Standardized IMGT® Nomenclature of Salmonidae IGH Genes, the Paradigm of Atlantic Salmon and Rainbow Trout: From Genomics to Repertoires


To cite this version:

Susana Magadan, Aleksei Krasnov, Saida Hadi-Saljoqi, Sergey Afanasyev, Stanislas Mondot, et al.. Standardized IMGT® Nomenclature of Salmonidae IGH Genes, the Paradigm of Atlantic Salmon and Rainbow Trout: From Genomics to Repertoires. Frontiers in Immunology, Frontiers, 2019, 10, 10.3389/fimmu.2019.02541 . hal-02867815

HAL Id: hal-02867815
https://hal.inrae.fr/hal-02867815
Submitted on 15 Jun 2020

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
In teleost fish as in mammals, humoral adaptive immunity is based on B lymphocytes expressing highly diverse immunoglobulins (IG). During B cell differentiation, IG loci are subjected to genomic rearrangements of V, D, and J genes, producing a unique antigen receptor expressed on the surface of each lymphocyte. During the course of an immune response to infections or immunizations, B cell clones specific of epitopes from the immunogen are expanded and activated, leading to production of specific antibodies. Among teleost fish, salmonids comprise key species for aquaculture. Rainbow trout (Oncorhynchus mykiss) and Atlantic salmon (Salmo salar) are especially important from a commercial point of view and have emerged as critical models for fish immunology. The growing interest to capture accurate and comprehensive antibody responses against common pathogens and vaccines has resulted in recent efforts to sequence the IG repertoire in these species. In this context, a unified and standardized nomenclature of salmonid IG heavy chain (IGH) genes is urgently required, to improve accuracy of annotation of adaptive immune receptor repertoire dataset generated by high-throughput sequencing (AIRRseq) and facilitate comparisons between studies and species. Interestingly, the assembly of salmonids IGH genomic sequences is challenging due to the presence of two large size duplicated IGH loci and high numbers of IG genes and pseudogenes. We used data available for Atlantic salmon to establish an IMGT standardized nomenclature of IGH genes in this species and then applied the IMGT rules to the rainbow trout IGH loci to set up a nomenclature, which takes into account
INTRODUCTION

Vertebrate species with jaws (Gnathostomata) that appeared more than 400 million years ago are all characterized by an adaptive immune system based on B and T cells along with the huge diversity and specificity of their antigen receptors, the immunoglobulins (IG) or antibodies and the T cell receptors (TR), respectively (1, 2). The analysis of the germline IG locus defines the genomic repertoire with the identification of the functional variable (V), diversity (D), and joining (J) genes that participate in the synthesis of VH domains. It also allows the identification of the functional constant (C) genes that encode the constant regions of the heavy chains and define their isotypes (3–7).

In teleost fish, B cell clonal responses are induced by infection or immunization, as described in humans or mice. Antibodies constitute a key factor for fish specific immunity and for the protection afforded by vaccines. As key species in aquaculture, Salmonids (family Salmonidae) including rainbow trout (Oncorhynchus mykiss; Oncmyk) and Atlantic salmon (Salmo salar; Salsal) constitute important models for the study of antibodies and B cell responses in fish.

Several groups started to clone and sequence IG cDNA from rainbow trout in the early 1990s (8–12). Comparison of VH domains (V-D-J-REGION) expressed in trout stocks from Sweden, France, and the US revealed differences in IGHV subgroup usage: subgroups named 8, 9, 10, and 11 were found only in Swedish stocks while subgroups 4 and 7 were only found in French stocks and subgroup 5 (now part of IGHV1) was found in Swedish, French, and US stocks. These observations suggested genetic differences between the IGHV gene germline repertoires of different populations, but this was not fully clear due to the very small numbers of sampled individuals. In 1996, expressed VH domain sequences were classified into a set of 11 IGHV subgroups, defining a first unified nomenclature for rainbow trout (13). A more extensive study performed in 2006 on American trout by the group of Steve Kaattari found all these subgroups expressed, indicating that IGHV subgroups may have a wider distribution than previously suggested. Two additional subgroups expressed at low frequency were also discovered in this study (14), leading to a repertoire of 13 IGHV subgroups. These subgroups were used for an IMGT gene table created in 2009, with a provisional gene nomenclature (letter S) for rainbow trout IGHV [path to access: IMGT Repertoire (IG and TR) > I. Locus and genes > Gene tables > IGHV > Rainbow trout (O. mykiss)]1.

1 http://www.imgt.org

In Atlantic salmon, Solem et al. described in 2001 nine IGHV subgroups (15), seven of which corresponded to IGHV subgroups defined in rainbow trout (1, 2, 3, 6, 8, 9, and 11). Southern blot experiments suggested that the number of genes per subgroup could vary between 1 and 7 ± 10. This work also clearly established that Atlantic salmon IGHV genes were rearranged and transcribed from both of the two Atlantic salmon IGH loci (IGH locus A on chromosome 6 and IGH locus B on chromosome 3), which were most likely produced by the salmonid whole genome duplication. These data actually suggested that genes from some subgroups could be expressed only from a single locus, while genes from other subgroups were expressed from both A and B loci. This analysis was later extended and refined in 2010 by Yasuie et al. from a complete assembly of the Atlantic salmon IGH A and B loci based on sequences of 24 bacterial artificial chromosomes (BAC) (16). This study provided a first map of the organization of the duplicated IGH loci of a salmonid species. Ninety-nine IGHV genes were found in locus A, and 103 in locus B; 23 IGHV genes are functional in locus A, and 32 in locus B. Using the IMGT threshold of 75% identity for the V-REGION, 18 IGHV subgroups were defined in this work (16). Subgroups that did fit with the IGHV subgroups established in rainbow trout were given a subgroup number consistent with the online 2009 IMGT gene table [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHV > Atlantic salmon (Salmo salar)]1.

As new genome assemblies of Atlantic salmon and rainbow trout have been recently made available, we decided to annotate the IGH locus of these species and to establish a common nomenclature of IGH genes based on IMGT rules. We used data previously published for Atlantic salmon (16) to develop a prototype for the Salmonid IMGT standardized nomenclature. We also applied the IMGT rules to the rainbow trout IGH loci as a novel example of IMGT genomic annotation. The objective was to take into account the specificities of the Salmonid loci and to develop a unique, consistent nomenclature, while respecting the IMGT Scientific chart rules and standards. These standards are based on the concepts of identification (keywords), classification (gene and allele nomenclature), description (labels), and numbering (IMGT unique numbering and IMGT Collier de Perles) (3). It is important to note that a consistent nomenclature is crucial to build IMGT reference directory sets that are constituted by the V-REGION, D-REGION, and J-REGION of each IMGT reference allele from IMGT/LIGM-DB (same accession numbers as GenBank, ENA, and DDBJ) (17). These reference directory sets are the fundamental basis for annotation of repertoire datasets produced by high-throughput AIRRseq approaches for the analysis of expressed repertoires.
in particular to define expressed clonotypes (18–20). The IMGT reference directories are built following the classification of the V, D, J, and C genes and alleles according to the IMGT rules and the assignment of the IMGT functionality: functional (F), open reading frame (ORF), or pseudogene (P) (IMGT Scientific chart > IMGT functionality) (3). These rules ensure that the nomenclature is consistent within and between species, and can be updated when more sequence data become available. Reference directory sets are used by IMGT/V-QUEST and IMGT/functionAnalysis (21, 22) for detailed analysis of nucleotide (nt) sequences of V domains [V-(D)-J-REGION]; by IMGT/DomainGapAlign, which provides alignments of amino acid (AA) sequences with the closest V and J regions for V domains and the closest C exons for C domains (23); by IMGT Collier de Perles based on the IMGT unique numbering for V and C domains (24, 25); and by IMGT/HighV-QUEST (26, 27) for high-throughput sequence analysis of expressed IGH repertoires and clonotype definition (18–20). Importantly, IMGT reference directory sets are freely available for the academic community and can be used by other programs developed for repertoire analysis.

In this work, we produced reference directory sets for IGH loci of Atlantic salmon and rainbow trout, based on a unique nomenclature developed for salmonids and following IMGT rules. We show how the particularities of salmonid IGH loci (duplicated loci in each haplotype, large number of genes and pseudogenes) were taken into account and how reference directory sets can be used for annotation of IGH expression datasets. We also discuss how the nomenclature and reference directories can be updated with new data and extended to other salmonid species.

**MATERIALS AND METHODS**

GU129139 and GU129140 from GenBank, ENA, and DDBJ, entered in IMGT/LIGM-DB (Rel. 201839-1) and IMGT annotated (GU129139 in Rel. 201923-5, Last updated, Version 11 and GU129140 in Rel. 201930-1, Last updated, Version 10), were selected as *S. salar* (*Salsal*) IMGT IGH locus prototypes. Sequences from these entries are from Atlantic salmon BAC library (CHORI-214), constructed from a Norwegian aquaculture strain male, from BACPAC Resources, Children’s Hospital Oakland Research Institute (CHORI) (16). GU129139 (931200 bp) (Salsal locus A, ssao6, IMGT locus ID: Salsal_IGH_1) is in reverse (REV) orientation on chromosome 6 whereas GU129140 (1063283 bp) (Salsal IGH locus B, ssao3, IMGT locus ID: Salsal_IGH_2) is in forward (FWD) orientation on chromosome 3.

For obtaining IMGT gene names, newly identified Atlantic salmon and rainbow trout IGH genes and alleles from genome assemblies were submitted to the IG, T cell receptors (TR), and major histocompatibility (MH) Nomenclature Sub-Committee (IMGT-NC) of the International Union of Immunological Societies (IUIS) Nomenclature Committee2,3. Two IMGT-NC reports #2019-5-0131 and #2019-7-02202 comprise the submission of 75 Atlantic salmon IGHV sequences from two accession numbers NC_027305.1 and NC_027302.1. These reports concern 75 different genes [35 Atlantic salmon IGHV on NC_027305.1 (Salsal locus A, ssao6) and 40 Atlantic salmon IGHV genes on NC_027302.1 (Salsal locus B, ssao3)] and correspond to 75 new alleles (61 of them are *01 and 14 are *02).

Two new entries were created in IMGT/LIGM-DB: IMGT000028 for Salsal locus A [S. salar (Atlantic salmon), taxon:8030, breed: double haploid, assembly GCF_000233375.1, GenBank assembly ID: GCA_000233375.4, chromosome 6, CM003284.1 (2050824–22238370, complement), IGH locus A] [this entry includes IMGT annotated genes from NC_027305.1 (Salsal ssao6)] and IMGT000029 for Salsal locus B [S. salar (Atlantic salmon), taxon:8030, breed: double haploid, assembly GCF_000233375.1, GenBank assembly ID: GCA_000233375.4, chromosome 3, CM003281.1 (7757817–79383607), IGH locus B] [this entry includes IMGT annotated genes from NC_027302.1 (Salsal ssao3)].

The rainbow trout genome (assembly: Omyk_1.0, June 2017; GenBank accession GCA_002163495.1) obtained from the homozygous Swanson clonal line was examined to locate IGH loci. Two IGH loci were identified, locus A on chromosome 13 (Oncmyk chr13) and locus B on chromosome 12 (Oncmyk chr12), both of them are in forward (FWD) orientation. The IMGT-NC Report #2019-10-0402 comprises the submission of 181 rainbow trout IGH gene sequences from NC_035089.1 (Oncmyk Omy13) and NC_035088.1 (Oncmyk Omy12). This IMGT-NC report concerns 181 different genes: 74 genes in locus A on Oncmyk chr 13 (49 IGHV, 11 IGHD, 10 IGHC) and 4 IGHC on NC_035089.1) and 107 genes in locus B on Oncmyk chr 12 (80 IGHV, 13 IGHD, 9 IGHC, and 5 IGHC on NC_035088.1) and corresponds to 181 new alleles *01. Two new entries were created in IMGT/LIGM-DB: IMGT000043 (IMGT/LIGM-DB) for Oncmyk locus A [O. mykiss (rainbow trout), taxon:8022, isolate: Swanson, assembly Omyk_1.0, GenBank assembly ID: GCF_002163495.1, chromosome 13: CM000794.1 (48012355–48422510), IGH locus A] [this entry includes IMGT annotated genes from NC_035089.1 (Oncmyk Omy13)] and IMGT000044 for Oncmyk locus B [O. mykiss (rainbow trout), taxon:8022, isolate: Swanson, assembly Omyk_1.0, GenBank assembly ID: GCF_002163495.1, chromosome 12: CM000794.1 (81302817–81805590), IGH locus B] [this entry includes IMGT annotated genes from NC_035088.1 (Oncmyk Omy12)].

**RESULTS**

The complete and correct assembly of the Salmonidae IGH loci is a significant challenge owing to (i) the existence of two duplicated loci due to the tetraploidization (named locus A and locus B), (ii) the large size of each locus, (iii) the high number of different IGHV subgroups compared to mammals, (iv) the internal amplification and potential gene conversion that occurred inside each locus during their evolution, and (v) the very high number of pseudogenes, many of them partial, relative to the functional genes.

We therefore explored how the standardized IMGT nomenclature could allow the identification and classification of genes and alleles in incomplete or not yet fully annotated genome assemblies.
assemblies. The IGH data published for Atlantic salmon (16), largely based on BAC sequencing, were used as a prototype for establishing the standardized IMGT nomenclature for salmonids and for dealing, by comparison, with newly identified IGH genes from both Atlantic salmon and rainbow trout genome assemblies. The particularities of these IGH loci (in particular the tetraploidization) were taken into consideration for consistency between salmonid species.

From IG Classes to IMGT Constant (C) Gene Names

Three antibody classes have been identified in fish, namely, IgM, IgD, and IgT, while IgG, IgA, and IgE are absent (28). IgM and IgD are generally co-expressed at the cell surface of the same B cells through alternative splicing, as in mammals. Soluble IgM are tetrameric and constitute the main antibody class in serum. A third class, IgT, is expressed in most fish groups including salmonids. Interestingly, the IG-Heavy-Tau chains of IgT have a VH domain that results from independent V-D-J rearrangements, and is not obtained by a switch process (29). IgT has been found only in bony fish and is particularly involved in mucosal immunity and protection (30). IGHD was cloned and characterized in rainbow trout and Atlantic salmon, in parallel to the discovery of IGHT encoding the third fish IG-Heavy-Tau isotype (28, 29) and then in Atlantic salmon (31).

By convention, IMGT groups are designated by the locus and gene type. Based on the four gene types, V (variable), D (diversity), J (joining), and C (constant), the IGH genes belong to four groups: IGHV, IGHD, IGJH, and IGHC. For the IGH locus, the constant genes are designated by the letter (and, if relevant, number) corresponding to the encoded isotype (IGHT, IGHM, and IGHD), instead of using the letter C.

The salmonid IGHC genes belong to three subgroups IGHM, IGHD, and IGHT and encode, when functional, the C-REGION of the heavy chain defining these three isotypes, IG-Heavy-Mu (heavy chain of the IgM class), IG-Heavy-Delta (heavy chain of the IgD class), and IG-Heavy-Tau (heavy chain of the IgT class) (Table 1). Salmonid locus A and locus B were assigned based on the literature, with the letter D (for “duplicated”) added to the conventional gene names for locus B.

Atlantic Salmon IGH Constant Genes and Associated D and J Genes

The Atlantic salmon IGH locus A, which is in a reverse (REV) orientation on chromosome 6 and spans 660 kilobases (kb) (with the V genes encompassing 600 kb) (Figure 1) includes 7 IGHC genes with 17 associated IGHD genes and 13 IGHJ genes. The Atlantic salmon IGH locus B, which is in forward (FWD) orientation on chromosome 3 and spans 720 kb (with the V genes encompassing 670 kb) (Figure 2) includes 5 IGHC genes with 11 associated IGHD genes and 8 IGHJ genes. The constant region of the IG-Heavy-Mu chain and of the IG-Heavy-Delta are encoded by a unique gene per locus (IGHM and IGHD for locus A and IGHM and IGHD for locus B) preceded by a D-J cluster. There are several IG-Heavy-Tau genes (IGHT), but the associated D-J cluster may be incomplete (lacking D and/or J genes). In Atlantic salmon, there is only one IGHT functional (F) gene per locus, IGHT4 for locus A and IGHT2D for locus B, each one having a complete D-J cluster (Table 2).

In the Atlantic salmon locus A, the D and J genes associated to IGHT genes comprise two D (IGHD1T2 and IGHD2T2) and two J (IGHJ1T2 and IGHJ2T2) upstream of the pseudogene (P) IGHT2, two J (IGHJ1T3 and IGHJ2T3) upstream of IGHT3 (P), five D (IGHD1T4 to IGHD5T4) and two J (IGHJ1T4 and IGHJ2T4) genes, all of them functional, upstream of IGHT4 (F) and one D (IGHD1T5) and two J (IGHJ1T5 and IGHJ2T5) upstream of IGHT5 (P). There is no IGHD or IGJ upstream of IGHT1 (P) (Table 2). The D and J associated to IGHM and IGHD comprise nine D (IGHD1 to IGHD9), all of them functional and five J genes, three of them functional (IGHJ1, IGHJ3, and IGHJ4), one with ORF, the IGHJ2, and one with alleles F or ORF (IGHJ5). They are located upstream of IGHM (F) and shared with the IGHD constant gene (F) (Table 2 and Figures S1, S2). Eleven IGHD not directly associated to constant genes are dispersed in locus A (IGHD-1 to IGHD-11).

In the Atlantic salmon locus B, the D and J genes associated to IGHT genes comprise one J (IGHJ1T1D) upstream of IGHT1D (P), two D (IGHD1T2D and IGHD2T2D), and two J (IGHJ1T2D and IGHJ2T2D) all functional upstream of IGHT2D (F) and three D (IGHD1T3D, IGHD2T3D, and IGHD3T3D) downstream of IGHT3D (P) (Table 2 and Figures S1, S2). The D

TABLE 1 | Salmonid Ig receptor classes, heavy chain types, and IGHC gene names.

<table>
<thead>
<tr>
<th>IG receptor class</th>
<th>IG heavy chain type</th>
<th>IG C-gene group</th>
<th>IG C-gene subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgM</td>
<td>IG-Heavy-Mu</td>
<td>IGHC</td>
<td>IGHM</td>
</tr>
<tr>
<td>IgD</td>
<td>IG-Heavy-Delta</td>
<td>IGHC</td>
<td>IGHD</td>
</tr>
<tr>
<td>IgT</td>
<td>IG-Heavy-Tau</td>
<td>IGHC</td>
<td>IGHT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locus A</th>
<th>Locus B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmo salar</td>
<td>Oncorhynchus mykiss</td>
</tr>
<tr>
<td>IGHM</td>
<td>IGMD</td>
</tr>
<tr>
<td>IGHD</td>
<td>IGHD</td>
</tr>
<tr>
<td>IGHT1</td>
<td>IGHT1D</td>
</tr>
<tr>
<td>IGHT2</td>
<td>IGHT2D</td>
</tr>
<tr>
<td>IGHT3</td>
<td>IGHT3D</td>
</tr>
<tr>
<td>IGHT4</td>
<td>IGHT4</td>
</tr>
<tr>
<td>IGHT5</td>
<td>IGHT5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locus A</th>
<th>Locus B</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHM</td>
<td>IGMD</td>
</tr>
<tr>
<td>IGHD</td>
<td>IGHD</td>
</tr>
<tr>
<td>IGHT1</td>
<td>IGHT1D</td>
</tr>
<tr>
<td>IGHT2</td>
<td>IGHT2</td>
</tr>
<tr>
<td>IGHT3</td>
<td>IGHT3</td>
</tr>
<tr>
<td>IGHT4</td>
<td>IGHT4</td>
</tr>
<tr>
<td>IGHT5</td>
<td>IGHT5</td>
</tr>
</tbody>
</table>
and J genes associated to IGHMD and IGHDD comprise six D (IGHD1D to IGHD6D, all functional) and five J genes (four functional, IGHJ1D to IGHJ4D) and one with alleles F or ORF (IGHJ5D). They are located upstream of IGHMD (F) and shared with the IGHD2 constant gene (F) (Table 2 and Figures S1, S2). Six IGHD not directly associated to constant genes are dispersed in locus B (IGHD-1D to IGHD-6D).

IGHD, IGHJ, and IGHC genes are reported in IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGH > IGHD > Atlantic salmon (S. salar); ibid., IGHJ > Atlantic salmon (S. salar); ibid., IGHC > Atlantic salmon (S. salar)]1.

Rainbow Trout IGH Constant Genes and Associated D and J Genes

Similar to the Atlantic salmon, the rainbow trout has one functional gene per IGH locus encoding the constant region of the IG-Heavy-Mu (IGHM gene in locus A and IGHMD gene in locus B), the constant region of the IG-Heavy-Delta (IGHD gene in locus A and IGHDD gene in locus B), and the constant region of the IG-Heavy-Tau (IGHT2 gene in locus A and IGHIT1D gene in locus B).

The rainbow trout IGH locus A, which spans 360 kb and is in a forward (FWD) orientation on chromosome 13, includes 11 IGHD genes, 10 IGHJ genes, and 4 IGHC genes (Table 3).
There are three D and two J genes upstream of IGHT1 (P), two D and two J genes upstream of IGHT2 (F), and six D and six J genes (all of them F) upstream of IGHM (F) and shared with the IGHD (F) constant gene (Figures S1, S2).

The rainbow trout IGH locus B, which spans 485 kb and is in a forward (FWD) orientation on chromosome 12, includes 13 IGHD genes, 9 IGHJ genes, and 3 IGHC genes (Table 3). There are four D genes (1 ORF and 3 F) and two J genes (both F) upstream of IGHT1D (F), and six D and seven J genes (all of them F) upstream of IGHMD (F) and shared with the IGHDD (F) constant gene (Figures S1, S2).

Sequences of rainbow trout IGHD and IGHJ genes and alleles are available in the downloadable IMGT reference directory sets from IMGT/GENE-DB (/download/GENE-DB) and from IMGT/V-QUEST (/download/V-QUEST/IMGT_V-QUEST_reference_directory/Oncorhynchus_mykiss/IG/IGHD.fasta; ibid., /IGHJ.fasta) and from IMGT/GENE-DB (/download/GENE-DB/IMGT_V-QUEST_reference_directory/Oncorhynchus_mykiss/IG/IGHD.fasta; ibid., /IGHJ.fasta). IGHD and IGHJ genes and alleles are reported in the IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHD > Rainbow trout (O. mykiss); ibid., IGHJ > Rainbow trout (O. mykiss)]

The demonstration that there is only one rainbow trout IGH-Heavy-Delta complete gene per locus, IGHD in locus A and
TABLE 2 | Atlantic salmon (Salmo salar) IGH constant C genes and associated D and J genes.

<table>
<thead>
<tr>
<th>IGHD genes</th>
<th>IGHJ genes</th>
<th>IGHC genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>–</td>
<td>–</td>
<td>IGHT1 P</td>
</tr>
<tr>
<td>IGH1T2 F</td>
<td>IGHJ1T2 F</td>
<td>IGHT2 P</td>
</tr>
<tr>
<td>IGH2T2 F</td>
<td>IGHJ1T3 F</td>
<td>IGHT3 P</td>
</tr>
<tr>
<td>IGH1T4 F</td>
<td>IGHJ1T4 F</td>
<td>IGHT4 P</td>
</tr>
<tr>
<td>IGH2T4 F</td>
<td>IGHJ2T4 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH3T4 F</td>
<td>IGHJ2T4 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH4T4 F</td>
<td>IGHJ2T4 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH5T4 F</td>
<td>IGHJIT5 F</td>
<td>IGHT5 P</td>
</tr>
<tr>
<td>IGH1T5 F</td>
<td>IGHJ1 F</td>
<td>IGHM F</td>
</tr>
<tr>
<td>IGH2 F</td>
<td>IGHJ2 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH3 F</td>
<td>IGHJ3 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH4 F</td>
<td>IGHJ4 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH5 F</td>
<td>IGHJ5 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH6 F</td>
<td>IGH6D F</td>
<td>–</td>
</tr>
<tr>
<td>IGH7 F</td>
<td>IGH7 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH8 F</td>
<td>IGH8 F</td>
<td>–</td>
</tr>
<tr>
<td>IGH9 F</td>
<td>IGH9 F</td>
<td>–</td>
</tr>
</tbody>
</table>

IGHDD in locus B, respectively, and that these two genes are functional, results from the analysis derived from applying the nomenclature of the Atlantic salmon IGH loci as well as the interpretation of expression data and published references (15, 16, 29, 31). The anomalies (partial IGHD and IGHDD genes with exons in aberrant localizations or in reverse-complementary orientation) are likely artifacts of the current genome assembly. For that reason, the functionality of the IGHD and IGHDD, deduced from literature data and supported by sequences external to the genome assembly, is shown in parentheses in Table 3.

Atlantic Salmon IGH Variable Genes

The Atlantic salmon IGH locus comprises a total of 303 IGH variable (IGHV) genes (145 IGHV in locus A on Salsal chromosome 6, spanning 600 kb, and 158 IGHV in locus B on Salsal chromosome 3, spanning 670 kb) (Figures 1, 2). There are a total of 67–69 functional genes, 12 ORF, and 222–224 pseudogenes (Table 4).

Based on the percentage of identity between nucleotide sequences of the V-REGION (threshold 75%), the Atlantic salmon 303 IGHV genes can be classified into 16 IGHV subgroups. IGHV genes are reported in IMGT Gene tables [IMGT Repertoire (IG and TR)]1 and the comparison of the CDR-IMGT lengths per subgroup (3) [IMGT Repertoire (IG and TR)] > 3. 2D and 3D structures > FR-IMGT and CDR-IMGT [IMGT Repertoire (IG and TR)] > 2. Proteins and alleles > Protein displays > IGHV > Atlantic salmon (S. salar)]1 and correspondence with previous gene names is indicated.

Translation of alleles *01 of F, ORF, and in-frame P are aligned according to the IMGT unique numbering in IMGT Protein display allowing the visualization of the FR-IMGT and CDR-IMGT [IMGT Repertoire (IG and TR)] > 2. Proteins and alleles > Protein displays > IGHV > Atlantic salmon (S. salar)]1 and the comparison of the CDR-IMGT lengths per subgroup (3) [IMGT Repertoire (IG and TR)] > 3. 2D and 3D structures > FR-IMGT and CDR-IMGT [IMGT Repertoire (IG and TR)] > 2. Proteins and alleles > Protein displays > IGHV > Atlantic salmon (S. salar)]1 and correspondence with previous gene names is indicated.

Rainbow Trout IGH Variable Genes

A total of 129 IGHV genes were identified in the rainbow trout genome, of which 57 can be considered fully functional or with an ORF without stop codon. A number of other sequences were identified as IGHV fragments in the assembly and were not included in the annotation. On chromosome 13 (locus A), 44 IGHV genes were found upstream of the functional IGHT2 gene, as well as 5 IGHV genes between the D-J-IGHM-IGHD cluster and the D-J-IGHM-IGHD cluster. Eighty IGHV genes were found on chromosome 12 (locus B): 70 IGHV were located upstream of the functional IGHT1D gene and 10 IGHV were found between
the D-J-IGT1D cluster and the D-J-IGHMD-IGHDD cluster. The 129 rainbow trout IGHV genes could be classified into the same 16 subgroups defined for the Atlantic salmon IGHV genes, containing from only 1 pseudogene (i.e., IGHV5, IGHV13, and IGHV14 subgroups) to 35 genes, i.e., IGHV1 subgroup, which includes 12 F, 2 ORF, and 21 P IGHV genes. Figure 3 shows a phylogenetic tree based on nucleotide sequences of IGHV genes (F and ORF) present in Atlantic salmon and rainbow trout IGH loci. While some IGHV subgroups are not represented in both species, as far as we know, this tree illustrates how rainbow trout IGHV genes nicely cluster with their Atlantic salmon counterparts.

**Expressed Repertoire Analysis**

IMGT/V-QUEST and its high-throughput version, IMGT/HighV-QUEST, can perform analysis of nucleotide sequences of the IG and TR variable domains (21, 22, 26, 27). These tools run against the IMGT/V-QUEST reference directory database that includes several sets (per group and per species) and are built based on the IMGT standards (3) (annotation in IMGT/LIGM-DB, Gene tables, Alignments of alleles, Protein display, entry in IMGT/GENE-DB). The IMGT/V-QUEST sets comprise IMGT reference sequences from all functional (F) and ORF genes and alleles (in Advanced parameters, Selection of IMGT reference directory set “F + ORF”). The sets also include IMGT reference sequences from pseudogenes (P) and alleles with an in-frame V-REGION for versatile genomic analysis (proposed by default, in Advanced parameters IMGT reference directory set “F + ORF + in-frame P”).

Altogether, IMGT/V-QUEST reference directory for Atlantic salmon IGHV contains 150 alleles that include 76 F, 15 ORF, and 59 P in-frame (release: 201931-4, 1st August 2019) (Table 5). The 76 F comprise, in addition to the 67 F alleles *01 (28 from locus A and 39 from locus B), 8 alleles *02 and 1 allele *03. The 15 ORF comprise, in addition to the 12 ORF alleles *01 (7 from locus A and 5 from locus B), 3 alleles *02. The 59 in-frame P comprise, in addition to the 54 P alleles *01 (26 from locus A and 28 from locus B), 5 alleles *02. Alleles of closely related duplicated genes are managed in the same Alignments of alleles, as shown, for example, for IGHV1-64*01 F and IGHV1-100*01 F, which have identical V-REGION nucleotide sequences [IMGT Repertoire (IG and TR) 2. Proteins and alleles > Alignments of alleles > IGHV > Atlantic salmon (S. salar)]).

IMGT/V-QUEST reference directory for rainbow trout IGHV contains sequences of 77 alleles that include 44 F, 11 ORF, and 22 P in-frame (download/V-QUEST/IMGT_V-QUEST_reference_directory/Oncorhynchus_mykiss/IG/IGHV.fasta). All the alleles are *01 (release: 201931-4, 1st August 2019) (Table 5). All IGHV genes and alleles (including the P out-of-frame) are reported in IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHV > Atlantic salmon (S. salar)]).

We then investigated the functionality and expression level of IGHV genes from the two species using the standardized nomenclature based on genomic annotation. To do so, adaptive immune receptor repertoire datasets generated by high-throughput sequencing (AIRRseq) were submitted to IMGT/HighV-QUEST analysis.

**Atlantic Salmon**

AIRRseq data from head kidney of Atlantic salmon were generated based on 5’RACE and specific primers for IGHM constant region [data from reference (32)]. Using the Atlantic salmon reference dataset updated in 2019, a total of 50 IGHV genes (42 functional “F”, 4 “ORF” and 4 pseudogenes “P”) were expressed in the dataset (Figure 4A). More than 80% of

### Table 3 | Rainbow trout (Oncorhynchus mykiss) IGH constant C genes and associated D and J genes.

<table>
<thead>
<tr>
<th>IGHD genes</th>
<th>IGHJ genes</th>
<th>IGHC genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHD1T1</td>
<td>F</td>
<td>IGHJ1T1</td>
</tr>
<tr>
<td>IGHD2T1</td>
<td>F</td>
<td>IGHJ2T1</td>
</tr>
<tr>
<td>IGHD3T1</td>
<td>F</td>
<td>IGHJ3T1</td>
</tr>
<tr>
<td>IGHD1T2</td>
<td>F</td>
<td>IGHJ1T2</td>
</tr>
<tr>
<td>IGHD2T2</td>
<td>F</td>
<td>IGHJ2T2</td>
</tr>
<tr>
<td>IGHD1</td>
<td>F</td>
<td>IGHJ1</td>
</tr>
<tr>
<td>IGHD2</td>
<td>F</td>
<td>IGHJ2</td>
</tr>
<tr>
<td>IGHD3</td>
<td>F</td>
<td>IGHJ3</td>
</tr>
<tr>
<td>IGHD4</td>
<td>F</td>
<td>IGHJ4</td>
</tr>
<tr>
<td>IGHD5</td>
<td>F</td>
<td>IGHJ5</td>
</tr>
<tr>
<td>IGHD6</td>
<td>F</td>
<td>IGHJ6</td>
</tr>
</tbody>
</table>

F, functional; ORF, open reading frame; P, pseudogene. The functionality is according to IMGT functionality (IMGT Scientific chart > IMGT functionality).
submitted sequences presented IGHV F genes. Interestingly, the majority of expressed V genes were from locus B (chromosome 3). This difference was reflected in the abundance of rearrangements (~66% from locus B) and in the diversity of IGHV genes expressed: 25 IGHV from locus B vs. 17 of locus A (Figure 4A). On average, IGHV1D-25*01, IGHV6D-18*01, IGHV6D-16*01, and IGHV1-73*01 were the most abundant IGHV functional genes, accounting for 30% of the expressed repertoire.

Rainbow Trout
In this species, we analyzed AIRRseq datasets from fish intraperitoneally immunized with a killed bacterial pathogen, *Yersinia ruckeri* [data from reference (33)]. 5′RACE PCR products were produced from spleen of immunized fish, using specific primers for IGHM constant region and with unique molecular identifiers (UIDs) for better data normalization (33). Only in-frame productive rearrangements (CDR3-IMGT without stop codons) were analyzed. Trout used in this study belonged to the isogenic line derived from Swanson strain that was selected for the rainbow trout genome project. Hence, these AIRRseq data express IGHV genes from the very same repertoire, which was annotated in the current IMGT reference directories. These data therefore provided a quantitative assessment of the expression of IGHV genes in the spleen of three genetically similar individuals responding to a pathogen.

In this dataset, IMGT/High V-QUEST unambiguously identified the IGHV gene in 94% of submitted sequences, 90% of them with at least 99% of sequence identity (52% with 100% of identity). A total of 55 IGHV genes (35 functional “F,” 9 “ORF,” and 7 pseudogenes “P”) were expressed. Interestingly, these rearrangements are from both IGH loci (A and B) in relatively similar proportions.

In each trout sample, about 17% of sequences corresponded to IGHV ORF genes and 1.7–4.7% corresponded to IGHV pseudogenes (most of them correspond to IGHV1D-12*01 P or IGHV1-21*01 P) involved in-frame junction rearrangements. This feature could be detected because we selected the IMGT/HighV-QUEST directory sets “F + ORF + in-frame P,” which also include pseudogenes with in-frame stop codon in V region or defect in the leader or recombination signal (RS) sequences (3). Although IGHV transcripts with stop codon are generally rare in mammals, they are typically much more frequent in fish, perhaps because nonsense-mediated mRNA decay (NMD) may work differently (28, 32, 34, 35).

Hence, about 80% of submitted rainbow trout sequences presented functional IGHV genes (Figure 4B). IGHV4D-24*01F, IGHV6D-40*01F, IGHV1-18*01F, and IGHV11-25*01F were the most expressed on average, with a limited interindividual variation as expected from the genetic constitution of the fish analyzed. In this dataset, for about 6% of submitted sequences, IMGT/HighV-QUEST provided two results assigned to distinct duplicated germline IGHV with alleles having identical or close sequences (for example, IGHV12D56*01/IGHV12D57*01, or IGHV8-30*01/IGHV8-40*01) owing to the gene duplication in salmonids.
Figure 3: Phylogenetic analysis of Atlantic salmon and rainbow trout IGHV gene sequences. A phylogenetic analysis of the V-REGION of Atlantic salmon (Salsal) and rainbow trout (Oncmyk) IGHV genes (F + ORF) was performed using the UPGMA method and a bootstrap based on 1,000 replicates. A total of 136 genes, 81 from Salsal (69 F + 12 ORF) and 55 from Oncmyk (44 F + 11 ORF) from the IMGT/V-QUEST reference directory sets (release 201901-4, 1st August 2019) (Table 5), were compared. Only one allele per gene was included in the analysis (allele *01 for all but two IGHV8-58*02 F and IGHV8-53*03 F). Nodes with a bootstrap support higher than 75% are indicated.

Although the datasets analyzed here for salmon and trout were not selected for direct comparison, it suggests that these two species (at least, the fish strains analyzed here) do not use the two loci in the same way (see above). A rigorous and comprehensive comparison of expressed repertoires between rainbow trout and Atlantic salmon will require a systematic comparison of AIRRseq data from multiple strains.

Genetic Variability of IG Genes in Salmonids

Making available a full annotation and versatile nomenclature also offers the possibility to better integrate new data about variability of IG (or TR) genes. This issue is of particular interest in Salmonids for two main reasons: (1) variations of IG gene sequences may affect the repertoire of specificities targeted by Abs, in turn impacting the quality and efficiency of responses against pathogens, and (2) salmonid IG loci are particularly complex with high numbers of functional genes and pseudogenes located in two regions; therefore, they constitute interesting models to understand mechanisms of short-term evolution of such loci and the potential importance of homogenization vs. diversification of IG sequences.

To get preliminary data about IGHV variation in a salmonid species, we took advantage of the full genome sequencing of 19 isogenic lines of rainbow trout. These lines were produced using a mitogynogenesis-based strategy by Quillet et al. (36, 37). They represent 19 haplotypes randomly picked from the so-called INRA-SY “synthetic” population. This population was created about 35 years ago by a planned random mating (i.e., panmictic) mixture of French, Danish, and American domestic populations, and has been maintained since without any voluntary selection. The 19 isogenic lines analyzed here do not appear to be closely related to the Swanson trout generated at Washington State University using androgenesis, which has been sequenced and constitutes the reference genome (38, 39).

The numbers of indel and SNP detected within IGHV genes and pseudogenes are indicated in Table 6. Genetic variation between isogenic lines overall appears to be relatively modest at this level. It seems to be more frequent in the locus located in chromosome 13 (67 SNP and 1 indel for 29 functional genes, 41 SNP and 3 indel for 20 pseudogenes) compared to chromosome 12 (23 SNP and no indel for 29 functional genes, and 53 SNP and 10 indel for 51 pseudogenes). The proportion of silent vs. non-silent mutations was not significantly different between the two regions (40NS/67 SNP for chromosome 13 and 13NS/23 SNP for chromosome 12), suggesting that these genes did not evolve under strong positive selection. Indel
TABLE 5 | Atlantic salmon (S. salar) and rainbow trout (O. mykiss) IGHV alleles included in the IMGT/V-QUEST reference directory sets (release 201931–4, 1st August 2019).

<table>
<thead>
<tr>
<th>IGHV subgroup</th>
<th>Atlantic salmon</th>
<th>Rainbow trout</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nb of genes</td>
<td>Nb of alleles</td>
</tr>
<tr>
<td></td>
<td>F*</td>
<td>ORF*</td>
</tr>
<tr>
<td>IGHV1</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>IGHV2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>IGHV3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>IGHV4</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>IGHV5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>IGHV6</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>IGHV7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>IGHV8</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>IGHV9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>IGHV10</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>IGHV11</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>IGHV12</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>IGHV13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IGHV14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IGHV15</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>IGHV16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
<td>150</td>
</tr>
<tr>
<td>Locus A + Locus B</td>
<td>135</td>
<td>150</td>
</tr>
</tbody>
</table>

*F, functional; ORF, open reading frame; P, pseudogene. Number of genes included in the IMGT reference directory, per subgroup and per functionality and total, are shown with, if relevant (more than one allele per gene), the corresponding number of alleles within brackets. The functionality is according to IMGT functionality (IMGT Scientific chart > IMGT functionality)\(^1\). and SNP were not significantly more frequent in pseudogenes. Variants were filtered to eliminate all assembly artifacts, but these data will have to be fully validated by resequencing, and the impact of variation on the gene status evaluated. We have indications that several new genes are present in productive and expressed rearrangements. This might be due to the absence of such genes in the genome of the Swanson strain or to gaps in the current reference genome assembly. In this context, it is of interest to evaluate the variability of IGHV gene numbers between the different haplotypes. Future assemblies will allow a more accurate description of the IGH diversity and variability. Incompleteness of the annotated repertoire may constitute a problem for repertoire analysis (for example, when a missing gene is used by a clonotype clonally selected in a response). Hence, sequences of genes that are not localized in the current assembly may be added to the IMGT Reference directories sets, providing that sufficient evidence is available to demonstrate their existence and expression. These sequences will be given a provisional name (with S) until their location and presence in the germline genomic sequence are validated. If new genes would appear, which do not belong to any of the IGHV subgroups identified and described in this work, a new subgroup may have to be defined. This is not impossible, but seems to be unlikely since we believe that the large set of IGHV sequences analyzed from Atlantic salmon and rainbow trout probably contains at least one representative of all subgroups. Such additions will be validated by the IG, TR, and MH Nomenclature Sub-Committee (IMGT-NC) \(^6\,\,7\) of the IUIS Nomenclature Committee\(^2,\,\,3\), following a procedure analogous to the one used for example for inferred alleles in human.

**CONCLUSION**

Genome assembly is available for both Atlantic salmon and rainbow trout, representing the two main genera of Salmonid (Salmo and Oncorhynchus). More genomic (and transcriptome) data are coming from a number of genomic backgrounds, which will provide a rich source of knowledge about variations of potential antibody repertoires in these species. We therefore revisited the description and annotation of the two IGH loci present in these two species, currently from cDNA and BAC clone sequences, based on the IMGT biocuration and nomenclature for Salmonid IGH genes that will facilitate the analysis of AIRR-seq data. The IG or antibody repertoire sequencing has started to develop both in rainbow trout and in Atlantic salmon, reflecting a growing interest for an accurate and comprehensive description of the response against common pathogens and vaccines. As full genome assemblies are now available for several salmonid species (Atlantic salmon, rainbow trout, coho salmon, and chinook salmon), comparative analysis of the IGH locus structure in
these closely related tetraploidized species is of great interest. It also appears very important to investigate the level of variation between germline repertoires of IG genes across commercial and wild salmonid stocks. This variation may have significant implications for practical issues in aquaculture and conservation; it will also be of significant interest for the basic comparative immunology community, in particular to address accurately the mechanisms of gene conversion, somatic hypermutation, and memory in these species and during vertebrate evolution.
### Table 6: Number of SNP and variants in IGHV genes and pseudogenes across 19 isogenic rainbow trout lines.

<table>
<thead>
<tr>
<th>Chrom</th>
<th>Start</th>
<th>Stop</th>
<th>Name</th>
<th>SNPs number (NS)</th>
<th>Indel number</th>
<th>Start</th>
<th>Stop</th>
<th>Name</th>
<th>SNPs number (NS)</th>
<th>Indel number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr12</td>
<td>81 322 385</td>
<td>81 322 680</td>
<td>IGHV6D-76</td>
<td>2(1)</td>
<td>0</td>
<td>81 312 817</td>
<td>81 313 128</td>
<td>IGHV6D-79</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 335 727</td>
<td>81 336 024</td>
<td>IGHV1D-73</td>
<td>2(2)</td>
<td>0</td>
<td>81 318 367</td>
<td>81 318 661</td>
<td>IGHV1D-78</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 339 680</td>
<td>81 339 683</td>
<td>IGHV12D-71</td>
<td>1(0)</td>
<td>0</td>
<td>81 320 821</td>
<td>81 321 111</td>
<td>IGHV1D-77</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 365 089</td>
<td>81 365 411</td>
<td>IGHV15D-69</td>
<td>0</td>
<td>0</td>
<td>81 332 711</td>
<td>81 333 058</td>
<td>IGHV2D-75</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 365 950</td>
<td>81 366 255</td>
<td>IGHV2D-68</td>
<td>0</td>
<td>0</td>
<td>81 333 678</td>
<td>81 333 966</td>
<td>IGHV1D-74</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 396 848</td>
<td>81 396 168</td>
<td>IGHV7D-62</td>
<td>0</td>
<td>0</td>
<td>81 336 977</td>
<td>81 337 276</td>
<td>IGHV4D-72</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 397 765</td>
<td>81 398 073</td>
<td>IGHV4D-60</td>
<td>0</td>
<td>0</td>
<td>81 383 887</td>
<td>81 384 159</td>
<td>IGHV1D-65</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 422 492</td>
<td>81 422 803</td>
<td>IGHV15D-54</td>
<td>0</td>
<td>0</td>
<td>81 384 684</td>
<td>81 384 968</td>
<td>IGHV6D-64</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 436 861</td>
<td>81 437 166</td>
<td>IGHV2D-50</td>
<td>0</td>
<td>0</td>
<td>81 388 213</td>
<td>81 388 533</td>
<td>IGHV1D-63</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 438 847</td>
<td>81 439 169</td>
<td>IGHV15D-49</td>
<td>0</td>
<td>0</td>
<td>81 396 766</td>
<td>81 397 086</td>
<td>IGHV1D-61</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 488 600</td>
<td>81 488 620</td>
<td>IGHV7D-45</td>
<td>0</td>
<td>0</td>
<td>81 398 323</td>
<td>81 398 601</td>
<td>IGHV1D-59</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 485 762</td>
<td>81 486 082</td>
<td>IGHV7D-44</td>
<td>0</td>
<td>0</td>
<td>81 399 513</td>
<td>81 399 812</td>
<td>IGHV4D-58</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 486 761</td>
<td>81 486 069</td>
<td>IGHV4D-43</td>
<td>0</td>
<td>0</td>
<td>81 401 688</td>
<td>81 402 032</td>
<td>IGHV12D-57</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 493 829</td>
<td>81 494 124</td>
<td>IGHV6D-40</td>
<td>0</td>
<td>0</td>
<td>81 402 044</td>
<td>81 402 727</td>
<td>IGHV12D-56</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 529 173</td>
<td>81 529 478</td>
<td>IGHV1D-35</td>
<td>0</td>
<td>0</td>
<td>81 421 382</td>
<td>81 421 688</td>
<td>IGHV1D-55</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 561 752</td>
<td>81 562 063</td>
<td>IGHV2D-30</td>
<td>0</td>
<td>0</td>
<td>81 434 530</td>
<td>81 434 810</td>
<td>IGHV2D-52</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 568 866</td>
<td>81 569 171</td>
<td>IGHV2D-28</td>
<td>0</td>
<td>0</td>
<td>81 435 683</td>
<td>81 435 988</td>
<td>IGHV2D-51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 595 836</td>
<td>81 595 138</td>
<td>IGHV10D-26</td>
<td>2(0)</td>
<td>0</td>
<td>81 447 972</td>
<td>81 448 244</td>
<td>IGHV1D-48</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 605 108</td>
<td>81 605 419</td>
<td>IGHV4D-24</td>
<td>1(0)</td>
<td>0</td>
<td>81 448 773</td>
<td>81 449 068</td>
<td>IGHV6D-47</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 618 281</td>
<td>81 618 689</td>
<td>IGHV4D-23</td>
<td>2(2)</td>
<td>0</td>
<td>81 453 306</td>
<td>81 453 632</td>
<td>IGHV1D-46</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 649 810</td>
<td>81 650 108</td>
<td>IGHV1D-17</td>
<td>5(3)</td>
<td>0</td>
<td>81 467 316</td>
<td>81 467 588</td>
<td>IGHV1D-42</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 653 372</td>
<td>81 653 678</td>
<td>IGHV1D-16</td>
<td>0</td>
<td>0</td>
<td>81 481 949</td>
<td>81 482 241</td>
<td>IGHV1D-41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 661 001</td>
<td>81 661 301</td>
<td>IGHV1D-15</td>
<td>0</td>
<td>0</td>
<td>81 511 235</td>
<td>81 511 555</td>
<td>IGHV1D-39</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 675 466</td>
<td>81 675 769</td>
<td>IGHV1D-12</td>
<td>2(2)</td>
<td>0</td>
<td>81 514 353</td>
<td>81 514 683</td>
<td>IGHV1D-38</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 696 523</td>
<td>81 696 828</td>
<td>IGHV2D-11</td>
<td>1(0)</td>
<td>0</td>
<td>81 516 973</td>
<td>81 517 272</td>
<td>IGHV4D-37</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 700 731</td>
<td>81 701 033</td>
<td>IGHV10D-9</td>
<td>0</td>
<td>0</td>
<td>81 519 248</td>
<td>81 519 565</td>
<td>IGHV12D-36</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 705 764</td>
<td>81 706 066</td>
<td>IGHV10D-7</td>
<td>2(1)</td>
<td>0</td>
<td>81 548 116</td>
<td>81 548 425</td>
<td>IGHV1D-34</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 717 198</td>
<td>81 717 494</td>
<td>IGHV6D-5</td>
<td>3(2)</td>
<td>0</td>
<td>81 550 863</td>
<td>81 551 152</td>
<td>IGHV1D-33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr12</td>
<td>81 737 673</td>
<td>81 737 973</td>
<td>IGHV1D-4</td>
<td>0</td>
<td>0</td>
<td>81 558 855</td>
<td>81 559 137</td>
<td>IGHV15D-32</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23(13)</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>53</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 6 | Continued

<table>
<thead>
<tr>
<th>Chrom</th>
<th>Start</th>
<th>Stop</th>
<th>Name</th>
<th>SNPs number (NS)</th>
<th>Indel number</th>
<th>Start</th>
<th>Stop</th>
<th>Name</th>
<th>SNPs number</th>
<th>Indel number</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 030 797</td>
<td>48 031 104</td>
<td>IGH</td>
<td>2(1)</td>
<td>0</td>
<td>48 138 071</td>
<td>48 138 427</td>
<td>IGHV8-29</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 034 515</td>
<td>48 034 814</td>
<td>IGHV6-46</td>
<td>1(0)</td>
<td>0</td>
<td>48 027 352</td>
<td>48 027 666</td>
<td>IGHV15-48</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 054 181</td>
<td>48 054 484</td>
<td>IGHV1-42</td>
<td>0</td>
<td>0</td>
<td>48 046 874</td>
<td>48 047 207</td>
<td>IGHV9-45</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 073 234</td>
<td>48 073 536</td>
<td>IGHV8-40</td>
<td>0</td>
<td>0</td>
<td>48 048 080</td>
<td>48 048 362</td>
<td>IGHV4-44</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 077 115</td>
<td>48 077 414</td>
<td>IGHV1-39</td>
<td>1(1)</td>
<td>0</td>
<td>48 051 342</td>
<td>48 051 671</td>
<td>IGHV1-43</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 082 080</td>
<td>48 082 391</td>
<td>IGHV16-37</td>
<td>3(3)</td>
<td>0</td>
<td>48 068 027</td>
<td>48 068 332</td>
<td>IGHV1-41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 093 298</td>
<td>48 093 597</td>
<td>IGHV1-36</td>
<td>2</td>
<td>0</td>
<td>48 079 966</td>
<td>48 080 277</td>
<td>IGHV16-38</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 104 897</td>
<td>48 105 217</td>
<td>IGHV6-35</td>
<td>0</td>
<td>0</td>
<td>48 108 554</td>
<td>48 108 889</td>
<td>IGHV9-34</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 109 683</td>
<td>48 109 994</td>
<td>IGHV14-33</td>
<td>0</td>
<td>0</td>
<td>48 116 928</td>
<td>48 117 231</td>
<td>IGHV2-27</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 122 499</td>
<td>48 122 783</td>
<td>IGHV6-32</td>
<td>0</td>
<td>0</td>
<td>48 147 810</td>
<td>48 148 106</td>
<td>IGHV6-26</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 127 168</td>
<td>48 127 466</td>
<td>IGHV6-31</td>
<td>0</td>
<td>0</td>
<td>48 157 816</td>
<td>48 158 168</td>
<td>IGHV9-24</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 135 329</td>
<td>48 135 631</td>
<td>IGHV6-30</td>
<td>0</td>
<td>0</td>
<td>48 211 869</td>
<td>48 212 178</td>
<td>IGHV7-17</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 145 742</td>
<td>48 146 047</td>
<td>IGHV2-28</td>
<td>7(3)</td>
<td>0</td>
<td>48 244 719</td>
<td>48 245 034</td>
<td>IGHV10-12</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 148 981</td>
<td>48 149 284</td>
<td>IGHV11-25</td>
<td>0</td>
<td>0</td>
<td>48 245 867</td>
<td>48 246 172</td>
<td>IGHV8-11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 164 983</td>
<td>48 165 317</td>
<td>IGHV9-23</td>
<td>4(3)</td>
<td>0</td>
<td>48 254 879</td>
<td>48 254 570</td>
<td>IGHV16-9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 166 898</td>
<td>48 167 189</td>
<td>IGHV4-22</td>
<td>9(7)</td>
<td>0</td>
<td>48 279 548</td>
<td>48 279 841</td>
<td>IGHV1-7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 168 162</td>
<td>48 168 465</td>
<td>IGHV1-21</td>
<td>5(3)</td>
<td>0</td>
<td>48 280 387</td>
<td>48 280 681</td>
<td>IGHV4-6</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 174 816</td>
<td>48 175 127</td>
<td>IGHV3-20</td>
<td>4(2)</td>
<td>0</td>
<td>48 316 059</td>
<td>48 315 761</td>
<td>IGHV6-3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 191 970</td>
<td>48 192 272</td>
<td>IGHV8-19</td>
<td>6(5)</td>
<td>0</td>
<td>48 339 577</td>
<td>48 339 885</td>
<td>IGHV4-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 201 668</td>
<td>48 201 973</td>
<td>IGHV1-18</td>
<td>2(0)</td>
<td>0</td>
<td>48 076 216</td>
<td>48 076 475</td>
<td>IGHV13-39</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 222 441</td>
<td>48 222 126</td>
<td>IGHV9-16</td>
<td>1(1)</td>
<td>1</td>
<td>48 222 441</td>
<td>48 222 126</td>
<td>IGHV9-16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chr13</td>
<td>48 223 844</td>
<td>48 223 544</td>
<td>IGHV9-15</td>
<td>4(2)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 237 588</td>
<td>48 237 899</td>
<td>IGHV16-14</td>
<td>6(6)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 243 688</td>
<td>48 243 987</td>
<td>IGHV1-13</td>
<td>3(0)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 250 487</td>
<td>48 250 789</td>
<td>IGHV1-10</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 257 525</td>
<td>48 257 828</td>
<td>IGHV2-8</td>
<td>2(1)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 307 972</td>
<td>48 307 670</td>
<td>IGHV8-5</td>
<td>7(2)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 312 160</td>
<td>48 311 865</td>
<td>IGHV6-4</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr13</td>
<td>48 327 417</td>
<td>48 327 761</td>
<td>IGHV1-2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study can be found in the www.imgt.org – accession numbers can be found within the manuscript. Any other data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

**AUTHOR CONTRIBUTIONS**

SMa, AK, M-PL, and PB conceived the project and wrote the manuscript. SMa, AK, IS, and PB designed experiments. SMa, AK, SH-S, SA, SMo, DL, RC, IS, OS, JH, BK, M-PL, and PB performed data analysis. SMa, AK, DL, and IS provided resources. All authors contributed to manuscript revision, and read and approved the submitted version.

**FUNDING**

This work was supported by Xunta de Galicia (ED431C 2016/041), by the Norwegian Research Council (ImCom Grant 267644), and by USDA AFRI Grant 2DN70-2DN7 (to IS). SMa also received funding from the People Programme Marie Curie Actions (FP7/2007-2013, REA Grant Agreement 600391). DL, RC, and PB were supported by institutional grants for Institut National de la Recherche Agronomique. RC and PB were supported by the ANR grant FishRNAvax.

**ACKNOWLEDGMENTS**

We are thankful to the entire IMGT® team in developing databases and tools. This work was granted access to the HPC@LR and to the High-Performance Computing (HPC) resources of the Centre Informatique National de l’Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat à l’Energie Atomique et aux Énergies Alternatives (CEA) under the allocation [036029] (2010–2019) made by GENCI (Grand Equipement National de Calcul Intensif). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.
SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02541/full#supplementary-material

Figure S1 | Alignment of the D-GENE-UNIT sequences of the IGHD (diversity) genes located upstream of the IGHM (focus A) and IGHMD (focus B) genes of Salmo salar (Salsa) and Oncorhynchus mykiss (On=myk). Genes of the locus B are identified by the letter D which follows the gene number. Labels are according to the D-GENE prototype (IMGT Scientific chart > 1. Sequence and 3D structure identification and description > IMGT prototypes table > D-GENE). The highly conserved FDYWGKGTXVT motif is pink highlighted and those residues that deviated from it are in red.

REFERENCES

31. Nordvik I. Identification of a novel immunoglobulin delta transcript and comparative analysis of the genes encoding IgD in Atlantic salmon.


**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.

Copyright © 2019 Magadan, Krasnov, Hadi-Saljoqi, Afanasyev, Mondot, Lallias, Castro, Salinas, Sunyer, Hansen, Koop, Lefranc and Boudinot. 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.