Evolution of duplicated \textit{IgH} loci in Atlantic salmon, \textit{Salmo salar}

Motoshige Yasuike\textsuperscript{1}, Johan de Boer\textsuperscript{1}†, Kristian R von Schalburg\textsuperscript{1}, Glenn A Cooper\textsuperscript{1}, Linda McKinnel\textsuperscript{1}, Amber Messmer\textsuperscript{1}, Stacy So\textsuperscript{1}, William S Davidson\textsuperscript{2}, Ben F Koop\textsuperscript{1}*

Abstract

\textbf{Background:} The Atlantic salmon (\textit{Salmo salar}) immunoglobulin heavy chain (\textit{IgH}) locus possesses two parallel \textit{IgH} isoloci (\textit{IGH-A} and \textit{IGH-B}), that are related to the genomic duplication event in the family Salmonidae. These duplicated \textit{IgH} loci in Atlantic salmon provide a unique opportunity to examine the mechanisms of genome diversity and genome evolution of the \textit{IgH} loci in vertebrates. In this study, we defined the structure of these loci in Atlantic salmon, and sequenced 24 bacterial artificial chromosome (BAC) clones that were assembled into the \textit{IGH-A} (1.1 Mb) and \textit{IGH-B} (0.9 Mb) loci. In addition, over 7,000 cDNA clones from the \textit{IgH} variable (VH) region have been sequenced and analyzed.

\textbf{Results:} The present study shows that the genomic organization of the duplicated \textit{IgH} loci in Atlantic salmon differs from that in other teleosts and other vertebrates. The loci possess multiple \textit{C\tau} genes upstream of the \textit{C\mu} region, with three of the \textit{C\tau} genes being functional. Moreover, the duplicated loci possess over 300 VH segments which could be classified into 18 families. This is the largest number of VH families currently defined in any vertebrate. There were significant structural differences between the two loci, indicating that both \textit{IGH-A} and \textit{-B} loci have evolved independently in the short time after the recent genome duplication approximately 60 mya.

\textbf{Conclusions:} Our results indicate that the duplication of the \textit{IgH} loci in Atlantic salmon significantly contributes to the increased diversity of the antibody repertoire, as compared with the single \textit{IgH} locus in other vertebrates.

Background

The adaptive immune system based on somatic recombination of immune receptor genes appeared in vertebrates some 500 million years ago (mya) [1,2]. While jawless vertebrates, such as lamprey and hagfish, assemble their variable lymphocyte receptors (VLRs) through recombination of leucine-rich repeat (LRR) modular units [3-5], jawed vertebrates generate their diverse repertoire of B and T cell antigen receptors through rearrangement of variable-(diversity)-joining (V-(D)-J) gene segments [6,7].

Immunoglobulins (Igs) are key molecules within the jawed vertebrate humoral immune system that are generated by the B cells for defence against a wide variety of pathogens. Igs are composed of two heavy (H) chains and two light (L) chains that are encoded by the \textit{IgH} locus and \textit{IgL} locus, respectively. Two different types of genomic rearrangement of the \textit{IgH} locus have evolved. In cartilaginous fishes, such as sharks and skates, closely linked individual clusters of VH-D-D-J-Constant (CH) gene segments are repeated 100 - 200 times [8,9]. In contrast, in most bony vertebrates (from teleost fishes to mammals), the VH, D, JH, and CH gene segments are in tandem arrays, also known as translocon organisation (VH)\textsubscript{n}-(D)\textsubscript{m}-(JH)\textsubscript{x}-(CH)\textsubscript{y} [10-13]. The contribution of multiple germ line VH, D and JH gene segments to antibody diversity is magnified by the random rearrangement of these segments in somatic cells [14]. In response to an antigen, mature B cells can change their expressed CH region genes, and the different CH region genes that possess different effector functions. In mammals, there are five Ig classes, named for their CH region component as IgM (\(\mu\) chain), IgD (\(\delta\) chain), IgG (\(\gamma\) chain), IgA (\(\alpha\) chain) and IgE (\(\epsilon\) chain) [9,15]. In teleosts, the predominant serum antibody is IgM, which

* Correspondence: bkoop@uvic.ca
† Contributed equally
\textsuperscript{1}Department of Biology, University of Victoria, PO Box 3020 STN CSC, Victoria, British Columbia, V8W 3N5, Canada
Full list of author information is available at the end of the article

© 2010 Yasuike et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
was the first Ig class identified [13]. Subsequently, IgD was also found, but the teleost IgD gene is expressed as a chimeric transcript that includes the first exon of the IgM gene (Cμ) [16-20]. It has long been believed that teleost fish possess only IgM and IgD. However, novel Ig classes (IgTs) have recently been found in zebrafish (Danio rerio) (IgZ) [21], rainbow trout (Oncorhynchus mykiss) [22], fugu (Fugu rubripes) (novel IgH) [23] and carp (Cyprinus carpio) (IgM-IgZ chimera) [24] as more sequences have become available.

In the zebrafish and rainbow trout IgH locus, a CH gene (Cτ) of the novel Ig class (IgT) exists upstream of Cμ and Cδ genes, possessing its own D and JH segments [21,22]. This organization of genes in the zebrafish and rainbow trout IgH loci resembles the mouse T cell receptor (TCR) α/δ locus (TRA/TRD) [21]. Similarly, the fugu ortholog of IgT (novel IgH) is also found upstream of Cμ and Cδ genes and this Ig has its own D and JH segments, but the gene organization of the fugu IgT differs significantly from zebrafish and rainbow trout IgTs [23]. However, in catfish, a CH region upstream of Cμ and Cδ genes similar to IgT has not been found [25]. In addition, the catfish IgH locus contains three linked pairs of Cμ and Cδ genes, but only one Cμ and possibly three Cδ genes are functional [25-27]. The three different Cδ gene regions encode heavy chains of membrane and secreted IgD, and although secreted IgD has so far been found only in catfish, it may not contain a functional V-region [26]. Recently, it has been reported that the stickleback (Gasterosteus aculeatus) IgH locus contains three tandem duplicated Cτ, Cμ and Cδ genes separated by VH, D, and JH segments, as well as a fourth Cτ gene in the 3′ end of the locus [28,29]. These findings indicate that there is a large amount of variability within the IgH loci among teleosts.

One interesting feature of the Atlantic salmon (Salmo salar) IgH locus is that it possesses two parallel IgH iso-loci (IGH-A and IGH-B) [12,30-32], that are related to the tetraploid ancestry of the family Salmonidae [30,33]. A recent study by Shihina et al. (2005) estimated the duplication event to have taken place approx. 60 mya based on sequence divergence of duplicated MHC class I regions of rainbow trout [34] Recently, the presence of two IgH loci have also been demonstrated in rainbow trout by in situ hybridization to rainbow trout chromosomal spreads with IgH-positive BAC clones [22]. However, only approximately 100 kb of the 3′ end of one rainbow trout IgH locus has been sequenced to date. Two IgM isotypes were found in Atlantic salmon and brown trout (Salmo trutta), while it has been suggested through gel filtration analysis that rainbow trout and Arctic char (Salvelinus alpinus) possess a single IgM [35]. Moreover, only one Cμ cDNA has been found from a single homozygous rainbow trout, whereas duplicated versions of the rainbow trout Cτ and Cδ genes have been suggested from cDNA variants [22]. These findings suggest that rainbow trout and Arctic char lost an intact IgM in evolution after the genera Salmo, Oncorhynchus and Salvelinus radiated (10 - 18 mya) [35,36]. Thus, determination of the structure of the loci in Atlantic salmon provides a unique opportunity for understanding the evolution of the IgH locus in salmonids.

In this study, to define the structure of the loci in Atlantic salmon, 24 bacterial artificial chromosome (BAC) clones were sequenced and complete IGH-A (1.1 Mb) and IGH-B (0.9 Mb) loci were assembled. In addition, over 7,000 clones from the IgH variable (VH) region cDNAs have been sequenced and analyzed. The present study shows that the Atlantic salmon IgH locus represents the most complex and diverse vertebrate IgH locus characterized to date.

Results

Overall organization of IGH-A and IGH-B

Two loci were assembled from overlapping BAC sequences. IGH-A was assembled from 7 BACs, and IGH-B was assembled using 8 BACs. A few contigs, internal to BACs could not be joined, resulting in two contigs for IGH-A and four contigs for IGH-B, that are separated by small gaps of unknown length in regions of repeated sequences. The sequences in the two IgH loci containing VH and CH regions cover approximately 670 kb (IGH-A) and 710 kb (IGH-B), respectively. Part of a second allele for IGH-A was identified in the assembly of a number of BAC sequences. We noted a similarity of 99.7% over contigs spanning 190 kb sequence (data not shown).

In each of the two IgH loci we identified numerous VH segments, many D and JH segments and several CH gene segments (Table 1, Figure 1, Additional file 1 and Additional file 2). The CH sequences are comprised of three classes, one Cμ and one Cδ in each locus and 3 and 5 complete or partial Cτ sequences, respectively. The Cμ and Cδ sequences are the most 3′ elements in both loci while the constant Cτ sequences are distributed throughout the loci. D and JH sequences are generally 5′ of the Cμ and Cτ sequences. Most interestingly, the region that contains the VH sequences coincides with the region that contains a large number of “Nhe I” elements, piggyBAC-like sequences that have also been recovered numerous times in the V region of the Atlantic salmon TRA/TRD locus [37].

The IGH-A and -B loci show 81-85% sequence identity surrounding the VH sequence region, but internally less identity than 81-85%, indicating significant reorganization involving the VH gene and Nhel-elements. Additional file 3 lists identified genes flanking the loci. A
dotter plot of IGH-A versus IGH-B is shown in Additional file 4. It is worth noting the similarity between the sequences flanking the two loci as this is in stark contrast to the lack of similarity between the loci themselves.

Table 1 Summary of CH, D, JH gene segments in the duplicated loci.

|       | CH | D  | JH   | in EST | comment                  |
|-------|----|----|------|--------|--------------------------|
| IGH-A | τA | -  | -    | no     | exon 4 only              |
|       | τA | 2  | 2    | no     | exon 1 and 2             |
|       | τA | 0  | 2    | no     | has FS                   |
|       | τA | 2 (5) 3 (3) | 2   | yes   | reverse orientation      |
| μA    | 1  | 2  | yes  |        |                          |
| δA    | 6 + 3 b | 5 | yes  |        |                          |
| IGH-B | τB | -  | 1    | no     | has FS                   |
|       | τB | 2  | 2    | yes   |                          |
|       | τB | 0  | 0    | no     | exon 1                   |
| μB    | 6  | 5  | yes  |        |                          |
| δB    | -  | -  | yes  |        |                          |

a 5' and 3' relative to the τA-4 gene orientation, which is in reverse orientation compared to the rest of the locus.
b The 3 D sequences are located approximately 20 kb 3' from the 6 D sequences and approximately 13 kb 5' of the J sequences.

Constant (CH) regions
Each locus contains several CH gene sequences, Cμ, Cδ, and Cτ. At the 3' end of each locus is one Cμ sequence followed by one Cδ sequence as previously reported [17,38]. The Cμ and Cδ sequences are approximately 98% similar between loci. Surprisingly, there are several Cτ sequences, 5 in IGH-A and 3 in IGH-B, spread out over each locus (Figure 1). The Cτ sequences in the IGH-A are as follows: starting from the 5' end; 1) CτA-1, partial (most of exon 4 only), 2) CτA-2, partial (5' start to approximately 40 base pairs into exon 3), 3) CτA-3, complete but has a frameshift, 4) CτA-4, complete but in reverse orientation, and 5) CτA-5, complete (Figure 1). In IGH-B there is 1) CτB-1, a complete sequence but with a frameshift, 2) CτB-2, complete, and 3) CτB-3, a partial (missing exon 1) (Figure 2). CτB-1 and CτB-2 are >99% identical. There are a total of two intact Cτ genes in IGH-A and one intact Cτ gene in IGH-B (Table 1 and Figure 2). The alignments of the intact CH gene amino acid sequences are available in Additional file 5. We constructed a phylogenetic tree from the translated sequences of the three intact Cτ genes and other CH genes (see Additional file 6). The three intact Cτ genes clustered within a branch containing teleost Cζ/τ sequences. Interestingly, CτA (CτA-4 and CτA-5) shared a branch with the rainbow trout Cτ, and CτB-2 branched basal to the CτA/rainbow trout Cτ clade. It has been reported that both duplicated IgD genes in Atlantic

```
Figure 1 Organization of the Atlantic salmon duplicated IgH loci. The sequences in the two IgH loci containing VH and CH regions cover approximately 670 kb (IGH-A) and 710 kb (IGH-B), respectively. The positions of the small gaps are indicated by triangles (IGH-A gap at position 760.2 kb, IGH-B gaps at position 282.8 kb, 372.8 kb and 555.4 kb). The numbers near VH genes indicate the VH genes family numbers. Transcriptional directions for the CH genes are shown by arrowheads. A numerated version of Figure 1 is available in Additional file 1 (IGH-A) and Additional file 2 (IGH-B).
```

Yasuike et al. BMC Genomics 2010, 11:486
http://www.biomedcentral.com/1471-2164/11/486
Page 3 of 16
salmon have a tandem duplication of $C_\delta_2-C_\delta_3-C_\delta_4$ [17,38]. However, our present study shows that the number of times exons ($C_\delta_2-C_\delta_3-C_\delta_4$) are repeated is different between loci, three times in $IGH-A$ and four times in $IGH-B$ (Figure 2). The sequence identity between these repeats of the two loci is very high (>98%), suggesting gene conversion events. The three functional $C_\tau$ genes ($C_\tau_B-2$, $C_\tau_A-4$ and $C_\tau_A-5$), the two $C_\mu$ genes ($C_\mu_A$ and $C_\mu_B$) and $C_\delta$ genes ($C_\delta_A$ and $C_\delta_B$) have been submitted to Genbank and the accession numbers are listed in Additional file 7.

**Diversity (D) segments**

D gene sequences were identified through the conserved pattern of their recombination signal sequences (RSS). A total of 25 D sequence genes were identified in the two $IgH$ loci (Figure 1). All segments are flanked on each end by a consensus nonamer-heptamer combination that is separated by 12 base pairs (see Additional file 8). Nine are located in the $IGH-A$ 5' of the $C_\mu$ gene, in a group of 6 and a group of 3. There are eight D segments associated with the three $C_\tau$ genes (0 for $C_\tau_A-1$ which has no 5’ part; 2 for $C_\tau_A-2$ which has exon 1 and 2 only; 0 for $C_\tau_A-3$; 2 for $C_\tau_A-4$ with also 3 on the 3’ side of the inverted $C_\tau_A-4$ (which is 5’ in the assembly) and 1 for $C_\tau_A-5$) (Table 1). In $IGH-B$ there are six D segments associated with the $C_\mu$ gene and two with one of the $C_\tau$ genes ($C_\tau_B-2$) (Table 1). Comparison of the sequences of the six D elements in each of the two groups associated with the $C_\mu$ genes in $IGH-A$ and $IGH-B$ indicates that the D elements occur in the same order in both loci. The additional group of three in $IGH-A$ is similar to the last three of the groups of six, indicating a duplication event (see Additional file 9).

**Joining (JH) segments**

Each of the two $C_\mu$ genes is preceded by five JH sequences as previously reported [12]. In addition, two JH segments are associated with each $C_\tau$ gene, except $C_\tau_A-1$ which has no 5’ region and $C_\tau_B-1$ which has only
a single JH segment, for a total of seven JH segments for the Cτ genes in IGH-A and three JH segments for the Cτ genes in IGH-B, located just 5’ upstream of their respective Cτ gene (Table 1). All segments have a fairly conserved 5’ RSS, a nonamer-heptamer combination separated by 24 base pairs (see Additional file 10). A 3’ AGGT splice site is found in 18 sequences and a TGGT site in 2 sequences. A translation of the coding sequence reveals a highly conserved FDYWGKTVTYS amino acid sequence. One of the JH sequences (JH-tq1-1) is a pseudogene as it is interrupted by a TAG stop codon.

The corresponding JH sequences in IGH-A and IGH-B are identical in coding sequence, except for JH-μA-3 and JH-μB-3 [12], and therefore cannot be distinguished in rearranged products.

Variable (VH) segments

Each locus contains a large number of VH genes and pseudogenes (Figure 1). A total of 153 sequences in IGH-A and 161 sequences in IGH-B were identified as matching VH gene sequences. 99 sequences in IGH-A and 103 sequences in IGH-B were characterized and named. Of these, 23 in IGH-A and 32 in IGH-B have a putative open reading frame (ORF). The deduced amino acid sequences of these VH genes have been submitted to Genbank and the accession numbers are listed in Additional file 7. The alignments of VH genes amino acid sequences are available in Additional file 11. Many other sequences are found only as fragments. VH genes start with a more or less consensus 5’-ATG(C/T)AAA (G/T)-3’ octamer sequence [39] located 5’ to the site of transcription initiation, and terminate at a nonamer-heptamer RSS. VH genes have a short exon 1 and a long exon 2 sequence. When complete exon 2 sequences (without the RSS) are aligned and 75% identity is applied, 18 families can be distinguished. Representative sequences from the 13 families that were identified in Oncorhynchus mykiss [40,41] align within 13 of these 18 families (Figure 3). The distribution of the VH families between IGH-A and -B is listed in Additional file 12. The number of sequences identified per family varies widely, from a single copy (in family 18) to 18 members (see Additional file 12). The orientation of the VH sequences indicates some rearrangement within the loci and an inversion event is evident when comparing IGH-A and IGH-B sequence (data not shown). This inversion event explains the inverse orientation of the CτA-4 sequence in IGH-A.

The use of VH sequences was grouped by family for analysis. VH families are used to a different extent by different constant genes. For example, family 8 is used by CμA much more frequently than by CμB, while the opposite is true for family 6 VH genes (Figure 4). The most commonly used gene families include families 1, 6,
and 8, members of which comprise 60% of the putative VH ORFs. Families that contain few members with putative ORFs are also rarely recovered in ESTs.

**Rearrangements**

A comprehensive set of VH specific primers (178 VH specific primers) were constructed to complement CH specific primers and ~12 clones from all positive products sequenced to identify expression and rearrangement patterns. More than 7,000 VH-D-JH-CH cloned PCR products amplified from the kidney and spleen of two healthy individuals. We found three main types of rearrangements with VH sequences; those with a Cμ gene, those with a Cτ gene, and those with both Cμ and Cδ exons. However, not all rearrangements involving Cδ include Cμ sequence.

Of 1,872 sequences generated from Cμ-specific primers, located in exon 2, 1,794 contained a conserved sequence in exon 1 of the Cμ sequence and were further analyzed. After removal of identical sequences, confirmation of an ORF in the amplified fragment, and a minimum of 98% match (BLAST) over 30 base pairs of variable sequences, a total of 140 unique sequences were obtained containing the CτB-2 gene, 284 sequences with the CτA-4 gene, and 39 sequences with the CτA-5 gene. Interestingly, these three genes are the putatively functional Cτ sequences, with the most frequently used CτA-4 gene which is in an inverted orientation. Two instances were observed where rearrangement took place with CτA-1, which has only the first 2 exons. Both occurred with
VH sequences that are 5’ of CrA-1. The inverted CrA-4 gene is found associated with VH sequences from both 5’ and 3’ of the constant sequence. In those instances where the joining sequence can be uniquely identified, we note that the Crb-2 gene is associated with JH-\(\text{t}\)-1, rather than JH-\(\text{t}\)-2-2, 95% of the time.

cDNA sequences containing C\(\delta\) were obtained with primers located in exon 3 of the C\(\delta\) sequence. There were 832 sequences of those amplified with the C\(\delta\) specific primers that contained C\(\delta\) sequences. Approximately 90% of the sequences are chimeric transcripts that contain the first exon of C\(\mu\) (C\(\mu\)1). However, in approximately 10% of the rearrangements involving C\(\delta\), no C\(\mu\) sequence and generally no use of the normal splice and recombination signals is evident. A JH and VH sequence is not obvious in all of these cases in the readable sequence and in many of these rearrangements joining takes place from somewhere inside or just after the VH sequence to various distributed sites inside exon 1, 2, or 3 of the C\(\delta\) sequence (Table 2). In one recovered rearrangement, the first RSS of the C\(\mu\)-D4 sequence and the second RSS of the next C\(\mu\)-D5 sequence is used, resulting in the use of two D sequences including a 335 bp intervening genomic sequence (est 007-171). Interestingly, in many of these atypical rearrangements, joining occurs at a short repeat sequence present at both joining ends (see Table 2 for examples), indicative of a homology-directed recombination event. Because many of the C\(\delta\) sequences are too similar in sequence to distinguish them unequivocally, we were unable to distinguish the distribution of JH sequences between the C\(\delta\)_A and C\(\delta\)_B. However, the data still shows a preference for the middle JH by C\(\delta\) as seen in C\(\mu\)_A and C\(\mu\)_B sequences (Figure 5).

In up to 20% of the rearrangements the variable sequence was identified as from one locus and the constant sequence from the other locus. For example, C\(\mu\)a was found rearranged with members of family 1 and 8 from the IGH-B, and C\(\mu\)b was found rearranged with members of family 1 and 6 from the IGH-A. However, twenty analyzed EST sequences in our EST database [42] that contained a C\(\mu\) sequence, contained a VH sequence from the same locus (data not shown). The locus origin of the JH sequences in the rearrangements could not be unambiguously established. Nevertheless, in a number of these rearrangements, the point of crossover appears to be located in the amplified part of the C\(\mu\) sequence based on the five single nucleotide differences between the two loci.

**Expression of the Atlantic salmon Ig genes**

The tissue distribution for four different forms of IgTs, IgT-B2, IgT-A3, IgT-A4 and IgT-A5, was examined by RT-PCR (Figure 6). Figure 6 represents the results from analysis of 12 tissues from 3 different adult individuals. The IgT genes were expressed at high levels in the kidney and spleen. It has been reported that other teleost IgM and IgD genes were also primarily expressed in kidney and spleen [16,18,22]. The kidney and spleen are the major lymphoid organs of teleosts [43,44]. The teleost anterior kidney is a main site for B lymphogenesis, while the teleost posterior kidney provides an environment capable of inducing B cell activation and differentiation into plasma cells [45]. The teleost spleen functions as a major secondary immune organ, as in mammalian species. Mature B cells are abundant at this site, and Ig-secreting cells have been detected from splenic B cells [46]. Interestingly, the IgT genes were also highly expressed in the mucosal tissues, such as the gut or gills (Figure 6). In other tissues, different expression patterns were observed among the different C\(\tau\) genes (Figure 6). In particular, the expression pattern of IgT-A3 was quite different from the other three IgTs (Figure 6). Interestingly, the constant region of the IgT-A3 (CrA-3) has a frameshift, and does not have any D segments. In addition, we could not find the IgT-A3 in VH cDNA clones. Therefore, the functionality of IgT-A3 must be questioned. It should also be noted that the expression of IgT-A4 was not detected in one fish. The expression of IgT-B2, IgT-A4 and IgT-A5 was highly expressed in the heart. Hansen et al. (2005) have suggested that the expression of Ig genes in the heart is

---

**Table 2 Examples of “atypical” rearrangements involving \(\delta\)**

| EST ID     | Sequence repeat | Rearrangement          | Join in \(\delta\) |
|------------|-----------------|------------------------|-------------------|
| 001-084    | CTAG            | V-(D)-J-\(\delta\)     | 641(e2)           |
| 002-029    | CTAC            | V-(D)-J-\(\delta\)     | 532(e2)           |
| 002-179    | GANCAG          | V(before RSS)-\(\delta\) | 1073(e3)         |
| 003-005    | CCA             | V(0.1)-J-V(0.1)-\(\delta\) | 562(e2)         |
| 3.007      | V-J-m-\(\delta\) | 1(e1)                  |                   |
| 3.169      | V-(D)-J-\(\delta\) | 1(e1)                 |                   |
| 004-006    | AGTG            | V-(D)-\(\delta\)      | 611(e2)           |
| 004-078    | C               | V(before RSS)-\(\delta\) | 236(e1)         |
| 006-069/55 | \(\delta\) only ? | 337(e1)             |                   |
| 006-078/82 | \(\delta\) only ? | 623(e2)           |                   |
| 006-106    | CAG             | V-(D)-J-\(\delta\)     | begin(e1)        |
| 007-025    | AGTGACAGACAG    | V(before RSS)-\(\delta\) | 506(e2)         |
| 007-171    | CATCAG          | D-genomic-D-J-\(\delta\) | 647(e2)         |
| 9.03       | V7-\(\delta\)   | 235(e1)               |                   |
| 009-052    | CCAC            | V-(D)-J-\(\delta\)     | 294(e1)           |
| 11.169     | CTG             | V-(D)-J-\(\delta\)     | 1089(e3)         |
| 12.059     | GAC             | V(before RSS)-\(\delta\) | 242(e1)         |
| 017-108    | ACACA           | V(before RSS)-\(\delta\) | 689(e2)         |
| 017-121    | CAGACAG         | V(before RSS)-\(\delta\) | 253(e1)         |

* Joining position in exon 1, 2, or 3.

b Joining does not involve RSS but initiates earlier.
Figure 6 Detection of Ig genes in various Atlantic salmon tissues by RT-PCR analysis. (A) Agarose gel electrophoresis with PCR products of the Ig genes. Ubiquitin is an internal control. (B) The level of expression is calculated relative to the ubiquitin (ub) expression level. Data are expressed as mean ± SE of three fish.
most likely due to circulating B cells, because salmonid blood is a rich source of leukocytes [22]. In the present study, fish were bled before isolation of tissues but were not exsanguinated. It is therefore assumed that blood remained in the heart.

We also examined the expression of Ig genes during three early developmental stages of Atlantic salmon by RT-PCR (Figure 7). IgM and IgD were weakly detected in the earliest stage (0.2 g/2.05 cm), and the expression of both genes increased at later stages of development. Similarly, the expression of both IgT-B2 and IgT-A5 was increased in these stages; however, the expression of these genes was negative or very weak in the earliest stage. Interestingly, the expression of IgT-A4 was not detected in the 16.2 g/11.5 cm fish, and at especially high levels in the 4.2 g/7.15 cm fish. Similarly, IgT-A3 was highly expressed in the 4.2 g/7.15 cm fish, and very weak expression of the gene was observed in only one individual of the 16.2 g/11.5 cm fish.

Four immune related genes, \(il-1\beta1\) (IL-1\(\beta\)1), \(tnf-\alpha\) (TNF-\(\alpha\)), \(mx\) (Mx) and \(cox-2\) (COX-2), were also examined (Figure 7). IL-1\(\beta\), TNF-\(\alpha\) and COX-2 are key mediators of the inflammatory response [47], while Mx proteins are members of the type I interferon (IFN)-inducible genes, and play a role in anti-viral defenses in teleosts [48]. The expression of IL-1\(\beta\)1 and TNF-\(\alpha\) was quite similar with strong expression observed in only one individual of the 4.2 g/7.15 cm fish. Only very weak expression of these genes was observed in other fishes at different stages. The expression of Mx increased during the three developmental stages examined, while the expression of COX-2 was quite variable among individual fishes and development stage.

**Discussion**

**Structure of the duplicated IGH loci, IGH-A and -B**

Mammalian IGH loci do not have any CH genes located upstream of C\(\mu\) genes. Recently, however, a novel CH...
gene located upstream of Cμ and Cδ genes has been found in the IgH locus of zebrafish, rainbow trout and fugu [21-23]. In the stickleback IgH locus, a cluster of Ct-Cμ-Cδ was found duplicated three times in tandem, with an additional Ct gene in the 3’ end of the locus [28,29]. Similarly, the catfish IgH locus contains three linked pairs of Cμ and Cδ genes, but a CH region upstream of Cμ and Cδ genes similar to IgT has not been found in that locus [25,26] (Figure 8). In this study we confirmed that two duplicated IgH loci can functionally coexist and further found that several novel CH (Ct) genes exist between the VH and Cμ region of the duplicated Atlantic salmon IgH loci, five in IGH-A and three in IGH-B. Therefore, the Atlantic salmon duplicated IgH loci are the only IgH loci so far known, in which multiple Ct genes are spread out over the region upstream of Cμ. Of these Ct genes, two Ct genes in IGH-A (CtA-4 and CtA-5) and one Ct gene in IGH-B (CtB-2) are functional (Figure 8). These three genes were recovered in cDNA clones associated with VH, D and JH sequences. Interestingly, the CtA-4 gene is in the inverse orientation as has also been observed with the Ca genes in duck and chicken [49,50]. In addition, the inverted CtA-4 gene is found associated with VH and D sequences both 5’ and 3’ of the constant sequence, and the CtA-4 gene was the most frequently used Ct gene in our present analysis of VH rearranged cDNAs. Thus, the Atlantic salmon expresses seven kinds of CH genes, three functional Ct genes (CtB-2, CtA-4 and CtA-5), in addition to the two previously known Cμ genes (CμA.
The amino acid sequence of C\textsubscript{\mu} gene exists in the IGH-A\textsubscript{\alpha} functional group of three D segments associated with the C\textsubscript{\mu} and 18 were found only in teleosts, the zebrafish locus contains 47 VH genes, of which 36 are presumed functional (77%) [21]. In catfish, there are 165 - 200 VH segments as estimated by Southern blotting, and the analysis of 10% of the germline VH genes suggested that approximately 50% are pseudogenes [53]. Although a large number of VH genes in the Atlantic salmon are pseudogenes, some of these pseudo VH genes were expressed. It is unclear whether these pseudo VH genes may play a functional role.

Structural differences between IGH-A and -B
We found some structural differences between the two parallel IgH loci in Atlantic salmon. The orientation of the VH sequences indicates some rearrangement within the loci and an approximately 200 kb inversion event is evident in the IGH-A locus when comparing the IGH-A and IGH-B sequence (Figure 1 and Additional file 4). This inversion event, which ranges from approximately position 480 kb to 680 kb (in Figure 1) explains the inverse orientation of the C\textsubscript{\tau\alpha}A-4 sequence in IGH-A. In addition, three VH families, families 13, 14 and 18 were found only in IGH-A (Figure 3). An additional group of three D segments associated with the C\textsubscript{\mu} gene exists in the IGH-A locus (see Additional file 9). The amino acid sequence of C\textsubscript{\mu} has an extra cysteine residue near the C-terminal end as described previously by Hordvik et al. (2002) [35] (see Additional file 5). This additional cysteine in C\textsubscript{\mu} is absent in C\textsubscript{\mu} of rainbow trout [35]. Moreover, the amino acid sequence identities between the rainbow trout C\textsubscript{\tau} and the Atlantic salmon C\textsubscript{\tau\alpha}s (C\textsubscript{\tau\alpha}A-4 and C\textsubscript{\tau\alpha}A-5) is higher (80 - 82%) than the similarity between the rainbow trout C\textsubscript{\tau} and the Atlantic salmon C\textsubscript{\tau\beta} (C\textsubscript{\tau\beta}-2) (75%). In addition, phylogenetic analysis indicates that C\textsubscript{\tau\alpha}s are more closely related to the rainbow trout C\textsubscript{\tau} than to C\textsubscript{\tau\beta} (see Additional file 6). It has been reported that both duplicated IgD genes in Atlantic salmon have a tandem duplication of C\textsubscript{\delta}2-C\textsubscript{\delta}3-C\textsubscript{\delta}4 [17,38]. While our present study shows that these three exons (C\textsubscript{\delta}2-C\textsubscript{\delta}3-C\textsubscript{\delta}4) are indeed repeated three times in IGH-A, they are repeated four times in IGH-B (Figure 2). These observations indicate that both IGH-A and -B loci have evolved independently in the short time after the recent genome duplication. Thus, the existence of two parallel IgH isoloci in Atlantic salmon has contributed to the extensive diversity of the antibody repertoire.

Atypical VH -D-JH -C\textsubscript{\delta} rearrangements in Atlantic salmon
Analysis of the RT-PCR products amplified using primers specific for C\textsubscript{\delta} and for VH genes showed that there were a large number of different (unexpected) band sizes and some multiple bands in PCR product from one set of primers (data not shown), indicating that various types of VH-D-JH-C\textsubscript{\delta} rearrangements exist in Atlantic salmon. The teleost IgD gene is expressed as a chimeric transcript that includes the first exon of IgM gene (C\textsubscript{\mu}1), because the teleost first exon of IgD (C\textsubscript{\delta}1) does not contain an appropriate cysteine expected to form the disulfide bond with the L chain [16-20]. Unexpectedly, analysis of the Atlantic salmon IgD VH cDNA clones revealed that approximately 10% of the IgD transcripts do not include the C\textsubscript{\mu}1 sequence, showing that the Atlantic salmon IgD can be expressed as both a chimer and without the inclusion of the C\textsubscript{\mu}1 sequence. Recently, it has been reported that either of the C\textsubscript{\mu}1 and C\textsubscript{\delta}1 exons could be observed in expressed porcine IgD cDNA sequences (VDJ)-C\textsubscript{\mu}1-hinge region (H)-C\textsubscript{\delta}2-C\textsubscript{\delta}3 or VDJ-C\textsubscript{\delta}1-H-C\textsubscript{\delta}2-C\textsubscript{\delta}3 [54]. The porcine genomic C\textsubscript{\delta}1 exon is highly similar to the C\textsubscript{\mu}1 exon with only 4 nucleotides difference. Both the C\textsubscript{\delta}3 and C\textsubscript{\mu}1 exons contain three cysteines, only one of which becomes part of the IgD transcript and interacts with the L chain. In contrast, the amino acid sequences of the Atlantic salmon C\textsubscript{\delta}1 exons are quite different from that of the C\textsubscript{\mu}1 exons. In addition, the C\textsubscript{\delta}1 exon lacks the cysteine, similar to other teleosts [17]. Therefore, the functionality of these non-chimeric IgD transcripts is questionable. In fact, these non-chimeric IgD transcripts generally do not use the normal splice and recombination signals. In many of these non-chimeric IgD transcripts, joining
takes place from somewhere inside or just after the VH sequence to various distributed sites inside exon 1, 2, or 3 of the Cδ sequence. In many of these instances, a short repeat sequence is present at the two joined ends (Table 2). We cannot at this time exclude the possibility of PCR artifacts contribute to these atypical VH-D-JH-Cδ sequences. However, because they are only observed with Cδ and not with Cr and Cμ, we assume that it is not very likely. It will be of interest, in future studies, to discover the functions of these unusual IgD transcripts.

Expression of the three novel IgT isotypes in Atlantic salmon

The expression of the four innate-immune related genes was variable among individual fishes and developmental stages, while the expression of Ig genes were quite similar among individual fishes and developmental stages (Figure 7). The expression of the IgM, IgD and four innate immune-related genes were detected in the earliest stage (0.2 g/2.05 cm fish). On the other hand, the expression of the three IgTs was negative or very weak in the earliest stage (Figure 7), suggesting that the Atlantic salmon IgTs are involved in the more mature developmental stage than the IgM and IgD.

As mentioned above, three intact Ct genes (CtA-4, CtA-5 and CtB-2) were found in Atlantic salmon. The Cμ and Cδ genes, respectively, show a high degree of amino acid identities between IGH-A and IGH-B (96% ~) [17,31]. In contrast, the three intact Ct genes (CtB-2, CtA-4 and CtA-5) exhibit significant sequence divergence at the amino acid level not only between loci but also within a locus (IGH-A) (see Additional file 5). The per cent amino acid identities between the CtA (CtA-4 and CtA-5) and CtB (CtB-2) sequences is 75-76, and between CtA-4 and CtA-5 is 87. RT-PCR analyses revealed different tissue distribution patterns among these three IgTs (Figure 6). In addition, the expression pattern of IgT-A4 was quite different from the other two IgTs during the three early developmental stages tested. The expression of IgT-A4 was not detected in the 16.2 g/11.5 cm fish (Figure 7). The high degree of amino acid diversity and the differential expression patterns among three Ct genes suggest that the three different IgTs may have different functions.

Interestingly, in our present RT-PCR analysis, the IgTs, especially IgT-A5, were highly expressed in the mucosal tissues, including gut or gills (Figure 6). The fugu IgT mRNA was also strongly expressed in goblet cells of the intestine and gill epithelium [23]. In mammals, the mucosal surfaces constitute the first defensive line against invading microbial pathogens. The specific immunological defense at this site is primarily mediated by IgA antibodies [59,60]. Although the structure of the *Xenopus laevis* IgX is quite different from the mammalian IgA, the IgX is considered an analog of IgA because its association with the mucosae of the intestine resembles that of IgA [61]. In teleosts, no typical mucosal Ig class such as the IgA and IgX has been identified, but small amounts of IgM is present in gut mucus of several teleosts [62-64]. However, Hatten et al. (2001) have
reported that IgM is not present in gut mucus of Atlantic salmon, and the gut mucus contains a large amount of proteolytic enzymes able to degrade serum IgM [65]. Their results suggested that antibodies related to the gut of the Atlantic salmon should be of another, yet unidentified, Ig class. Because the three new IgTs identified in our present study were found to be highly expressed in mucosal tissues, these IgTs might form the mucosal Ig class in Atlantic salmon.

Conclusions

The present study shows that the genomic organization of the duplicated IgH loci in Atlantic salmon differs from that in other teleosts. The loci possess multiple Cr genes upstream of the Cμ region, with three of the Cr genes being functional. Moreover, the duplicated loci possess over 300 VH segments which could be classified into 18 families. This is the largest number of VH families currently defined in any vertebrate. Our results indicate that the duplication of the IgH loci in Atlantic salmon contributes heavily to the increases in diversity of the antibody repertoire, as compared with the single IgH locus in other vertebrates. Previous studies of the Atlantic salmon TRA/TRD and TRG loci revealed that Atlantic salmon clearly has one of the largest TCR repertoires known for any vertebrate [66,67].

Methods

Cloning and sequencing of VH cDNAs

Adult Atlantic salmon (Mowi stock) tissues were obtained from the Department of Fisheries and Oceans (Robert Devlin, WestVan Lab., West Vancouver, British Columbia). Adult fish were euthanized, followed by rapid dissection of tissues. Tissues were flash frozen in liquid nitrogen or dry ice and stored at -80°C until RNA extraction. Total RNA was extracted from the kidney and spleen of two healthy individuals using TRIzol reagent (Invitrogen). Purified total RNA (1.0 μg) was reverse transcribed with SuperScript™ II (Invitrogen) using oligo (dT)15 primer as described in the manufacturer’s protocol. The cDNAs were synthesized in 25 μl reactions incubated at 42°C for 90 min and the transcriptase heat-inactivated at 70°C for 30 min.Equal amounts of each cDNA were combined and the mixture used as PCR template.

One hundred seventyeight (178) forward primers were designed from VH sequences identified for one or several VH sequences per primer (generally as part of a family). The reverse primers were designed from the consensus sequence of Cμ, Cδ, and Cτ. For increased specificity, nested primers were also designed from the position of 2 x SSC and 0.1% SDS, and the second and third each consisting of 1 x SSC and 0.1% SDS. Filters were visualized using BioMax film (Kodak). BAC clones were chosen based on the physical BAC fingerprint map for Atlantic salmon [70] that is publicly available on the internet Contig Explorer (iCE) version 3.5 [71]. The BAC end sequence information, that is available in ASalBase [72], was also used for selection of the BAC clones. BAC shotgun libraries were constructed and sequenced on an ABI 3730 DNA sequencer, each of which was assembled using PHRED and PHRAP [73,74] and Consed [75].

The sequence of the IGHA and IGHB loci were deposited in Genbank under accession numbers, IGHA [GenBank:GU129139], IGHB [GenBank:GU129140] and the other IGHA allele [GenBank:GU321975 - GU321980]. The nucleotide sequences for the Cμ, Cδ and Cτ probes were based on EST sequence data and are provided in Additional file 13.

Sequencing of Atlantic salmon IGH-A and -B loci

The Atlantic salmon IGH-A and -B loci were isolated and sequenced as previously described [66,67]. An Atlantic salmon BAC library (CHORI-214), constructed from a Norwegian aquaculture male strain, was obtained from BACPAC Resources, Children’s Hospital Oakland Research Institute (CHORI) [69]. Six BAC library filters were hybridized with three (Cμ, Cδ and Cτ) 70-mer oligo probes (Integrated DNA Technologies) that were 5’-end-labeled with 32P-ATP using T4 polynucleotide kinase (Invitrogen). The labeled probes were added to BAC filters that had been pre-hybridized at 65°C for 4 h (5 x SSC, 5 x Denhardt’s, 0.1% SDS). The hybridization was carried out overnight at 65°C. Three washes were performed, each for 30 min at 50°C; the first consisting of 2 x SSC and 0.1% SDS, and the second and third each consisting of 1 x SSC and 0.1% SDS. Filters were visualized using BioMax film (Kodak). BAC clones were chosen based on the physical BAC fingerprint map for Atlantic salmon [70] that is publicly available on the internet Contig Explorer (iCE) version 3.5 [71]. The BAC end sequence information, that is available in ASalBase [72], was also used for selection of the BAC clones. BAC shotgun libraries were constructed and sequenced on an ABI 3730 DNA sequencer, each of which was assembled using PHRED and PHRAP [73,74] and Consed [75].
consensus sequence of IGH-A and -B locus for Cμ and Cδ genes, and from the four different forms of Cr genes. These nested primers were located in the first exon of τ (Ct1), and in the second exon of μ (Cμ2) and δ (Cδ2). The PCR primers used in this study are shown in Additional file 13. PCR products were used to conduct Dotter plots (Promega) with an initial denaturation of 2 min at 95°C and then 30 cycles at 30 s of denaturation at 95°C, 30 s of annealing at 55°C, and 1 min of extension at 72°C. PCR products were cloned into pCR2.1 (TA Cloning Kit, Invitrogen) according to the manufacturer’s protocol. Twelve (12) clones from each positive PCR product were sequenced as described above. Sequences obtained from Cμ and Cr genes were sequenced in single forward direction, while the longer sequences that include the Cδ gene were sequenced in both forward and reverse directions. We only analyzed the results where the two reads could form a contig with a 100% overlap. The cDNA sequences were BLAST searched against the VH, JH and CH gene sequences to identify their presence in the clones.

RT-PCR analysis of Ig genes expression

Twelve different Atlantic salmon tissues from three different adult fish were provided by the Department of Fisheries and Oceans. Immature and juvenile stages of fish were provided by Marine Harvest Canada (Big Tree Creek Hatchery, Sayward, B.C.) Total RNA was extracted and reverse transcribed using the method described above for tissues (kidney and spleen) from two individuals for three different development stages, as well as tissues from three individual adult fish (kidney, muscle, skin, gut, gill, spleen, brain, heart, gonad, liver, eye and pyloric caeca). PCR primers were designed from the consensus sequence of IGH-A and IGH-B locus for Cμ and Cδ genes, and four different forms of Cr genes. The primers correspond to sequences in the first and third exon for Cμ and Cδ genes, and in the fifth and seventh exon for Cδ gene. Ubiquitin was used as an internal positive control. In addition, four immune related genes, interleukin (IL-1)β1, tumour necrosis factor (TNF)-α, Mx and cyclooxygenase (COX)-2, were also examined. The PCR primers used in this study are shown in Additional file 13. PCR was performed using GoTaq DNA polymerase (Promega) with an initial denaturation step of 2 min at 95°C and then 30 or 35 cycles as follows: 30 s of denaturation at 95°C, 30 s of annealing at 55°C and 1 min of extension at 72°C. The PCR products derived from each primer set were TA-cloned and confirmed by sequencing. The PCR products were electrophoresed on a 1.0% agarose gel. The intensity of the amplification bands was semi-quantitatively measured using ImageJ software [81], and divided by the intensity of the respective ubiquitin signals.

Additional material

Additional file 1: Features of the IGH-A. Table listing genes and pseudogenes identified in the IGH-A.

Additional file 2: Features of the IGH-B. Table listing genes and pseudogenes identified in the IGH-B.

Additional file 3: The identified genes flanking the loci. Table listing genes flanking the loci identified with the Digit Web Server (http://syntheticbiology.pw/pic/en/cnbi015158b151572i/).

Additional file 4: Dotter plot of locus A (IGH-A) versus locus B (IGH-B). This file contains a dotter plot of IGH-A versus IGH-B.

Additional file 5: Alignment of amino acid sequences encoded by (A) Cτ, (B) Cμ and (C) Cδ. This file contains multiple sequence alignments of amino acid sequences encoded by (A) Cτ, (B) Cμ and (C) Cδ obtained from ClustalW. Identical residues are shown as dots (.) and gaps are shown as hyphens (-).

Additional file 6: Phylogenic relationships for the CH genes in various species. Phylogenetic tree showing the relationship of the CH genes amino acid sequences of CH2 and CH3 domains of α, human and rat; CH3 and CH4 of μ; rat, α and duck CH4 and CH5 of new antigen receptor (NAR); CH5 and CH6 of α, NARC and teleost &. The tree was constructed with the MEGA 4 package by neighbor-joining (NJ) method and bootstrap values for replicated 1,000 were represented by percentages on the edge of node. The bootstrap values greater than 50% are presented. The scale bar indicates the branch length. Genbank accession numbers are as follows: α duck [GenBank:AAA68660], human [GenBank:AAC82528], Atlantic salmon δ, and α [GenBank:AF278717, AF141605], catfish [GenBank:IT8537], fugu [GenBank:BAD34542], zebrasfish [GenBank:CA11477], Xenopus [GenBank:DO350886], human [GenBank:AAA52771], e. human [GenBank:AA593935], opossum (Monodelphis domestica) [GenBank:AA79674], γ human [GenBank:AAC82527]; γ; mouse [GenBank:AA596971], μ. Atlantic salmon μ and μ [GenBank:AA824064, AA664980], bowfin (Amia calva) [GenBank:ACU12468], carp [GenBank:AB004105], catfish [GenBank:MC72303], gar (Lepisosteus osseus) [GenBank:U14355], ladyfish (Elops saurus) [GenBank:MD6182], lungfish (Protopterus aethiopicus) [GenBank:AF377224] nurse shark (Ginglymostoma cirratum) [GenBank:MK92851], rainbow trout [GenBank:X83372], skate (Leucoraja eniscus) [GenBank:MK26979], sturgeon (Acipenser baeri) [GenBank:Y11253], zebrasfish [GenBank:Y643753], Xenopus [GenBank:MD0849], chicken [GenBank:J001613], mouse [GenBank:J00443], human [GenBank:X19490]. u. Xenopus [GenBank:x15114] ex. lungfish [GenBank:AF437277], sandbar shark (Carcharhinus plumbeus) [GenBank:CPU04560]. NARC nurse shark [GenBank:GCU15405]; NARC nurse shark [GenBank:GCU18701]; ω: grass carp (Ctenopharyngodon idella) [GenBank:DQ489733], rainbow trout ω1 and ω2 [GenBank:AAW66978] and [GenBank:AAW66981], perch (Sparus sarus), [GenBank:DO316666], zebrasfish [GenBank:AY643752].

Additional file 7: Genbank accession numbers for deduced amino acid sequences of CH and VH domains. Table listing the accession numbers for deduced amino acid sequences of CH and VH domains.

Additional file 8: Alignment of D sequences. This file contains a multiple sequence alignment of D sequences obtained from ClustalW.

Additional file 9: Phylogenic trees showing the relationship between the (A) D and (B) JH sequences. This file contains phylogenic trees for (A) D and (B) JH genes.

Additional file 10: Alignment of JH sequences. This file contains a multiple sequence alignment of JH sequences obtained from ClustalW.

Additional file 11: Alignment of amino acid VH sequences. This file contains a multiple sequence alignment of amino acid VH sequences obtained from ClustalW. Identical residues are shown as dots (.) and gaps are shown as hyphens (-).

Additional file 12: Distribution of variable (VH) families in the two IgH loci. Table showing the number of sequences identified per family.

Additional file 13: PCR primers and oligo probes. Table listing the PCR primers and oligo probes used in this study.
Abbreviations
igHt: immunoglobulin heavy chain locus, igL: Immunoglobulin light chain locus, IGHA: Immunoglobulin heavy chain A, IGH-B: Immunoglobulin heavy chain B, VH: Variable region of immunoglobulin heavy chain, D: Diverse region of immunoglobulin heavy chain, CH: Constant region of immunoglobulin heavy chain, Cµ: µ chain constant region, Cδ: δ chain constant region, Cγ: γ chain constant region, Cα: α chain constant region, Cε: ε chain constant region, Cζ: ζ chain constant region, TRB/TRA: T cell receptor (TCR) α/β locus, TRC, TRγ: locus.

Acknowledgements
This work was supported by Genome Canada, Genome BC and by the Natural Sciences and Engineering Research Council of Canada (B.K., W.D.). We would like to thank Robert Devlin (Department of Fisheries and Oceans) for providing adult salmon tissue samples; Marine Harvest Canada for providing immature and juvenile stages of fish; Krzysztof P. Lubieniecki (Simon Fraser University) for his assistance in construction of BAC shotgun libraries; Eric Rondeau (University of Victoria) for his assistance in obtaining tissues for RT-PCR analysis; Dan Sandersen (University of Victoria) for his help in preparing the manuscript.

Author details
1. Department of Biology, University of Victoria, PO Box 3020 STN CSC, Victoria, British Columbia, V8W 3N5, Canada. 2. Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A1S6, Canada.

Authors’ contributions
My performed the BAC library preparation, VH cDNA cloning, RT-PCR studies and drafted the manuscript. JdB performed the data analysis and drafted the manuscript. WSD contributed to the project planning and direction. BFK, LM, AM, SS performed the BAC library preparation and DNA sequencing for the project. WSD contributed to the project planning and direction. BFR contributed to the planning, design, and direction of the project. All authors read and approved the final manuscript.

Received: 15 February 2010 Accepted: 2 September 2010
Published: 2 September 2010

References
1. Cannon PJ, Haire NR, Rast PJ, Litman GW. The phylegetic origins of the antigen-binding receptors and somatic diversification mechanisms. Immuno Rev 2004, 2001(12):12-22.
2. Pancer Z, Cooper MD. The evolution of adaptive immunity. Annu Rev Immunol 2006, 24(1):497-518.
3. Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z. Diversity and function of adaptive immune receptors in a jawless vertebrate. Science 2005, 310(5756):1970-1973.
4. Pancer Z, Amemiya CT, Ethhardt GPA, Cetin J, Larry Garriland G, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 2004, 430(6996):174-180.
5. Pancer Z, Saha NR, Kasimatis J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD. Variable lymphocyte receptors in hagfish. Proc Natl Acad Sci USA 2005, 102(28):9224-9229.
6. Basias CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. Cell 2002, 109(2):545-555.
7. Schlissel MS. Regulating antigen-receptor gene assembly. Nat Rev Immunol 2003, 3(11):890-899.
8. Dooley H, Flajnik MF. Antibody repertoire development in cartilaginous fish. Dev Comp Immunol 2006, 30(1-2):43-56.
9. Hsu E, Pullham N, Rumfelt LL, Flajnik MF. The plasticity of immunoglobulin gene systems in evolution. Immuno Rev 2006, 210:8-26.
10. Amemiya CT, Litman GW. Complete nucleotide sequence of an immunoglobulin heavy-chain gene and analysis of immunoglobulin gene organization in a primitive teleost species. Proc Natl Acad Sci USA 1990, 87(2):811-815.
11. Ghaffari SH, Lobb CJ. Organization of immunoglobulin heavy chain constant and joining region genes in the channel catfish, Mol Immunol 1992, 29(2):151-159.
12. Hordvik I, De Vries Lindstrom C, Voie AM, Jacob AJL, Endresen C. Structure and organization of the immunoglobulin m heavy chain genes in Atlantic Salmon, Salmo salar. Mol Immunol 1997, 34(8-9):631-639.
13. Warr GW. The immunoglobulin genes of fish. Dev Comp Immunol 1995, 19(1):1-12.
14. Abbas KA, Lichtman HA, Pillai S. Lymphocyte development and the rearrangement and expression of antigen receptor genes. Cellular and Mol Immunol Philadelphia: WB Saunders, 6 2007, 153-187, 6th edition edn.
15. Snapper CM, Marcu KB, Zelazowski P. The immunoglobulin class switch: beyond “accessibility”. Immunity 1997, 6(3):217-223.
16. Hirono I, Nam BH, Enomoto J, Uchino K, Aoki T. Cloning and characterisation of a CDNA encoding Japanese flounder Paralichthys olivaceus lgD. Fish Shellfish Immunol 2003, 15(1):63-70.
17. Hordvik I, Thevarajan J, Samdal I, Bastani N, Krossøy B. Molecular cloning and phylogenetic analysis of the Atlantic salmon immunoglobulin D gene. Stand J Immunol 1999, 50(2):202-210.
18. Saha NR, Suetake H, Kikuchi K, Suzuki Y. Fugu immunoglobulin D: a highly unusual gene with unprecedented duplications in its constant region. Immunogenetics 2004, 56(6):438-447.
19. Stenvik J, Jørgensen TD. Immunoglobulin D (lgD) of Atlantic cod has a unique structure. Immunogenetics 2000, 51(6):452-461.
20. Wilson M, Bengtgen E, Miller NW, Clem LW, Du Pasquier L, Warr GW. A novel chimeric Ig heavy chain from a teleost fish shares similarities to lgD. Proc Natl Acad Sci USA 1997, 94(9):4593-4597.
21. Danilova N, Bussmann J, Jekoek K, Steiner LA. The immunoglobulin heavy-chain locus in zebrasfin: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol 2005, 6(3):295-302.
22. Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain isotype (lgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. Proc Natl Acad Sci USA 2005, 102(19):6919-6924.
23. Svan R, Arman A, Sato K, Yamaguchi R, Sakai M. Discovery of a new class of immunoglobulin heavy chain from fugu. European Journal of Immunobiology 2005, 35(11):320-331.
24. Svan R, Arman A, Nakao M, Watanuki H, Sakai M. Discovery of a novel immunoglobulin heavy chain gene chimera from common carp (Cyprinus carpio L). Immunogenetics 2005, 57(6):458-463.
25. Bengtén E, Quiniou S, Kimika J, Waldbieser G, Warr G, Miller N, Wilson M. Structure of the catfish IgH locus: analysis of the region including the single functional IGHM gene. Immunogenetics 2006, 58(10):831-844.
26. Bengtén E, Quiniou SMA, Stage TB, Katagiri T, Miller NW, Clem LW, Warr GW, Wilson M. The IgH locus of the channel catfish, Ictalurus punctatus, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgG. J Immunol 2002, 169(5):2488-2497.
27. Bengtén E, Clem LW, Miller NW, Warr GW, Wilson M. Channel catfish immunoglobulins: Repertoire and expression. Dev Comp Immunol 2006, 30(1-2):77-92.
28. Bao Y, Wang T, Guo Y, Zhao Z, Li N, Zhao Y. The immunoglobulin gene loci in the teleost Gasterosteus aculeatus. Fish Shellfish Immunol 2010, 28:40-8.
29. Gamblín-Deza E, Sánchez-Espinel C, Magadán-Mompó S. Presence of an unique IgT on the IgH locus in three-spined stickleback fish (Gasterosteus aculeatus) and the very recent generation of a repertoire of VH genes. Dev Comp Immunol 2010, 34:114-22.
30. Hordvik I. The impact of ancestral tetraploidy on antibody heterogeneity in salmonid fishes. Immunol Rev 1998, 166(1):153-157.
31. Hordvik I, Voie AM, Lette J, Male H, Endresen C. Cloning and sequence analysis of two isotypic IgM heavy chain genes from Atlantic salmon, Salmo salar L. Eur J Immunol 1992, 22(11):2957-2962.
32. Solem ST, Hordvik I, Killie JEA, Warr GW, Jørgensen TD. Diversity of the immunoglobulin heavy chain gene in the Atlantic salmon (Salmo salar L) is contributed by genes from two parallel IgH isoforms. Dev Comp Immunol 2007, 25(6):403-417.
33. Ohno’s Evolution by gene duplication, New York Springer-Verlag 1970.
34. Shima T, Dijkstra JM, Shimizu S, Watanabe A, Yangaya K, Kiryu I, Fujwara A, Nishida-Umemura C, Kaba Y, Hirono I, et al. Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (Oncorhynchus mykiss), a species with a presumably recent tetraploid ancestry. Immunogenetics 2003, 56(12):878-893.
35. Hordvik I, Berven FS, Solem ST, Hatten F, Endresen C. Analysis of two IgM isotypes in Atlantic salmon and brown trout. Mol Immunol 2002, 39(5-6):313-321.
36. Oakley TH, Phillips RB. Phylogeny of Salmonine fishes based upon growth hormone introns: Atlantic (Salmo) and Pacific (Oncorhynchus) salmon are not sister taxa. Mol Phylogenet Evol 1999, 11:381-393.
37. de Boer J, Yazawa R, Davidson WS, Koop BF: Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 2007, 8(1):422.
38. Hordvik I: Identification of a novel immunoglobulin δ transcript and comparative analysis of the genes encoding IgD in Atlantic salmon and Atlantic halibut. Mol Immunol 2002, 39(1-2):85-91.
39. Landolfi NF, Yin XM, Capra JD, Tucker PW: A conserved heptamer upstream of the IgH promoter region octamer can be the site of a coordinate protein-DNA interaction. Nucl Acids Res 1988, 16(12):5033-5114.
40. Brown G, Kaatari L, Kaatari S: Two new Ig VH gene families in Oncorhynchus mykiss. Immunogenetics 2006, 58(11):933-936.
41. Roman T, Charlemagne J, Matsumura T, Anderson E, Bengt E, Pilstrom L, Hansen J, Kaatari S: Unified nomenclature of Ig VH genes in rainbow trout (Oncorhynchus mykiss): definition of eleven VH families. Immunogenetics 1996, 43(5):325-326.
42. Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA, Robb A, Beetz-Sargent M, Holt R, Moore R, et al: A salmonid ST genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics 2008, 9:645.
43. Press CM, Evesen Ø: The morphology of the immune system in teleost fishes. Fish Shellfish Immunol 1999, 9(4):309-319.
44. Zapata A, Deez B, Cejalvo T, Guitierrez-de Frías C, Cortés A: Ontogeny of the immune system of fish. Fish Shellfish Immunol 2006, 20(2):263-136.
45. Zwollo P, Cole S, Bromage E, Kaatari S, et al: A salmonid ST genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics 2008, 9:645.
46. Zwollo P, Haines A, Rosato P, Gumulak-Smith J: Molecular and cellular analysis of B-cell populations in the rainbow trout using Pax5 and immunoglobulin markers. Dev Comp Immunol 2008, 32(12):1482-1496.
47. Seecombe CJ, Wang T, Hong S, Peddle S, Stamp M, Laing KJ, Cunningham C, Zhou J: Cytokines and innate immunity of fish. Dev Comp Immunol 2001, 25(8a):719-723.
48. Leong CA, Trobridge DG, Kim HYC, Johnson M, Simon B: Interferon-inducible Mx proteins in fish. Immunol Rev 1998, 166(1):349-363.
49. Lundqvist ML, Middleton DL, Hazard S, Wan GW: The morphology of the immune system in teleost fishes. Fish Shellfish Immunol 2006, 20(2):263-136.
50. Zhao Y, Babbani H, Shimizu A, Hammarstrom L: Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted u gene upstream of a condensed uis gene. Immunology 2000, 101(3):348-353.
51. Johnston CM, Wolf AL, Bolland DJ, Corcoran AE: Complete sequence assembly and characterization of the CSB/L6 mouse Ig heavy chain V region. J Immunol 2006, 177(6):4221-4234.
52. Matsuda F, Ishii K, Bourvagner P, Kuma K, Hayashida H, Miyata T, Honyo T: The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med 1996, 188(11):2151-2162.
53. Yang T, Venturino-Holman T, Lobb CJ: Structure, genomic organization, and phylogenetic implications of six new VH families in the channel catfish. Mol Immunol 2003, 40(5):247-260.
54. Zhao Y, Pan-Hammarstrom Q, Kacksvick J, Hammarstrom L: The porcine Ig δ gene: Unique chimeric splicing of the first constant region domain in its heavy chain transcripts. J Immunol 2003, 171(3):1312-1318.
55. David AR, Melanie RW, Norman WM, Clem LW, Gregory WW, David AR, Melanie RW, Norman WM, Clem LW, Gregory WW: Evolutionary variation of immunoglobulin μ heavy chain RNA processing pathways: origins, effects, and implications. Immunol Rev 1996, 166(1):143-151.
56. Miller N, Wilson M, Bluffton E, Stuge T, Warr G, Clem W: Functional and molecular characterization of teleost leukocytes. Immunol Rev 1998, 166(1):187-199.
57. Vret C, Gurr W: The origin of the “one cell-one antibody” rule. J Immunol 2009, 182(3):1229-1230.
58. Li J, Barreda DR, Zhang YA, Boshia H, Gelman AE, LaPatsa S, Tott L, Sunyer JD: B lymphocytes from early vertebrates have potent phagocytic and microbial abilities. Nat Immunol 2006, 7(10):1116-1117.
59. Jenny MW, Jiri M: Mucosal immunoglobulins. Immunol Rev 2005, 206(1):64-82.
60. Underdown BJ, Schiiff JM: Immunoglobulin A: Strategic defensive initiative at the mucosal surface. Annu Rev Immunol 1986, 4(1):389-417.
61. Rainer M, Louis Du P, Ellen H: Is Xenopus Ig an analog of IgM? Eur J Immunol 1996, 26(12):2823-2830.
62. Solomon ST, Stervik J: Antibody repertoire development in teleosts-a review with emphasis on salmonids and Gadus morhua L. Dev Comp Immunol 2006, 30(1-2):537-76.
63. Maki JL, Dickerson HW: Systemic and cutaneous mucus antibody responses of channel catfish immunized against the protozoan parasite Ichthyophthirius multifilis. Clin Diag Lab Immunol 2003, 10(5):876-881.
64. Zhao X, Findly RC, Dickerson HW: Cutaneous antibody-secreting cells and B cells in a teleost fish. Dev Comp Immunol 2008, 32(5):500-508.
65. Hatten F, Fredriksen A, Hordvik I, Endresen C: Presence of IgM in cutaneous mucus, but not in gut mucus of Atlantic salmon. Salmo salar: Serum IgM is rapidly degraded when added to gut mucus. Fish Shellfish Immunol 2001, 11(3):357-358.
66. Yazawa R, Cooper GA, Hunt P, Beetz-Sargent M, Robb A, Conrad M, McKinnel L, So S, Jantzen S, Phillips RB, et al: Striking antigen recognition diversity in the Atlantic salmon T-cell receptor αβ locus. Dev Comp Immunol 2008, 32(5):204-212.
67. Yazawa R, Cooper GA, Beetz-Sargent M, Robb A, McKinnel L, Davidson WS, Koop BF: Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus. Mol Immunol 2008, 45(8):2150-2157.
68. Warren RL, Nelson BH, Holt RA: Profiling model T-cell metagenomes with short reads. Bioinformatics 2009, 25(4):458-464.
69. Thorsen J, Zhu B, Frengen E, Ozogawa K, de Jong P, Koop BF, et al: Evolutionary variation in the Atlantic salmon T-cell receptor αβ locus. Dev Comp Immunol 2009, 23(5-6):899-902.
70. Cameron CD, Boddet I, Schein JL, Jones SJM, Mann MA: Internet Contig Explorer (ICE) - A Tool for Visualizing Clone Fingerprint Maps. Genome Res 2003, 13(6a):1244-1249.
71. AsABase. [http://www.asaas.org].
72. Ewing B, Green P: Base-calling of automated sequencer traces using PHRED. II. Error probabilities. Genome Res 1998, 8:186-194.
73. Ewing B, Hillier L, Wendel V, Green P: Base-calling of automated sequencer traces using PHRED. I. Accuracy assessment. Genome Res 1998, 8:175-185.
74. Gordon D, Abajian C, Green P: A graphical tool for sequence finishing. Genome Res 1998, 8(3):195-202.
75. Sonnhammer ELL, Durbin R: A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 1995, 167(1-2):1-C1-C10.
76. Higgins DG, Fleaxby AJ, Fuchs R, CLUSTAL V: Improved software for multiple sequence alignment. Comput Appl Biosci 1992, 8(2):189-191.
77. Kumar S, Tamura K, Nei M: MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 2004, 5(2):150-156.
78. Lefranc MP, WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report. Immunogenetics 2007, 59(12):899-902.
79. Genes identified with Digit Web Server. [http://synthetic-biology.jp/sw/ pic/en/cnr15121nhr15157/2f/]
80. ImageJ. [http://rsweb.nih.gov/j/]
81. Ventura-Holman T, Lobb CJ: Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the Iglc locus represents a composite of two gene clusters. Mol Immunol 2002, 38(7):557-564.