BMC Genomics

Research article

Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (*Salmo salar*)

Håvard Harstad, Morten F Lukacs, Hege G Bakke and Unni Grimholt*

Address: Department of Basic Science and Aquatic Medicine, Norwegian School of Veterinary Science, Oslo, Norway

Email: Håvard Harstad - havard.harstad@veths.no; Morten F Lukacs - morten.lukacs@veths.no; Hege G Bakke - hege.bakke@veths.no; Unni Grimholt* - unni.grimholt@veths.no

* Corresponding author

Abstract

Background: In teleosts, the Major Histocompatibility Complex (MHC) class I and class II molecules reside on different linkage groups as opposed to tetrapods and shark, where the class I and class II genes reside in one genomic region. Several teleost MHC class I regions have been sequenced and show varying number of class I genes. Salmonids have one major expressed MHC class I locus (UBA) in addition to varying numbers of non-classical genes. Two other more distant lineages are also identifyed denoted L and ZE. For class II, only one major expressed class II alpha (DAA) and beta (DAB) gene has been identified in salmonids so far.

Results: We sequenced a genomic region of 211 kb encompassing divergent MHC class II alpha (*Sasa-DBA*) and beta (*Sasa-DBB*) genes in addition to NRGN, TIPRL, TBCEL and TECTA. The region was not linked to the classical class II genes and had some synteny to genomic regions from other teleosts. Two additional divergent and expressed class II sequences denoted DCA and DDA were also identified in both salmon and trout. Expression patterns and lack of polymorphism make these genes non-classical class II analogues. *Sasa-DBB*, *Sasa-DCA* and *Sasa-DDA* had highest expression levels in liver, hindgut and spleen respectively, suggestive of distinctive functions in these tissues. Phylogenetic studies revealed more yet undescribed divergent expressed MHC class II molecules also in other teleosts.

Conclusion: We have characterised one genomic region containing expressed non-classical MHC class II genes in addition to four other genes not involved in immune function. Salmonids contain at least two expressed MHC class II beta genes and four expressed MHC class II alpha genes with properties suggestive of new functions for MHC class II in vertebrates. Collectively, our data suggest that the class II is worthy of more elaborate studies also in other teleost species.

Background

Teleost fish are the largest group of vertebrates comprising almost half of the total living vertebrates. In tetrapods and sharks the major histocompatibility genes are linked in a complex on a single chromosome [1]. In all teleosts studied so far, including salmonids, the MHC class I and class II regions reside on different linkage groups [2-4]. Extensive studies of the genomic class I region has been conducted in several fish species including rainbow trout and Atlantic salmon [5-9]. The Atlantic salmon and rainbow trout genomes encode one major MHC class I locus designated *UBA* with additional non-classical MHC class I

genes in two duplicated MHC class I regions [7,9]. Both regions also harbour genes involved in the antigen presentation pathway, including proteosome subunits and the transporter for antigen processing. These genes all reside in the class II region in mammals [10].

Genomic class II regions are described in detail in zebrafish [11] and stickleback [12]. In zebrafish, one class II alpha locus and a number of class II beta loci residing on two different linkage groups have previously been identified, where only the *DAA* and *DAB* loci are known to be expressed [13]. Analysis of the functional class II region in zebrafish showed close linkage of the *DAA* and *DAB* loci on chromosome 8, but lack of other genes residing in the human MHC region [14]. In stickleback, a 99.5 kb genomic segment contained a tandem duplicate of expressed MHC class II alpha and beta genes, designated *Gaac-DAA/DAB* and *Gaac-DBA/DBB*. In Atlantic salmon, the major MHC class II alpha and beta genes are designated *DAA* and *DAB* respectively. They are closely linked and cosegregate as functional haplotypes [3,15-18]. Both the class II alpha as well as the class II beta chains have polymorphic alpha 1 and beta 1 domains [15,19], although much less polymorphic than the class I alpha 1 domain [3].

In humans, there are three classical expressed class II loci denoted HLA-DP, DQ and DR. Two additional nonclassical expressed MHC class II molecules, HLA-DM and HLA-DO, are found to control the composition of the peptide repertoire displayed by MHC class II molecules on the cell surface of antigen presenting cells [20]. The nonclassical class II molecule HLA-DM regulates unloading of CLIP and loading of peptide onto classical MHC class II molecules [21]. CLIP is derived from the invariant chain (Ii) and functions as a chaperone for class II molecules as it mediates and maintains correct assembly of alpha beta dimers. HLA-DM also serves as a peptide editor in early endocytic compartments [22]. HLA-DO preferentially promotes peptide loading of MHC class II molecules that are dependent on the chaperone activity of DM, and influences editing in a positive way for some peptides and negatively for others [23]. In terms of polymorphism, HLA-DO and HLA-DM are generally unpolymorphic in contrast to the three classical expressed loci HLA-DP, DQ and DR [24,25].

Reports on non-classical MHC class II molecules in teleosts have so far been scarce, although sequences with low identity to their classical DAB counterparts have been described in *Xiphophorus* fishes and in the guppy *Poecilia reticulata* [26]. The aim of this study was to analyse the MHC class II situation in Atlantic salmon through low stringency screening of available BAC libraries using classical *DAA* and *DAB* probes. Here we describe the genomic organization and expression patterns of two Atlantic salmon nonclassical MHC class II genes. Through phylogenetic and comparative analyses we also uncover additional nonclassical MHC class II molecules.

Results and discussion

An Atlantic salmon BAC library [27] screened at low stringency with radioactive labeled probes for MHC class IIα (*DAA*) and class IIβ (*DAB*) hybridized to 20 BAC clones (Table 1). The clones were ordered into one DAA and one DAB contig by restriction fragment analysis and southern hybridization. Some DAB positive BACs contained unstable inserts, displaying several deletions during restriction mapping. GRASP *Hin*dIII fingerprint information [28] confirmed our DAA contig, while DAB clones were placed in different contigs or identified as singletons potentially due to their unstable nature.

To decipher between classical DAA/DAB-positive BACs and other potential class II BACs we tested both genomic DNA from the library fish as well as MHC class II positive BACs for presence of a class II alpha minisatellite marker residing in the 3'UTR of *Sasa-DAA*. Classical MHC class II alpha and beta alleles cosegregate as haplotypes in Atlantic salmon [15] and most haplotypes show a stable linkage between haplotype and class II alpha 3'UTR marker. The BAC library fish was homozygous for a marker previously shown to segregate with the *DAA*0101-DAB*0801* haplotype [15]. None of the DAA or DAB positive BACs were positive for the marker, suggesting they potentially represented new MHC class II loci while the classical MHC class II region was not present in the BAC library.

Due to the confusing fingerprints analysis and deletions in the DAB positive BACs, they were not considered for full-sequencing. Among the DAA positive BACs, clone 630N19 had a centered DAA-positive fragment (data not shown), and was thus subjected to shotgun sequencing. Readings were assembled with a redundancy greater than

Table 1: Atlantic salmon BAC library screen.

Probe(s)	No. of positive clones	BAC clone identities
DAA	8	139H17, 311J12, 327D01, 430J21, 485N02, 579A17, 630N19, 650J16
DAB	12	113K11, 121O12, 332F21, 416L17, 417A07, 429J19, 525D04, 537K01, 612D01, 619A16, 642F05, 667J06

Positive Atlantic salmon BAC clones from initial library screen using probes for *Sasa-DAA* and *Sasa-DAB*.

10 over 211,190 bp [GenBank: <u>EU008541</u>]. One gap in the BAC clone 630N19 sequence still remains due to a small repeat region (~100 bp) eluding sequencing. The gap is located 2 kb in from the T7 end and has been verified by PCR using flanking primers.

Dotplot analysis of 630N19 against itself showed no extended regions of local similarity (data not shown). The overall G+C content of the complete sequence was 43.5%. Analysis of the complete sequence with Repeatmasker unveiled 2 SINEs and 69 other repeats. Repeat elements occupied 2.08% of the entire genomic sequence. Using a salmonid repeat-specific database, appr. 20% of the BAC contained repeats [29]. A total of 20 ORFs were identified within the 211 kb sequence where all but six had sequence similarity to transposable elements or other repeats.

Linkage analysis using the 3'UTR marker of classical MHC class II gene *Sasa-DAA* and a marker identified in the 630N19 BAC sequence showed that the two regions are unlinked with *Sasa-DAA* residing on linkage group 6 and the 630N19 marker on linkage group 5 (T.Moen, Akvaforsk, pers.com).

Gene Content and Organization
The BAC encompassed six genes with homology to MHC class II alpha, MHC class II beta, NRGN (Neurogranin), TIPRL (TIP41, TOR signalling pathway regulator-like), TBCEL (Tubulin folding cofactor E-like) and TECTA (Tectorin alpha) (Fig. 1). The MHC class II alpha and beta related loci had low sequence identity to classical *Sasa-DAA* and *Sasa-DAB* sequences, and were thus designated *Sasa-DBA* and *Sasa-DBB* as suggested by R.J.M.Stet (The MHC Nomenclature Committee, pers.com). The Neurogranin domain family is the human homolog of the neuron-specific rat RC3/neurogranin gene. This gene encodes a postsynaptic protein kinase substrate that binds calmodulin in the absence of calcium [30]. The TIP41 domain family belongs to the TOR signalling pathway that activates a cell-growth program in response to nutrients in *Saccharomyces cerevisiae* [31], where TIP41 interacts with TAP42 and negatively regulates the TOR signalling pathway. TIP41 homologs are found in vertebrates, but very little information on its function is described. All the above mentioned genes have matching Atlantic salmon ESTs (Table 2).

Only rainbow trout EST matches were found for the TBCEL gene (Table 2). TBCEL is a novel regulator of tubulin stability with overexpression causing depolymerization of microtubules and suppression resulting in an increase in the number of stable microtubules [32]. No perfect EST match was found for TECTA, but mouse TECTA is expressed during early ontogeny and such libraries are not available for salmonids yet. The Atlantic salmon TECTA protein contained several conserved domains with identical organization to mammalian TECTA molecules (Fig. 2) [33,34]. An extracellular domain (NIDO) of unknown function is found in nidogen (entactin). The von Willebrand factor type D domains (VWD) are blood glycoprotein domains that are required for normal hemostasis. The domain of unknown function denoted as C8 contains 8 conserved cysteine residues and is found in disease-related proteins such as von Willebrand factor, Alpha tectorin, Zonadhesin and Mucin molecules. The trypsin inhibitor-like cysteine rich domains (TIL) are found in many extracellular proteins. The zona pellucida (ZP) domain is found responsible for sperm-adhesion and is present in multidomain transmembrane proteins such as glycoprotein GP2, uromodulin and TGF-beta receptor type III.

Comparison of salmonid MHC class II sequences
The *DBA* sequence showed 46% amino acid (aa) sequence identity when compared to *Sasa-DAA*0101*. Further UniGene and EST database searches identified one partial rainbow trout *DBA* candidate in addition to two other Atlantic salmon and rainbow trout class II
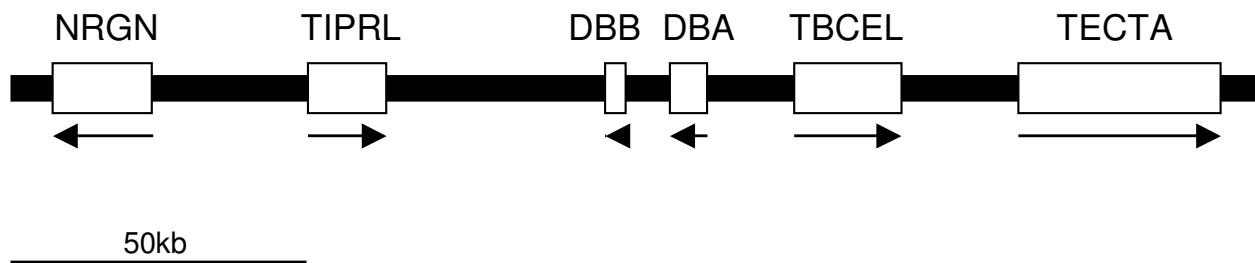

NRGN **TIPRL** **DBB** **DBA** **TBCEL** **TECTA**

50kb

Figure 1
Gene organization in the 630N19 BAC sequence. Black line (plus boxes) shows regions of sequence information, 211190 bp in total. Locus designation is based on sequence identity to matching ESTs and human nomenclature is used. Genes are depicted by white boxes and the arrows indicate gene orientation. Sequence is drawn to scale.

Table 2: Nucleotide sequence identities between the 630N19 BAC-encoded genes and matching Atlantic salmon EST sequences.

BAC gene	Abbrevation	EST Acc.No.	Identity (%)
Neurogranin	NRGN	DY707147	100
TIP41 TOR signalling pathway regulator-like	TIPRL	DY719987	99
MHC class II beta	DBB	DY726096	100
MHC class II alpha	DBA	EG757342	100
Tubulin folding cofaktor E-like	TBCEL	CA368403	91
Tectorin alpha	TECTA	N/A	NA

alpha-like sequences which we denoted *DCA* and *DDA* (Fig. 3). Amino acid sequence identity within tentative locus ranged from 85 to 93% across species while identity between all salmonid class II alpha sequences ranged from 39 to 61% both across loci and species.

All Atlantic salmon alpha 1 domains, except *Sasa-DDA*, contained cysteines potentially involved in formation of disulphide bridges (Fig. 3). The rainbow trout *Onmy-DBA* and *Onmy-DDA* sequences are partial ESTs disrupting further comparison. A distinct feature of the *Sasa-DBA* sequence is a unique 7 aa insertion at aa position 41–47, not seen in any of the other salmonid alpha 1 domains. This insertion probably influences which cysteines are involved in formation of disulphide bridges. The salmonid alpha 2 domains contained cysteines at positions 121 and 177 potentially involved in formation of disulphide bridges. Putative N-linked glycosylation sites were found in the *DAA* (aa 141–143), *DCA* (aa 134–136) and *DDA* (aa 116–118) sequences, but not in the *DBA* sequence.

The *Sasa-DBB* locus showed 49% amino acid sequence identity when compared to *Sasa-DAB*0201*. Full-length sequencing of the *DBB*-matching EST clone [Genbank: DV106186] was performed by primer walking. Rainbow trout sequences denoted *Onmy-DBB* as locus definition name was available in Genbank, but sequence identity in the beta2 domain suggested this may be an *Onmy-DAB* allele rather than a new locus (Fig. 4). EST database searches for other salmonid *DBB*-like sequences were not successful. All salmonid beta-1 domains contained cysteines at positions 10 and 75 potentially involved in formation of a disulphide bridge (Fig. 4) and have a puta-

tive N-linked glycosylation site at aa 37–39. The *Sasa-DBB* sequence had an additionally N-linked glycosylation site at position 84–86. The salmonid beta-2 domains contained cysteines at positions 113 and 169, while the *Sasa-DBB* sequence had an additional N-linked glycosylation site at positions 94–96. Looking at the CP/CYT/TM region, the *Sasa-DBB* sequence had a 12 aa additional C-terminal sequence when compared to the other salmonid MHC class II beta sequences.

In summary, Atlantic salmon harbours at least two divergent MHC class II beta sequences (*Sasa-DAB* and *Sasa-DBB*) and four divergent MHC class II alpha sequences (*Sasa-DAA*, *Sasa-DBA*, *Sasa-DCA* and *Sasa-DDA*), which all share most of the characteristics typical for vertebrate MHC class II molecules.

Polymorphism and expression of Atlantic salmon MHC class II molecules

To evaluate the polymorphic content of the BAC genes *Sasa-DBA* and *Sasa-DBB* we compared exon 2 sequences from 7 unrelated Atlantic salmon for both genes and to all available matching ESTs. One aa substitution was found in the *Sasa-DBB* exon 2 (G/D at pos.72), while no substitutions were found in the *Sasa-DBA* sequences. As the included sequences were derived from both Norwegian as well as Canadian waters, one may assume these genes have little or no polymorphism analogous to the nonclassical MHC class II genes identified in higher vertebrates. As for *Sasa-DCA* and *Sasa-DDA*, we have not performed any extensive polymorphic studies, but based on available EST information these loci contain little or no polymorphism
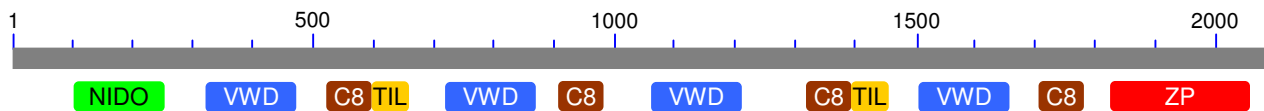

Figure 2
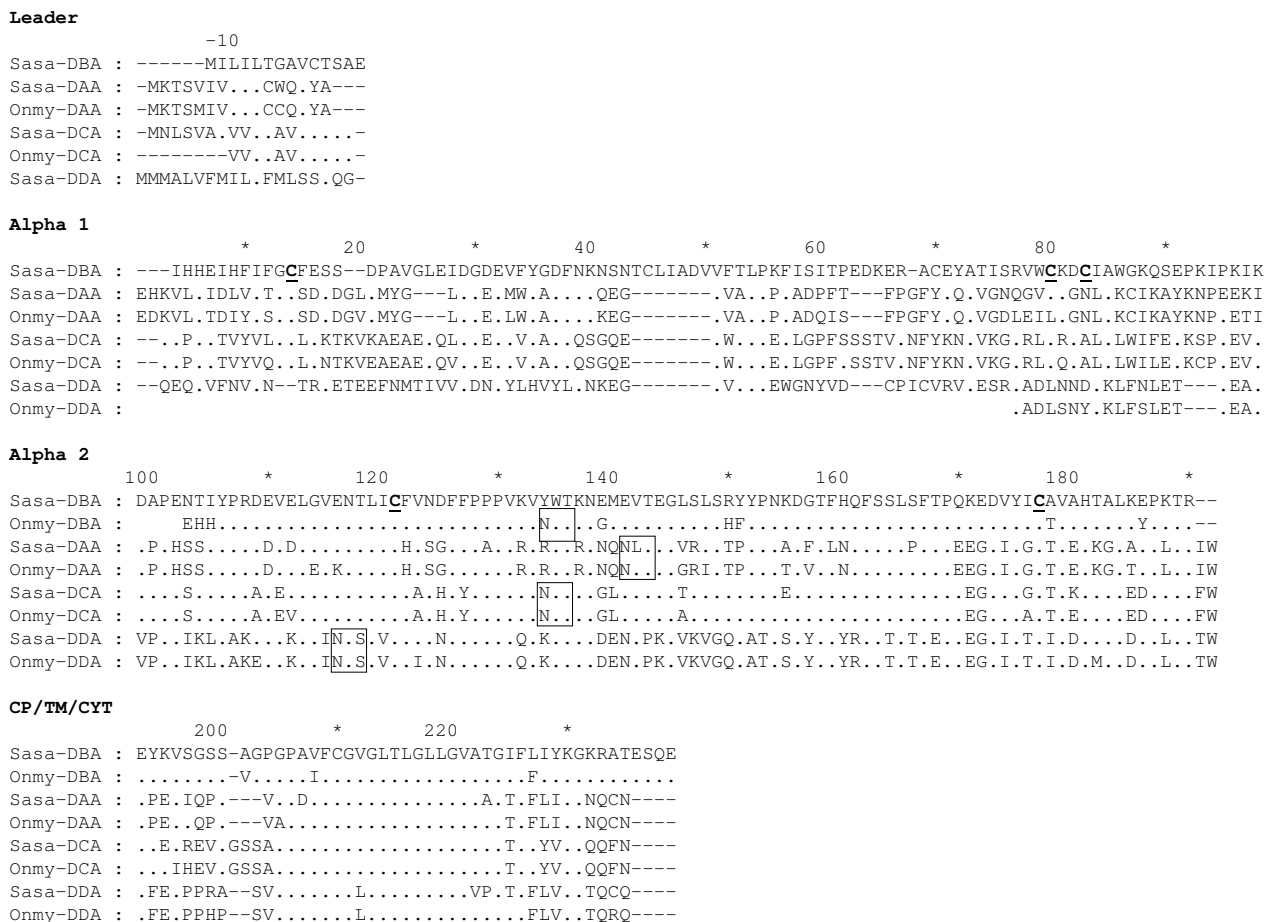**Screenshot of Blastp Domain hits for the Atlantic salmon tectorin alfa aa sequence.**

```
Leader
                    -10
Sasa-DBA : ------MILILTGAVCTSAE
Sasa-DAA : -MKTSVIV...CWQ.YA---
Onmy-DAA : -MKTSMIV...CCQ.YA---
Sasa-DCA : -MNLSVA.VV..AV.....-
Onmy-DCA : -------VV..AV.....-
Sasa-DDA : MMMALVFMIL.FMLSS.QG-


Alpha 1
                    *        20        *        40        *        60        *        80        *
Sasa-DBA : ---IHHEIHFIFGCFESS--DPAVGLEIDGDEVFYGDFNKNSNTCLIADVVFTLPKFISITPEDKER-ACEYATISRVWCKDCIAWGKQSEPKIPKIK
Sasa-DAA : EHKVL.IDLV.T..SD.DGL.MYG---L..E.MW.A....QEG-------.VA..P.ADPFT---FPGFY.Q.VGNQGV..GNL.KCIKAYKNPEEKI
Onmy-DAA : EDKVL.TDIY.S..SD.DGV.MYG---L..E.LW.A....KEG-------.VA..P.ADQIS---FPGFY.Q.VGDLEIL.GNL.KCIKAYKNP.ETI
Sasa-DCA : --..P..TVYVL..L.KTKVKAEAE.QL..E..V.A..QSGQE-------.W...E.LGPFSSSTV.NFYKN.VKG.RL.R.AL.LWIFE.KSP.EV.
Onmy-DCA : --..P..TVYVQ..L.NTKVEAEAE.QV..E..V.A..QSGQE-------.W...E.LGPF.SSTV.NFYKN.VKG.RL.Q.AL.LWILE.KCP.EV.
Sasa-DDA : --QEQ.VFNV.N--TR.ETEEFNMTIVV.DN.YLHVYL.NKEG-------.V...EWGNYVD---CPICVRV.ESR.ADLNND.KLFNLET---.EA.
Onmy-DDA :                                                                    .ADLSNY.KLFSLET---.EA.


Alpha 2
            100        *        120        *        140        *        160        *        180        *
Sasa-DBA : DAPENTIYPRDEVELGVENTLICFVNDFFPPPVKVYWTKNEMEVTEGLSLSRYYPNKDGTFHQFSSLSFTPQKEDVYICAVAHTALKEPKTR--
Onmy-DBA :     EHH.............................N...G.........HF..........................T.......Y....--
Sasa-DAA : .P.HSS.....D.D.........H.SG...A..R.R..R.NQNL..VR..TP...A.F.LN.....P...EEG.I.G.T.E.KG.A..L..IW
Onmy-DAA : .P.HSS.....D...E.K.....H.SG.....R.R..R.NQN...GRI.TP...T.V..N.........EEG.I.G.T.E.KG.T..L..IW
Sasa-DCA : ....S......A.E.........A.H.Y......N...GL.....T.......E............EG...G.T.K....ED....FW
Onmy-DCA : ....S......A.EV.........A.H.Y......N...GL.....A............EG...A.T.E....ED....FW
Sasa-DDA : VP..IKL.AK...K..IN.S.V....N.......Q.K....DEN.PK.VKVGQ.AT.S.Y..YR..T.T.E..EG.I.T.I.D....D..L..TW
Onmy-DDA : VP..IKL.AKE..K..IN.S.V..I.N.......Q.K....DEN.PK.VKVGQ.AT.S.Y..YR..T.T.I.D.M..D..L..TW


CP/TM/CYT
            200        *        220        *
Sasa-DBA : EYKVSGSS-AGPGPAVFCGVGLTLGLLGVATGIFLIYKGKRATESQE
Onmy-DBA : .........-V.....I.................F...........
Sasa-DAA : .PE.IQP.---V..D................A.T.FLI..NQCN----
Onmy-DAA : .PE..QP.---VA..................T.FLI..NQCN----
Sasa-DCA : ..E.REV.GSSA...................T..YV..QQFN----
Onmy-DCA : ...IHEV.GSSA...................T..YV..QQFN----
Sasa-DDA : .FE.PPRA--SV.......L.........VP.T.FLV..TQCQ----
Onmy-DDA : .FE.PPHP--SV.......L..............FLV..TQRQ----
```

Figure 3
Alignment of salmonid MHC class II alpha sequences. Genbank accession numbers are as follows: *Sasa-DAA*0101* [Genbank: L77086], *Onmy-DAA*02* [Genbank: CAB96451], *Sasa-DBA* (630N19) [Genbank: EG757343], *Sasa-DCA* [Genbank: DW549478], *Onmy-DCA* [Genbank: CR376525], *Sasa-DDA* [Genbank: DW557800], *Onmy-DDA* [Genbank: BX085673]. Dots indicate identities, dashes indicate gaps or missing sequence information. Cysteine residues involved in putative Ig fold are bold and underlined, and the unique N-linked glycosylation sites found in salmonids are boxed. Individual domains and regions are defined based on mammalian class II sequences.

A common feature for all the above described salmonid MHC class II sequences is their relatively low expression, based on hits in EST databases, when compared against their classical expressed counterpart. To investigate the MHC class II expression patterns, various tissues of unstimulated Atlantic salmon were analyzed for gene expression by means of real-time PCR (Fig. 5). Foregut, hindgut, head kidney, gills, spleen, heart, liver and muscle tissues were taken from one Atlantic salmon individual, where muscle represented a non-immunologically active tissue. EF1A served as the reference gene. The highest *Sasa-DAB* expression was detected in spleen, followed by gills, hindgut and head kidney. Lower *Sasa-DAB* levels were observed in heart, liver and foregut with muscle showing the lowest expression levels. This *Sasa-DAB* expression

pattern fits with the observations done by Koppang et al. [35]. *Sasa-DBB* showed highest expression in liver, followed by spleen, head kidney and heart. Our DBB-matching EST clone [Genbank: DV106186] descend from a liver cDNA library, confirming the expression of this gene in liver. Both macrophage and interhepatocytic cell populations have been found in liver of Atlantic salmon [36], suggestive of an immune function. Lower *Sasa-DBB* levels were observed in hindgut, gills and foregut with no detectable expression in muscle. *Sasa-DBA* showed highest expression levels in spleen and heart, followed by gills, muscle, hindgut, head kidney, foregut and liver. The expression patterns for the *DBA* and *DBB* genes were supported with additional semi-quantitative PCR (Fig. 6) using primers for different regions of the genes. *Sasa-DCA*

```
Leader
                       -10
Sasa-DBB : MYVLNCFSIHLLLLFSSLSEVVDSSD
Onmy-DBB : .SKSIR.Y.C.AVAL.----TLYET.
Sasa-DAB : .SMSI-.CVS.T.VL.----IFSGT.
Onmy-DAB : .SMPIA.Y.C.T..W.----IFSGT.


Beta 1
                   *         20         *         40         *         60         *         80         *
Sasa-DBB : EDFAHDDAWCRFSSRDLHNMEYILEHHFNKILVAQYNSTTERWTGYTAWGVISAEKWNEDPDEIPRRRTDMGVLCKPYANRIYNATEMFM
Onmy-DBB : GYASDVVTR.LY..I.M.GA.F.QSYT...VEHLRF...VGEFV...EL.LKN.KRL.RGQ-.VVQM.GELER....N.DIH.R.ILDKT
Sasa-DAB : GY.EQVVRQ..Y..K..QGI.F.DSYV...AEYVRF...VGKYV...EL..KN..A..KG.-.LAVELGELERF..HN.AIY.S.ILDKT
Onmy-DAB : GY.HQRVTQ.....E...G....DSYV..QAEYIRF...VG.FV...EH.LKN..A..S.AGILGQEQVQLESY..HN.AID.S.ILDKT


Beta 2
                   100         *         120         *         140         *         160         *         180
Sasa-DBB : VEPNVTLRLEGPSSD---SSLVCSVHFFYPKHIRVTWLRNGEEVTSDVTSTDVLANGLWSYQIQSYLKYTPTTGERITCMVEHISQTEPKLYYW
Onmy-DBB : ...H.R.SSVT.P.GRHPAM.M..AYD....P.......D.R..K......EE...D.Y...H.H.E...KS..K.S.......L...MM.H.
Sasa-DAB : ...H.R.SSVA.P.GRHPAM.M..AYD....P.......D.R..K......EE...D.Y...H.H.E...RS..K.S.......L...MV.H.
Onmy-DAB : ...H.R.SSVT.P.GRHPAM.M..AYD....Q.......D.R..K......EE...D.Y...H.H.E...KS..K.S.......L...MM.H.


CP/TM/CYT
                   *         200         *         220         *
Sasa-DBB : DPSLPKSEKNKIVIGVCGLLLGVVFVVAGLIYWAKSTGRLLGLIGERDYGTCD
Onmy-DBB : .....EA.R...A..AS..V..TILAL.....YK.KSSGV.------------
Sasa-DAB : .....EA.R...A..AS..V..AILAL.....YK.KSSGV.------------
Onmy-DAB : .....EA.R...A..AS..V..TILAL.....YK.KSSGV.------------
```

Figure 4
Alignment of salmonid MHC class II beta sequences. Sequence references are as follows: *Sasa-DBB* (630N19) [Genbank: DY726096], *Sasa-DAB* [Genbank: CAD27784], *Onmy-DBB* [Genbank: AAD53026], *Onmy-DAB* [Genbank: AAA79133]. Dots indicate identities, dashes indicate gaps or missing sequence information. Cysteine residues involved in putative Ig fold are bold and underlined, and the unique N-linked glycosylation sites found in salmonids are boxed. Individual domains and regions are defined based on mammalian class II sequences.

showed high expression in hindgut with low expression levels in gills and liver and none in remaining tissues. Hindgut of teleosts have a well recognised capacity to take up antigen [37], an ability that has been related to the existence of a mucosal immune system in fish. *Sasa-DDA* showed a strikingly high expression in spleen. Spleen is a central organ for antigen display and response, and a major component of the fish lymphoid system. Lower *Sasa-DDA* levels were observed in heart, head kidney, gills, liver and hindgut with foregut and muscle showing the lowest expression levels. Based on raw C_T values, *DDA* expression level in spleen was the only gene with an expression level comparable to classical DAB expression.

Based on lack of polymorphism and expression patterns deviating from their classical counterpart *Sasa-DAB*, the *Sasa-DBA*, *Sasa-DBB*, *Sasa-DCA* and *Sasa-DDA* molecules can be classified as non-classical MHC molecules similar to mammalian non-classical HLA-DM or HLA-DO molecules and can be involved in catalysis of peptide loading and stabilization of the classical *Sasa-DAA/DAB* molecules. However, highest expression of *Sasa-DBB* and *Sasa-DCA* in liver and hindgut respectively, and a *Sasa-DDA* expression level in spleen comparable to *Sasa-DAB*, may

indicate that these molecules have a unique and yet unidentified function in these organs in salmonids.

Phylogenetic analysis of MHC class II related molecules
The phylogenetic relationship between salmon and other expressed teleosts MHC class II sequences were subjected to phylogenetic analyses. For MHC class II beta, the majority of sequences cluster according to phylogeny with zebrafish, medaka, and salmon as representatives for cypriniformes, acanthopterygii and salmonids respectively (Fig. 7). The connection between locus and expressed beta sequences are known for a few species. For medaka and stickleback, the DAB and DBB represent two very similar sequences from two different loci [12] suggestive of a recent gene duplication. The rainbow trout DBB sequence is published as *Onmy-DBB* although proof of a duplicated class II beta locus in rainbow trout is lacking. In zebrafish (Dare), authors have reported a varying number of class II loci [11,14] although the assumption has been that *Dare-DAA/-DAB* was the major expressed molecule [13]. Based on sequence divergence, the *Dare-DCB* and *-DDB* sequences most likely represent two additional expressed zebrafish class II beta loci. However, lacking knowledge on polymorphic content and expression
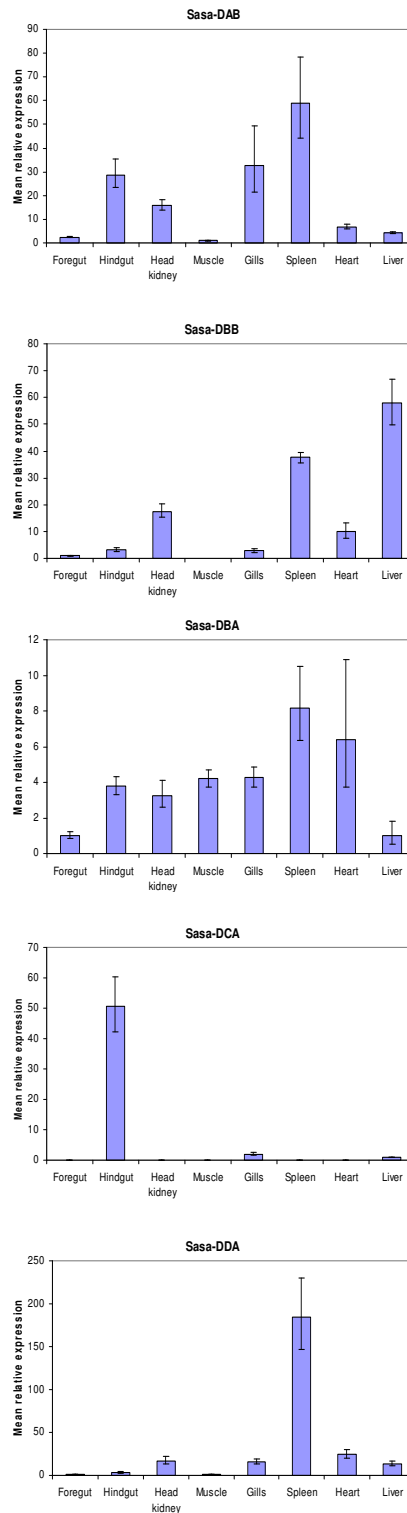
Figure 5
Expression analysis of Atlantic salmon MHC class sequences. Relative expression of *Sasa-DAB*, *Sasa-DBB*, *Sasa-DBA*, *Sasa-DCA* and *Sasa-DDA* in various tissues of Atlantic salmon using EF1A as reference gene. Data are mean $\Delta\Delta C_T$ levels ± SEM values relative to the tissue with lowest expression set to one.
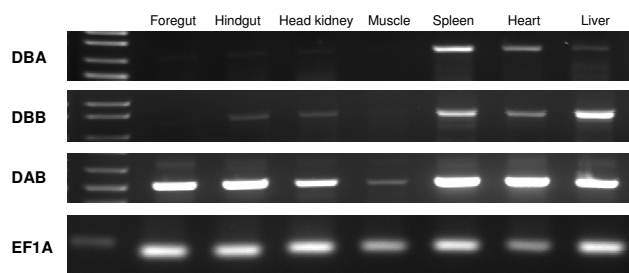
Figure 6
Expression analysis of *Sasa-DBA* and *Sasa-DBB*. Semi-quantitative RT-PCR of *Sasa-DBA* (DBA19F/DBA660R) and *Sasa-DBB* (DBB98F/DBB825R) in various tissues of Atlantic salmon. Primers are located at different regions of the genes when compared to real time PCR analysis. *Sasa-DAB* was included as positive control, while EFIA was internal house-keeping control.

profiles hinders definition of these genes as classical or non-classical. Fathead minnow (Pipr) contains four very divergent class II beta sequences, but their genomic organization remains to be characterized. Although difficult, some lineage-specific residues/motifs are noticeable (the sequence alignments are found in Additional File 1). The acanthopterygii class II beta sequences contain lineage specific residues such as the L215, M223, I242, L245, S248, F252, R257, R260 aa's motif in the COOH- end. Fewer distinct lineage-spesific residues exist for the cypriniformes sequences, but the S51, M53 or the I245, I247 motifs are potential candidates allowing for some exceptions. Cypriniformes as well as acanthopterygii sequences share a RILV motif at aa position 262–266.

Six teleost class II beta sequences from zebrafish, stickle-back, fathead minnow, xiphophorus and Atlantic salmon (*Sasa-DBB*) identified through a thorough database search show low sequence identity to the above described sequences. No lineage specific residues are apparent between these and the above lineages although some residues shared between *Sasa-DBB* and Ximu-DXB suggest they have a common ancestor. Both these sequences share a 132–134 RLE, Y191 and T203 motif in the beta2 domain. The xiphophorus DXB sequence has been reported as a non-classical sequence [26] while none of the other EST sequences have been defined as MHC sequences in the database or in publications. Identifying expressed MHC class II molecules ease with growing number of ESTs available as exemplified by zebrafish. A total of 1.3 million zebrafish EST sequences are available in Genbank, probably explaining why we now identify four expressed MHC class II beta loci.

For MHC class II alpha, the majority of sequences also cluster according to phylogeny with medaka, zebrafish and salmon as representatives for the acanthopterygii, cypriniformes, and salmonid clusters respectively (Fig. 8). Stickleback has two reported class II alpha loci, denoted as *Gaac-DAA* and *-DAB* [12]. If the third stickleback EST sequence represents a third expressed locus or is an allelic version of the two reported remains to be established. As for the class II beta sequences, zebrafish also contains at least three different expressed class II alpha sequences most likely representing three different loci. The three different class II alpha sequences for fathead minnow is also comparable to its four different class II beta sequences. In addition to their classical DAA sequences, Atlantic salmon and rainbow trout contain three other class II alpha sequences, which are much more divergent than the different sequences identified in zebrafish or fathead minnow. The close relationship between the *Sasa-DBA* and *Sasa-DCA* sequences suggest a common ancestral gene. If *Sasa-DDA* is a more ancient duplication of the same ancestral locus, which has evolved away from the *Sasa-DBA* and *Sasa-DCA* sequences, remains to be established. The genomic location and gene surroundings of the *Sasa-DCA* and *-DDA* loci is currently unknown. As for the beta sequences, some lineage-specific residues/motifs are noticeable for the alpha sequences as well (the sequence alignments are found in Additional File 2). The acanthopterygii class II alpha sequences contain lineage specific residues such as the P82, Q107, I108, S131, V252, E269, S271 residues shared by most sequences. Lineage specific residues for the cypriniformes sequences are the D85, T154, P159, Q180, R182, V225 aa's with a few exceptions. Lacking lineage specific residues for the duplicated salmonid sequences may suggest ancient duplications or different origin.

As more ESTs are sequenced from teleosts and we move deeper into each library, more class II loci from more fish will most likely appear enabling a broader understanding of their origin and evolution.

Genome evolution and paralagous regions
Beside the major human MHC region on chromosome 6, paralogous copies of genes from the MHC region have been found on chromosomes 1, 9 and 19 which are thought to be derived from two whole genome duplications that occurred in a common ancestor of all vertebrates [38]. Evolutionary analysis of complete genome sequences from the pufferfish species *Tetraodon* [39] and *Takifugu* [40] suggest that an additional genome duplication occurred early in the teleost lineage [350 million years ago (Mya)], close to the origin of the teleost fish themselves. All the fish of the family *Salmonidae* apparently descended from yet another unique genome duplication event that occurred in the lineage leading to
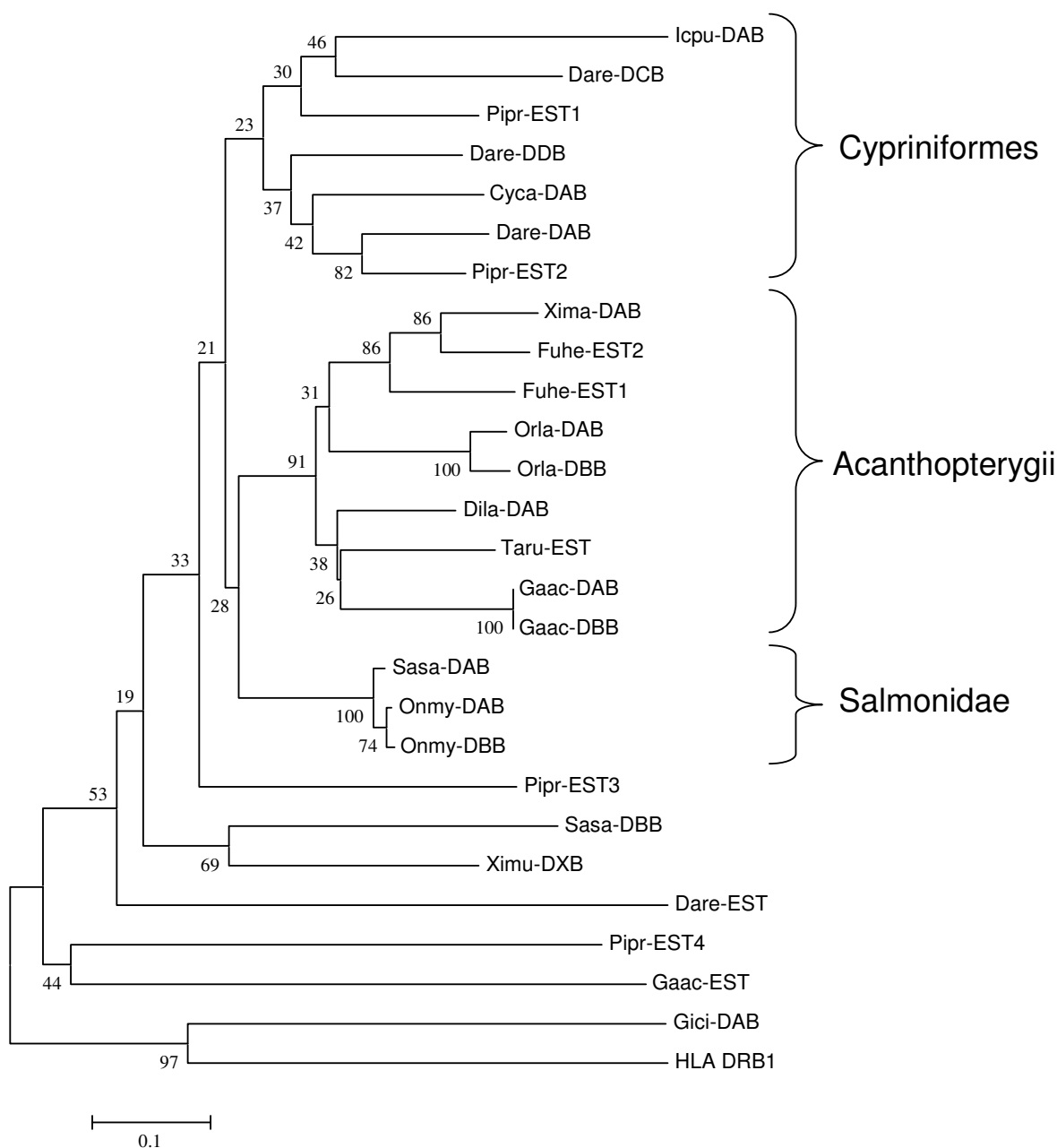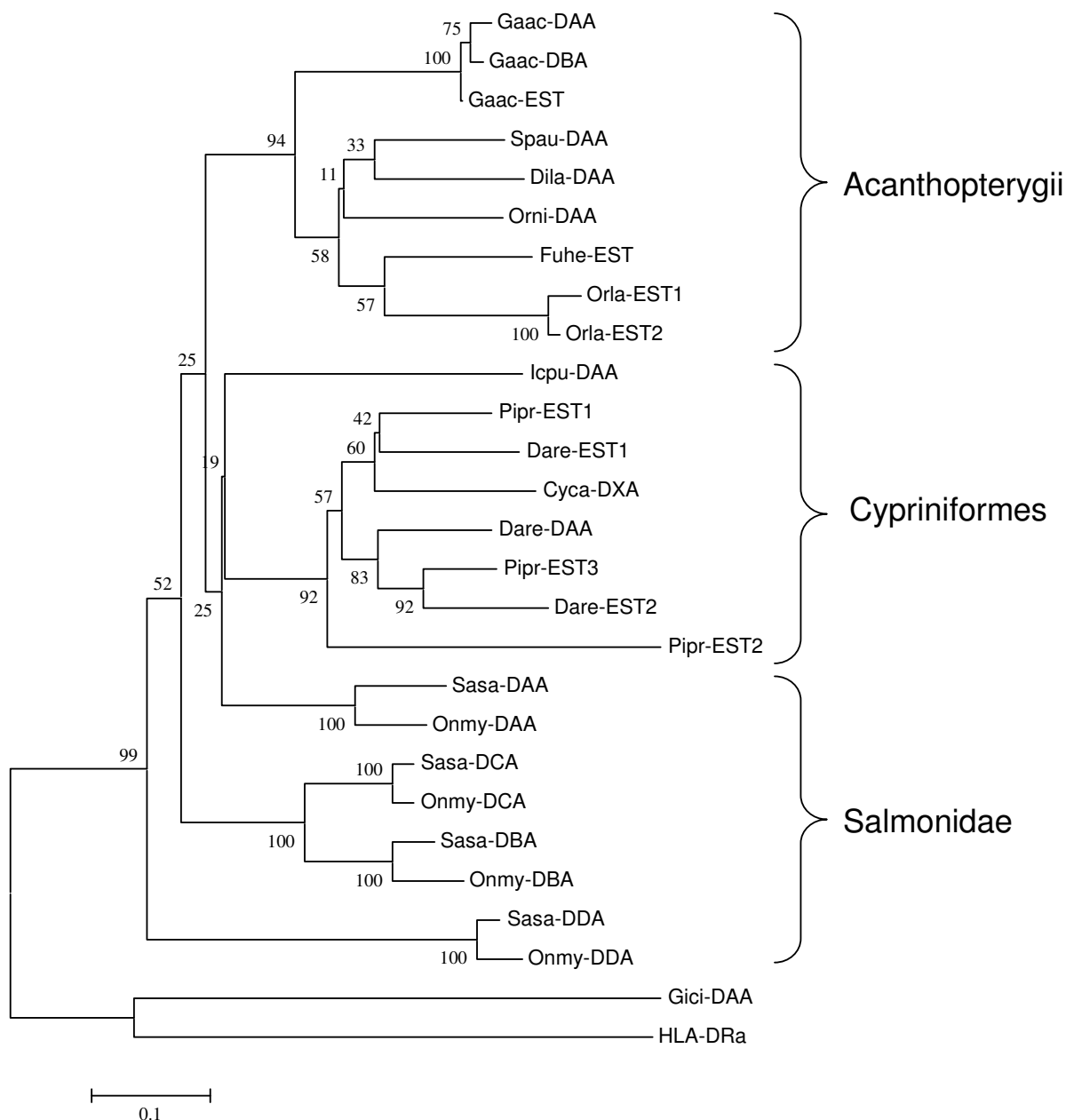
Figure 7
Phylogeny of teleost MHC class II beta sequences. Phylogenetic tree analysis by NJ method for full-length amino acid sequences. Consensus trees were based on 1000 bootstrap replications and reported with the bootstrap support values (in percent) indicated above the respective nodes. Sequence references are as follows: Sasa-DAB [GenBank: CAD27784], Sasa-DBB [GenBank: DY726096], Onmy-DAB [GenBank: AAA79133], Onmy-DBB [GenBank: AAD53026], Orla-DAB [GenBank: BAA94279], Orla-DBB [GenBank: BAA94280], Xima-DAB [GenBank: AAC05652], Gaac-DAB [GenBank: AAU01918], Gaac-DBB [GenBank: AAU01920], Gaac-EST [Genbank: DN681207], Dila-DAB [GenBank: ABH09450], Fuhe-EST1 [GenBank: CN976662], Fuhe-EST2 [GenBank: CN984097], Taru-EST [GenBank: CA846190], Icpu-DAB [GenBank: AAB67871], Cyca-DAB [GenBank: CAA64709], Dare-DAB [GenBank: NP_571551], Dare-DCB [GenBank: CAD56804], Dare-DDB [GenBank: AAA87893], Dare-EST [Genbank: CK126567], Pipr-EST1 [GenBank: DT084791], Pipr-EST2 [GenBank: DT351641], Pipr-EST3 [GenBank: DT139435], Pipr-EST4 [GenBank: DT355684], Ximu-DXB [GenBank: AAS55041], Gici-DAB [GenBank: AAF82681] and HLA_DRB1 [GenBank: AAA59781].

Figure 8
Phylogeny of teleost MHC class II alpha sequences. Phylogenetic tree analysis by NJ method for full-length amino acid sequences. Consensus trees were based on 1000 bootstrap replications and reported with the bootstrap support values (in percent) indicated above the respective nodes. Sequence references are as follows: Sasa-DAA [GenBank: AAL40122], Sasa-DBA [GenBank: EG757342], Sasa-DCA [GenBank: DW549478], Sasa-DDA [GenBank: DW557800], Onmy-DAA [GenBank: CAB96451], Onmy-DBA [GenBank: CX137594], Onmy-DCA [GenBank: CR376525], Onmy-DDA [GenBank: BX085673], Gaac-DAA [GenBank: AAU01917], Gaac-DBA [GenBank: AAU01919], Gaac-EST [GenBank: DN737221], Spau-DAA [Gen-Bank: AAY42849], Orla-EST1 [GenBank: DC261023], Orla-EST2 [Genbank: BJ884671], Fuhe-EST [Genbank: CV816904], Dila-DAA [Genbank: ABH09446], Orni-DAA [Genbank: AAF66843], Cyca-DXA [Genbank: CAA64707], Pipr-EST1 [Genbank: DT253073], Pipr-EST2 [Genbank: DT092734], Pipr-EST3 [Genbank: DT311896], Dare-DAA [Genbank: NP_571565], Dare-EST1 [Genbank: CK018982], Dare-EST2 [Genbank: CO928661], Icpu-DAA [Genbank: AAD39865], Gici-DAA [GenBank: AAA49310], HLA-DRA [Genbank: NP_061984]

salmonid fishes 25–100 Mya [41]. Evidence supporting the unique and relatively "young" salmonid genome duplication event was found by studying MHC class I regions in rainbow trout [9] and Atlantic salmon [7]. Here, class IA and class IB regions are found on separate chromosomes with a sequence identity of 85% (rainbow trout) and 82,5% (Atlantic salmon), suggestive of the duplication occurring approx. 60 Mya. Two other more distant MHC class I-like genes defined as ZE [42] and L [43] might be remnants from two of the earlier whole-genome duplication events.

The ancestral MHC must have contained both class I as well as class II. Shark and frog both contain MHC regions resembling the human MHC region containing both MHC class I as well as MHC class II molecules [1]. The MHC class II region is assumed to have been translocated to another area early in the teleost branch as teleosts are the only phylogenetic group with no linkage of class I and class II. In contrast to teleosts, shark seems to have chosen a different paralogue as its major MHC class II locus being more similar to the tetrapod MHC class II molecules than to the classical teleost class II. The salmonid DBA/DCA and DDA sequences cluster closer to the tetrapod/shark sequences and could also be a version of this paralogue. Based on the phylogenetic clustering, the salmonid DBA and DCA genes most likely originated from the whole-genome duplication event that occurred in salmonids 60 Mya, similar to the class I IA and IB region duplication. DDA and DAA might be remnants of two older whole-genome duplication events potentially analogous to the human Chromosome 1, 6, 9, 19 paralogues. The human TIPRL homolog [Genbank: NM_152902] is located on chromosome 1q23.2, a region that also encompasses MHC paralogous genes such as CD1 and COL11A1 [44] while the human homologs of NRGN, TECTA and TBCEL all reside on chromosome 11. If the salmon TIPRL and class II genes originate from the primordial human chromosome 1 remains to be established.

Other teleost genomes show some synteny to the molecules encoded on the Atlanitc salmon BAC clone. A tblastn search against Ensembl genomic databases of tetraodon, stickleback, medaka and zebrafish (Table 3) showed a tight linkage between the TBCEL and TECTA genes in all species except tetraodon. Stickleback and

medaka also harbour a TIPRL homolog on the same chromosome as TBCEL and TECTA, although at a distance of approximately 12 Mb and 17 Mb away respectively. The zebrafish NRGN homolog is located on the same chromosome as TBCEL and TECTA approximately 12 Mb in distance. Homologs to the *Sasa-DBB* and *Sasa-DBA* genes were not found in any other available teleost genomes. In summary, the *Sasa-DBB* and *Sasa-DBA* genes are most likely a duplication of a primordial class II region that has been translocated into the NRGN-TIPRL-TBCEL-TECTA region. The duplication seems to be salmon specific although more teleost genome sequences may prove us wrong.

Conclusion
In this study we describe a novel MHC class II region in Atlantic salmon consisting of two closely linked MHC class II alpha and beta loci denoted *Sasa-DBA* and *Sasa-DBB* respectively. The genes are not linked to the classical class II loci and have low sequence identity to classical sequences. The region shows some conserved synteny towards other described teleost genomic regions. Two additional divergent class II alpha sequences were also identified and represent two additional salmonid class II alpha loci. Expression patterns and lack of polymorphism make these genes non-classical class II analogues indicating that these molecules may have a unique and yet unidentified function in salmonids. In conclusion, we have identified a more complex picture for MHC class II in teleost fishes, with most species containing multiple expressed loci. These newly discovered genes may expand our understanding of MHC class II function in vertebrates.

Methods
BAC library screening
An Atlantic salmon (*Salmo salar*) CHORI-214 bacterial artificial chromosome (BAC) library was obtained from BACPAC Resources, Children's Hospital Oakland Research Institute [45]. The library consisted of approximately 299,000 recombinant clones, representing 18-fold genome coverage and an average insert size of 188 kb [27]. Probes specific for the *S. salar* MHC class II were PCR amplified (primers listed in Table 4) from cDNA clones (Grimholt pers. com.) and purified from agarose gel slices with the GenClean III Kit (Qbiogene). Probes were radioactive labelled with α^{32}P-CTP (Amersham) using

Table 3: Tblastn search of Atlantic salmon 630N19 BAC clone genes against available teleost genomes.

	NRGN	TIPRL	DBB	DBA	TBCEL	TECTA
Tetraodon	Chr.16 (0,8 Mb)	Chr.Un_random (138,0 Mb)	N/A	N/A	Chr.16 (8,0 Mb)	Chr.Un_random (82,6 Mb)
Stickleback	Chr.1 (8,4 Mb)	Chr.7 (12,6 Mb)	N/A	N/A	Chr.7 (24,9 Mb)	Chr.7 (24,9 Mb)
Medaka	Chr.13 (15,3 Mb)	Chr.14 (13,1 Mb)	N/A	N/A	Chr.14 (30,4 Mb)	Chr.14 (30,4 Mb)
Zebrafish	Chr.5 (69,7 Mb)	Chr.10 (24,8 Mb)	N/A	N/A	Chr.5 (57,4 Mb)	Chr.5 (55,7 Mb)

Rediprime Random Labelling Kit (Amersham), including spermine precipitation of labelled DNA (Feinberg and Vogelstein, 1984). Filter hybridizations were conducted as described by CHORI. Probed BAC library filters were stored in Phospho-image cassettes for 2–24 hours and hybridizations visualized by Typhoon Phospho Image Scanner (Amersham).

Characterization of positive BAC clones
DNA from positive BAC clones was isolated from 3 ml LB overnight cultures, supplemented with 20 µg/ml chloramphenicol, as described by CHORI. Mapping of BAC clones was performed by restriction fragment analysis with *Not*I and *Nru*I (NEB). Secondary screening of positive BACs were done by Southern blot hybridization of digested DNAs using *DAA* and *DAB* cDNA probes, in addition to end-labelled Sp6 and T7 oligos. End-labelling was performed with γ^{32}P-ATP (Amersham) using T4 Polynucleotide Kinase (NEB) and hybridized overnight at 42°C with Church hybridization buffer. Filters were washed 3 × 30 min with 3 × SSC/0.1%SDS at 42 °C and visualized by Typhoon Phospho Image Scanner (Amersham).

BAC subclone library and sequencing
The 630N19 BAC clone positive for DAA was subjected to nucleotide sequencing using a shotgun strategy. BAC DNA was isolated and purified by an alkaline lysis procedure using Nucleobond columns (BD Biosciences ClonTech). 15 microgram (or more) of isolated DNA was nebulized (Invitrogen) (20PSI/15s), trimmed and end filled with Mung Bean Nuclease (NEB), T4 DNA polymerase (NEB) and Klenow Fragment (NEB). The blunt-ended DNA was size fractioned by electrophoresis and the fragments corresponding to 1–4 kb were excised and gel purified using GeneClean III (Qbiogene). Fragments were ligated into a pUC19/*Sma*I vector (Fermentas) and transformed into XL-Gold competent cells (Stratagen) using the manufacturer's protocol. DNA plasmid from BAC subclones were isolated using standard alkaline lysis method [46]. More than 3000 BAC subclones were sequenced both using universal forward and reverse M13 primers with Big Dye Terminator Kit version 3.1 (Applied Biosystems) and analysed on ABI 3100 DNA automated sequencer. Sequences were screened for *E. coli* content using Pregap4 [47,48]. The resulting ~5500 high quality sequence reads

Table 4: Primers used for probes, sequencing and gene expression.

Primer	Sequence (5'-3')	Position	Comments
oOV4F	GCGCCATACTGGACAAGACAG	Exon 2	cDNA amplification *Sasa-DAB*
oOV6R	CTGCTGCAGATTCAGCAACAT	3' UTR	cDNA amplification *Sasa-DAB*
DAB40F	ATGTCGATGTCTATCTTCTG	Exon 1	cDNA amplification *Sasa-DAB*
DAB489R	GTACCAGTCCCCGTTAGCCAG	Exon 3	cDNA amplification *Sasa-DAB*
DAA66F	TGCTGGCAGGTGTATGCAGAA	Exon 1	cDNA amplification *Sasa-DAA*
DAA660R	GAATGTTCCGGCAGCCACTCC	Exon 4	cDNA amplification *Sasa-DAA*
DBA19F	TCACTGGAGCCGTTTGCACTT	Exon 1	cDNA/gDNA amplification *Sasa-DBA*
DBA51F	CATCATGAAATTCACTTTATC	Exon 2	cDNA/gDNA amplification *Sasa-DBA*
DBA304R	CTTTGATTTTTGGTATTTTCG	Exon 2	cDNA/gDNA amplification *Sasa-DBA*
DBA660R	CCCCAGCAGCCCAAGAGTGAG	Exon 4	cDNA/gDNA amplification *Sasa-DBA*
DBB98F	CTCTGAGACTTGATTATGTAT	Exon 1	cDNA/gDNA amplification *Sasa-DBB*
DBB177F	TTGATTCCTCAGATGAAGATT	Exon 2	cDNA/gDNA amplification *Sasa-DBB*
DBB441R	CCATAAACATCTCTGTTGCGT	Exon 2	cDNA/gDNA amplification *Sasa-DBB*
DBB825R	CAGTAGATTTCGCCCAGTAGA	Exon 4	cDNA/gDNA amplification *Sasa-DBB*
DA847F	GATGGCAAAGAGGAAAGTGAG	3' UTR	FAM labeled minisatellite primer for *Sasa-DAA*
DA1054R	TTGTTATGCTCTACCTCTGAA	3' UTR	Minisatellite primer for *Sasa-DAA*
Sp6	ATTTAGGTGACACTATA	pTARBAC2.1	End sequencing BAC vector Sp6 end
T7	TAATACGACTCACTATAGGG	pTARBAC2.1	End sequencing BAC vector T7 end
M13F	GTTTTCCCAGTCACGAC	pUC19	DNA plasmid sequencing primer
M13R	CAGGAAACAGCTATGAC	pUC19	DNA plasmid sequencing primer
EF1A_F	CACCACCGGCCATCTGATCTACAA	Acc. No. AF321836	Real-time PCR Elongation factor 1-alpha
EF1A_R	TCAGCAGCCTCCTTCTCGAACTTC	Acc. No. AF321836	Real-time PCR Elongation factor 1-alpha
DAB566F	ATGGTGGAGCACATCAGCC	Exon 3	Real-time PCR *Sasa-DAB*
DAB620R	CTCAGCCTCAGGCAGGGAC	Exon 4	Real-time PCR *Sasa-DAB*
DBB418F	GCCATATGCTAATCGGATCTAC	Exon 3	Real-time PCR *Sasa-DBB*
DBB500R	ACACACAAGGCTGGAGTCACT	Exon 4	Real-time PCR *Sasa-DBB*
DBA538F	GCATCGCATGGGGCAAACAG	Exon 3	Real-time PCR *Sasa-DBA*
DBA641R	ATTCACAAAGCAGATGAGGGTG	Exon 4	Real-time PCR *Sasa-DBA*
DCA277F	GATTGTGTCGGGATGCCTTG	Exon 2	Real-time PCR *Sasa-DCA*
DCA357R	CTCCTCAGCCCTGGGGTAGAT	Exon 3	Real-time PCR *Sasa-DCA*
DDA225F	AGTTGCAGAGAGCCGAAGAGC	Exon 2	Real-time PCR *Sasa-DDA*
DDA317R	TTCACTTCATCCTTGGCATACA	Exon 3	Real-time PCR *Sasa-DDA*

were basecalled using PHRED [49,50], assembled using PHRAP [51] and then viewed and edited using Consed [52]. Remaining gaps or areas of uncertainty were analyzed by designing new internal sequencing primers via Autofinish [53] within Consed, followed by direct sequencing of shotgun clones containing segments of interest. Restriction digest of the isolated BAC were compared to in silico digests for assembly conformation. BAC 630N19 was deposited in Genbank under the accession number [Genbank: EU008541].

Gene predictions and phylogenetic analysis
DIGIT [54] and GENSCAN [55] were used to predict novel genes and to identify open reading frames. Dotter [56] was used to compare the BAC sequence to itself as well as to other BACs and to identify duplicated regions. Blast searches identified possible functions of predicted genes [57,58]. EST cluster consensus files were obtained from GRASP EST database [59,60] and other EST were identified with UniGene [61,62]. Full-length sequencing of the EST cDNA clone DV106186 was performed by primer walking. Sim4 [63] and Spidey [64] were used to adjust exon and intron boundaries aligning EST/cDNA sequences to the BAC sequence. Repeatmasker [29,65] was used to identify repeats. Multiple sequence alignments of the assumed or verified expressed exons were done using ClustalX [66] followed by manual inspection.

Phylogenetic trees were created using neighbour-joining method in MEGA3 [67]. Consensus trees were based on 1000 bootstrap replications and reported with the bootstrap support values (in percent) indicated above the respective nodes. Gaps were removed and phylogenetic data reported using the Poisson correction model with uniform rates across all sites.

Gene expression analysis
mRNA was extracted from Atlantic salmon tissues (foregut, hindgut, head kidney, muscle, gills, spleen, heart and liver) using QuickPrep *micro* mRNA Purification Kit (GE Healthcare Life Science). 1 μl of mRNA sample was used for quantification with Nanodrop spectrometer (Nanodrop Technologies, DE). All samples were DNase treated using Turbo DNA-free™ (Ambion, Austin, TX, USA). Gene specific PCR primers (Table 4) were designed manually for the amplification of approximately 100 bp fragments and synthesized by ProOligo (Paris, France). mRNA was treated with DNase and amplicons were, when possible, placed over introns and product size and specificity was confirmed by agarose gel electrophoresis (Gel logic 200 Imaging system, Kodak) and sequencing. Quantitative real-time PCR was conducted on a Chromo 4 Real-time Detection System (MJ Research, USA). Reactions were performed in 25 μl including ~10 ng of mRNA with a one-step real-time RT-PCR kit according to the manufac-

turer's instructions (RealQ RT-PCR Master Mix One Step, Ampliqon). PCR parameters were 55°C for 30 min, 95°C for 15 min, followed by 40 cycles consisting of 95°C for 15s, 60°C at 30s and 72°C at 30s. A melting-curve analysis was performed for each sample to check for unspecific amplification. Baseline level and cycle threshold (C_T) were set manually and C_T values > 37 were rejected. Relative expression of mRNA in relation to the housekeeping gene elongation factor 1α (EF1A) was calculated using the $\Delta\Delta C_T$ method [68]. Data from real-time RT-PCR are presented as the mean ± SEM of at least three independent samples and the tissue with the lowest expression was used as calibrator.

Tissue-specific expression of selected genes *Sasa-DBA* and *Sasa-DBB* were determined by gene-specific amplification including MHC class II beta (*Sasa-DAB*) as a positive reference gene and EF1A as internal housekeeping control. The primer sequences are presented in Table 4. The PCR reactions (25 μl total volume) included 20 ng cDNA, 0,4 μM of each primer, 200 μM dNTPs, 1 × buffer (Herculase 10 × PCR buffer), and 0.5 units of Herculase polymerase (Stratagene). Amplifications were conducted as follows: initial denaturation at 95°C for 3 min, 30 cycles consisting of 95°C for 20 s, 54°C for 30 s, 72°C extension for 90 s; followed by a final extension of 72°C for 10 min. PCR products were separated on 2% agarose gel stained with EtBr for visualization with Gel logic 200 Imaging system (Kodak).

Authors' contributions
HH performed library screening, sequencing and BAC restriction mapping, annotations, gene expression analysis, sequence data analysis and drafted the manuscript. MFL performed sequencing, annotations, sequence data analysis and revision of manuscript. HGB performed library screening and sequencing. UG contributed to planning, design, direction, analysis and revision of manuscript. All authors read and approved the final manuscript.

Additional material

References

1. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF: **Primitive synteny of vertebrate major histocompatibility complex class I and class II genes.** *Proc Natl Acad Sci U S A* 2000, **97:**4712-4717.
2. Sato A, Figueroa F, Murray BW, Malaga-Trillo E, Zaleska-Rutczynska Z, Sultmann H, Toyosawa S, Wedekind C, Steck N, Klein J: **Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes.** *Immunogenetics* 2000, **51:**108-116.
3. Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJ: **The major histocompatibility class I locus in Atlantic salmon (Salmo salar L.): polymorphism, linkage analysis and protein modelling.** *Immunogenetics* 2002, **54:**570-581.
4. Phillips RB, Zimmerman A, Noakes MA, Palti Y, Morasch MR, Eiben L, Ristow SS, Thorgaard GH, Hansen JD: **Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region.** *Immunogenetics* 2003, **55:**561-569.
5. Michalova V, Murray BW, Sultmann H, Klein J: **A contig map of the Mhc class I genomic region in the zebrafish reveals ancient synteny.** *J Immunol* 2000, **164:**5296-5305.
6. Clark MS, Shaw L, Kelly A, Snell P, Elgar G: **Characterization of the MHC class I region of the Japanese pufferfish (Fugu rubripes).** *Immunogenetics* 2001, **52:**174-185.
7. Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS, Koop BF: **Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (Salmo salar).** *BMC Genomics* 2007, **8:**251.
8. Matsuo MY, Asakawa S, Shimizu N, Kimura H, Nonaka M: **Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (Oryzias latipes).** *Immunogenetics* 2002, **53:**930-940.
9. Shiina T, Dijkstra JM, Shimizu S, Watanabe A, Yanagiya K, Kiryu I, Fujiwara A, Nishida-Umehara C, Kaba Y, Hirono I, Yoshiura Y, Aoki T, Inoko H, Kulski JK, Ototake M: **Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (Oncorhynchus mykiss), a species with a presumably recent tetraploid ancestry.** *Immunogenetics* 2005, **56:**878-893.
10. Flajnik MF, Kasahara M: **Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system.** *Immunity* 2001, **15:**351-362.
11. Kuroda N, Figueroa F, O'hUigin C, Klein J: **Evidence that the separation of Mhc class II from class I loci in the zebrafish, Danio rerio, occurred by translocation.** *Immunogenetics* 2002, **54:**418-430.
12. Reusch TB, Schaschl H, Wegner KM: **Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback.** *Immunogenetics* 2004, **56:**427-437.
13. Sultmann H, Mayer WE, Figueroa F, O'hUigin C, Klein J: **Organization of Mhc class II B genes in the zebrafish (Brachydanio rerio).** *Genomics* 1994, **23:**1-14.
14. Sambrook JG, Figueroa F, Beck S: **A genome-wide survey of Major Histocompatibility Complex (MHC) genes and their paralogues in zebrafish.** *BMC Genomics* 2005, **6:**152.
15. Stet RJ, de VB, Mudde K, Hermsen T, van HJ, Shum BP, Grimholt U: **Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (Salmo salar) give rise to diverse class II genotypes.** *Immunogenetics* 2002, **54:**320-331.
16. Grimholt U, Getahun A, Hermsen T, Stet RJ: **The major histocompatibility class II alpha chain in salmonid fishes.** *Dev Comp Immunol* 2000, **24:**751-763.
17. Aoyagi K, Dijkstra JM, Xia C, Denda I, Ototake M, Hashimoto K, Nakanishi T: **Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (Oncorhynchus mykiss).** *J Immunol* 2002, **168:**260-273.
18. Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, Stet RJ, Secombes C, Parham P: **Modes of salmonid MHC class I and II evolution differ from the primate paradigm.** *J Immunol* 2001, **166:**3297-3308.
19. Consuegra S, Megens HJ, Leon K, Stet RJ, Jordan WC: **Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus.** *Immunogenetics* 2005, **57:**16-24.

20.   Kropshofer H, Hammerling GJ, Vogt AB: **The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules.** *Immunol Rev* 1999, **172:**267-278.
21.   Vogt AB, Kropshofer H, Moldenhauer G, Hammerling GJ: **Kinetic analysis of peptide loading onto HLA-DR molecules mediated by HLA-DM.** *Proc Natl Acad Sci U S A* 1996, **93:**9724-9729.
22.   Kropshofer H, Hammerling GJ, Vogt AB: **How HLA-DM edits the MHC class II peptide repertoire: survival of the fittest?** *Immunol Today* 1997, **18:**77-82.
23.   Kropshofer H, Vogt AB, Thery C, Armandola EA, Li BC, Moldenhauer G, Amigorena S, Hammerling GJ: **A role for HLA-DO as a co-chaperone of HLA-DM in peptide loading of MHC class II molecules.** *EMBO J* 1998, **17:**2971-2981.
24.   Hermel E, Yuan J, Monaco JJ: **Characterization of polymorphism within the H2-M MHC class II loci.** *Immunogenetics* 1995, **42:**136-142.
25.   Servenius B, Rask L, Peterson PA: **Class II genes of the human major histocompatibility complex. The DO beta gene is a divergent member of the class II beta gene family.** *J Biol Chem* 1987, **262:**8759-8766.
26.   McConnell TJ, Godwin UB, Norton SF, Nairn RS, Kazianis S, Morizot DC: **Identification and mapping of two divergent, unlinked major histocompatibility complex class II B genes in Xiphophorus fishes.** *Genetics* 1998, **149:**1921-1934.
27.   Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Hoyheim B: **A highly redundant BAC library of Atlantic salmon (Salmo salar): an important tool for salmon projects.** *BMC Genomics* 2005, **6:**50.
28.   Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ, de Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell LA, Mathewson C, Osoegawa K, Parisotto SE, Phillips RB, Rise ML, von Schalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N, Yang G, Zhu B: **A physical map of the genome of Atlantic salmon, Salmo salar.** *Genomics* 2005, **86:**396-404.
29.   **Salmon Specific Repeat Masker** *Internet* 2007 [http://grasp.mbb.sfu.ca/GRASPRepetitive.html].
30.   Martinez De AC, Perez JL, Bernal J, Coloma A: **Structure, organization, and chromosomal mapping of the human neurogranin gene (NRGN).** *Genomics* 1997, **41:**243-249.
31.   Jacinto E, Guo B, Arndt KT, Schmelzle T, Hall MN: **TIP41 interacts with TAP42 and negatively regulates the TOR signaling pathway.** *Mol Cell* 2001, **8:**1017-1026.
32.   Bartolini F, Tian G, Piehl M, Cassimeris L, Lewis SA, Cowan NJ: **Identification of a novel tubulin-destabilizing protein related to the chaperone cofactor E.** *J Cell Sci* 2005, **118:**1197-1207.
33.   Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004, **32:**W327-W331.
34.   Marchler-Bauer A, Anderson JB, Derbyshire MK, Weese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH: **CDD: a conserved domain database for interactive domain family analysis.** *Nucleic Acids Res* 2007, **35:**D237-D240.
35.   KOPPANG EO, LUNDIN M, PRESS CML, RONNINGEN K, LIE O: **Differing levels of Mhc class II [beta] chain expression in a range of tissues from vaccinated and non-vaccinated Atlantic salmon (Salmo salarL.).** *Fish & Shellfish Immunology* 1998, **8:**183-196.
36.   Speilberg L, Evensen O, Nafstad P: **Liver of juvenile Atlantic salmon, Salmo salar L.: a light, transmission, and scanning electron microscopic study, with special reference to the sinusoid.** *Anat Rec* 1994, **240:**291-307.
37.   Rombout JHWM, Berg AA: **Immunological importance of the second gut segment of carp. I. Uptake and processing of antigens by epithelial cells and macrophages.** *Journal of Fish Biology* 1989, **35:**13-22.
38.   Kasahara M: **The chromosomal duplication model of the major histocompatibility complex.** *Immunol Rev* 1999, **167:**17-32.
39.   Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De B V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra

G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest CH: **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431:**946-957.
40.   Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297:**1301-1310.
41.   Allendorf FW, Thorgaard GH: **Tetraploidy and the evolution of salmonid fishes.** In *Evolutionary Genetics of Fishes* Edited by: Turner BJ. New York, Plenum; 1984:1-53.
42.   Miller KM, Li S, Ming TJ, Kaukinen KH, Schulze AD: **The salmonid MHC class I: more ancient loci uncovered.** *Immunogenetics* 2006, **58:**571-589.
43.   Dijkstra JM, Katagiri T, Hosomichi K, Yanagiya K, Inoko H, Ototake M, Aoki T, Hashimoto K, Shiina T: **A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes.** *Immunogenetics* 2007, **59:**305-321.
44.   Katsanis N, Fitzgibbon J, Fisher EM: **Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci.** *Genomics* 1996, **35:**101-108.
45.   **Children's Hospital Oakland Research Institute (CHORI)** *Internet* 2007 [http://bacpac.chori.org/].
46.   Birnboim HC, Doly J: **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** *Nucleic Acids Res* 1979, **7:**1513-1523.
47.   **The Staden Package** *Internet* 2007 [http://staden.sourceforge.net/].
48.   Staden R, Beal KF, Bonfield JK: **The Staden package, 1998.** *Methods Mol Biol* 2000, **132:**115-130.
49.   Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8:**175-185.
50.   Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8:**186-194.
51.   **Phrap; a program for assembling shotgun DNA sequences** *Internet* 2007 [http://www.phrap.org/].
52.   Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8:**195-202.
53.   Gordon D, Desmarais C, Green P: **Automated finishing with autofinish.** *Genome Res* 2001, **11:**614-625.
54.   **Digit Web Server** *Internet* 2007 [http://digit.gsc.riken.go.jp/].
55.   Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268:**78-94.
56.   Sonnhammer EL, Durbin R: **A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis.** *Gene* 1995, **167:**GC1-10.
57.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
58.   Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
59.   **Consortium for Genomics Research on All Salmon** *Internet* 2007 [http://web.uvic.ca/cbr/grasp/].
60.   Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SR, Smailus DE, Jones SJ, Schein JE, Marra MA, Butterfield YS, Stott JM, Ng SH, Davidson WS, Koop BF: **Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics.** *Genome Res* 2004, **14:**478-490.
61.   **UniGene: An Organized View of the Transcriptome** *Internet* 2007 [http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene].
62.   Schuler GD: **Pieces of the puzzle: expressed sequence tags and the catalog of human genes.** *J Mol Med* 1997, **75:**694-698.

63. Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W: **A computer program for aligning a cDNA sequence with a genomic DNA sequence.** *Genome Res* 1998, **8:**967-974.
64. Wheelan SJ, Church DM, Ostell JM: **Spidey: a tool for mRNA-to-genomic alignments.** *Genome Res* 2001, **11:**1952-1957.
65. **Repeatmasker** *Internet* 2007 [http://www.repeatmasker.org/].
66. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25:**4876-4882.
67. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5:**150-163.
68. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25:**402-408.