, Head kidney (HK) and spleen from three pre-smolt Atlantic Salmon were determined by comparing the normalized percentage of total productive TRAV rearrangements in each tissue among the three individuals, statistical significance was determined by one way ANOVA followed by Tukey's multiple comparisons test with **p*<0.05 and ***p*<0.01.

0.078%), TRAV7-6 (0.089, 0.088 and 0.090%), TRAV8-1 (0.124, 0.124 and 0.126%), TRAV21-2 (0.0447, 0.0449 and 0.0446%) and TRAV22 (0.447, 0.449 and 0.446). Certain V genes, including TRAV6-5, TRAV6-6, TRAV7-5, TRAV7-7, TRAV10-6, TRAV18-2, TRAV18-3 and TRAV20-1 as well as all members

of the TRAV4 subgroup, which consists of two functional (TRAV4-1, TRAV4-3) and two ORF (TRAV4-2 and TRAV4-4) genes were absent from the data set. For the remaining genes annotated as ORF three (TRAV5-7, TRAV8-5 and TRAV12-3) were expressed and detected at varying degrees ranging from an

average of 0.89, 1.25 and 0.18% in fish 1, 2 and 5 respectively. Overall subgroups TRAV18, which consists of six functional genes, TRAV16 (3F genes), TRAV20 (3F genes) and TRAV21 (2 F genes) were consistently underrepresented across tissues in all three fish. Highly variable TRAV frequencies within given subgroups was also revealed (**Figure 7D**). For example, TRAV2-1, which represented one of the most frequently used V genes in all three individuals represents 34% of all TRAV2 subgroup containing productive rearrangements with TRAV2-4 accounting for 24% followed by TRAV2-3 (10%), TRAV2-2 (9.9%) and TRAV2-6 (6.6%). Similar patterns with specific members of a given subgroup dominating was observed for subgroup TRAV6, TRAV7, TRAV8, TRAV10, TRAV13, TRAV14 and TRAV19. Since the V-HEPTAMER and V-NONAMER of all functional members of these TRAV subgroups are similar, and their V-SPACER has an expected length of 22 bp there is no obvious explanation for these recombination biases and may point to additional constraints related either to pairing with the beta chain, or to selection. Comparably among the TRAV3 genes, all but one gene (TRAV3-7) were expressed at a relatively uniform, moderately high frequency representing about 8.5% of the total TRAV usage in the 3 individuals. Similarly, members of the TRAV9 subgroup, which consists of four functional TRAV, are all expressed at similar levels and represented an average of about 5% of the total TRAV usage in the 3 individuals.

Similar to the pattern observed for TRAV expression, the four most frequently expressed TRAJ genes were shared among the three fish (**Figure 7D**). Most commonly used are TRAJ70 (1.54, 1.51 and 1.40% in fish 1, 2 and 5 respectively) followed by TRAJ107 (1.53, 1.45 and 1.41%), TRAJ116 (1.46, 1.41 and 1.43%) and TRAJ89 (1.43, 1.41 and 1.40%). Several TRAJ genes annotated as functional were either absent from the dataset or expressed at extremely low frequencies (normalized frequency <0.05%) including TRAJ6, TRAJ8, TRAJ20, TRAJ22, TRAJ26, TRAJ30, TRAJ42, TRAJ47, TRAJ52, TRAJ62, TRAJ63, TRAJ81, TRAJ84, TRAJ86, TRAJ90, TRAJ91, TRAJ93, TRAJ113, TRAJ114 and TRAJ118.

Notably, while the majority of TRAV genes were used at similar frequency across all three tissues examined TRAV24, TRAV19-5, TRAV5-1 and TRAV5-3 predominated in the spleen, while TRAV8-3 was most common in HK and TRAV8-4 had the highest expression frequency in the gill (**Figure 7E**).

Expression of Atypical Atlantic Salmon TRAV Genes

To gain insight into the expression profiles and potential preferential TRAJ usage of the structurally unique TRAV genes (*i.e.*, TRAV1, TRAV3 and TRAV22) we determined the relative frequency of TRAV/J combinations among productive transcripts identified for each of these groups and compared this to transcripts containing two structurally conventional TRAV genes (TRAV19-5 and TRAV19-9). As an additional contrast to TRAV3, we also analyzed TRAV16 containing transcripts, which represent the three V genes with the longest

CDR2 region (10 AA) in both Atlantic salmon and rainbow trout. For the representative TRAV1-1 and TRAV3-1 the overall J usage pattern was comparable to that of TRAV19. TRAV1-1 transcripts were found rearranged to 72 of the 118 functional TRAJ compared to 78 for TRAV3-1, 79 for TRAV19-5 and 90 for the highly expressed TRAV19-9. A similar pattern was obtained for the other members of the TRAV1 (TRAV1-2) and TRAV3 (TRAV3-2, TRAV3-3, TRAV3-4, TRAV3-5, TRAV3-6 and TRAV3-7) groups (**Figures 5G, 6D**). While the individual expression profiles for TRAV1-1 and TRAV3-1 indicate some preferential J usage the overall TRAV-J combinations for both of these subgroups is diverse indicating that no strong subgroup specific selection constraints are placed on either of these TRAV groups. In contrast, sequence profiles for TRAV16-2 and TRAV22 containing transcripts were biased and displayed prominent inter-individual variation (**Figures 5G, 6D**). On average, TRAV16-2 transcripts were significantly more likely to be found rearranged to either TRAJ70 or TRAJ37 compared to any other J gene in all three individuals. However, the majority of TRAV16-2 containing productive rearrangements in the gill of fish 1 and fish 2 were combined with TRAJ53 representing 10,3 and 29,2% of total productive rearrangements in this tissue. However, in the same fish this specific combination was significantly less abundant in head kidney (1,6 and 1,01%) and spleen (4,1 and 2,31%) and overall represented less than 2% of TRAV16-2 containing rearrangements in fish 5. The remaining two TRAV16 genes, TRAV16-1 and TRAV16-3 were both expressed at very low frequencies representing less than 0,0066% of the total productive sequences and, in the majority of transcripts TRAV genes were combined with TRAJ68_or TRAJ75. Concerning TRAV22, an even more restricted profile was revealed with no TRAV22-TRAJ combinations shared between the three fish. In fish 1 all TRAV22 containing transcripts were combined with either TRAJ 37, 53 or 107, while for fish 2 TRAV22-TRAJ74 and TRAV22-TRAJ128 combinations made up 100% of the productive transcripts, and in fish 5 TRAV22 was combined with TRAJ50, 79, 94 or 116.

DISCUSSION

New genome assemblies from teleost species provide a high quality source of knowledge to analyze repertoire sequencing data and decipher adaptive immune responses to pathogens or vaccines. However, the full assembly of IG and TR loci remain challenging, and more efforts are required to improve their annotation and generate standardized sequence database allowing automated annotation of AIRRseq (Adaptive Immune Receptor Repertoire sequencing) data. Standardization is critical to warrant comparison across studies and species. Atlantic salmon and rainbow trout are key species for fish farming globally, and they have become important models in fish immunology. High-quality genome assemblies are now available for several salmonid species and strains. We therefore decided to revisit the description and annotation of TRA/TRD locus present in Atlantic salmon and two strains of rainbow trout

(Swanson and Arlee) with the aim of establishing a common salmonid nomenclature.

The TRA/TRD gene linkage is conserved in teleosts (including salmonids) as in other vertebrates; however, the TRD genes are not embedded between the TRAV and TRAJ genes as in mammals (15, 20–23). The genomic organization of TRA/TRD locus in salmonids is similar to that described in other teleosts, with a locus in reverse orientation on the chromosome, as defined by the transcriptional orientation of the TRDC and TRAC. This TRA/TRD locus comprise at its 5' end (chr FWD3' side) the cluster TRDD-TRDJ-TRDC-TRAJ-TRAC followed by the translocated downstream TRAV genes at the 3' end of the locus (chr FWD5' side) in opposite orientation (17, 20, 26). This structure implies that rearrangements of TRAV genes occur by inversion that, as opposed to deletion, would preserve the elements between the two rearranged genes. The gene rearrangement by inversion has been also described in teleost immunoglobulin light chains (49–51). This maximizes the number of genes available for secondary rearrangements, which may lead to a higher diversity and confer selective advantage through editing of self-reactive receptors. In addition, on the 3' end of the Atlantic salmon TRA/TRD locus we could also identify a block of 29 TRAV genes in direct orientation as the TRDD-TRDJ-TRDC-TRAJ-TRAC cluster although downstream of it. Our results indicate that most of these TRAV are expressed in Atlantic salmon mucosal (gill) and systemic immune tissues (HK and spleen), suggesting different possibilities, or a complex rearrangement process between chromatids or a mistake in the Atlantic salmon genome assembly. Further analysis of new salmonid genomes will be helpful to solve this question and to verify if this block of TRAV genes is also present in other salmonid species.

A whole genome duplication (WGD) event occurred in the common ancestor of extant salmonids species (52–55). This WGD involved spontaneous doubling of all chromosomes (autotetraploidization), and is named “Ss4R” to account for additional rounds of WGD at the base of teleost fishes and vertebrate lineage (54). The analysis of the rainbow trout genome revealed stability of the two ancestral genome copies suggesting a slow rediploidization process (52). Two immunoglobulin heavy chain loci are present in the Atlantic salmon and rainbow trout (38, 56, 57). In contrast, there is only one TRA/TRD locus in both species, located on Atlantic salmon chromosome 14 and rainbow trout chromosome 8. The presence of two functional IGH loci significantly contributes to the increased diversity of antibodies. While antibodies recognize antigens in native form, $\alpha\beta$ TR sense processed peptide antigens bound to MH or MH-like molecules (6). This mode of MH-restricted interaction not only imposes strong structural constraints, but also leads to highly degenerated peptide recognition by TR. This might explain at least in part the loss of one TRA/TRD locus after duplication, as the large number of TRA/TRD V, D and J genes present in one locus might be sufficient to produce a fully diversified repertoire given the number of available T lymphocytes.

Another important aspect is the impact of locus structure and/or the presence of multiple loci on the development of salmonid lymphocytes. The differentiation of T and B lymphocytes in human

and mice is a highly orchestrated process, wherein the generation of productive specific antigen receptors (IG or TR), and the expression of an unique clonal receptor by each lymphocyte are key aspects (58–61). In humans and mice, early steps of $\alpha\beta$ T lymphocyte differentiation leads to rearrangement of TRB locus, then the expression of the pre-T-cell receptor (pre-TR), which comprises a TRB chain covalently linked to the invariant pre-T alpha (pT α) chain encoded by a unrearranged gene (62, 63). Signaling through the pre-TR is involved in the induction of CD4 and CD8 expression, pre-T cell proliferation and differentiation, further TRA rearrangement and allelic exclusion (63). The allelic exclusion allows only one TRB chain to be expressed at the surface of a cell (64). By contrast, both TRA can be rearranged on both haplotypes through receptor editing (65). The potential diversity available to receptor editing is increased by the very high numbers of TRAV and TRAJ genes, a feature conserved in vertebrates and most apparent in salmonids. Interestingly, the surrogate pT α chain has been described in most mammals (human, mouse, opossum, platypus and so on), bird and lizard (66), but no pT α has been identified in teleosts and frog to date. Although a recent study with CD79-GFP transgenic zebrafish lines suggests that both IGH and IGL loci could rearrange simultaneously, without requirement of surrogate L chain (67), nothing is known about T cell maturation in teleosts. The mechanisms underlying the allelic exclusion required in salmonid TR loci to ensure clonal responses, and the possible role of the particular structure of the salmonid TRA/TRD locus implying V-(D)-J rearrangements by inversion, remain to be clarified.

While the organization of the salmonid TRA/TRD locus is conserved, the total number of TRAV genes is different in the two studied species, being 239 genes in Atlantic salmon and 163 in rainbow trout (Arlee line). Similar to what we previously described in salmonid IGH loci (38, 39), this large difference is mainly due to pseudogenes, with 123 annotated in Atlantic salmon and only 56 in rainbow trout (Arlee line). Furthermore, since there are no significant gaps (≥ 1000 bp) in the current annotated Atlantic salmon TRA/TRD locus, the higher number of TRAV genes identified by Yazawa et al. (23) maybe the result of sequence assembly from BAC clones or structural genome variations between different genetic backgrounds. Further comparisons with other de-novo genome assemblies derived from other Atlantic salmon genetic backgrounds are required to verify this hypothesis. The analysis of functional genes defined 25 subgroups shared by both salmonid species, with one single group identified only in *Oncorhynchus*, supporting a fair conservation of TRA/TRD potential repertoire across salmonids. Additionally, deep sequencing analysis data of Atlantic salmon TRA in different tissues also show that, except TRAV4, all subgroups are expressed indicating a balanced usage of the potential diversity.

Among functional salmonid TRAV genes we identified some members with nucleotide insertions and deletions that likely impact the structure of TRA. Related sequences without 1st-CYS 23, like the TRAV1 subgroup, are present in other salmonids including brown trout, Coho, Chinook, Chum, Sockeye, Arctic char, and Lake trout. Interestingly, other species belonging to distinct families of teleosts also have related sequences without 1st-CYS 23, including stickleback, lump, goby, flounder and

Atlantic cod. However, TRAV1 sequences could not be detected in the genome of many other fish families. No obvious evolutionary pattern could be defined, suggesting that TRAV1 like sequences might have been lost independently in multiple branches of the teleost tree.

The mode of interaction of human and murine TR with a classical pMH complex, places the germline encoded CDR1 and CDR2 loops of the TRA over the α helices of the MH protein, whereas the encoded CDR3 loops over the peptide (68–71). Thus, all members of salmonid TRAV3 subgroup that completely lack the CDR2, are unlikely to recognize pMH complexes in this mode. The lack of CDR2 loops owing to short CDR2 sequences has been also identified in trout TRBV6 (72) and in bovine TRAV genes (22). In addition, the CDR3 stretch encoded by genes belonging to salmonid TRAV22 subgroup are longer than usual and present particular structural elements that likely directly affect the antigen recognition. Analysis of expressed TRA repertoire indicates that all these particular TRAV genes are expressed at mucosal and systemic level; additionally, compared to other TRAV genes, TRAV22 showed a more restricted TRAJ usage. These structural features and restrict TRAJ suggest a possible involvement in invariant-like TR or the specialization to recognize free antigen with no antigen processing or presentation as described in other vertebrates (73, 74).

The proposed TRA/TRD locus annotation reflects the phylogenetic relationships between genes from *Oncorhynchus* (rainbow trout) and *Salmo* (Atlantic salmon). This resource will facilitate understanding diversification mechanisms, and should contribute to assess somatic hypermutation of TR sequences, a process recently reported (75). Hypermutation has been observed in dromedary (*Camelus dromedarius*) TRG/TRD (3, 76, 77), in sandbar shark (*Carcharhinus plumbeus*) TRG (78, 79), in nurse shark (*Ginglymostoma cirratum*) TRA (75, 80) and in the teleost ballan wrasse (*Labrus bergylta*) TRA (81).

In conclusion, we have provided an annotation for the complete salmonid TRA/TRD locus that represent an important step to better define the expression pattern of these TR isotypes among salmonids and through vertebrate evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at NCBI under BioProject ID: PRJNA765183.

REFERENCES

1. Flajnik MF, Du Pasquier L. Evolution of Innate and Adaptive Immunity: Can We Draw a Line? *Trends Immunol* (2004) 25:640–4. doi: 10.1016/j.it.2004.10.001
2. Flajnik MF, Du Pasquier L. Evolution of the Immune System. In: WE Paul, editor. *Fundamental Immunology*. New York: Wolters Kluwer & Lippincott Williams & Wilkins (2013). p. 67–128.
3. Ciccarese S, Vaccarelli G, Lefranc MP, Tasco G, Consiglio A, Casadio R, et al. Characteristics of the Somatic Hypermutation in the Camelus Dromedarius T Cell Receptor Gamma (TRG) and Delta (TRD) Variable Domains. *Dev Comp Immunol* (2014) 46:300–13. doi: 10.1016/j.dci.2014.05.001
4. Lefranc M-P, Lefranc G. *The T Cell Receptor Factsbook*. London, UK: Academic Press (2001). 398 pages.

ETHICS STATEMENT

The authors confirm that the experimental protocols used for the live fish experiments were based on the Animal Welfare Act (<https://www.regjeringen.no/en/dokumenter/animal-welfare-act/id571188/>) and performed in accordance with relevant guidelines and regulations given by the Norwegian Animal Research Authority.

AUTHOR CONTRIBUTIONS

E-SE, PB, and SuM conceived the project, E-SE, CF, StM, M-PL, PB, and SuM performed data analysis. E-SE, M-PL, PB, and SuM wrote the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

E-SE was supported by the Tromsø Research Foundation Starting Grant and by the Aquaculture program of The Research Council of Norway (Grant No. 295036). PB was supported by INRAE, by the European Commission (Grant Agreement 311993 TARGETFISH), and by the Agence Nationale de la Recherche (ANR-16-CE20-0002-01 FishRNAVax). SuM was supported by UVIGO and by Xunta de Galicia (GRC-ED431C 2020/02).

ACKNOWLEDGMENTS

The authors thank Lars Daniel Håland at the Genomics Resource Centre Tromsø (GSCT; <http://www.uit.no/helsefak/genomics>) for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.753960/full#supplementary-material>

5. Lefranc M-P, Lefranc G. Immunoglobulins or Antibodies: IMGT® Bridging Genes, Structures and Functions. *Biomedicines* (2020) 8:1–117. doi: 10.3390/BIOMEDICINES8090319
6. Davis MM, Bjorkman PJ. T-Cell Antigen Receptor Genes and T-Cell Recognition. *Nature* (1988) 334:395–402. doi: 10.1038/334395a0
7. Rast JP, Litman GW. T-Cell Receptor Gene Homologs are Present in the Most Primitive Jawed Vertebrates. *Proc Natl Acad Sci U.S.A.* (1994) 91:9248–52. doi: 10.1073/pnas.91.20.9248
8. Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW. α , β , γ , and δ T Cell Antigen Receptor Genes Arose Early in Vertebrate Phylogeny. *Immunity* (1997) 6:1–11. doi: 10.1016/S1074-7613(00)80237-X
9. Antonacci R, Massari S, Linguitti G, Jambrenghi AC, Giannico F, Lefranc MP, et al. Evolution of the T-Cell Receptor (TR) Loci in the Adaptive Immune

- Response: The Tale of the TRG Locus in Mammals. *Genes (Basel)* (2020) 11:624. doi: 10.3390/genes11060624
10. Antonacci R, Giannico F, Ciccarese S, Massari S. Genomic Characteristics of the T Cell Receptor (TRB) Locus in the Rabbit (*Oryctolagus cuniculus*) Revealed by Comparative and Phylogenetic Analyses. *Immunogenetics* (2014) 66:255–66. doi: 10.1007/s00251-013-0754-1
 11. Linguiti G, Antonacci R, Tasco G, Grande F, Casadio R, Massari S, et al. Genomic and Expression Analyses of Tursiops Truncatus T Cell Receptor Gamma (TRG) and Alpha/Delta (TRA/TRD) Loci Reveal a Similar Basic Public $\gamma\delta$ Repertoire in Dolphin and Human. *BMC Genomics* (2016) 17:634. doi: 10.1186/s12864-016-2841-9
 12. Mondot S, Lantz O, Lefranc MP, Boudinot P. The T Cell Receptor (TRA) Locus in the Rabbit (*Oryctolagus cuniculus*): Genomic Features and Consequences for Invariant T Cells. *Eur J Immunol* (2019) 49:2146–58. doi: 10.1002/eji.201948228
 13. Pégrier P, Bertignac M, Chentli I, Nguefack Ngoune V, Folch G, Jabado-Michaloud J, et al. IMGT[®] Biocuration and Comparative Study of the T Cell Receptor Beta Locus of Veterinary Species Based on Homo Sapiens TRB. *Front Immunol* (2020) 11:821. doi: 10.3389/fimmu.2020.00821
 14. Pégrier P, Bertignac M, Ngoune VN, Folch G, Jabado-Michaloud J, Giudicelli V, et al. IMGT[®] Biocuration and Comparative Analysis of Bos Taurus and Ovis Aries TRA/TRD Loci. *Genes (Basel)* (2021) 12:1–20. doi: 10.3390/genes12010030
 15. Piccinni B, Massari S, Caputi Jambrenghi A, Giannico F, Lefranc M-P, Ciccarese S, et al. Sheep (*Ovis Aries*) T Cell Receptor Alpha (TRA) and Delta (TRD) Genes and Genomic Organization of the TRA/TRD Locus. *BMC Genomics* (2015) 16:709. doi: 10.1186/s12864-015-1790-z
 16. Moulana M, Taylor EB, Edholm ES, Quiniou SMA, Wilson M, Bengtén E. Identification and Characterization of Tcr γ and Tcr δ Chains in Channel Catfish, *Ictalurus punctatus*. *Immunogenetics* (2014) 66:545–61. doi: 10.1007/s00251-014-0793-2
 17. Seelye SL, Chen PL, Deiss TC, Criscitiello MF. Genomic Organization of the Zebrafish (*Danio rerio*) T Cell Receptor Alpha/Delta Locus and Analysis of Expressed Products. *Immunogenetics* (2016) 68:365–79. doi: 10.1007/s00251-016-0904-3
 18. Gerritsen B, Pandit A, Zaaaroui-Boutahar F, van den Hout MCGN, van IJcken WFJ, de Boer RJ, et al. Characterization of the Ferret TRB Locus Guided by V, D, J, and C Gene Expression Analysis. *Immunogenetics* (2020) 72:101–8. doi: 10.1007/s00251-019-01142-9
 19. Mineccia M, Massari S, Linguiti G, Ceci L, Ciccarese S, Antonacci R. New Insight Into the Genomic Structure of Dog T Cell Receptor Beta (TRB) Locus Inferred From Expression Analysis. *Dev Comp Immunol* (2012) 37:279–93. doi: 10.1016/j.dci.2012.03.010
 20. Fischer C, Bouneau L, Ozouf-Costaz C, Crnogorac-Jurcovic T, Weissenbach J, Bernot A. Conservation of the T-Cell Receptor α/δ Linkage in the Teleost Fish *Tetraodon nigroviridis*. *Genomics* (2002) 79:241–8. doi: 10.1006/geno.2002.6688
 21. Liu F, Li J, Lin IYC, Yang X, Ma J, Chen Y, et al. The Genome Resequencing of TCR Loci in *Gallus gallus* Revealed Their Distinct Evolutionary Features in Avians. *ImmunoHorizons* (2020) 4:33–46. doi: 10.4049/immunohorizons.1900095
 22. Reinink P, Van Rhijn I. The Bovine T Cell Receptor Alpha/Delta Locus Contains Over 400 V Genes and Encodes V Genes Without CDR2. *Immunogenetics* (2009) 61:541–9. doi: 10.1007/s00251-009-0384-9
 23. Yazawa R, Cooper GA, Hunt P, Beetz-Sargent M, Robb A, Conrad M, et al. Striking Antigen Recognition Diversity in the Atlantic Salmon T-Cell Receptor α/δ Locus. *Dev Comp Immunol* (2008) 32:204–12. doi: 10.1016/j.dci.2007.05.002
 24. Parra ZE, Ohta Y, Criscitiello MF, Flajnik MF, Miller RD. The Dynamic Tcr δ : Tcr δ Chains in the Amphibian *Xenopus tropicalis* Utilize Antibody-Like V Genes. *Eur J Immunol* (2010) 40:2319–29. doi: 10.1002/eji.201040515
 25. Wang X, Huang J, Wang P, Wang R, Wang C, Yu D, et al. Analysis of the Chinese Alligator Tcr α/δ Loci Reveals the Evolutionary Pattern of Atypical Tcr δ /Tcr μ in Tetrapods. *J Immunol* (2020) 205:637–47. doi: 10.4049/JIMMUNOL.2000257
 26. Wang K, Gan L, Kunisada T, Lee I, Yamagishi H, Hood L. Characterization of the Japanese Pufferfish (*Takifugu rubripes*) T-Cell Receptor α Locus Reveals a Unique Genomic Organization. *Immunogenetics* (2001) 53:31–42. doi: 10.1007/s002510100298
 27. Tian JY, Qi ZT, Wu N, Chang MX, Nie P. Complementary DNA Sequences of the Constant Regions of T-Cell Antigen Receptors α , β and γ in Mandarin Fish, *Siniperca chuatsi* Basilewsky, and Their Transcriptional Changes After Stimulation With *Flavobacterium columnare*. *J Fish Dis* (2014) 37:89–101. doi: 10.1111/jfd.12042
 28. Buonocore F, Castro R, Randelli E, LeFranc MP, Six A, Kuhl H, et al. Diversity, Molecular Characterization and Expression of T Cell Receptor γ in a Teleost Fish, the Sea Bass (*Dicentrarchus labrax*, L.). *PLoS One* (2012) 7:e47957. doi: 10.1371/journal.pone.0047957
 29. Shang N, Sun XF, Hu W, Wang YP, Guo QL. Molecular Cloning and Characterization of Common Carp (*Cyprinus carpio* L.) Tcr γ and CD3 γ/δ Chains. *Fish Shellfish Immunol* (2008) 24:412–25. doi: 10.1016/j.fsi.2007.12.007
 30. Nam B-H, Hirono I, Aoki T. The Four TCR Genes of Teleost Fish: The cDNA and Genomic DNA Analysis of Japanese Flounder (*Paralichthys olivaceus*) TCR α -, β -, γ -, and δ -Chains. *J Immunol* (2003) 170:3081–90. doi: 10.4049/jimmunol.170.6.3081
 31. Partula S, de Guerra A, Fellah JS, Charlemagne J. Structure and Diversity of the TCR Alpha-Chain in a Teleost Fish. *J Immunol* (1996) 157:207–12.
 32. Partula S, de Guerra A, Fellah JS, Charlemagne J. Structure and Diversity of the T Cell Antigen Receptor Beta-Chain in a Teleost Fish. *J Immunol* (1995) 155:699–706.
 33. Lefranc M-P. Immunoglobulin and T Cell Receptor Genes: IMGT[®] and the Birth and Rise of Immunoinformatics. *Front Immunol* (2014) 5:22. doi: 10.3389/fimmu.2014.00022
 34. Robison BD, Wheeler PA, Sundin K, Sikka P, Thorgaard GH. Composite Interval Mapping Reveals a Major Locus Influencing Embryonic Development Rate in Rainbow Trout (*Oncorhynchus mykiss*). *J Hered* (2001) 92:16–22. doi: 10.1093/jhered/92.1.16
 35. Gao G, Magadan S, Waldbieser GC, Youngblood RC, Wheeler PA, Scheffler BE, et al. A Long Reads-Based De-Novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout. *G3 Genes[Genomes]Genetics* (2021) jkab052:1–11. doi: 10.1093/g3journal/jkab052
 36. Hordvik I, Torvund J, Moore L, Endresen C. Structure and Organization of the T Cell Receptor Alpha Chain Genes in Atlantic Salmon. *Mol Immunol* (2004) 41:553–9. doi: 10.1016/j.molimm.2004.03.006
 37. Yazawa R, Cooper GA, Beetz-Sargent M, Robb A, McKinnel L, Davidson WS, et al. Functional Adaptive Diversity of the Atlantic Salmon T-Cell Receptor Gamma Locus. *Mol Immunol* (2008) 45:2150–7. doi: 10.1016/j.molimm.2007.12.007
 38. Magadan S, Krasnov A, Hadi-Saljoqi S, Afanasiev S, Mondot S, Lallias D, et al. Standardized IMGT[®] Nomenclature of Salmonidae IGH Genes, the Paradigm of Atlantic Salmon and Rainbow Trout: From Genomics to Repertoires. *Front Immunol* (2019) 10:2541. doi: 10.3389/fimmu.2019.02541
 39. Magadan S, Mondot S, Palti Y, Gao G, Lefranc MP, Boudinot P. Genomic Analysis of a Second Rainbow Trout Line (Arlee) Leads to an Extended Description of the IGH VDJ Gene Repertoire. *Dev Comp Immunol* (2021) 118:103998. doi: 10.1016/j.dci.2021.103998
 40. Lefranc M. Nomenclature of the Human T Cell Receptor Genes. *Curr Protoc Immunol* (2000) 40(1):42. doi: 10.1002/0471142735.ima010s40
 41. Lefranc MP, Rabbitts TH. A Nomenclature to Fit the Organization of the Human T-Cell Receptor γ and δ Genes. *Res Immunol* (1990) 141:615–8. doi: 10.1016/0923-2494(90)90068-A
 42. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. *Mol Biol Evol* (2018) 35:1547–9. doi: 10.1093/molbev/msy096
 43. Magoč T, Salzberg SL. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* (2011) 27:2957–63. doi: 10.1093/BIOINFORMATICS/BTR507
 44. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* (2016) 11:e0163962. doi: 10.1371/JOURNAL.PONE.0163962
 45. Kaas Q, Ehrenmann F, Lefranc M-P. IG, TR and IgSF, MHC and MhcSF: What do We Learn From the IMGT Colliers De Perles? *Briefings Funct Genomics Proteomics* (2008) 6:253–64. doi: 10.1093/bfpg/elm032
 46. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT Unique Numbering for Immunoglobulin and T Cell Receptor Variable

- Domains and Ig Superfamily V-Like Domains. *Dev Comp Immunol* (2003) 27:55–77. doi: 10.1016/S0145-305X(02)00039-3
47. Bernard D, Six A, Rigottier-Gois L, Messiaen S, Chilmoneczyk S, Quillet E, et al. Phenotypic and Functional Similarity of Gut Intraepithelial and Systemic T Cells in a Teleost Fish. *J Immunol* (2006) 176:3942–9. doi: 10.4049/jimmunol.176.7.3942
48. Pasqual N, Gallagher M, Aude-Garcia C, Loiodice M, Thuderoz F, Demongeot J, et al. Quantitative and Qualitative Changes in V-J α Rearrangements During Mouse Thymocytes Differentiation: Implication for a Limited T Cell Receptor α Chain Repertoire. *J Exp Med* (2002) 196:1163–73. doi: 10.1084/jem.20021074
49. Hsu E, Criscitiello MF. Diverse Immunoglobulin Light Chain Organizations in Fish Retain Potential to Reverse B Cell Receptor Specificities. *J Immunol* (2006) 177:2452–62. doi: 10.4049/jimmunol.177.4.2452
50. Zimmerman AM, Yeo G, Howe K, Maddox BJ, Steiner LA. Immunoglobulin Light Chain (IgL) Genes in Zebrafish: Genomic Configurations and Inversal Rearrangements Between (VL-JL-CL) Gene Clusters. *Dev Comp Immunol* (2008) 32:421–34. doi: 10.1016/j.dci.2007.08.005
51. Timmusk S, Partula S, Pilsström L. Different Genomic Organization and Expression of Immunoglobulin Light-Chain Isotypes in the Rainbow Trout. *Immunogenetics* (2000) 51:905–14. doi: 10.1007/s002510000221
52. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, et al. The Rainbow Trout Genome Provides Novel Insights Into Evolution After Whole-Genome Duplication in Vertebrates. *Nat Commun* (2014) 5:3657. doi: 10.1038/ncomms4657
53. Christensen KA, Davidson WS. Autopolyploidy Genome Duplication Preserves Other Ancient Genome Duplications in Atlantic Salmon (*Salmo Salar*). *PLoS One* (2017) 12:e0173053. doi: 10.1371/journal.pone.0173053
54. Houston RD, Macqueen DJ. Atlantic Salmon (*Salmo salar* L.) Genetics in the 21st Century: Taking Leaps Forward in Aquaculture and Biological Understanding. *Anim Genet* (2019) 50:3–14. doi: 10.1111/age.12748
55. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic Salmon Genome Provides Insights Into Rediploidization. *Nature* (2016) 533:200–5. doi: 10.1038/nature17164
56. Hordvik I. Immunoglobulin Isotypes in Atlantic Salmon. *Salmo salar*. *Biomolecules* (2015) 5:166–77. doi: 10.3390/biom5010166
57. Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Messmer A, et al. Evolution of Duplicated IgH Loci in Atlantic Salmon. *Salmo salar*. *BMC Genomics* (2010) 11:486. doi: 10.1186/1471-2164-11-486
58. Von Boehmer H, Karjalainen K, Pelkonen J, Borgulya P, Rammensee H-G. The T-Cell Receptor for Antigen in T-Cell Development and Repertoire Selection. *Immunol Rev* (1988) 101:21–37. doi: 10.1111/j.1600-065X.1988.tb00731.x
59. Kondo K, Ohigashi I, Takahama Y. Thymus Machinery for T-Cell Selection. *Int Immunol* (2019) 31:119–25. doi: 10.1093/intimm/dxy081
60. Ritter MA, Larché M. T and B Cell Ontogeny and Phylogeny. *Curr Opin Immunol* (1988) 1:203–9. doi: 10.1016/0952-7915(88)90002-7
61. Melchers F, ten Boekel E, Seidl T, Kong XC, Yamagami T, Onishi K, et al. Repertoire Selection by Pre-B-Cell Receptors and B-Cell Receptors, and Genetic Control of B-Cell Development From Immature to Mature B Cells. *Immunol Rev* (2000) 175:33–46. doi: 10.1111/j.1600-065x.2000.imr017510.x
62. Fehling HJ, Krotkova A, Saint-Ruf C, von Boehmer H. Crucial Role of the Pre-T-Cell Receptor Alpha Gene in Development of Alpha Beta But Not Gamma T Cells. *Nature* (1995) 375:795–8. doi: 10.1038/375795a0
63. Von Boehmer H. Unique Features of the Pre-T-Cell Receptor α -Chain: Not Just a Surrogate. *Nat Rev Immunol* (2005) 5:571–7. doi: 10.1038/nri1636
64. Brady BL, Steinel NC, Bassing CH. Antigen Receptor Allelic Exclusion: An Update and Reappraisal. *J Immunol* (2010) 185:3801–8. doi: 10.4049/jimmunol.1001158
65. McGargill MA, Derbinski JM, Hogquist KA. Receptor Editing in Developing T Cells. *Nat Immunol* (2000) 1:336–41. doi: 10.1038/79790
66. Smelty P, Marchal C, Renard R, Sinzelle L, Pollet N, Dunon D, et al. Identification of the Pre-T-Cell Receptor α Chain in Nonmammalian Vertebrates Challenges the Structure-Function of the Molecule. *Proc Natl Acad Sci U.S.A.* (2010) 107:19991–6. doi: 10.1073/pnas.1010166107
67. Liu X, Li Y-S, Shinton SA, Rhodes J, Tang L, Feng H, et al. Zebrafish B Cell Development Without a Pre-B Cell Stage, Revealed by CD79 Fluorescence Reporter Transgenes. *J Immunol* (2017) 199:1706–15. doi: 10.4049/jimmunol.1700552
68. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the Complex Between Human T-Cell Receptor, Viral Peptide and HLA-A2. *Nature* (1996) 384:134–41. doi: 10.1038/384134a0
69. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, et al. An $\alpha\beta$ T Cell Receptor Structure at 2.5 Å and its Orientation in the TCR-MHC Complex. *J Immunol* (2010) 185:209–19. doi: 10.1126/science.274.5285.209
70. Reinherz EL, Tan K, Tang L, Kern P, Liu JH, Xiong Y, et al. The Crystal Structure of a T Cell Receptor in Complex With Peptide and MHC Class II. *Sci* (80-) (1999) 286:1913–21. doi: 10.1126/science.286.5446.1913
71. Rudolph MG, Stanfield RL, Wilson IA. How TCRs Bind MHCs, Peptides, and Coreceptors. *Annu Rev Immunol* (2006) 24:419–66. doi: 10.1146/annurev.immunol.23.021704.115658
72. Boudinot P, Boubekeur S, Benmansour A. Primary Structure and Complementarity-Determining Region (CDR) 3 Spectratyping of Rainbow Trout Tcr β Transcripts Identify Ten V β Families With V β 6 Displaying Unusual CDR2 and Dsferrently Spliced Forms. *J Immunol* (2002) 169:6244–52. doi: 10.4049/jimmunol.169.11.6244
73. Boudinot P, Mondot S, Jouneau L, Teyton L, Lefranc MP, Lantz O. Restricting Nonclassical MHC Genes Coevolve With TRAV Genes Used by Innate-Like T Cells in Mammals. *Proc Natl Acad Sci USA* (2016) 113:E2983–92. doi: 10.1073/pnas.1600674113
74. Robert J, Edholm ES. A Prominent Role for Invariant T Cells in the Amphibian. *Xenopus laevis tadpoles Immunogenet* (2014) 66:513–23. doi: 10.1007/s00251-014-0781-6
75. Ott JA, Ohta Y, Flajnik MF, Criscitiello MF. Lost Structural and Functional Inter-Relationships Between Ig and TCR Loci in Mammals Revealed in Sharks. *Immunogenetics* (2021) 73:17–33. doi: 10.1007/s00251-020-01183-5
76. Vaccarelli G, Antonacci R, Tasco G, Yang F, Giordano L, El Ashmaoui HM, et al. Generation of Diversity by Somatic Mutation in the *Camelus dromedarius* T-Cell Receptor Gamma Variable Domains. *Eur J Immunol* (2012) 42:3416–28. doi: 10.1002/eji.201142176
77. Antonacci R, Minecchia M, Lefranc MP, Ashmaoui HME, Lanave C, Piccinni B, et al. Expression and Genomic Analyses of *Camelus dromedarius* T Cell Receptor Delta (TRD) Genes Reveal a Variable Domain Repertoire Enlargement Due to CDR3 Diversification and Somatic Mutation. *Mol Immunol* (2011) 48:1384–96. doi: 10.1016/j.molimm.2011.03.011
78. Chen H, Bernstein H, Ranganathan P, Schluter SF. Somatic Hypermutation of TCR γ V Genes in the Sandbar Shark. *Dev Comp Immunol* (2012) 37:176–83. doi: 10.1016/j.dci.2011.08.018
79. Chen H, Kshirsagar S, Jensen I, Lau K, Covarrubias R, Schluter SF, et al. Characterization of Arrangement and Expression of the T Cell Receptor γ Locus in the Sandbar Shark. *Proc Natl Acad Sci USA* (2009) 106:8591–6. doi: 10.1073/pnas.0811283106
80. Ott JA, Castro CD, Deiss TC, Ohta Y, Flajnik MF, Criscitiello MF. Somatic Hypermutation of T Cell Receptor α Chain Contributes to Selection in Nurse Shark Thymus. *Life* (2018) 7:e28477. doi: 10.7554/eLife.28477
81. Bilal S, Lie KK, Sæle Ø, Hordvik I. T Cell Receptor Alpha Chain Genes in the Teleost Ballan Wrasse (*Labrus bergylta*) are Subjected to Somatic Hypermutation. *Front Immunol* (2018) 9:1101. doi: 10.3389/fimmu.2018.01101

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Edholm, Fenton, Mondot, Paulssen, Lefranc, Boudinot and Magadan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.