

RESEARCH ARTICLE

Open Access



Whole genome duplications have provided teleosts with many roads to peptide loaded MHC class I molecules

Unni Grimholt

Abstract

Background: In sharks, chickens, rats, frogs, medaka and zebrafish there is haplotypic variation in MHC class I and closely linked genes involved in antigen processing, peptide translocation and peptide loading. At least in chicken, such MHCIIa haplotypes of MHCIIa, TAP2 and Tapasin are shown to influence the repertoire of pathogen epitopes being presented to CD8+ T-cells with subsequent effect on cell-mediated immune responses.

Results: Examining MHCII haplotype variation in Atlantic salmon using transcriptome and genome resources we found little evidence for polymorphism in antigen processing genes closely linked to the classical MHCIIa genes. Looking at other genes involved in MHCII assembly and antigen processing we found retention of functional gene duplicates originating from the second vertebrate genome duplication event providing cyprinids, salmonids, and neoteleosts with the potential of several different peptide-loading complexes. One of these gene duplications has also been retained in the tetrapod lineage with orthologs in frogs, birds and opossum.

Conclusion: We postulate that the unique salmonid whole genome duplication (SGD) is responsible for eliminating haplotypic content in the paralog MHCIIa regions possibly due to frequent recombination and reorganization events at early stages after the SGD. In return, multiple rounds of whole genome duplications has provided Atlantic salmon, other teleosts and even lower vertebrates with alternative peptide loading complexes. How this affects antigen presentation remains to be established.

Keywords: MHC, Antigen presentation, Peptide loading complex, Phylogeny, Teleosts, Salmonids

Background

Currently our view of antigen processing, peptide loading and peptide presentation in teleosts mainly relies on extrapolation of functional data generated in humans, a few other mammals and chickens. In these species Major histocompatibility complex class I (MHCII) molecules are key players in discriminating self from non-self. Classical MHCII (MHCIIa) molecules, consisting of an alpha chain and a non-covalently associated beta2-microglobulin (b2m) chain, are displayed at the surface of most cells where they normally present peptides derived from self-molecules chopped into smaller fragments in the cytoplasm. If foreign elements such as viruses are present in the cytoplasm, they are prone for degradation and presentation by MHCIIa molecules. After

peptide loading, the MHCIIa molecules are transported to the cell surface for recognition by CD8+ T cells, thereby initiating an immune response when the peptide originates from a non-self protein.

Newly synthesized MHCIIa molecules are processed in the endoplasmic reticulum (ER) to become mature properly folded molecules [reviewed in [1]]. This transition to maturity is aided by the molecular chaperones calnexin (CANX), calreticulin (CALR) and heat shock protein family A member 5 (HSPA5 alias BiP) that assist in correct folding and assembly of the MHCIIa molecule with the beta2-microglobulin chain. Subsequent association with ERp57 (protein disulphide isomerase family A member 3 alias PDIA3), an enzyme that catalyzes disulfide bond formation, finalizes maturation of the MHCIIa molecule. To form the peptide-loading complex, the CALR and ERp57-stabilized MHCIIa molecule then associates with tapasin (TAPBP) and the transporters

Correspondence: Unni.Grimholt@vetinst.no
Fish Research Group, Norwegian Veterinary Institute, Ullevaalsveien 68, 0454 Oslo, Norway

associated with antigen processing (TAP1- TAP2) heterodimer [2].

Peptide loading of MHCIIa molecules is a multi-step process starting with the degradation of proteins in cytosol. Proteins tagged for destruction by ubiquitin are generally degraded by a proteasome consisting of an inner core with seven alpha and seven beta subunits associated with one or two regulatory particles [3]. Upon stimulation such as an infection, a MHCIIa specific proteasome defined as the immune-proteasome is induced by interferon gamma where three of the seven proteasome beta-components PSMB8, PSMB9 and PSMB10 replace the constitutive components PSMB5, PSMB6 and PSMB7. This change in components produces peptides preferably with hydrophobic or positively charged residues at the C terminus, which are optimally suited for binding to the antigen transporter TAP and to MHC Ia molecules [4]. Two additional interferon-induced subunits called PSME1 and PSME2 provide an added regulatory element to the core immunoproteasome but how they influence the MHCII peptide repertoire is debated [5]. It should be noted that the interferon-inducible immunoproteasome components have not been identified in birds [6].

The protein degradation products generated in cytosol are then transported into the ER through an MHCIIa specific transporter consisting of the molecules TAP1 and TAP2. This heterodimer preferentially translocates peptides with a length of 8–12 residues into the ER in an ATP-dependent process [7, 8]. Longer peptides may be transported, but then in a kinked conformation as TAP has a length restriction for the peptide N and C-terminal distance [9]. The three N-terminal and the last C-terminal residues of the peptide are critical for binding to TAP [10–12].

In the ER lumen, the MHCIIa-b2m-CALR-ERp57 complex associates with tapasin (TAPBP) and transporter associated with antigen processing TAP1/TAP2 into what is defined as the peptide-loading complex (PLC) [13]. TAPBP acts as a bridge between TAP and MHCIIa enhancing TAP stability and peptide translocation. TAPBP also stabilizes empty MHCIIa molecules and optimizes MHCIIa loading with peptides. MHCIIa alleles have a variable TAPBP dependency where some alleles do not require TAPBP for loading with high affinity peptides [14]. TAPBP is also linked to the protein disulphide isomerase ERp57, contributing to structural stability of the PLC [15]. To finalize the functional PLC complex, ERp57 is linked to calreticulin, while calreticulin holds on to the MHCIIa glycan located at the C-terminal end of the alpha 1 domain [16].

As the length of peptides translocated into the ER varies, some need N-terminal trimming to fit properly into the MHCIIa groove as the C-terminus is already

MHCIIa compatible. ERAP1/ ERAP2 can trim both free peptides as well as MHCIIa-bound peptides [17]. Optimal MHCIIa peptide binding may also be influenced by the tapasin-related molecule (TAPBP) although this molecule is not a member of the PLC [18, 19]. Similar to TAPBP, MHCIIa alleles vary in their association with TAPBP [19]. Recently, Neerinx et al. [20] showed that TAPBP associates with UDP-glucose:glycoprotein glucosyltransferase 1 (UGT1), a folding sensor in the calnexin/calreticulin quality control cycle that is known to regenerate the Glc1Man9GlcNAc2 moiety on glycoproteins. Thus, TAPBP could serve a dual role also routing empty or low-affinity bound MHCIIa molecules back to the PLC for refolding.

In Chicken, there is a classical dominantly expressed and highly polymorphic MHCIIa (BF2) gene residing within the major MHC region. This gene is flanked by TAP1 and TAP2 genes while a single TAPBP gene is located less than 40 kb away [21–23]. The chicken tapasin and TAP genes are polymorphic with each unique BF2, TAP1, TAP2 and TAPBP combination segregating as a stable functional haplotype. TAPBP and TAP polymorphism govern the peptides available for binding to the BF2 allele [24, 25]. Thus BF2 alleles in heterozygous animals can in theory bind peptides pumped by both haplotypes as opposed to alleles in homozygous animals, thus broadening the MHCIIa peptide repertoire. The importance of peptide binding can be exemplified by Mareks disease in chicken, where MHCIIa alleles binding a large variety of peptides induce protection while those binding a restricted number of peptides are linked to susceptibility. Presumably, presenting a larger repertoire of pathogen epitopes activate a wider range of T cell clones, a response needed for protection against this pathogen. To counteract against inducing autoimmune reactions, the MHCIIa molecules with a wide peptide binding repertoire are only present in low copy numbers on the cell surface while MHCIIa molecules with a narrow peptide repertoire are much higher expressed on the cell surface [24, 25]. A similar picture was described in humans, where also here lower surface expression levels corresponded with a broader peptide binding ability. Although TAPBP and TAP are not polymorphic in humans, allelic variation in how the molecules associate with TAPBP was suggested to influence the peptide binding repertoire.

Polymorphism in genes closely linked to MHCIIa is also reported in other tetrapods. In rats, two allelic variants of TAP2 linked to specific subsets of MHCIIa were shown to deliver different spectrum of peptides [26, 27]. In frogs, linked and highly divergent biallelic PSMB8, TAP1, TAP2 and MHCIIa sequence variants of ancient origin have been described [28, 29]. Unfortunately the

functional difference between these frog gene variants has not yet been investigated.

Similar to frogs both the PSMB8, TAP2 and tapasin molecules are encoded within the major MHC class I regions of teleosts [30]. This region also contains a teleosts specific duplication of the PSMB9 gene denoted PSMB12 and a gene duplicate of the human PSMB10 gene denoted PSMB13. The teleost PSMB10 gene is for some species located within the extended MHCIa region while for others translocated elsewhere [30, 31]. Polymorphism in the proteasome component PSMB8 has been reported [32, 33] where two distinct lineages of PSMB8 denoted PSMB8A and PSMB8F were found in sharks, cyprinids and salmonids [34]. The two variants existed as duplicate genes in shark, but were defined as alleles in cyprinids and salmonids by Tsukamoto et al. [34], while McConnell et al. defined them as paralogs [31]. When comparing gene sequences from three medaka MHCIa haplotypes, sequence polymorphism was observed both within the PSMB8A gene as well as in the PSMB10 gene [33]. Recently haplotypic variation was also reported in zebrafish where haplotypes contained a varying number of MHCIa genes were linked to polymorphic TAP2, tapasin and PSMB8 molecules [31]. Thus, functional haplotypes linked to the classical MHCIa genes may be a common trait in teleosts.

Atlantic salmon is a species with a single classical MHCIa gene [35] similar to chicken, but with a duplicated MHCI region due to the unique salmonid whole genome duplication (SGD) event that occurred 94 million years ago [30, 36]. The MHCIa region on chromosome 27 harbors the single classical MHCIa locus denoted UBA, while the paralog MHCIb region on chromosome 14 has one expressed non-classical MHCI gene denoted UDA and one or two pseudogenes denoted UCA [37, 38]. The MHCI genes in both regions are flanked by tapasin, the proteasome components PSMB8, PSMB9, PSMB12, PSMB13 and TAP2 alongside several other genes also present in the MHCI region of other teleosts. Similar to other teleosts, the TAP1 gene is located outside the major MHCI region. Based on the reported MHCIa haplotypes in sharks, rats, chicken,

frogs, zebrafish and medaka, we set out to investigate the peptide loading machinery and functional MHCIa haplotypes in salmonids compared to other teleosts.

Results

The fact that sequence variation has been found in genes closely linked to MHCI that influence peptide processing, transport and loading in a diverse range of species including the teleosts medaka and zebrafish (See Table 1) could indicate that this is a general trend. To investigate functional MHCIa-linked polymorphism in salmonids we looked at available genomes from Atlantic salmon [39], rainbow trout ([40] and unpublished assembly) and coho salmon (unpublished assembly) as well as Northern pike [41], a species basal to the salmonids that has not experienced the salmonid whole genome duplication event. For Atlantic salmon and trout there are also some BAC sequences covering the paralog MHCI regions [37, 42]. See Fig. 1 for phylogenetic relationships between the various teleost species.

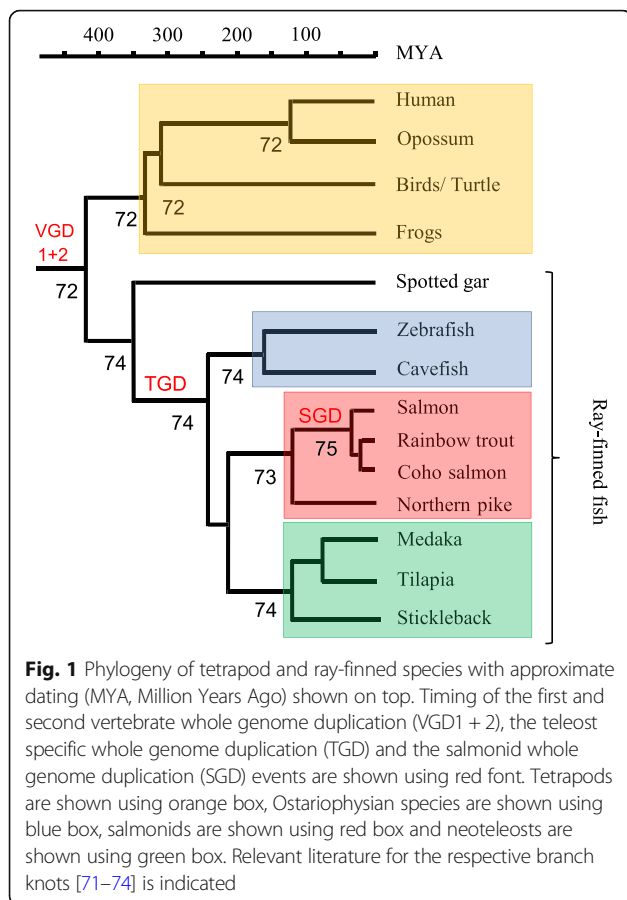
PSMB8F linked haplotypes

Based on data from medaka [33, 43, 44] and zebrafish [31] (summarized in Additional file 1; Fig. S1), a good marker for functional teleost MHCIa haplotypes is the PSMB8F gene variant. The term haplotype here defines a group of closely linked gene variants (alleles) residing within one chromosomal region that are inherited together from a single parent. Thus to investigate polymorphism in genes involved in antigen processing in salmonids, we first searched salmonid genome resources for presence of the PSMB8a and PSMB8F variants reported to exist as allelic variants in salmonids [34]. In the following we use -a and -b extensions for paralog genes originating from the salmonid whole genome duplication. In the case of MHCI, Ia and -a extensions refer to genes originating from the salmonid classical UBA regions and Ib to genes in the paralog region containing non-classical MHCI genes respectively. Each chromosomal region containing MHCI and physically linked genes is in the following shown as e.g. Atlantic salmon Ia_#A referring to a specific collection of gene

Table 1 Polymorphism in MHC class I linked genes

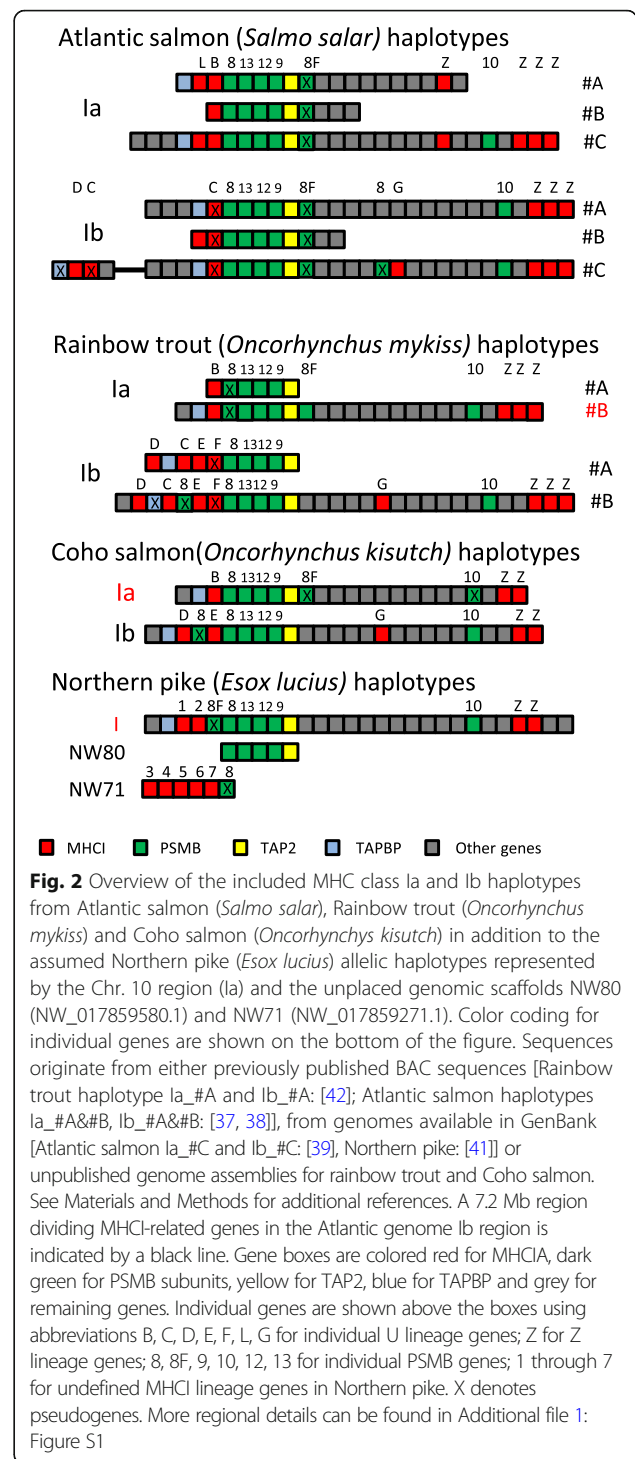
Species/ Gene	MHCI	PSMB8	PSMB9	PSMB10	PSMB12	PSMB13	TAP2	TAPBP
Human								
Rat								
Chicken								
Frog								
Medaka								
Zebrafish								
Shark								

Orange cells indicate reported gene polymorphism in MHCI-linked genes from rats [26, 27], chicken [22, 23], frogs [29, 78], medaka and sharks [34] and zebrafish [31]. Grey cells indicate lack of immunoproteasome components [6]. White cells indicate no reported polymorphisms

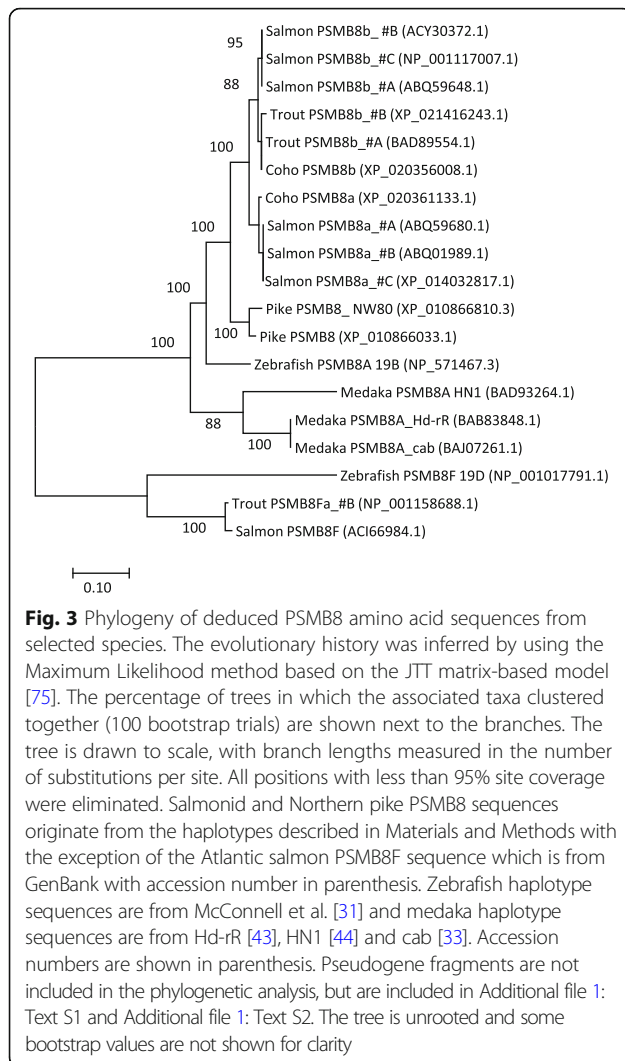


variants in the Ia region while Ia_#B refers to another collection of gene variants in this same region.

None of the previously published MHCla or MHClb regions from Atlantic salmon [37, 38] nor those in rainbow trout [42], determined from sequencing of BAC clones, contained a bona fide PSMB8F gene (Figs. 2 and 3, Additional file 1: Figure S1, Additional file 1: Text S1). Neither was this presumed allelic variant present in the paralog MHCi regions of the Atlantic salmon genome originating from a double haploid [39]. However, in both assemblies of the rainbow trout genome [[40] & new unpublished GenBank assembly GCA_002163495.1) we found the PSMB8F gene located in between the TAP2 and the BRD2 loci in the Ia region (Fig. 2, trout haplotype Ia_#B). Although automatically annotated as two separate open reading frames in the most recent genome assembly (70,115+ CDQ70116) this could well constitute a functional locus in trout as supported by the GenBank transcriptome shotgun assembly sequences GBTD01219587.1 and EZ768376.1. This Ia_#B haplotype also contained a PSMB8 pseudogene in between the UBA and the PSMB13a genes, similar to the location of the PSMB8 gene in the majority of other analysed teleosts [30]. The pseudo-nature of this locus is supported by a PSMB8 pseudogene in this position also in the previously



published trout Ia_#A haplotype [42]. Looking back at the published trout MHCi BAC sequences, both these Ia_#A and Ib_#A haplotypes terminated immediately following the TAP2 locus so the PSMB8F gene could potentially also have been present in these haplotypes. Based on the pseudogene nature of the PSMB8 gene in both the trout Ia_#A and Ia_#B haplotypes, perhaps this is compensated



for by a functional PSMB8F gene. For the coho (*Oncorhynchus kisutch*) genome, a PSMB8F gene is also located in between the TAP2 and BRD2 genes of the Ia region in this species (Fig. 2, Additional file 1: Figure S1, Additional file 1. Text S2), but based on the genome assembly this is a pseudogene.

As opposed to rainbow trout, there seems to be a bona fide coho PSMB8 gene in the expected position in between the UBA and PSMB13a loci. Thus, the trout and coho data do not support the proposed allelic nature of the PSMB8A and PSMB8F sequence variants suggested by Tsukamoto et al. [34], but rather suggests that they are two different loci residing at different locations within the salmonid MHCIIa region. This is further supported by small PSMB8F pseudogene fragments in between the TAP2 and the BRD2 gene in all Atlantic salmon Ia and Ib haplotypes (Fig. 2, Additional file 1: Figure S1, Additional file 1: Text S1, Additional file 1: Text S2), supporting the location of this gene in this

position in a predecessor to Atlantic salmon, coho salmon and rainbow trout. There are also PSMB8 fragments elsewhere in salmonid genomes such as in between the trout Ib UCA and UEA genes and the coho Ib region UEA and UDA genes suggesting this gene has been shuffled around during salmonid evolution. Northern pike, a diploid species basal to the salmonids [41], does not add any clarity to the evolution of these genes, as it contains a PSMB8F pseudogene in between a seemingly functional PSMB8 gene and two duplicate MHCII genes. We did however find expressed PSMB8F match for both Northern pike (GATF01015276.1) as well as Atlantic salmon (ACI66984.1) (Additional file 1: Text S2), suggesting this gene may be functional in some haplotypes.

In zebrafish the PSMB8F gene is located in a unique haplotype on chromosome 19 denoted 19D alongside a single classical MHCIIa gene denoted UGA and two highly divergent TAP2 genes [31]. The two other haplotypes A and B described by McConnell et al. [31] contain the PSMB8A variant only in a location different from the PSMB8F gene in the 19D haplotype (Additional file 1; Figure S1). It should be noted that this zebrafish UGA gene is not identical to the non-classical gene denoted UGA in salmonids. Even if the PSMB8F gene has been silenced in most trout, coho and Northern pike haplotypes, there should still be remnants of other gene polymorphisms in these species. We thus contrasted closely linked gene sequence variants in PSMB8F positive regions against those in Atlantic salmon regions without the PSMB8F variant (Fig. 2). Based on data from zebrafish and medaka suggesting that TAPBP, TAP2 and PSMB variants have evolved to serve specific MHCIIa alleles [31, 34], we first investigated the phylogenetic relationship amongst the MHCIIa gene sequences. As teleosts have alpha 1 domain lineages that cluster into lineages shared between distantly related species [30, 33, 45] this analysis was performed based on individual alpha 1–3 domains.

Alpha I domain sequence of the rainbow trout UBA locus, the second pike UBA locus and the zebrafish UGA locus all cluster together alongside other alpha 1 domain sequences from lineage V (Additional file 1: Figure S2) [30, 45]. The two other PSMB8F containing regions did not comply with linkage of the MHCIIa alpha 1 domain region and the PSMB8F gene sequence, as the coho a1 domain sequence clusters alongside lineage I sequences and the pike UBA1 sequence clusters alongside lineage III sequences. Comparing pike with the Atlantic salmon MHCIIa region could suggest that the pike UBA1 locus is an ortholog of the Atlantic salmon ULA locus while the pike UBA2 locus is the ortholog of the salmon UBA locus, although the pike UBA1 locus still contains a transmembrane region as opposed to the

salmon ULA gene. If so, the only deviation from a functional haplotype including MHCIIa gene variants and the PSMB8F gene is coho based on the alpha 1 domain analysis. However, the sequenced Northern pike animal was not a double haploid as seen in the two unassembled scaffolds containing five additional U lineage loci and a duplicate PSMB-TAP2 region (Additional file 1: Text S1, scaffolds NW_017859271.1 and NW_017859580.1) which most likely represents an allelic variant. There is currently no data available enabling definition of these pike genes as classical or non-classical referring to polymorphic content, peptide binding ability and tissue expression distribution.

We know that the alpha 1 domain in salmonids fluctuate between different alpha 2 and downstream domains [30, 45–48] so potentially there has been a translocation or recombination of the coho alpha 1 domain, losing the alpha 1 domain V lineage in this haplotype. This is consistent with the clustering of the alpha 2 domain of coho and rainbow trout alpha 2 domains with other lineage I alpha 2 domains (Additional file 1: Figure S3). But for the alpha 2 domain zebrafish UGA clusters with alpha 2 domain lineage II sequences with a convincing bootstrap value and the pike sequences form a separate clade unlinked to the other PSMB8F-linked haplotype sequences. For the alpha 3 domain, the sequences cluster in a species specific manner providing no further clues as to functional haplotypes. Thus, the MHCII gene sequence analyses does not convincingly support functional haplotypes represented by the PSMB8F gene sequence in salmonids.

Zebrafish did not contain any polymorphism in the PSMB12 gene sequences, but some sequence variation in the PSMB9 gene [31]. For salmonids, the salmonid PSMB9 and PSMB12 genes had more species specific sequence variation than haplotypic variation and none of the variable sites in the PSMB9 sequences matched the sequence variation seen in zebrafish (Additional file 1: Figure S3).

Of the phylogenetically related zebrafish PSMB7, PSMB10 and PSMB13 genes, only the PSMB13 gene displayed haplotypic sequence variation. Zebrafish PSMB7 and PSMB10 genes were unlinked to any MHC genes residing on chromosome 21 and 4 respectively [31]. In salmonids both the PSMB13 and the PSMB10 genes reside within both the paralog MHCII region where the PSMB10 gene is located closer to the MHCII Z lineage genes than to the U lineage loci (Fig. 2 and Additional file 1: Figure S4). The Atlantic salmon PSMB7 gene resides on linkage group 1 with a pseudogene copy on linkage group 11 (Additional file 1: Text S1). As opposed to zebrafish, there is no indication of any haplotypic variation in any of these PSMB8F linked genes in any salmonid species. As an example the rainbow trout Ia_#A

haplotype contains a UBA*0501 allele while the trout Ia_#B haplotype contains a new yet unnamed allele, but their PSMB13 sequences are identical. Ditto for Atlantic salmon where the PSMB13 sequences are identical in the Ia_#A and Ia_#C haplotypes containing UBA*0201 and UBA*0301 alleles respectively.

In zebrafish, the TAP2 alleles linked to the B and D haplotypes are highly divergent [31]. In salmonids the most prominent sequence variation is seen in the Atlantic salmon TAP2a_#C sequence where there are 11 amino acid differences within the first 31 amino acids in the N-terminal region when compared to TAP2 variants of the other two MHCIIa haplotypes (Fig. 4 and Additional file 1: Text S3). This TAP2 sequence resembles both those in the rainbow trout Ia region as well as those in the paralog Ib region, but expressed support (transcriptome shotgun assembly sequence GBRB01034973.1) suggests it is a true sequence and not a genome assembly artefact. Northern pike on the other hand, does display considerable variation in the TAP2 gene sequences located in the main Ia region and the NW80 scaffold (Figs. 2 and 4). Sequence identity between the two pike TAP2 gene sequence variants is 87% which is intermediate in comparison to the 98%, 96%, 71% and 61% amino acid sequence identities

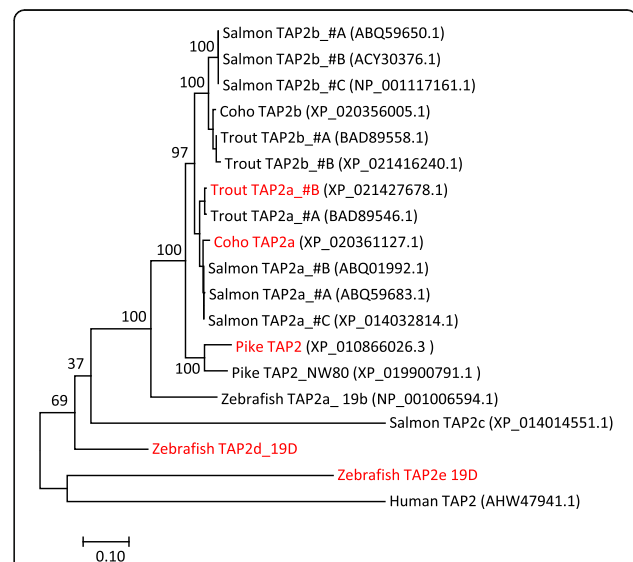


Fig. 4 Phylogeny of deduced TAP2 amino acid sequences from selected species. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le and Gascuel 2008 model [76]. The percentage of trees in which the associated taxa clustered together (100 bootstrap trials) are shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Positions with less than 95% site coverage were eliminated. For sequences not originating from the selected haplotypes accession numbers are shown in parenthesis. Gene sequences linked to haplotypes containing the PSMB8F gene variant (Fig. 2) are shown using red font. Sequence references not shown in figure can be found in [31]. The tree is unrooted and some bootstrap values are not shown for clarity

observed between the chicken, rat, frog and zebrafish TAP2 sequence variants respectively. Nine residue positions coincide with polymorphic residues in rat or chicken and many variable positions are located within and surrounding the first transmembrane domain known to interact with tapasin [49].

For the salmonid tapasin (TAPBP) gene, there is less sequence information available, as many of the sequenced BACs did not cover this locus (Fig. 2). There are a few amino acid differences between salmon, trout and coho, but no apparent haplotypic variation. Atlantic salmon also has a third TAPBP gene linked to a duplicate UDA locus 7 Mb upstream from the major Ib region, but this TAPBP gene seems to be a pseudogene (Additional file 1: Figure S5).

Based on the above analyses of salmonid MHCI haplotype gene sequences, we did not find evidence for polymorphism in antigen processing genes with the exception of a TAP2 sequence variant in Atlantic salmon. This Atlantic salmon TAP2a_#C sequence differs in many of the first 31 amino acid residues of the N-terminal region a variation that could influence binding affinity for TAPBP [49], but does not match the sequence variation seen in zebrafish, rat or chicken TAP2 sequences (Additional file 1: Text S3). Lack of polymorphism in the selected genes is supported by current and previous sequence analyses using other GenBank resources. With Northern pike displaying a TAP2 polymorphism matching many of the polymorphic residue positions found in other species, the lack of functional haplotypes in salmonids may be unique.

What about other genes influencing MHCI assembly and peptide loading?

A paralog MHCI region with functional gene copies of PSMB8, PSMB9, PSMB12, PSMB13 and TAP2 (Table 2) provides salmonids with an added complexity in comparison with zebrafish and medaka. This complexity of having many duplicate immune genes which potentially have acquired slightly new functions as suggested by Lien et al. [39] could also relate to other genes involved in antigen degradation, transport and loading. This spurred us to investigate a broader set of genes influencing antigen processing, transport and loading. Calnexin, which binds newly synthesized MHCI prior to assembly with b2m, is encoded by two duplicate loci in Atlantic salmon (Fig. 5, Additional file 1: Figure S6). Both genes display fair expression in immunologically important tissues such as head kidney and spleen as well as other organs such as brain and ovary (Table 2). This gene duplication has previously been identified in trout [50] and is a remnant of the SGD displaying high sequence identity between paralogs. If one or both copies serve the classical MHCIa molecules remains to be established.

The calnexin MHCI heavy chain complex then associates with beta2-microglobulin (b2m) (Fig. 5). This b2m gene is encoded by 3–10 loci in trout [51, 52] while the Atlantic salmon genome contains 13 loci where at least one seems to be a pseudogene (Additional file 1: Text S1). These 12–13 Atlantic salmon gene sequences cluster into two distinct clades with 82% amino acid identity (data not shown). Due to high sequence identity within each clade, it is difficult to assess the expressed status of each Atlantic salmon locus (Table 2). Many other teleosts also have two clades of b2m gene sequences [53], but here the amino acid sequence identity between clades is mostly lower than between the salmonid clades (Additional file 1: Figure S7). The gene duplications have occurred individually in many species with for instance unique gene duplications in Ostariophysi represented by zebrafish and Mexican tetra and another unique duplication in neoteleosts represented by tilapia, medaka and stickleback. Which molecules these divergent teleost b2m molecules support is unknown, but at least in medaka and stickleback the two b2m sequence variants can only serve the U and Z lineages as the remaining MHCI lineages S, L and P are not present in these species [30]. A more speculative idea would be that both b2m sequence groups bind to the classical MHCIa molecule and as such influence the peptide repertoire as previously reported for mouse MHCI molecules [54].

Once the b2m molecule is associated with the MHCI alpha chain, calnexin is replaced by calreticulin (Fig. 5). In Atlantic salmon there are three CALR genes, but also three calreticulin-like (CALRL) genes (Additional file 1: Figure S6) with amino acid sequence identity between the two groups ranging from 67 to 70% (data not shown). Phylogenetically, the gene sequences here denoted Atlantic salmon CALR1a, CALR.1b, CALR1.2 cluster with the human CALR gene sequence and also the previously published catfish sequence denoted CALRL2 [55]. The remaining three Atlantic salmon gene sequences form a separate cluster alongside the two other catfish sequences denoted catfish CALRL and CALR, clustering with a 99% bootstrap value to the other CALR sequences. Sequences belonging to the CALR and CALRL clades also exist in spotted gar, a species that split off from the teleost lineage prior to the teleost whole genome duplication (TGD) [56] (Fig. 1, Additional file 1: Figure S6), suggesting these genes are remnants of the second vertebrate genome duplication (VGD2). The genes represented by the CALRL1 and CALRL2 clades originated before Ostariophysi branched off from the remaining teleosts having orthologs also in catfish. They are thus a product of the TGD event that occurred approximately 350 million years ago after spotted gar split off from the main teleost lineage [56]. Based on sequence identity, phylogenetic clustering and

Table 2 Expression levels of Atlantic salmon genes in various tissues

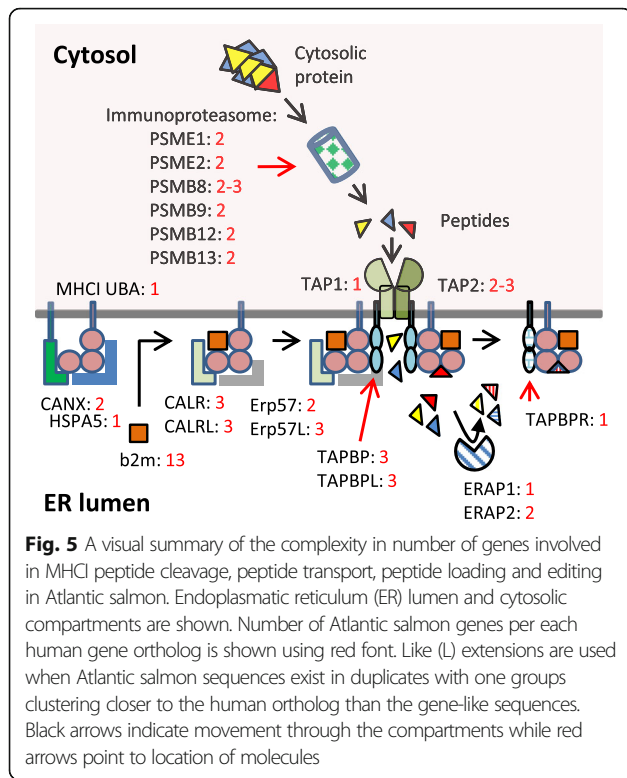
Gene	HK	Gills	Spleen	Heart	Gut	Muscle	Brain	Liver	Nose	Skin	Ovary	Testis	Length bp
UBA*0301	61,85	126,09	146,46	8,87	198,82	10,67	10,09	6,64	32,64	3,58	0,68	38,91	1068
B2m3	241,00	569,57	479,95	56,97	639,00	64,70	44,04	45,12	190,62	29,28	179,04	398,30	351
B2m11	543,57	580,91	832,41	70,09	571,84	107,32	37,12	78,03	266,89	29,99	7,24	256,98	351
PSMB8a	2,00	1,44	1,82	0,02	1,71	0,18	0	0,02	1,15	1,79	0,03	71,90	804
PSMB8b	17,14	13,85	15,00	1,28	15,49	1,66	0,17	1,08	7,58	1,05	0,03	1,89	804
PSMB9a	17,49	15,11	15,82	3,29	12,78	3,16	0,86	2,78	12,07	3,41	0,09	0	654
PSMB9b	8,02	9,05	9,24	0,21	8,21	1,12	0,57	0,16	5,37	0,20	0,17	5,85	654
PSMB12a/b	22,41	20,48	18,08	0,92	19,71	2,34	0,70	0,49	15,23	2,79	0,04	2,27	654
PSMB13a/b	10,78	8,26	9,26	0,33	13,73	0,96	0,31	0,12	3,83	0,96	0	3,10	834
TAPBPpa	19,33	16,51	19,42	3,72	11,64	3,04	2,40	2,91	7,28	1,78	0,19	8,50	1188
TAPBPpb	1,82	2,63	1,93	0,27	0,95	0,43	0,19	0,49	1,11	0,60	0,15	1,79	1314
TAPBPL1a	4,29	3,99	4,02	1,28	2,23	0,74	0,38	1,20	1,63	0,15	0,32	6,01	1302
TAPBPL1b	11,05	9,96	9,44	2,23	10,09	2,40	1,66	2,24	5,10	1,24	0,45	4,23	1308
TAPBPL2	4,30	6,63	3,22	0,33	1,65	0,25	0,45	0,39	1,65	0	0,41	1,13	1233
TAPBPR	11,76	12,92	16,86	1,42	8,26	1,07	0,92	0,93	4,45	1,04	0,56	2,21	1305
CANX1a	49,66	41,56	33,33	19,81	32,56	6,83	31,13	16,27	20,01	4,10	45,86	23,93	1869
CANX1b	26,84	45,90	22,80	26,67	34,29	6,78	39,16	22,50	19,56	1,00	36,13	17,87	1806
CALR1a	0,27	0,11	0,03	0,03	0,11	0,13	26,25	0	0,98	0	0,44	9,24	1266
CALR1b	2,01	1,72	1,85	0,73	0,83	0,69	25,33	0,24	3,68	1,66	1,47	5,33	1272
CALR1.2	7,60	6,03	8,09	11,92	9,98	2,38	10,38	16,60	6,40	0,96	7,86	3,26	1251
CALRL1	90,52	59,46	78,99	46,96	107,53	19,86	35,14	44,47	43,42	12,19	161,40	52,69	1242
CALRL2a	29,51	28,11	27,36	44,78	40,98	9,29	11,87	43,77	27,14	2,52	14,18	8,36	1341
CALRL2b	61,91	56,49	69,51	150,7	115,6	14,67	20,38	189,45	45,01	8,73	19,04	22,47	1257
ERP57a	75,11	57,87	68,47	64,06	106,6	24,76	28,67	70,15	60,39	19,46	48,66	54,01	1482
ERP57b	31,61	24,51	23,76	25,75	67,19	6,95	5,23	31,96	26,31	4,94	24,37	17,47	1482
ERP57L1	2,33	2,57	2,38	1,60	3,48	2,81	3,03	0,17	3,08	1,11	6,77	4,35	1482
ERP57L2a	0,14	0,36	1,36	0,07	0,12	0,05	5,19	0,07	0,96	0	1,20	5,04	1485
ERP57L2b	3,83	1,88	3,11	1,35	2,61	1,17	8,96	0,55	1,49	0,10	0,75	4,98	1485
ERAP1	9,98	9,74	6,87	1,88	12,56	1,54	1,02	1,41	3,97	0,36	1,67	2,35	2874
ERAP2a	2,69	3,02	2,48	0,40	2,38	0,75	0,12	0,28	1,26	0,55	0,31	0,50	2829
ERAP2b	7,84	4,25	6,57	2,92	6,51	0,37	0,13	3,23	2,27	0,59	0,03	0,38	2835
PSME1a	51,88	34,54	57,89	17,40	76,45	8,38	4,82	17,65	26,87	8,24	7,97	18,11	750
PSME1b	55,19	37,30	52,94	9,11	30,20	11,31	16,69	9,23	21,16	12,77	16,13	22,93	1211
PSME2a	60,61	44,08	53,22	12,50	63,67	13,98	14,00	8,67	42,65	8,99	12,96	39,77	1206
PSME2b	67,46	34,63	46,10	10,68	51,22	6,61	5,32	7,52	22,59	5,29	2,47	11,70	1135

Read mapping was performed with the stringency of 98% identity and 95% coverage. Tissues with RPKM values below 1 are shaded grey, those with RPKM values between 10 and 50 are shaded blue, those with RPKM values above 50 are shaded pink, and those with levels above 100 are shaded yellow. Nucleotide sequence lengths used in the analysis is indicated in base pairs where only open reading frame sequences were included. Gene names with a and b extensions indicate SGD paralog genes. Sequence identity between PSMB12a/b and PSMB13a/b paralogs was too high to separate between duplicates. The b2m3 and b2m11 sequences represent each of the two sequence groups as sequence identity within group was too high to provide reliable data

chromosomal location, the CALR1a and CALR1b genes as well as the CALRL2a and CALRL2b genes originate from the SGD duplication event [39]. Human and chicken CALR3 sequences form an outgroup and have no detectable orthologs in Atlantic salmon.

Both the CALR as well as the CALR-like molecules display fair sequence identity as well as conservation of residues known to influence human CALR glycan binding suggesting both groups bind glycan moieties (Additional file 1: Figure S6). Both teleost CALR and CALRL sequence groups also have an ER retention signal (KDEL) and an acidic C-terminal region known to affect ER-retention and recycling [57], further supporting their

functional resemblance of human CALR. However, the CALRL genes are generally expressed approximately 10 times higher than the CALR genes (Table 2). And CALRL expression levels are not restricted to immunologically important tissues as seen with high expression levels also in heart, liver and ovary for some of these genes. In catfish, they found both CALR and CALRL genes to be induced upon infection [55]. In rainbow trout, calreticulin was reported as a single copy gene [58], but this sequence is an orthologue of the Atlantic salmon gene here denoted CALRL2b. A later functional study of this trout CALRL2b orthologue reported limited response to endoplasmic reticulum stress and also little response to stimulation

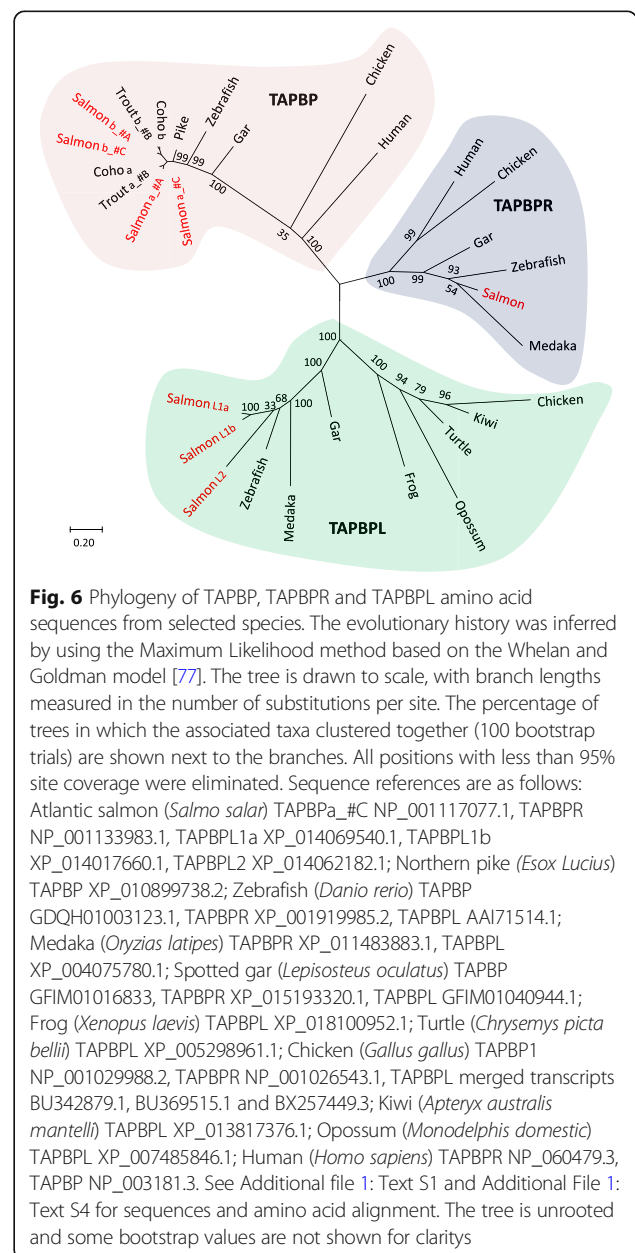


with the viral mimicry stimulant polyI:C [59]. Future studies are needed to clarify the functional distinction of teleost CALR and CALRL genes.

The next molecule to associate with the MHC I/b2m/CALR complex is ERp57 alias PDIA3 (Fig. 5). As seen for the CALR and CALRL genes, a similar picture emerges for ERp57. Here there are two Atlantic salmon ERp57 genes, but also three additional ERp57-like (ERp57L) genes (Additional file 1: Figure S8). The Atlantic salmon ERp57a/b genes cluster with human and zebrafish ERp57 sequences while the ERp57L sequences form a separate clade. Amino acid sequence identity between the Atlantic salmon ERp57 and ERp57L groups is approximately 60% (data not shown). Spotted gar has both an ERp57 sequence as well as an ERp57L sequence, suggesting this duplication originates from the second VGD2 event (Fig. 1) [56]. Both the ERp57L1 sequence and the ERp57L2a/-L2b sequences have orthologs in zebrafish and is most likely a remnant of the TGD event where also here the duplicates have been retained as expressed copies in Atlantic salmon (Table 2). Both Atlantic salmon ERp57a and -b genes and the two ERp57L2a and -L2b genes originate from the unique SGD event with amino acid sequence identities of 95 and 92% respectively. An ERp57 ortholog to the gene sequence here defined as Atlantic salmon ERp57a has been described both in rainbow trout [60] as well as in seabass [61]. This trout gene was induced upon stimulation by a viral mimicry molecule, but also by ER stressors.

How the ERp57L genes would respond to similar stimulations remains unknown but the ERp57a/b genes display a much higher expression level in most tissues compared to the ERp57L genes (Table 2).

Following ERp57 association, the complex then associates with Tapasin (Fig. 5). The TAPBP and TAPBPR gene sequences have previously been described in trout [62] where both genes were shown to respond to viral infection. In addition to the tapasin sequences linked to the MHC I a and I b regions in Atlantic salmon and a single TAPBPR gene, there are also three additional gene sequences with blast match to TAPBPR and TAPBP here denoted TAPBP-like or TAPBPL (Fig. 6, Additional file 1: Text S4). Sequence identity



between Atlantic salmon TAPBP, TAPBPR and TAPBPL is similar to the 22% amino acid identity found between the human TAPBP and TAPBPR sequences [63]. Both zebrafish and spotted gar only have one TAPBPL gene suggesting the TAPBPL duplications are unique to salmonids. This previously undefined TAPBPL gene is also present in frogs, turtles, alligators, birds and marsupials, but seems to have been lost in the lineage leading to placental mammals (Fig. 6). The TAPBPL gene duplication thus occurred prior to the split between the tetrapod and the bony fish lineages i.e. is also a remnant of the VGD2 event (Fig. 1). The two TAPBPL1a and TAPBPL1b genes originate from the SGD with an amino acid sequence identity of 88% but these two gene sequences only have 53% identity to the third TAPBPL2 sequence which then seems of a more ancient origin.

Many of the TAPBP and TAPBPR residues known to interact with MHC class I [18, 19, 64] are also conserved in the TAPBPL sequences (Additional file 1: Text S4). The human TAPBP C95 residue known to bind ERp57 [64], is not conserved in any of the other TAPBP, TAPBPR or TAPBPL sequences questioning the relevance of this residue in other species. However, a unique cysteine in the TAPBPL sequences located 11 amino acids further downstream could have a similar function as the human TAPBP C95 residue or it could resemble the unique C94 residue in TAPBPR known to interact with UGT1 [20]. The N-linked glycosylation site at N233 known to interact with CALR [65] is preserved in some of the TAPBP, TAPBPR and TAPBPL sequences. The single TAPBP lysine residues in the transmembrane region associating with TAP [2] is not conserved in the TAPBPL sequences, while some TAPBPL sequences display an ER retention signal as found in mammalian TAPBP [66]. Thus, based on conservation of many MHCI-interacting residues, the glycosylation site used to interact with CALR and the ER retention signal, TAPBPL sequences share more structural similarities with TAPBP than with TAPBPR supporting their name as TAPBPL and not TAPBPRL. Although we can only speculate as to the TAPBPL function, all three Atlantic salmon genes are expressed and their expression profiles resemble that of the TAPBP and TAPBPR genes with some tissue specific patterns (Table 2).

In addition to paralog functional copies of the immunoproteasome components PSMB8–13 [Fig. 2, [37]], Atlantic salmon also has duplicate copies of the interferon-inducible PSME1 and PSME2 regulatory subunits. These duplicates originate from the SGD with amino acid sequence identities of 93 and 91% respectively (Additional file 1: Figure S9) and display fair expression in a wide variety of organs (Table 2). As with the majority of paralogs, we do not know how these paralogs influence the peptide repertoire available for MHCI binding. Atlantic salmon also has four genes for the PSME3 subunit (Additional file 1: Figure S9),

but at least in humans this subunit is not interferon inducible and thus not a part of the immunoproteasome. If this also holds true for Atlantic salmon remains to be established.

Once transported inside the ER, the N-terminal end of peptides are further trimmed by ERAP1 and ERAP2 molecules [17]. For ERAP1 there is only one gene, while there are two copies of the ERAP2 gene (Fig. 5, Additional file 1: Figure S10). These two ERAP2 genes originate from the SGD with 93% amino acid sequence identity and both gene copies display low expression levels mainly in immunologically important tissues (Table 2).

Discussion

As opposed to what has been found in sharks, rats, frogs, chicken, medaka and zebrafish and allelic polymorphism in genes closely linked to the MHCIa gene was not evident in salmonids.

Northern pike, a diploid species basal to the salmonids [41], does display TAP2 polymorphism where many of the variable amino acid residues coincide with variable residue positions found in other species suggesting this polymorphism has been lost en route to salmonids.

It seems odd that there are functional MHCIa haplotypes in zebrafish and medaka but not in salmonids. One explanation for the loss of functional Ia haplotypes is the unique salmonid genome duplication that occurred approximately 94 million years ago [39]. Medaka and zebrafish have not experienced an additional WGD event after the third WGD that occurred in a ancestor to all teleost fish approximately 320 million years ago [56]. WGD provides raw material for evolutionary diversification, but must balance against negative dosage-effects, regulatory errors, negative protein-protein interactions and mitotic mistakes. Most WGDs are followed by a reduplication phase with extensive reorganizations to balance against the negative effects of duplications mentioned above. As seen in Atlantic salmon, a burst of transposon-mediated repeat expansions most likely triggered this reduplication phase resulting in increased homeologue sequence divergence and large chromosomal rearrangements such as fusions, fissions, deletions and inversions [39] that disrupted the possibility for homeologous pairing during mitosis. Potentially this extensive rearrangement and sequence divergence eradicated the functional haplotypes found in other teleost species in part including Northern pike. However, it should be noted that this study only includes a few haplotypes so there may be other haplotypes with polymorphic genes influencing MHCIa peptide processing, transport and loading.

Having paralog MHCI regions, the PSMB, TAP and TAPBP genes in the Atlantic salmon Ia and Ib regions could have evolved to serve MHCI molecules encoded by individual regions where genes within the Ia region serve the UBA molecule while genes within the Ib

region serve the UDA molecule. Without functional data, these paralog genes could of course also serve MHC I molecules originating from both regions.

The SGD also provided salmonids with many paralog genes that are retained as extant expressed and presumably functional copies. Lien et al. [39] found that SGD duplicates tended to belong to closely related but still different co-expression clusters suggesting far more instances of neo-functionalization than subfunctionalization. Many Atlantic salmon genes involved in peptide processing, transport and loading exist in multiple expressed copies with sequence identity reflecting which of the three whole genome duplication events they originated from. Atlantic salmon has CALR gene duplicates originating from both the SGD as well as the TGD in addition to a gene duplication that originates from the second vertebrate whole genome duplication (VGD2, Fig. 1). This provides Atlantic salmon with three CALR genes and three CALRL genes. As SGD paralogs tend to acquire novel functions based on transcript expression patterns [39], the CALR and CALRL gene duplicates may provide Atlantic salmon with a variety of CALR-like functions. A similar picture is seen for ERp57 and TAPBP, where gene duplications have provided Atlantic salmon with five expressed ERp57/ ERp57L genes and six expressed TAPBP/ TAPBPL genes. As spotted gar also contains these CALR/ CALRL, ERp57/ ERp57L and TAPBP/ TAPBPL gene duplications, they originate from the VGD2 event and not from the teleost specific genome duplication event. This is further supported by the fact that a TAPBPL gene is also present in the tetrapod lineage with what seems to be a bona fide gene even in opossum. How this TAPBPL molecule affects MHC I peptide loading or editing is unclear, but it may hold new and intriguing functions as seen for the TAPBP gene [18–20].

It is tempting to speculate that there exists alternative peptide-loading complexes in teleosts based on the presence of duplicate CALR/ CALRL, ERp57/ERp57L and TAPBP/ TAPBPL genes in both zebrafish, salmonids and medaka. If this holds true one may only speculate on which MHC molecules these complexes associate with and how that difference in PLC affects the peptides transported and loaded into the peptide binding groove of these MHC I molecules. Expression data support the CALR, ERp57L and TAPBP/ TAPBPL genes to be functionally related as they all display fairly low expression levels in identical tissues. Expression levels of CALRL and ERp57 are much higher and more diversified suggesting they may have different or additional roles.

Conclusion

We found no evidence pointing to functional polymorphism in TAP, TAPBP, PSMB8, PSMB9, PSMB12 and PSMB13 genes closely linked to classical UBA alleles in

salmonids as opposed to what has been found previously in zebrafish, medaka and several other species. The unique salmonid whole genome duplication has most likely disrupted such haplotypes with frequent recombination between chromosomal paralogs prior to their diversification. However, several whole genome duplications have provided salmonids with many duplicated genes involved in peptide generation, loading and editing most likely broadening their biological function compared to what is found in mammals. A surprise is the functional retention of many genes originating from the second vertebrate whole genome duplication event providing both spotted gar as well as teleosts with potential alternative versions of the peptide-loading complex. Future studies are needed to understand the functional relevance of these gene duplications including the alternative PLCs.

Methods

Data mining and bioinformatics

MHC class I haplotype gene sequences originate either from genomes available in GenBank or previously sequenced BAC clones as follows: **Atlantic salmon** (*Salmo salar*) BACs [37, 38]: Haplotype **Ia_#A** BACs 868O01 (EF441211), 92I04 (EF427384.1), 539 M19 (EF427383) and 129P21 (GQ505858); Haplotype **Ia_#B** BAC 714P22 (EF210363); Haplotype **Ib_#A** BAC 8I14 (EF427379); Haplotype **Ib_#B** BAC 438 J08 (FJ969490); Genome assembly GCA_000233375.4 [39] Haplotype **Ia_#C** Chr.27 NC_027326: 10.000.000–10.656.000; Haplotype **Ib_#C** genome Chr.14 region NC_027313:50.800.000–59.500.000. **Rainbow trout** (*Oncorhynchus mykiss*) BACs [42] Haplotype **Ia_#A** BAC AB162342; Haplotype **Ib_#A** BAC AB162343; Trout genome (GCA_002163495.1) Haplotype **Ia_#B** Chr.18 (CM007952.1); Haplotype **Ib_#B** Chr.14 (CM007948.1). **Coho salmon** (*Oncorhynchus kisutch*) genome GCA_002021735.1 haplotype **Ia** region Chr.17 NC_034190: 25300000–26,000,000; Haplotype **Ib** region Chr.14 NC_034187: 22.700.000–23.500.000. **Northern pike** (*Esox Lucius*) genome (GCA_000721915.3) Haplotype **Ia** region Chr.10 (NC_025977.3), and presumed allelic haplotype represented by the unplaced genomic scaffolds NW_017859580.1 (**NW80**) continued with NW_017859271.1 (**NW71**). Some genomic unannotated gene sequences as well as unlinked gene sequences were identified using various blastN and TblastN searches of Ensembl and NCBI databases and evolutionary diverged as well as species-specific sequences. Open reading frames were predicted using FGENESH [67] and verified using available expressed resources. Potential genomic assembly errors could influence the analyses. Some smaller pseudogene remnants that did not contribute to evolutionary understanding were neglected. Expressed match was either identified through TblastN search against EST,

GenBank nucleotide (cDNA) and available TSA/SRA resources. RPKM values shown in Table 2 were defined using CLC Genomic Workbench 6.0.5 [68] and SRR transcriptome runs from a single individual [39] as follows: HK (head kidney) SRR1422860, gills SRR1422858, spleen SRR1422870, heart SRR1422862, gut SRR1422859, muscle SRR1422866, brain SRR1422856, liver SRR22865, nose SRR1422867, skin SRR1422869, ovary SRR1422871, testis SRR1422872. 1.

Phylogenetic analysis

All amino acid sequence alignments were performed using Clustal X [69]. The phylogenetic trees were inferred using best-fit models calculated by MEGA7 [70] and bootstrapped using 100 replicates.

Additional files

Additional file 1: Figure S1. Teleost MHCI haplotypes. **Text S1.** Deduced amino acid gene sequences. **Text S2.** Alignment of deduced PSMB8 amino acid sequences. **Figure S2.** MHCI data. **Figure S3.** PSMB9 and PSMB12 data. **Figure S4.** PSMB7, PSMB10, PSMB13 data. **Text S3.** Alignment of deduced TAP2 amino acid sequences. **Figure S5.** TAPBP data. **Figure S6.** CANX, CALR and CALRL data. **Figure S7.** B2m data. **Figure S8.** ERp57 and ERp57L data. **Text S4.** TAPBP; TAPBPR and TAPBPL data. **Figure S9.** PSME data. **Figure S10.** ERAP data (PDF 4922 kb)

Abbreviations

B2m: Beta2-microglobulin; CALR: Calreticulin; CANX: Calnexin; MHC: Major histocompatibility complex; MYA: Million years ago; PSMB: Proteasome subunit beta; SGD: Salmonid-specific whole genome duplication; TAP: Transport associated protein; TAPBP: Tapasin; TGD: Teleost-specific whole genome duplication; VGD: Vertebrate genome duplication

Acknowledgements

Not applicable.

Funding

This study was funded by the Norwegian Veterinary Institute who had no role in the the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All data supporting the conclusions of this article are referred to or included within the article and its additional files. The datasets used and/or analysed during the current study are available from the corresponding author on request.

Authors contributions

UG was the only scientist involved in data gathering, analyses and manuscript drafting. The author read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares that she has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 October 2017 Accepted: 15 February 2018

Published online: 23 February 2018

References

1. Neefjes J, Jongstra ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol.* 2011;11(12):823–36.
2. Sadasivan B, Lehner PJ, Ortmann B, Spies T, Cresswell P. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity.* 1996;5(2):103–14.
3. Adams J. The proteasome: structure, function, and role in the cell. *Cancer Treat Rev.* 2003;29(1):3–9.
4. Ferrington DA, Gregerson DS. Immunoproteasomes: structure, function, and antigen presentation. *Prog Mol Biol Transl Sci.* 2012;109:75–112.
5. Cascio P. PA28alpha: the enigmatic magic ring of the proteasome? *Biomol Ther.* 2014;4(2):566–84.
6. Erath S, Groettrup M. No evidence for immunoproteasomes in chicken lymphoid organs and activated lymphocytes. *Immunogenetics.* 2015; 67(1):51–60.
7. van Ender PM, Tampe R, Meyer TH, Tisch R, Bach JF, McDevitt HO. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity.* 1994;1(6):491–500.
8. Nijenhuis M, Hammerling GJ. Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. *J Immunol.* 1996;157(12):5467–77.
9. Herget M, Baldauf C, Scholz C, Parcej D, Wiesmuller KH, Tampe R, Abele R, Bordignon E. Conformation of peptides bound to the transporter associated with antigen processing (TAP). *Proc Natl Acad Sci U S A.* 2011;108(4):1349–54.
10. Peters B, Tong W, Sidney J, Sette A, Weng Z. Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics.* 2003;19(14):1765–72.
11. Uebel S, Kraas W, Kienle S, Wiesmuller KH, Jung G, Tampe R. Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proc Natl Acad Sci U S A.* 1997;94(17):8976–81.
12. van Ender PM, Riganelli D, Greco G, Fleischhauer K, Sidney J, Sette A, Bach JF. The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med.* 1995;182(6):1883–95.
13. Cresswell P, Bangia N, Dick T, Diedrich G. The nature of the MHC class I peptide loading complex. *Immunol Rev.* 1999;172:21–8.
14. Peh CA, Burrows SR, Barnden M, Khanna R, Cresswell P, Moss DJ, McCluskey J. HLA-B27-restricted antigen presentation in the absence of tapasin reveals polymorphism in mechanisms of HLA class I peptide loading. *Immunity.* 1998; 8(5):531–42.
15. Dick TP, Bangia N, Peaper DR, Cresswell P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. *Immunity.* 2002; 16(1):87–98.
16. Kozlov G, Pocanschi CL, Rosenauer A, Bastos-Aristizabal S, Gorelik A, Williams DB, Gehring K. Structural basis of carbohydrate recognition by calreticulin. *J Biol Chem.* 2010;285(49):38612–20.
17. Chen H, Li L, Weimershaus M, Evnouchidou I, van Ender P, Bouvier M. ERAP1-ERAP2 dimers trim MHC I-bound precursor peptides; implications for understanding peptide editing. *Sci Rep.* 2016;6:28902.
18. Hermann C, Trowsdale J, Boyle LH. TAPBPR: a new player in the MHC class I presentation pathway. *Tissue Antigens.* 2015;85(3):155–66.
19. Morozov GI, Zhao H, Mage MG, Boyd LF, Jiang J, Dolan MA, Venna R, Norcross MA, McMurtrey CP, Hildebrand W, et al. Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *Proc Natl Acad Sci U S A.* 2016;113(8):E1006–15.
20. Neerinx A, Hermann C, Antrobus R, van Hateren A, Cao H, Trautwein N, Stevanovic S, Elliott T, Deane JE, Boyle LH. TAPBPR bridges UDP-glucose: glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *elife.* 2017;6:e23049.
21. Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, Salomonsen J. Gene organisation determines evolution of function in the chicken MHC. *Immunol Rev.* 1999;167:101–17.
22. van Hateren A, Carter R, Bailey A, Kontouli N, Williams AP, Kaufman J, Elliott T. A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. *J Biol Chem.* 2013;288(45): 32797–808.

23. Walker BA, Hunt LG, Sowa AK, Skjodt K, Gobel TW, Lehner PJ, Kaufman J. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proc Natl Acad Sci U S A*. 2011;108(20):8396–401.
24. Chappell P, Meziane El K, Harrison M, Magiera L, Hermann C, Mears L, Wrobel AG, Durant C, Nielsen LL, Buus S, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *elife*. 2015;4:e05345.
25. Tregaskes CA, Harrison M, Sowa AK, van Hateren A, Hunt LG, Vainio O, Kaufman J. Surface expression, peptide repertoire, and thermostability of chicken class I molecules correlate with peptide transporter specificity. *Proc Natl Acad Sci U S A*. 2016;113(3):692–7.
26. Rudolph MG, Stevens J, Speir JA, Trowsdale J, Butcher GW, Joly E, Wilson IA. Crystal structures of two rat MHC class Ia (RT1-a) molecules that are associated differentially with peptide transporter alleles TAP-A and TAP-B. *J Mol Biol*. 2002;324(5):975–90.
27. Joly E, Le Rolle AF, Gonzalez AL, Mehling B, Stevens J, Coadwell WJ, Hunig T, Howard JC, Butcher GW. Co-evolution of rat TAP transporters and MHC class I RT1-a molecules. *Curr Biol*. 1998;8(3):169–72.
28. Namikawa C, Salter-Cid L, Flajnik MF, Kato Y, Nonaka M, Sasaki M. Isolation of Xenopus LMP-7 homologues. Striking allelic diversity and linkage to MHC. *J Immunol*. 1995;155(4):1964–71.
29. Ohta Y, Powis SJ, Lohr RL, Nonaka M, Pasquier LD, Flajnik MF. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in Xenopus: further evidence for co-evolution among MHC class I region genes. *Eur J Immunol*. 2003;33(11):3017–27.
30. Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM. A comprehensive analysis of teleost MHC class I sequences. *BMC Evol Biol*. 2015;15:32–49.
31. McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, Yoder JA, Andrade J, de Jong JL. Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. *Proc Natl Acad Sci U S A*. 2016;113(34):E5014–23.
32. Miura F, Tsukamoto K, Mehta RB, Naruse K, Magtoon W, Nonaka M. Transspecies dimorphic allelic lineages of the proteasome subunit beta-type 8 gene (PSMB8) in the teleost genus *Oryzias*. *Proc Natl Acad Sci U S A*. 2010;107(50):21599–604.
33. Nonaka MI, Nonaka M. Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics*. 2010;62(5):319–32.
34. Tsukamoto K, Miura F, Fujito NT, Yoshizaki G, Nonaka M. Long-lived dichotomous lineages of the proteasome subunit beta type 8 (PSMB8) gene surviving more than 500 million years as alleles or paralogs. *Mol Biol Evol*. 2012;29(10):3071–9.
35. Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A, Saebo S, Stet RJ. MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics*. 2003;55(4):210–9.
36. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc R Soc B*. 2013;281:20132881.
37. Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, Koop BF, Grimholt U. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. *BMC Genomics*. 2010;11:154.
38. Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS, et al. Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). *BMC Genomics*. 2007;8:251.
39. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, et al. The Atlantic salmon genome provides insights into rediploidization. *Nature*. 2016;533(7602):200–5.
40. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, Bento P, Da Silva C, Labadie K, Alberti A, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Comm*. 2014;5:3657.
41. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schallburg KR, Lemon C, Bird NH, Koop BF. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS One*. 2014;9(7):e102089.
42. Shiina T, Dijkstra JM, Shimizu S, Watanabe A, Yanagiya K, Kiryu I, Fujiwara A, Nishida-Umehara 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*. 2005;56(12):878–93.
43. Matsuo M, 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;53:930–40.
44. Tsukamoto K, Hayashi S, Matsuo M, Nonaka M, Kondo M, Shima MI, Asakawa S, Shimizu N, Nonaka M. Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. *Immunogenetics*. 2005;57:420–31.
45. Kiryu I, Dijkstra JM, Sarder RI, Fujiwara A, Yoshiura Y, Ootake M. New MHC class Ia Domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. *Fish Shellfish Immunol*. 2005;18:243–54.
46. Hansen JD, Strassburger P, Du PL. Conservation of an alpha 2 domain within the teleostean world, MHC class I from the rainbow trout *Oncorhynchus mykiss*. *Dev Comp Immunol*. 1996;20(6):417–25.
47. Nonaka MI, Aizawa K, Mitani H, Bannai HP, Nonaka M. Retained orthologous relationships of the MHC class I genes during euteleost evolution. *Mol Biol Evol*. 2011;28(11):3099–112.
48. 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(5):3297–308.
49. Koch J, Guntrum R, Tampe R. The first N-terminal transmembrane helix of each subunit of the antigenic peptide transporter TAP is essential for independent tapasin binding. *FEBS Lett*. 2006;580(17):4091–6.
50. Sever L, Vo NT, Bols NC, Dixon B. Rainbow trout (*Oncorhynchus mykiss*) contain two calnexin genes which encode distinct proteins. *Dev Comp Immunol*. 2014;42(2):211–9.
51. Magor KE, Shum BP, Parham P. The beta 2-microglobulin locus of rainbow trout (*Oncorhynchus mykiss*) contains three polymorphic genes. *J Immunol*. 2004;172(6):3635–43.
52. Shum BP, Azumi K, Zhang S, Kehrler SR, Raison RL, Detrich HW, Parham P. Unexpected beta2-microglobulin sequence diversity in individual rainbow trout. *Proc Natl Acad Sci U S A*. 1996;93(7):2779–84.
53. Kondo H, Darawiroj D, Gung YT, Yasuike M, Hirono I, Aoki T. Identification of two distinct types of beta-2 microglobulin in marine fish, *Pagrus major* and *Seriola quinqueradiata*. *Vet Immunol Immunopathol*. 2010;134(3–4):284–8.
54. Perarnau B, Siegrist CA, Gillet A, Vincent C, Kimura S, Lemonnier FA. Beta 2-microglobulin restriction of antigen presentation. *Nature*. 1990;346(6286):751–4.
55. Liu H, Peatman E, Wang W, Abernathy J, Liu S, Kucuktas H, Lu J, Xu DH, Klesius P, Waldbieser G, et al. Molecular responses of calreticulin genes to iron overload and bacterial challenge in channel catfish (*Ictalurus punctatus*). *Dev Comp Immunol*. 2011;35(3):267–72.
56. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427–37.
57. Raghavan M, Wijeyesakere SJ, Peters LR, Del Cid N. Calreticulin in the immune system: ins and outs. *Trends Immunol*. 2013;34(1):13–21.
58. Kales S, Fujiki K, Dixon B. Molecular cloning and characterization of calreticulin from rainbow trout (*Oncorhynchus mykiss*). *Immunogenetics*. 2004;55(10):717–23.
59. Kales SC, Bols NC, Dixon B. Calreticulin in rainbow trout: a limited response to endoplasmic reticulum (ER) stress. *Comp Biochem Physiol B Biochem Mol Biol*. 2007;147(4):607–15.
60. Sever L, Bols NC, Dixon B. The cloning and inducible expression of the rainbow trout ERp57 gene. *Fish Shellfish Immunol*. 2013;34(2):410–9.
61. Pinto RD, Moreira AR, Pereira PJ, dos Santos NM. Two thioredoxin-superfamily members from sea bass (*Dicentrarchus labrax*, L.): characterization of PDI (PDI A1) and ERp57 (PDI A3). *Fish Shellfish Immunol*. 2013;35(4):1163–75.
62. Landis ED, Palti Y, Dekoning J, Drew R, Phillips RB, Hansen JD. Identification and regulatory analysis of rainbow trout tapasin and tapasin-related genes. *Immunogenetics*. 2006;58(1):56–69.
63. Teng MS, Stephens R, Du Pasquier L, Freeman T, Lindquist JA, Trowsdale J. A human TAPBP (TAPASIN)-related gene, TAPBP-R. *Eur J Immunol*. 2002;32(4):1059–68.
64. Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity*. 2009;30(1):21–32.

65. Rizvi SM, Del Cid N, Lybarger L, Raghavan M. Distinct functions for the glycans of tapasin and heavy chains in the assembly of MHC class I molecules. *J Immunol*. 2011;186(4):2309–20.
66. Paulsson KM, Jevon M, Wang JW, Li S, Wang P. The double lysine motif of tapasin is a retrieval signal for retention of unstable MHC class I molecules in the endoplasmic reticulum. *J Immunol*. 2006;176(12):7482–8.
67. Salamov AA, Solovyev VV. Ab initio gene finding in drosophila genomic DNA. *Genome Res*. 2000;10(4):516–22.
68. CLC Genomics Workbench 6. <https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>.
69. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8.
70. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
71. Benton MJ, Donoghue PC. Paleontological evidence to date the tree of life. *Mol Biol Evol*. 2007;24(1):26–53.
72. Campbell MA, Lopez JA, Sado T, Miya M. Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. *Gene*. 2013;530(1):57–65.
73. Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A*. 2012;109(34):13698–703.
74. Crete-Lafreniere A, Weir LK, Bernatchez L. Framing the Salmonidae Family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS One*. 2012;7(10):e46662.
75. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutations data matrixes from protein sequences. *CABIOS*. 1992;8:275–82.
76. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008;25(7):1307–20.
77. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001;18(5):691–9.
78. Nonaka M, Yamada-Namikawa C, Flajnik MF, Du Pasquier L. Trans-species polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus. *Xenopus Immunogenetics*. 2000;51(3):186–92.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

