MHC class I evolution; from Northern pike to salmonids

Unni Grimholt (Unni.Grimholt@vetinst.no)
Norwegian Veterinary Institute https://orcid.org/0000-0002-0202-8061
Morten Lukacs
Norwegian Veterinary Institute

Research article

Keywords: MHC class I, antigen presentation, phylogeny, Northern pike, salmonids

Posted Date: June 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36819/v1

Version of Record: A version of this preprint was published on January 21st, 2021. See the published version at https://doi.org/10.1186/s12862-020-01736-y.

Abstract

Background

Salmonids are of major importance both as farmed and wild animals. The financial importance to aquacultural industry as well as to rural communities that profit from wild salmon fishing is substantial. With the changing environment comes changes in pathogenic pressures so understanding the immune system of all salmonid species is of essence. Major histocompatibility complex (MHC) genes are key players in the adaptive immune system signalling infection to responding T-cells populations. Their key role in inducing immunity and their link to disease resistance instigate studies on structure, function and evolution. Northern Pike, a basal sister clade to salmonids, represent a species which has not experienced the fourth salmonid specific whole genome duplication.

Results

Comparing the gene organization and evolution of MHC class I gene sequences in Northern pike versus salmonids displays a complex picture of how many of these genes evolved. Salmonid Ia and Ib Z lineage genes are not orthologs to the Northern pike Z lineage sequences. Instead, the salmonid genes have experienced unique gene duplications in the two duplicated regions as well as in the Salmo and Oncorhynchus branch. Potentially, transposase elements enabling these duplications were already present in the Northern pike Z lineage gene.

Conclusions

Although both Northern pike as well as salmonids have expanded their U and Z lineage genes, these gene duplications have occurred separately in pike and in a salmonid ancestor. However, the similarity between these duplications suggest the transposable machinery was present in a common ancestor. The salmonid MHCIa and MHCIb regions were mostly formed during the 80MYA since the split from pike and before the Oncorhynchus and Salmon branch separated. As seen in tetrapods, the non-classical U lineage genes are diversified duplicates of their classical counterpart. One MHCI lineage, the L lineage, experienced massive species-specific gene duplications after Oncorhynchus and Salmo split approximately 25 MYA. Based on what we currently know about L lineage genes, this diversity will most likely affect immune responses in individual species.

Introduction

Major histocompatibility complex (MHC) molecules are involved in protection against invading pathogens. The term MHC represents two different classes of molecules with many different genes within each class. Briefly, MHC class II (MHCII) molecules are composed of an alpha as well as a beta chain each with two extracellular domains. For classical MHCII molecules, polymorphism primarily

resides in the alpha 1 and beta 1 domains, which bind and present peptides originating from the extracellular space to CD4 + T-cells. The alpha 2 and beta 2 domains contribute with molecular structure and CD4 association. Mammalian non-classical MHC class II molecules assist in the peptide loading of their classical counterpart [1] while teleost non-classical MHCII molecules have as yet unknown functions [2].

Most MHC class I (MHCI) molecules are composed of an alpha chain non-covalently linked to a beta2-microglobulin (b2m) molecule. Classical MHCI molecules are defined by their polymorphic content, their expression in most tissues and their ability to bind and present peptides to CD8 + T-cells. Here, the two extracellular alpha 1 and alpha 2 domains of the alpha chain are highly polymorphic and responsible for binding peptides from self and non-self proteins to CD8 + T-cells. The alpha 3 domain and b2m contribute with structural stability and CD8 binding. In humans the classical genes are denoted HLA-A, HLA-B and HLA-C and have 3720, 4604 and 3470 protein alleles registered in the IPD-HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/stats.html).

Non-classical MHCI molecules have more restricted expression patterns, lower levels of polymorphism and most bind non-peptide ligands. One exceptions is the human HLA-E molecule that binds and presents signal peptide sequences from classical MHCI molecules to natural killer (NK) cells [3]. Non-classical MHCI molecules come in various formats, with or without b2m association, and some also lack the alpha 3 domain. Examples of non-peptide ligands are for instance the tetrapod CD1 molecule that binds glycolipids and the mammalian MR1 molecule that binds vitamin metabolites [3]. MHC molecules with three extracellular alpha domains, but no b2m association, are MICA and MICB which interact with an NKG2D receptor expressed on NK and γδT-cells. Another MHCI molecule that interacts with the NKG2D receptor is ULBP, but this molecule lacks the alpha 3 domain and does not associate with b2m.

Some of the human MHC class I molecules can be traced as far back as chicken such as the CD1 molecule, but no lineages can be traced as far back as sharks [4]. Teleost fish are too phylogenetically distant and share no MHCI orthology with human MHCI lineages, although they both originated from a common ancestor 450 million years ago. The only lineage shared between the sarcopterygian and actinopterygian lineages is the teleost MHCI Z lineage that is also present in lungfish [5, 6].

One major difference between the mammalian and teleost MHC is the regional organization. Humans have a 4 Megabase (Mb) region containing both the classical MHCI and MHCII genes in addition to some non-classical MHC genes. Some genes involved in generating and transporting peptides also reside in this region, such as the proteasome component beta genes PSMB8, PSMB9 and the antigen transporter TAP2. In teleosts, the classical MHCI and MHCII genes have separated with class I being linked to genes involved in peptide generation and transport while MHC class II genes reside elsewhere [7, 8].

Salmonids experienced a whole genome duplication 94 million years ago where many of the duplicated regions are retained [9, 10]. This includes a duplicate version of the entire MHCI region with one region containing the classical MHCI UBA gene denoted the MHCIa region. The duplicate homeolog region harbours several non-classical U lineage genes as seen in Rainbow trout and Atlantic salmon and is

denoted the MHCIb region [11, 12]. The genomic organization of MHCI regions in other salmonids is currently unknown. Northern pike and salmonids are sister taxa where pike is basal to the salmonids and has not experienced the fourth salmonid specific whole genome duplication event (4WGD). Northern pike therefore enables studies of how the 4WGD affected evolution of genes and gene duplicates.

In addition to the teleost specific Z lineage mentioned above, we have defined five other teleost MHCI lineages denoted U, L, S, P and H [5, 13, 14]. The U lineage is composed of both classical as well as non-classical peptide-binders. Most teleosts studied so far only have one to possibly three classical U lineage genes. Atlantic salmon and Rainbow trout both have only one classical MHCI gene denoted UBA, while Medaka has two classical MHCI genes denoted UAA and UBA [11, 12, 15]. Zebrafish seems to have varying haplotypes with one to three classical MHCI genes [16]. A species that chose a different approach is Atlantic cod, which has expanded the MHCI lineage with 100 genes or more, potentially compensating for the lack of MHC class II molecules [17]. This massive gene expansion may be unique to Gadidae species. Number and polymorphic content of classical U lineage genes in other teleost and ray-finned species are currently not well defined.

Previous studies have shown that U lineage domains have different evolutionary histories with alpha 1 domain sequences segregating as distinct lineages shared between distantly related species [18–20]. Also alpha 2 domain sequences display some evolutionary conserved lineages, although this pattern is less pronounced than for the alpha 1 domain. Alpha 3 domains on the other hand, seem more structurally constrained potentially due to adaptation to species-specific b2m and CD8 association.

One additional MHCI lineage is most likely a peptide-binder, i.e. the Z lineage, which we found to have a completely conserved peptide-binding motif in all studies ray-finned fishes [5]. These Z lineage genes reside in both the MHCIa and MHCIb regions in Atlantic salmon [5]. A complete conservation of the peptide binding residues suggest an intriguingly conserved, but yet undefined, function.

Peptides bound by the MHCI molecules are generated by a cytosolic proteasome and then transported into the endoplasmatic reticulum through a MHC-peptide specific transporter consisting of the TAP1 and TAP2 subunits. Upon stimulation, the multi-subunit proteasome becomes an immunoproteasome complex through replacement of the three constitutive subunits PSMB5-7 with the subunits PSMB8-10. Due to the third teleosts specific whole genome duplication event, teleosts have a duplicate PSMB9 subunit denoted PSMB12 in addition to a duplicate PSMB10 subunit denoted PSMB13. The PSMB8-13 subunits all reside in the duplicate MHCIa and MHCIb region in Atlantic salmon and Rainbow trout [21]. As opposed to Zebrafish [22], the PSMB8 and PSMB13 subunits were not polymorphic in these salmonids [21]. The MHC-peptide specific transporter subunit TAP2 also resides in these duplicate MHC regions. This subunit is not polymorphic in Atlantic salmon and Rainbow trout either, as opposed to that found in Zebrafish.

None of the four remaining teleost MHCI lineages seem like peptide binders. The L lineage molecules most likely binds hydrophobic ligands, and can be traced back to spotted gar, a species that separated from teleosts before the teleost specific third whole genome duplication event (3WGD)[5]. Different

Atlantic salmon L lineage genes were recently shown to vary in their response to pathogen stimulations [23], suggesting they have different roles in defence against pathogens.

The function of the remaining three teleost MHCI lineages is currently unknown. The P lineage can also be traced back to spotted gar and has greatly expanded inn some species such as Fugu [5]. Also the H lineage is found in spotted gar. Sequences from this H lineage show unprecedented deterioration of its extracellular domains, where teleosts have lost the alpha 3 domain as compared to their spotted gar orthologue. Also the alpha 1 and alpha 2 domains of teleost H lineage molecules is shorter in some species while the cytoplasmic tail has been conserved across divergent species [13]. The S lineage is only found in teleosts.

As mentioned above, salmonids experienced a whole genome duplication approximately 94 million years ago (MYA)[10] where many of the duplicated genes and regions are retained. At least in Atlantic salmon, duplicated genes have taken on new functions rather than sub-functionalization [9]. Access to many new salmonid genomes now open for investigations on how the MHC genes and regions have evolved in this complex duplicated landscape. As a reference, we use Northern pike, a species basal to the salmonids, a species that has not had the additional 4WGD event.

Results And Discussion

The results presented below are based on the genomes of the salmonids Atlantic salmon, Brown trout, Rainbow trout, Sockeye salmon, Coho salmon, Chinook salmon and Charr (see Material and Methods for details). All genomes, apart from Charr and Northern pike, originated from completely homozygous or so-called double haploid animals thus eliminating the added confusion of allelic gene variants. To understand the evolution of genes, the salmonid data are compared against results from the Northern pike genome, a species that is basal to salmonids, but lacks the 4WGD [22](Fig. 1). Genomes from the three Salmonidae genomes Coregonus, *Hucho hucho* and *Thymallus thymallus* were ignored as they contained un-annotated or incomplete genomic regions, thus not enabling informative comparisons.

There seems to be some confusion as to the origin of the NCBI Salvelinus genome, now annotated as *Salvelinus* in NCBI, which may potentially be *Salvelinus malma malma* and not Salvelinus alpinus as presented in the original article [24, 25]. Using standardised nomenclature exemplified by Sasa for *Salmo salar* and Eslu for *Esox lucius*, we also used Saal for *Salvelinus alpinus* although it may be Sama.

Orthology between salmonid regions is a summary of data obtained from Christensen et al. and Sutherland et al. [24, 26] presented in Additional file 1. For Brown trout, the linkage groups presented by Leitwein and coworkers [27] do not match the chromosome numbers in the NCBI genome, so regional orthology is currently based on blast match with region specific genes from other salmonids when this was informative.

Evolution of salmonid MHCIa and MHCIb regions

Based on previous data we define the genomic region containing the classical UBA locus as the MHCIa region and the duplicate region containing non-classical genes as the MHCIb region [11, 12]. Genes residing within these two regions also have an -a or -b extension. All salmonid genomes analysed in this study contained well-defined and annotated duplicated MHCIa and MHCIb regions (Additional file 2). The Ia region, containing the UBA locus, was overall identical for all species with a few exceptions. Brown trout has a CD5-like gene in between the SLC39A7a and RING2a gene, not present in any of the other species.

In Zebrafish, there are functional MHCI haplotypes with polymorphism in both proteasome subunits PSMB8, PSMB13 as well as TAP2 [28]. In Rainbow trout, the two allelic PSMB8 variants found in Zebrafish are encoded by two different genes in the Onmy-MHCIa region [21]. Here, the *Onmy-PSMB8a* gene is a pseudogene while the *Onmy-PSMB8F* gene is functional. PSMB8F pseudogenes have previously been found in the duplicate Atlantic salmon MHCIa and MHCIb regions [21]. However, there is a bona fide Atlantic salmon *Sasa-PSMBF* sequence in Genbank (ACI66984.1), suggesting some Atlantic salmon haplotypes may have a functional variant of this gene. Neither pike nor other salmonids have an annotated PSMB8F gene in the MHCIa region, but Charr has a PSMB8F gene on an unplaced scaffold (XP_023998549.1).

The duplicate MHCIb region was also mostly identical in all analysed species. The LHX9-like gene found in Northern pike is present in all salmonid MHCIb regions with the exception of Salvelinus. All but Salvelinus and Northern pike also have a varying number of chitin synthase-like (CHS2) genes in between the RXRB and SLC39A7 genes. Chitin synthase is a well-known molecule in fungi and invertebrates, but also seems to have functional roles in fish and amphibians [29]. In Chinook salmon there is a duplicate of the entire MHCIb region (Genbank NW_020128813), which most likely is an assembly artefact as the sequenced animal was a double haploid.

Evolution of U lineage genes

Six Northern pike U lineage genes reside on chromosome 10 here defined as *Eslu-UAA* through *Eslu-UFA* (Fig. 2, Additional files 2–4). Based on phylogeny, there seems to be three original genes where each of the three genes have duplicated into *Eslu-UAA* and *Eslu-UBA*, *Eslu-UCA* and *Eslu-UDA* and *Eslu-UEA* and *Eslu-UFA* (Fig. 3). *Eslu-UCA* is only a partial sequence and may be a pseudogene. The polymorphic content of these genes remains undefined, but there is one EST and one TSA matching the Eslu-UAA/UBA genes (Genbank GH268323 and TSA GATF010284) and one EST originating from one of the *Eslu-UEA* or *Eslu-UFA* loci (EV373903). The seventh pike U lineage gene is located on an unplaced scaffold (*Eslu-UGA*, NW_022995044), and seems to be a pseudogene duplicate of the *Eslu-UDA* gene.

As previous studies have shown that the three extracellular alpha domains of U lineage sequences display different evolutionary patterns [5, 18–20], we made phylogenetic trees of both entire mature extracellular amino acid sequences as well as trees of individual alpha 1, alpha 2 and alpha 3 domain sequences to identify orthology (Fig. 3, Additional file 3).

Phylogenies of alpha 1 domain sequences shared by distantly related teleost species, show that also non-classical genes share these lineages (Fig. 2)[5, 18–20]. Non-classical UEA gene sequences share the alpha 1 domain lineage Va, UGA gene sequences share the alpha 1 domain lineage II and most UCA and UDA gene sequences cluster with the alpha 1 domain lineage I. Also Northern pike U lineage genes share these alpha 1 domain lineages. *Eslu-UAA* and *Eslu-UBA* alpha 1 domain sequences cluster with alpha 1 domain lineages Vb, *Eslu-UDA* clusters with lineage IIIa and *Eslu-UEA* and *Eslu-UFA* cluster with lineage IIIb sequences. In the alpha 2 domain analysis, all Northern pike sequences cluster together, although the bootstrap value is only 31 percent (Additional file 3). A similar clustering is also seen for all Northern pike alpha 3 domain sequences, with a somewhat higher bootstrap value.

Only one salmonid U lineage gene, UHA, resides outside of the two duplicated MHCIa and MHCIb regions (Table 1, Additional file 2). Sequences from this gene display strongly supported clusters in all phylogenies. Northern pike and Sockeye salmon did not display any UHA gene sequences, but the remaining salmonids all have UHA lineage genes on one homeolog of Northern pike Chr. 16 (Additional file 1). Atlantic salmon and Charr have regionally duplicated UHA lineage genes where at least the duplicate *Sasa-UHA2* gene is a pseudogene (Additional file 4). Although the two Charr UHA gene sequences are incomplete, there is an expressed UHA1/2-like sequence in *Salvelinus malma* (Genbank AYG86905.1), suggesting at least one of these UHA loci are functional also in Charr. Overall, UHA gene sequences are very different from other U lineage sequences (Fig. 3, Additional file 3), suggesting an ancient origin. However, we have not been able to find orthologs in any other teleost, so these genes may have evolved fast in salmonids.

Table 1
Number of MHC class I lineage genes in salmonids and Northern pike

	U	Z	L	S	H	P
Northern pike	7	5	4	1	1	0
Atlantic salmon	8	7	12	6	2	1
Brown trout	7	7	25	3	2	1
Rainbow trout	7	6	14	2	2	1
Coho salmon	6	5	14	2	2	1
Sockeye salmon	5	5	14	2	2	1
Chinook salmon	9	7	16	2	2	1
Charr	11	4	13	2	2	1

Only Atlantic salmon has a duplicate annotated U lineage gene in the MHCIa region denoted ULA, a gene that lacks the transmembrane domain (Additional file 2–4)[30]. We know that the UBA loci from Atlantic salmon, Rainbow trout, Brown trout and Sockeye salmon are classical MHCI loci with considerable

polymorphism [18, 20, 31–34]. There are currently 48 Atlantic salmon and Rainbow trout UBA alleles registered in the IPD-MHC database [35] while 31 and 34 alleles have been defined in Brown trout and Sockeye salmon. The polymorphic content of UBA loci from Coho, Chinook and Charr remains undetermined.

MHC class I gene richness is most profound in the salmonid MHCIb regions, with Brown trout and Salvelinus having four U lineage genes surrounding the TAPBPb and PSMB8b genes (Fig. 2). Rainbow trout has three annotated U lineage genes in this region with an additional fourth *Onmy-UFA* pseudogene reported previously [12]. Previous studies have shown that Rainbow trout and Atlantic salmon MHCIb regions contain non-classical MHC genes, displaying low polymorphism and more restricted expression patterns than their classical UBA counterparts [11, 12]. Sockeye, Chinook and Coho salmon all have two annotated U lineage genes in this region. This region then resembles the three original MHCI genes found on Northern pike Chr.10.

Salvelinus has two additional unplaced scaffolds containing U lineage genes, all clustering with UBA sequences (UXA, UZA1/2; Additional files 3 & 4). Their origin and location is unknown, but as the sequenced genome does not originate from a double haploid animal, they could be allelic variants of non-classical U lineage genes or assembly artefacts. Chinook salmon also has two additional U lineage genes residing on unplaced scaffolds here denoted Onts-U1 and Onts-U2. Onts-U1 seems like a pseudogene with sequence identity to *Onts-UCA*. Onts-U2 is a duplicate of the *Onts-UEA* gene sequence, and most likely represents an assembly artefact as the Chinook salmon genome originates from a double haploid.

MHCIb regions also contain a unique UGA gene that is present in all analysed salmonids, located in between the SLC39A7b and RING2Ab genes (Fig. 2). Chinook salmon lacks an annotated UGA gene, although there are expressed Chinook sequences supporting a functional UGA locus (e.g. GGDU01219126.1). The gene denoted UGA in Northern pike (Additional file 4) is not an ortholog to the salmonid UGA genes, so UGA is a gene duplication that translocated to the MHCIb region after salmonids split from pike. UGA lineage sequences show strongly supported clusters in alpha 1 and alpha 2 domain phylogenies, while the alpha 3 domain sequences are more dispersed.

Based on location and phylogenetic clustering, the UEA gene seems to have existed in a primordial salmonid, but was then lost in Atlantic and Sockeye salmon (Fig. 2&3, Additional files 3 & 4). All UEA alpha domain phylogenies show strongly supported clusters. Salmonid UCA and UDA gene sequences also form strongly supported clusters in the alpha 1 and alpha 2 domain sequence phylogenies, suggesting they originate from a salmonid ancestor. Duplications from a single primordial UC/DA gene to multiple UCA and UDA genes seem to have occurred individually in the Oncorhynchus and Salmo lineages based on the alpha 2 domain phylogenies, as well as in each individual species (Fig. 3, Additional file 3). The gene sequences defined as UFA in Charr and Brown trout do not cluster in phylogenies, so they represent within species gene duplications. However, the UFA pseudogene previously reported in Rainbow trout, clusters with the UFA sequence from Brown trout (data not shown), so this gene originated in a salmonid ancestor.

We have previously shown that the Atlantic salmon MHCIb region contains haplotypes with varying number of non-classical *Sasa-UCA* and *Sasa-UDA* genes [36]. In the Atlantic salmon genome, this haplotypic variation is more pronounced with an 8.6 Mb region separating the *Sasa-UCA* pseudogene from two additional UCA and UDA genes as opposed to the 30 Kb separating the UDA and UCA genes in the previously sequenced BAC (Genbank FJ969490). If this represents haplotypic variation or is a study artefact remains unknown, but Brown trout, the closest relative to Atlantic salmon, does not show this UCA/UDA gene duplication 10 Mb upstream suggesting this MHCIb haplotype may be Atlantic salmon specific.

Oncorhynchus species also have a unique U lineage pseudogene located approximately 10 Mb upstream of their UCA genes, which we here name UMA. These Oncorhynchus regions are unique as they do not contain the same genes as those surrounding the Atlantic salmon genome UDA gene 8.6 Mb upstream of the major MHCIb region. Nor does this region resemble the UIA region found in Medaka, where there is approximately 14 Mb between the classical UAA/UBA genes and a UIA gene [19]. This Oncorhynchus UMA gene then seems to be a unique gene duplication that occurred on the Oncorhynchus lineage only. Looking at other teleost species, two S lineage genes in *Astyanax mexicanus* are located in an unplaced region in between MYO1G and SGK1 genes, a region also containing the Oncorhynchus UMA region genes EYA3, CDK5R1 and XKR8.3. However, phylogeny does not support any relationship between these Mexican tetra S lineage sequences and the Oncorhynchus UMA sequences (data not shown). A plausible explanation would then be that Atlantic salmon and an ancestral Oncorhynchus species have experienced unique but similar translocations of the UCA/UDA and UMA genes.

To summarize, evolutionary orthology between individual Northern pike and salmonid MHCI gene sequences is not apparent in our phylogenies. The seven U lineage pike genes occurred through duplications in pike after the split from salmonids. A similar gene expansion of U lineage genes in the MHCIb region has occurred in a salmonid ancestor, where a primordial UBA gene has duplicated and diversified into the non-classical genes found in the MHCIb regions today. Such a species-specific duplication of classical genes into diversified non-classical genes has also occurred in some tetrapod species [4, 37].

Z lineage evolution

In addition to the six U lineage genes, Northern pike also has five Z lineage genes on linkage group 10 (Fig. 2, Table 1, Additional files 2–4). In comparison, the salmonid MHCIa and Ib regions all have from two to four Z lineage genes per region. Due to the unique position of the Salmo ZAA gene residing in the MHCIa region, we chose to reserve this ZAA name to reflect a location in between the VHSVa induced protein and ATF6a. The remaining sequences are named ZBA through ZDA regardless of phylogenetic clustering. Of pike and salmonid Z lineage genes, only *Onmy-ZDAb* and *Satr-ZDAb* seem like pseudogenes.

Phylogenetic trees of the entire mature extracellular amino acid Z lineage sequences display two well-supported clades, each with two sub-clades. Surprisingly, all Northern pike Z lineage gene sequences

cluster together with a strong bootstrap support, suggesting they are within species gene duplications (Fig. 4). A similar strongly supported clustering of pike Z lineage sequences is also seen when we perform phylogenies of individual extracellular domains (Additional file 3). Based on the two to four Z lineage gene duplicates identified in salmonid MHCIa and MHCIb regions (Fig. 2), one would have expected some orthology between pike and salmonid gene sequences.

The first clade (Fig. 4, clade 1) consists of MHCIa region sequences, while the second clade (Fig. 4, clade 2) consist of MHCIb region sequences, suggesting the Z lineage genes evolved independently in the MHCIa and MHCIb regions (Fig. 4, Additional file 2 &3). Clade 1 gene sequences are further divided into two subclades, one containing Oncorhynchus gene sequences (clade 1.1) and the other with Salmo and Salvelinus gene sequences (clade 1.2). Clade 1.1 suggests that one original Oncorhynchus gene expanded to the three ZBAa, ZCAa and ZDAa genes present in this region today where ZDAa is a more recent duplicate of ZBAa. Although not as strongly supported, Salmo and Salvelinus Z lineage Ia genes within clade 1.2 also seem like within region duplicates of one common ancestor. It seems that the evolutionary process has repeated itself with the ZBAa and ZDAa genes being a more recent gene duplications while ZCAa is an older gene duplication. The Salmo ZAAa gene is also a more recent duplication of the ZBA or ZDA gene. Charr MHCIa Z lineage sequences show a dual clustering, with the *Saal-ZBAa* sequence clustering with Oncorhynchus while the *Saal-ZCAa* sequence clusters with Salmo ZCAa sequences.

Sequences originating from the MHCIb region split into two strongly supported subclusters (Fig. 4, subclades 2.1 and 2.2) and in this region Oncorhynchus and Salmo Z lineage genes share an evolutionary history. The subclade 2.1 contains ZCAb sequences while subclade 2.2 contains ZBAb sequences. The only exception is Atlantic salmon sequences where *Sasa-ZBAb* and *Sasa-ZCAb* represents a more recent gene duplication (Figs. 2 & 4). The Sasa-ZBAb seems to b the only soluble Z lineage molecule, lacking the transmembrane region [36].

To summarize, the salmonid MHCIa and MHCIb Z lineage genes are not orthologs of the Northern pike Z lineage genes. Instead, the salmonid Z lineage genes have experienced unique gene duplications in the two duplicated regions sharing an evolutionary history in the MHCIb region, but evolving independently for Oncorhynchus and Salmo species in the MHCIa region. Potentially, transposable elements enabling these duplications were already present in Northern pike. As seen in other teleosts [5], the eight peptide anchoring residues are also conserved in salmonid Z lineage sequences (data not shown).

Evolution Of L Lineage Genes

Northern pike has four L lineage genes dispersed on Chr.2, 15 and 20 where salmonids have orthologs to the pike genes on Chr.02 and Chr.20 based on phylogeny and regional orthology (Table 1, Fig. 5, Additional file 1 & 4). Nomenclature is mostly based on phylogenetic clustering with previously identified L lineage gene sequences [5, 14], as exemplified by the LGA gene sequences, which form a strongly supported phylogenetic cluster. L lineage genes have exploded in salmonids with 12–13 genes in Charr

and Atlantic salmon and 25 genes in Brown trout. Many Charr L lineage genes seem like pseudogenes, while the remaining species have 9–21 bona fide genes.

The previously published Rainbow trout LAA gene [14], is also found in salmonid species, whereas this gene was lost in Northern pike (Additional file 2). Fragments of this gene is found on Atlantic salmon homeolog chromosomes 13 and 15 and ortholog regions in the other salmonids (Additional file 1), flanked by ANKS1A and SARG genes. Only Rainbow trout and Sockeye salmon seem to have bona fide LAA genes, where the LAA genes from the other species are pseudogenes. The sequences are quite distant from the remaining L lineage sequences and form the base of the phylogenetic tree (Additional file 3).

Another older L lineage gene previously described in Atlantic salmon, LIA, [5] has orthologs in all species including Northern pike (Additional file 1–4). LIA gene sequences are also quite old forming a strongly supported branch quite basal in the phylogenetic tree. Only the Charr LIA gene seems to be a pseudogene with an internal stop codon. This LIA gene is flanked by VWA8 and F5 in all species. Although salmonid LIA regions are orthologous to Northern pike Chr. 16, the pike LIA gene resides on Chr. 20, suggesting a translocation in a salmonid ancestor. The salmonid homeolog chromosome also hold L lineage genes in most species represented by the LLA and LJA genes, being mostly pseudogenes. Although not strongly supported, the Northern pike L lineage region on Chr. 15, here called *Eslu-LPA* clusters with the LIA gene sequences and is most likely a gene duplication specific for Northern pike. A similar unique gene duplication is seen for the Atlantic salmon *Sasa-LKA* gene with no orthologous region in other salmonids or Northern pike.

Salmonid LDA gene sequences represent another strongly supported clade, but also clusters with the remaining gene sequences from Northern pike and salmonids (Additional file 2–4). This LDA gene is not present in pike, and also only located on one of the salmonid homeologs, suggesting it translocated to this region after the 4WGD event. The LDA gene is flanked by IRAK1BP1 and IL17RD genes in all species.

Salmonid orthologs to the Northern pike Chr. 2 genes here defined as *Eslu-LBA* and *Eslu-LCA* have expanded a lot with Brown trout being the most extreme with fifteen L lineage genes on Chr.12 (Fig. 5) (Additional file 1). FAH and CTXND1/ ARNT2 genes, flanking the two pike L lineage genes on chromosome 2, are also present in orthologous regions represented by Atlantic salmon Chr.11 and Chr.26 [26]. Most likely due to regional complexity, clustering genes from Coho, Chinook, Sockeye and Charr all reside on unplaced scaffolds. Gene expansions have occurred locally after the 4WGD. For instance, Atlantic salmon Chr.11 with the two duplicate *Sasa-LCA* genes is a homeolog of Rainbow trout Chr. 26 containing nine L lineage genes. Brown trout Chr. 12 and an unplaced Chinook scaffold both display a similar L gene expansion with twelve and eight L lineage genes respectively. Most of the Chinook genes on this unplaced scaffold seem like pseudogenes with internal stop codons while most of those on Rainbow trout Chr.26 seem like bona fide genes. Also LEA/LMA genes as well as LFA genes reside in strongly supported clusters displaying a shared evolutionary history within each of these clusters.

To summarize, the L lineage genes have exploded in some salmonids with Brown trout being the most extreme with 25 L lineage genes. The other salmonids seem to have between five and fourteen functional L lineage genes. A structural investigation of L lineage sequences found them to be able to bind quite hydrophobic structures, possibly analogue to mammalian CD1 molecules [5]. Our understanding of the L gene function has since advanced with the study by Edholm and co-workers [23] showing that L lineage genes display different responses upon stimulation. Six Atlantic salmon L lineage genes were included in their study where *Sasa-LIA* responded to a single-stranded RNA virus but not when challenged with a bacteria. *Sasa-LIA* and *Sasa-LGA* both responded to stimulation by type I interferon A, while *Sasa-LHA* did not. Instead, *Sasa-LHA* responded to a variety of viral and bacterial TLR ligands. These results show that duplicate L lineage genes have acquired a variety of functional roles in protection against pathogens. In particular Brown trout with 21 potentially bona fide genes may hold more surprises when it comes to functional diversity of this MHCI lineage.

Evolution Of S, H And P Lineage Genes

S lineage genes have previously been described in many teleosts [5]. This gene is also present in Northern pike on Chr. 1 (Table 1, Additional file 2 and 3). Most salmonids have duplicate S lineage genes on both homeologs, where the SBA gene has been silenced in a primordial salmonid (Table 1, Additional files 2 &4). The gene is mostly flanked by VWA5A and CIPC/ AKT2 genes in both regions. Atlantic salmon has six S lineage genes, all residing on unplaced scaffolds. Three of these six Atlantic salmon genes seem to be pseudogenes. In a previous study, we sequenced a bacterial artificial chromosome (BAC) clone originating from Chr. 09, which contained one SAA gene in addition to the flanking VWA5 and AKT2 genes [36]. We did not find other BACs positive for the SAA probe, so potentially there are individual differences in the number of SAA genes in Atlantic salmon.

A fifth MHCI lineage described in teleosts is the P lineage, which has expanded to 24 genes in the pufferfish Fugu [5]. Remnants of this P lineage is lacking in Northern pike while all salmonid P lineage genes have been silenced (Table 1, Additional file 2 & 4). Only one homeolog has remnants of this P lineage gene, suggesting it has been deleted in the duplicated region. The PAA gene is surrounded by PPP1R12A_like and Immunoglobulin light chain (Ig-L) genes. We previously found an IgL gene linked to a UIA gene in Medaka and to Z lineage genes in stickleback [5]. IgL genes are also found linked to the shark MHC region, suggesting it was present in the primordial MHC region [38].

We recently found a sixth MHC class I lineage in teleosts which we denoted the H lineage [13]. One HAA lineage gene is present in Northern pike and all salmonids studied here have HAA and HBA genes on homeologs to this pike HAA gene on Chr. 3 (Table 1, Additional files 1, 2, 4). All regions have TOX and PPP1R7 genes flanking the H lineage gene. The HAA genes seem functional in all species, while the HBA gene is a pseudogene at least in Salmo species. In Coho and Chinook, there are expressed reads matching the HBA gene (GGDU01537164.1, GDQG01022515.1), suggesting the homeolog HBA gene has retained a function in some species. The fact that H lineage sequences lack the alpha 3 domain, and has

a cytoplasmic domain highly conserved also between distant teleost species, suggests that teleost MHCI may have a broader functional diversity than previously envisioned. Mammalian equivalents with such a molecular structure are the ULBP/ RAET genes, which interact with the NKG2D receptor upon stress or infection [39]. If the salmonid H lineage molecules have a similar function remains to be seen.

Conclusion

Although both Northern pike as well as salmonids have expanded their U and Z lineage genes, these gene duplications have occurred separately in pike and in a salmonid ancestor. However, the similarity between these duplications suggest the transposable machinery was present in a common ancestor. The salmonid MHCIa and MHCIb regions were mostly formed during the 69 MYA since the split from pike and before the Oncorhynchus and Salmon branch separated. As seen in tetrapods, the non-classical U lineage genes are diversified duplicates of their classical counterpart. One MHCI lineage, the L lineage, experienced massive species-specific gene duplications after Oncorhynchus and Salmo split approximately 25 MYA. Based on what we currently know about L lineage genes, this diversity will most likely affect immune responses in individual species.

Material And Methods

Material

Genomes used in this study are as follows: *Salvelinus alpinus/malma* GCA_002910315.2 (Charr; [24]), *Salmo trutta* GCA_901001165.1 (Brown trout, unpublished), *Oncorhynchus nerka* GCA_006149115.1 (Sockeye salmon; unpublished), *Oncorhynchus tshawytscha* GCA_002872995.1 (Chinook salmon [40]), *Oncorhynchus kisutch* GCA_002021735.2 (Coho salmon; unpublished), *Oncorhynchus mykiss* GCA_002163495.1 (Rainbow trout; [41]), *Salmo salar* GCA_000233375.4 (Atlantic salmon, [9]), and *Esox Lucius* GCA_004634155.1 (Northern pike; [22]).

Data mining

Genome searches were performed using previously identified Atlantic salmon MHC gene sequences [5, 13, 36] and tblastn against selected salmonid genomes. Genomic regions identified through these searches were screened for annotated genes and sometimes regional genes were identified using additional blast searches. Presented data rely on the NCBI genomes where both regional assembly and gene predictions may contain some errors.

Sequence alignments and Phylogenies

Amino acid sequences were aligned using ClustalX [42] with manual corrections for some predicted sequences. Individual domain sequences used in phylogenies were extracted using Jalview [43]. The evolutionary history of selected amino acid sequences was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [44]. Additional phylogenetic trees were also tested using

the Neighbor-Joining method [45](data not shown). The percentage of trees in which the associated taxa clustered together are shown next to the branches. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA7 [46]. Pseudogene sequences were mostly omitted from the analyses.

Declarations

Acknowledgements:

Not applicable

Funding:

This study was funded by the Norwegian Research Council project #274635 who had no role in 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.

Author contribution:

UG was responsible for study design, most data gathering, analyses and manuscript drafting. ML assisted in data gathering, analyses and writing of the manuscript.

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Competing interests:

The authors declare they have no competing interests.

References

1. Klein J. The natural history of the Major Histocompatibility complex. In. New York: John Wiley & Sons; 1986.
2. Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC evolutionary biology. 2013;13:260.
3. Adams EJ, Luoma AM. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol. 2013;31:529–61.
4. Dijkstra JM, Yamaguchi T, Grimholt U. Conservation of sequence motifs suggests that the nonclassical MHC class I lineages CD1/PROCR and UT were established before the emergence of tetrapod species. Immunogenetics. 2018;70(7):459–76.
5. Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM. **A comprehensive analysis of teleost MHC class I sequences**. *BMC evolutionary biology* 2015, 15(32).
6. Stet RJ, Kruiswijk CP, Dixon B. Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Crit Rev Immunol. 2003;23(5–6):441–71.
7. Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J: **Mapping of mhc class I and class II regions to different linkage groups in the zebrafish**, *Danio rerio*. *Immunogenetics* 1997, **46**(2):129–134.
8. Grimholt U. **MHC and Evolution in Teleosts**. *Biology (Basel)* 2016, 5(1).
9. 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.
10. Macqueen DJ, Johnston IA: **A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification**. *Proceedings Biological sciences / The Royal Society* 2014, **281**(1778):20132881.
11. 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.
12. 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.
13. Grimholt U, Tsukamoto K, Hashimoto K, Dijkstra JM. **Discovery of a Novel MHC Class I Lineage in Teleost Fish which Shows Unprecedented Levels of Ectodomain Deterioration while Possessing an Impressive Cytoplasmic Tail Motif**. *Cells* 2019, 8(9).
14. 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(4):305–21.

15. 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.
16. McConnell SC, Restaino AC, de Jong JL. Multiple divergent haplotypes express completely distinct sets of class I MHC genes in zebrafish. Immunogenetics. 2014;66(3):199–213.
17. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, et al. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 2011;477(7363):207–10.
18. 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(1):260–73.
19. Nonaka MI, Aizawa K, Mitani H, Bannai HP, Nonaka M. Retained orthologous relationships of the MHC Class I genes during euteleost evolution. Molecular biology evolution. 2011;28(11):3099–112.
20. 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.
21. Grimholt U. Whole genome duplications have provided teleosts with many roads to peptide loaded MHC class I molecules. BMC evolutionary biology. 2018;18(1):25.
22. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg 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.
23. Svenning S, Gondek-Wyrozemska AT, van der Wal YA, Robertsen B, Jensen I, Jorgensen JB, Edholm ES. Microbial Danger Signals Control Transcriptional Induction of Distinct MHC Class I L Lineage Genes in Atlantic Salmon. Front Immunol. 2019;10:2425.
24. Christensen KA, Rondeau EB, Minkley DR, Leong JS, Nugent CM, Danzmann RG, Ferguson MM, Stadnik A, Devlin RH, Muzzerall R, et al. **The Arctic charr (***Salvelinus alpinus***) genome and transcriptome assembly**. PloS one. 2018;13(9):e0204076.
25. Shedko S: **Assembly ASM291031v2** (**Genbank**: **GCA_002910315.2**) **identified as assembly of the Northern Dolly Varden (**Salvelinus malma malma**) genome, and not the Arctic char** (**S. alpinus**) **genome**. *arXivorg* 2019.
26. Sutherland BJG, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, Bernatchez L. Salmonid Chromosome Evolution as Revealed by a Novel Method for Comparing RADseq Linkage Maps. Genome Biol Evol. 2016;8(12):3600–17.
27. Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire PA. **A Dense Brown Trout (***Salmo trutta***) Linkage Map Reveals Recent Chromosomal Rearrangements in the Salmo Genus and the Impact of Selection on Linked Neutral Diversity**. G3 (Bethesda). 2017;7(4):1365–76.
28. 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 USA. 2016;113(34):E5014–23.

29. Tang WJ, Fernandez J, Sohn JJ, Amemiya CT. Chitin is endogenously produced in vertebrates. Curr Biol. 2015;25(7):897–900.
30. Miller KM, Li S, Ming TJ, Kaukinen KH, Schulze AD. The salmonid MHC class I: more ancient loci uncovered. Immunogenetics. 2006;58(7):571–89.
31. 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.
32. Kiryu I, Dijkstra JM, Sarder RI, Fujiwara A, Yoshiura Y, Ototake M. **New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species**. *Fish Shellfish Immunol* 2005:243–254.
33. O'Farrell B, Benzie JA, McGinnity P, de Eyto E, Dillane E, Coughlan J, Cross TF. **Selection and phylogenetics of salmonid MHC class I: wild brown trout (*Salmo trutta*) differ from a non-native introduced strain**. PloS one. 2013;8(5):e63035.
34. McClelland EK, Ming TJ, Tabata A, Miller KM. **Sequence analysis of MHC class I alpha2 from sockeye salmon (*Oncorhynchus nerka*)**. Fish Shellfish Immunol. 2011;31(3):507–10.
35. Maccari G, Robinson J, Bontrop RE, Otting N, de Groot NG, Ho CS, Ballingall KT, Marsh SGE, Hammond JA. IPD-MHC: nomenclature requirements for the non-human major histocompatibility complex in the next-generation sequencing era. Immunogenetics. 2018;70(10):619–23.
36. 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.
37. Hughes AL, Nei M. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Molecular biology evolution. 1989;6(6):559–79.
38. Ohta Y, Kasahara M, O'Connor TD, Flajnik MF. Inferring the "Primordial Immune Complex": Origins of MHC Class I and Antigen Receptors Revealed by Comparative Genomics. J Immunol. 2019;203(7):1882–96.
39. Bacon L, Eagle RA, Meyer M, Easom N, Young NT, Trowsdale J. Two human ULBP/RAET1 molecules with transmembrane regions are ligands for NKG2D. J Immunol. 2004;173(2):1078–84.
40. Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, Withler RE, Rondeau EB, Koop BF, Devlin RH. **Chinook salmon (*Oncorhynchus tshawytscha*) genome and transcriptome**. PloS one. 2018;13(4):e0195461.
41. 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. Nature communications. 2014;5:3657.
42. 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.

43. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.
44. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8(3):275–82.
45. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology evolution. 1987;4(4):406–25.
46. Kumar S, Stecher G, Tamura K: **MEGA7**: **Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Molecular biology and evolution* 2016, **33**(7):1870–1874.
47. 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.
48. Macqueen DJ, Primmer CR, Houston RD, Nowak BF, Bernatchez L, Bergseth S, Davidson WS, Gallardo-Escarate C, Goldammer T, Guiguen Y, et al. Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. BMC Genomics. 2017;18(1):484.

Figures

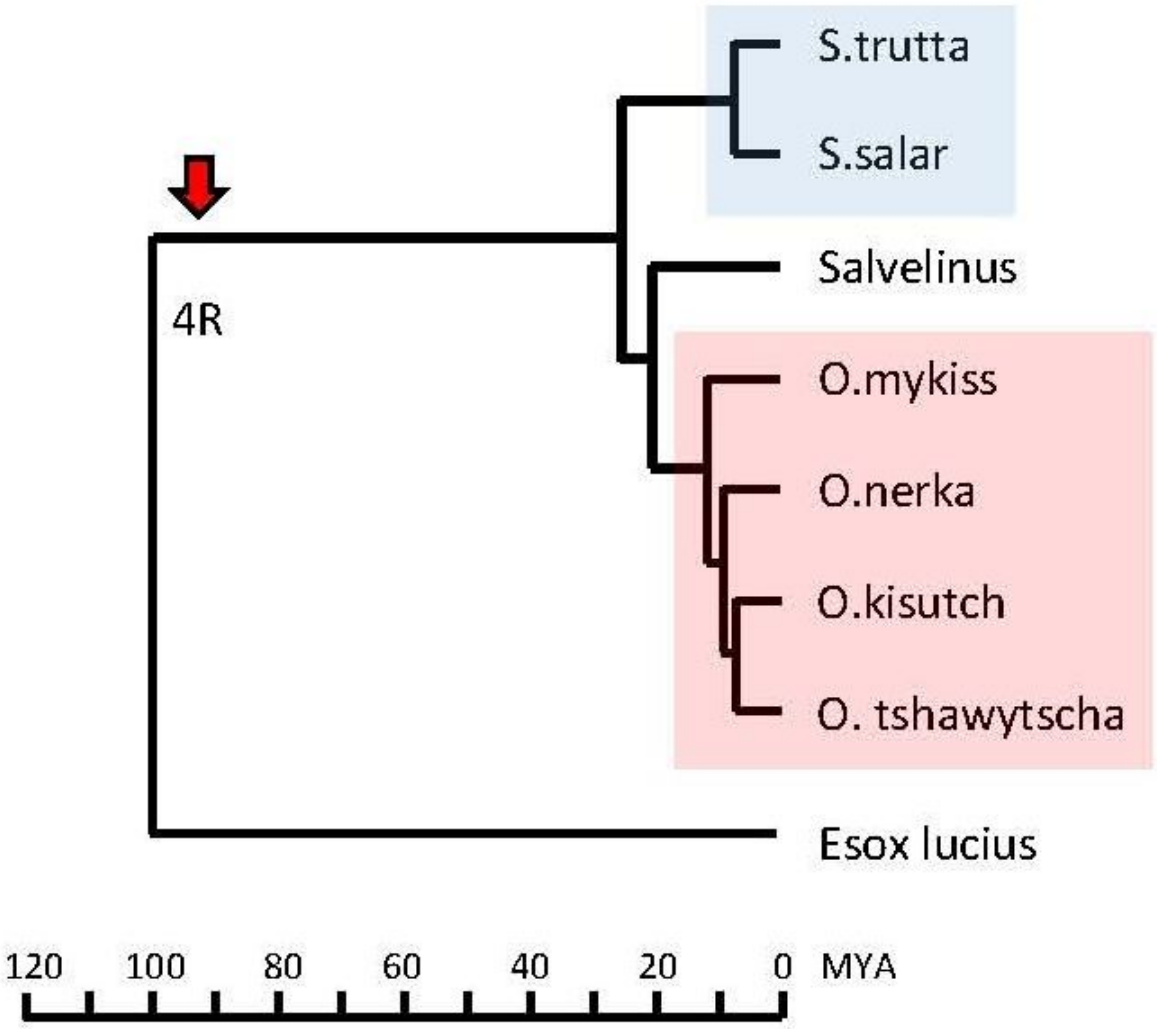

Figure 1

Phylogeny of Salmonidae and Northern pike Phylogenetic relationship between included species. Dating of individual events are based on data from [47, 48]. Salmo and Oncorhynchus species are shown using a blue and red box respectively. The unique salmonid whole genome duplication event that occurred approximately 94 million years ago (MYA)[10] is shown using a red arrow.

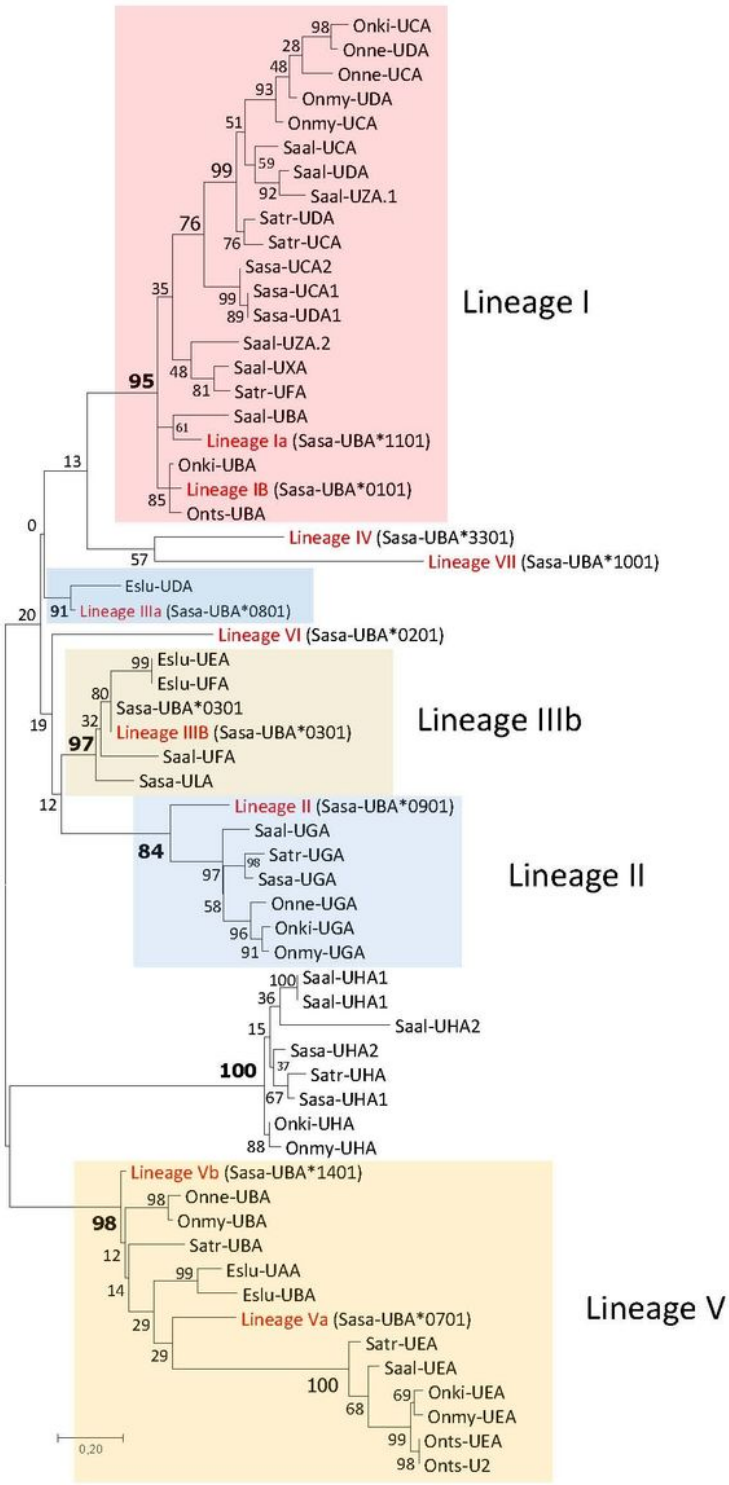

Figure 2

Salmonid MHClb regions compared against the MHCI region of Northern pike. The genomic surroundings of MHC class I genes in selected fishes are shown with genes represented by blocks. Data originate from the NCBI genome database. Chromosomal location of each region is shown below as well as on the left hand side of each region. Gene boxes are colour shaded as follows: red boxes are MHCI genes, green boxes represents a unique MHClb region chitin synthase gene that is present in various numbers (shown

by x), and a LHX9 gene present in Northern pike and the MHCIb region. Remaining genes are shown using grey boxes. The black line in the Northern pike region is introduced due to space restrictions, but represents a continuous region.

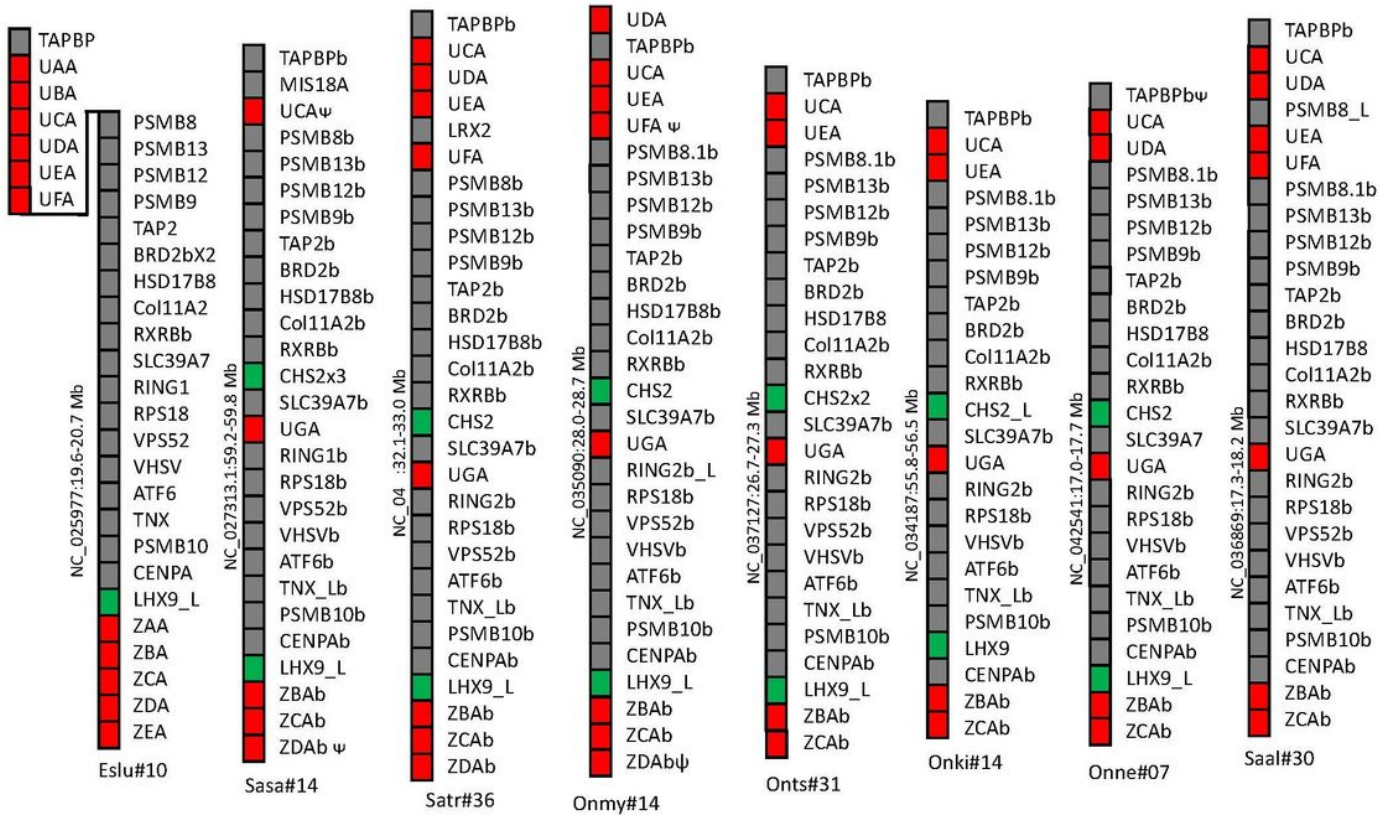

Figure 3

Phylogeny of deduced U lineage alpha 1 domain amino acid sequences. The tree with the highest log likelihood (-3715.54) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5333)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 60 amino acid sequences. There were a total of 85 positions in the final dataset. Atlantic salmon lineage defining sequences originate from [5].

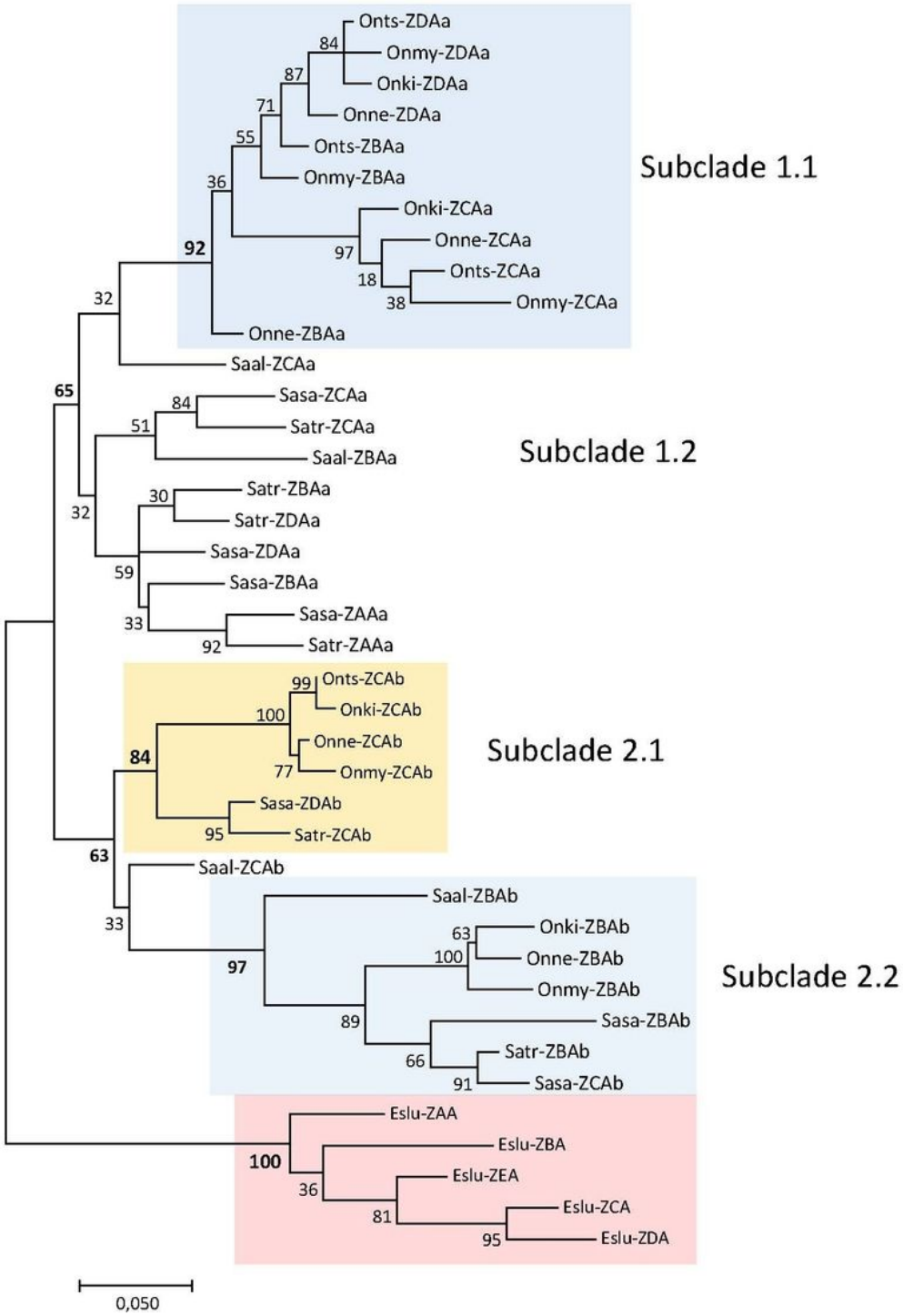

Figure 4

Phylogeny of deduced extracellular Z lineage amino acid sequences. The tree with the highest log likelihood (-3771,17) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0,3726)). The tree is drawn to scale, with branch

lengths measured in the number of substitutions per site. The analysis involved 40 amino acid sequences. There were a total of 282 positions in the final dataset.

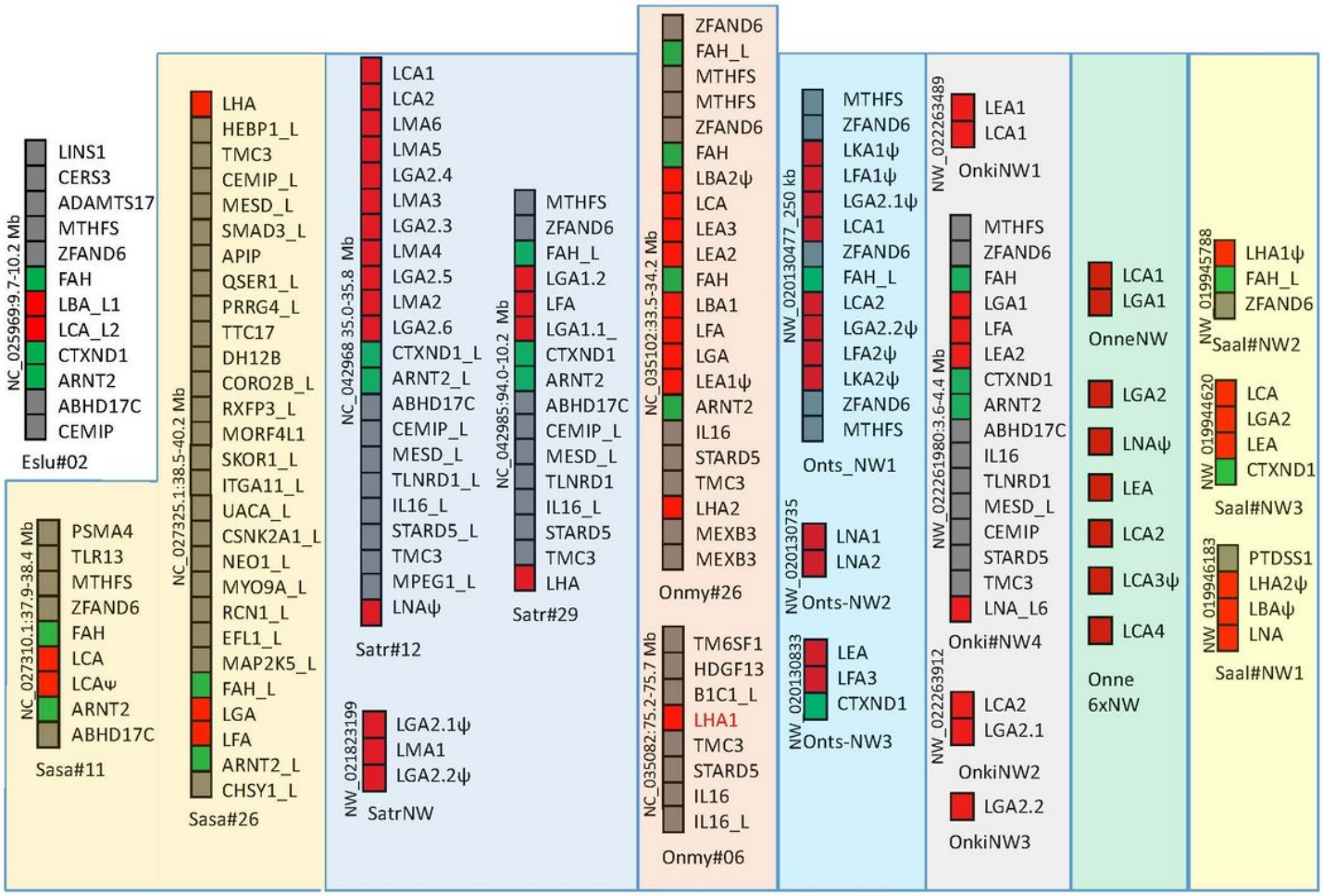

Figure 5

Comparison of L lineage regions from salmonids and Northern pike Genomic regions containing L lineage genes clustering in phylogenetic analyses and based on regional orthology. Genes represented by boxes are colour shaded as follows: red boxes are L lineage genes, green boxes are flanking genes found in most regions and grey boxes are other genes. Additional colour shading is used for regions from each species. Regional location is shown on the side of each region and species and chromosome when available is shown below. Details of unplaced scaffolds can be found in Additional file 3. Atlantic salmon and Rainbow trout genes are on homeolog chromosomes (see Additional file 1), orthology to Brown trout chromosomes is undefined and regions from the remaining species are all unplaced scaffolds (NW), thus proving no informative on orthology. Pseudogenes are shown using ψ. Many genes have the extension _L for _like as they need further phylogenetic and functional studies to warrant a definite gene name.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile4.DeducedMHCIaminoacidsequences.pdf
- AdditionalfileЗ.PhylogenyofMHCIsequences.pdf
- Additionalfile2.ComparedMHCIregions.pdf
- Additionalfile1.Chromosomeorthology.pdf